Economics Letters 261 (2026) 112833

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

An improved density approximation for the Zivot–Andrews test

Riccardo (Jack) Lucchetti

Dipartimento di Scienze Economiche e Sociali (DiSES) Università Politecnica delle Marche, Italy

ARTICLE INFO

Keywords:
Unit root testing
Structural breaks
Simulation methods

ABSTRACT

The unit-root test by Zivot and Andrews(1992) for series with possible structural breaks has been the industry standard for over 30 years. All available software reports the critical values presented in the original article, which were computed with the technology available at the time.

By a much larger simulation exercise, the relevant distributions are approximated by Gaussian mixtures. This makes the computation of p-values straightforward and handles finite-sample issues very naturally.

We show that the discrepancies between the original critical values and ours are relatively minor, but not negligible in some cases.

1. Introduction

The issue of testing for unit roots in time series potentially subject to structural breaks was first thoroughly investigated by Perron (1989), who proposed a strategy based on three possible specifications for the type of break. In Perron's contribution, the break date was assumed known; shortly afterwards, this limitation was overcome by Zivot and Andrews (1992), who proposed a simple adaptation of Perron's setup to accommodate cases when the break date is unknown. The literature has then picked up from this intuition and extended it in various directions: for example, Gregory and Hansen (1996) generalised the idea to a cointegration context and Lumsdaine and Papell (1997) extended the analysis to allow for two possible breakpoints instead of one.

The success of the Zivot–Andrews test has been enormous: at the time of this writing, Google Scholar reports 11,311 citations for the 1992 article. Unsurprisingly, the test is implemented in all major software packages, either natively or via add-on packages. All these implementations use the critical values that were tabulated in the original paper by computing empirical quantiles from 5000 realisations of the asymptotic distribution.

In this paper, I perform a much larger simulation exercise, which makes it possible to approximate the desired densities very accurately in a computationally cheap way via normal mixtures.

2. Notation and computational procedure

The unit-root test under examination comes in three different flavours: model (A), in which the series may be subject to an abrupt change in level (the so-called "crash" model), model (B), in which the structural break affects the slope of a deterministic linear trend and (C), which combines the two previous ones. Here I follow the notation in Perron (1989) and Zivot and Andrews (1992) and indicate with T_B the observation at which the structural break occurs. The three models corresponding to cases (A), (B) and (C) are simple variations on the Augmented Dickey–Fuller theme:

$$y_t = \mu + \theta\ DU_t + \beta t + d\ D(T_B)_t + \alpha y_{t-1} + \sum_{j=1}^{k} c_j \Delta y_{t-j} + \varepsilon_t \tag{1}$$

$$y_t = \mu + \beta t + \gamma\ DT_t^* + \alpha y_{t-1} + \sum_{j=1}^{k} c_j \Delta y_{t-j} + \varepsilon_t \tag{2}$$

$$y_t = \mu + \theta\ DU_t + \gamma\ DT_t^* + \beta t + d\ D(T_B)_t + \alpha y_{t-1} + \sum_{j=1}^{k} c_j \Delta y_{t-j} + \varepsilon_t \tag{3}$$

where $D(T_B)_t = \mathbb{1}\left[t = T_B + 1\right]$, $DU_t = \mathbb{1}\left[t > T_B\right]$, $DT_t^* = \mathbb{1}\left[t > T_B\right](t - T_B)$, and $\mathbb{1}[\cdot]$ is the indicator function. As the in ADF test, the relevant test statistic is the t-ratio for the parameter α.

The limit distributions of the corresponding three test statistics are functionals of suitably modified Brownian motions and lack a closed-form representation. Moreover, it is well known that in some cases the finite-sample distribution of unit-root tests may be quite different from the asymptotic one: Sephton (2017) and Sephton (2022) provide adjustments for the KPSS and the ADF-GLS tests, respectively. In the original article, a simulation exercise (quite comprehensive at the time) was used to investigate the consequences on the test distribution of short-term dependence and non-normality in the innovations ε_t. That approach, however, is quite expensive in practice and is very seldom used in actual applied work. Therefore, we do not pursue this avenue here, but we do provide a finite-sample correction for the test statistic in the standard Gaussian case, explained below.

E-mail address: r.lucchetti@univpm.it.

https://doi.org/10.1016/j.econlet.2026.112833
Received 29 July 2025; Received in revised form 22 December 2025; Accepted 22 January 2026
Available online 27 January 2026

The procedure we followed is loosely based on the one in Appendix B in Zivot and Andrews (1992): for a given sample size T, a random-walk process y_t is generated by cumulating standard normal pseudo-random variates.[1] Then, one of the regression Eqs. (1)–(3) is estimated by OLS (with $k = 0$) and for each T_B the t statistic for $\alpha = 1$ is computed ($t_\alpha = \frac{\hat{\alpha}-1}{se_\alpha}$). The simulated test statistic is the minimum t_α across all values of T_B in an integer grid between h and $T-h+1$, where $h = \max(T/20, d+1)$ is a trimming parameter used to ensure that the regressions have a sensible number of degrees of freedom. The scalar d is the number of deterministic terms included in Eqs. (1)–(3), so $d = 3$ for models A and B and $d = 4$ for model C. This choice does not mirror exactly the one originally made in Zivot and Andrews (1992), since it makes trimming dependent on the sample size T, which was fixed at 1000 in the original article. Experimentation with other trimming choices yielded qualitatively identical results.

For each of the three cases, this procedure is repeated 60,000 times on a grid where $T = 2^H$, where $H = 6 \dots 12$, so $T = \{64, \dots, 4096\}$.[2]

The estimated density plots of the test statistic for the three cases are shown in Fig. 1, which are, unsurprisingly, very similar to the thick lines in Fig. 1 of the original article. As can be seen, however, for models A and C the distribution of the test for $T = 64$ is quite different from the asymptotic one. On the contrary, the shape of the density for model B looks rather stable as a function of T, but it seems to be much more skewed towards the left.

Those densities were then approximated via three different models, which can be generally described as

$$f(x|T) \simeq \sum_{i=1}^{K} \lambda_i \frac{1}{\sigma_i(T)} \varphi\left[\frac{x-\mu_i(T)}{\sigma_i(T)}\right], \tag{4}$$

where $\varphi[\cdot]$ is the standard normal density. In practice: the density of the test statistic x for a given sample size T is approximated by a mixture of K normal densities, whose mean and variance may be a function of the sample size T.[3] Of course, the mixture weights λ_i sum to 1.[4] This choice makes it easy to compute approximate p-values for one-tailed tests simply by

$$\hat{F}(x|T) \simeq \sum_{i=1}^{K} \hat{\lambda}_i \Phi\left[\frac{x-\hat{\mu}_i(T)}{\hat{\sigma}_i(T)}\right] \tag{5}$$

where $\Phi(\cdot)$ is the standard normal CDF. Appendix section contains an example implementation in Python. Critical values, if needed, can be computed easily via standard root-finding methods such as the one by Ridders (1979).

It should be noted that the p-values calculated via Eq. (5) correspond to the design described above, that is a Gaussian random-walk null, with the test regressions estimated using $k = 0$. In empirical implementations, however, k may have to be chosen by a data-dependent rule and non-iid issues, such as serial correlation and conditional heteroskedasticity may also have to be taken into account. In those cases, the p-values calculated by the present approach may be nothing more than a convenient approximation.

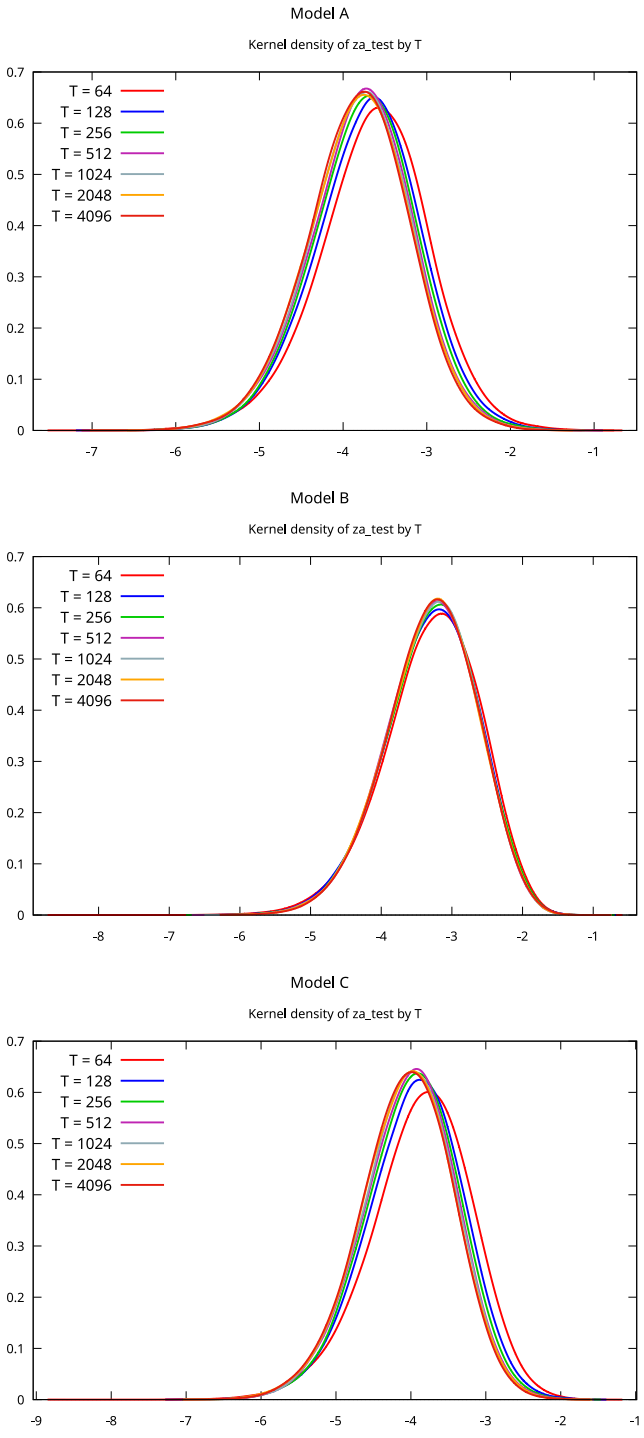

Fig. 1. Kernel density estimates.

3. Results

Since 60,000 simulated values were generated for 7 different values of T for each of the three models, the parameters of Eq. (4) were estimated by maximum likelihood with 420,000 simulated observations per model. On the basis of the Akaike information criterion, $K = 3$ was chosen for all models.

In practice, for models A and C, I chose the following specification for mean and standard deviation:

$$\mu_i(T) = a_{0,i} + \frac{a_{1,i}}{\sqrt{T}}; \tag{6}$$

[1] The algorithm used is a slight refinement of the Ziggurat method by Marsaglia and Tsang (2000); the uniform generator is xoshiro256+ by Blackman and Vigna (2021).

[2] The minimum value of 64 was chosen as representative of what could be considered a relatively small sample in a typical modern time-series empirical exercise. Applying the results of the present paper to much smaller datasets is, in my opinion, quite inadvisable.

[3] This idea is similar in spirit to the one used in Doornik (1998) and Boswijk and Doornik (2005), who use the gamma distribution to approximate cointegration tests.

[4] In principle, the mixture weights λ_i could also be a function of T, but this was found to be unnecessary in the present case.

Table 1
Parameter estimates.

	Model A		Model B		Model C	
	Estimate	s.e.	Estimate	s.e.	Estimate	s.e.
$a_{0,1}$	−3.5997	0.1044	−3.2953	0.0510	−3.8208	0.0212
$a_{0,2}$	−3.7138	0.0097	−3.8443	0.1354	−4.2469	0.0557
$a_{0,3}$	−4.0897	0.0554	−2.6643	0.0399	−4.2597	0.0721
$a_{1,1}$	−2.0437	1.0233			2.2648	0.3175
$a_{1,2}$	1.8127	0.1422			−3.5956	1.2792
$a_{1,3}$	3.1183	0.4317			2.5215	0.3276
$b_{0,1}$	0.5870	0.0256	0.5049	0.0372	0.4634	0.0169
$b_{0,2}$	0.4722	0.0083	0.6543	0.0237	0.7595	0.0338
$b_{0,3}$	0.6571	0.0121	0.3818	0.0185	0.5962	0.0174
$b_{1,1}$	2.2515	0.4373			0.3643	0.0816
$b_{1,2}$	0.1151	0.0651			−0.5742	0.4592
$b_{1,3}$	0.0720	0.2405			−0.0502	0.1951
λ_1	0.1688	0.0595	0.5764	0.1357	0.3636	0.0851
λ_2	0.4127	0.0341	0.2233	0.0732	0.1090	0.0409
λ_3	0.4185	0.0685	0.2002	0.0727	0.5274	0.0959

Table 2
Descriptive statistic on the estimated CDFs.

	Model A	Model B	Model C	Theoretical
Mean	0.5000477	0.5000531	0.5000424	0.5
Median	0.5007546	0.5015293	0.4993337	0.5
Minimum	5.8413786E−07	1.0024051E−14	8.4269109E−11	0
Maximum	0.9999995	0.9999592	0.999995	1
Std. Dev.	0.2887865	0.2886433	0.288747	0.2886751
Skewness	−0.0023257	−0.0017289	0.0004929	0
Ex. kurtosis	−1.2051838	−1.1953614	−1.2034184	−1.2
5% perc.	0.0507296	0.0501873	0.0502872	0.05
95% perc.	0.949595	0.9510948	0.9493854	0.95
IQ range	0.5016917	0.4988682	0.5011459	0.5

Table 3
Probability integral transform for $T = 192$: descriptive statistics.

	Model A	Model B	Model C	Theoretical
Mean	0.503527	0.505386	0.498989	0.5
Median	0.505731	0.508330	0.498173	0.5
Minimum	0.000104	0.000007	0.000064	0
Maximum	0.999883	0.999688	0.999963	1
Std. Dev.	0.288159	0.293013	0.289570	0.2886751
Skewness	−0.022284	−0.017505	0.017344	0
Ex. kurtosis	−1.191198	−1.233505	−1.208682	−1.2
5% perc.	0.048214	0.051328	0.050896	0.05
95% perc.	0.948233	0.954988	0.952115	0.95
IQ range	0.499418	0.515154	0.502203	0.5

$$\sigma_i(T) = b_{0,i} + \frac{b_{1,i}}{\sqrt{T}}. \tag{7}$$

Experiments with additional terms using different powers of T yielded inferior results and are not reported. In this formulation, the $a_{0,i}$ and $b_{0,i}$ parameters describe the asymptotic distribution while the $a_{1,i}$ and $b_{1,i}$ parameters provide the necessary small-sample correction. For model B, the relative stability of the density with respect to T made all the estimated parameters $a_{1,i}$ and $b_{1,i}$ not significant, so they were omitted from the final specification. Parameter estimates are displayed in Table 1.

In order to check the quality of the approximation, Table 2 reports summary statistics for the probability integral transform $\hat{F}(x)$ (see Eq. (5)) computed on the 420,000 simulated draws using the estimates in Table 1. Of course, the distribution of $F(x)$ should be $U(0,1)$. The rightmost column reports the theoretical values for a $U(0,1)$ distribution. As can be seen from the table, the approximation is perfectly satisfactory for practical purposes.

A further check can be carried out by simulating the statistic with a value of T not included in the grid used for the simulation above, and therefore interpolated: Table 3 is similar to Table 2 but was calculated on 10000 simulations for $T = 192$, which was not used for the computation of the mixture coefficients and can be considered representative of a fairly typical macroeconomic application. As can be seen, the approximation seems to be sufficiently accurate in practice.

Table 4
Original and re-computed critical values.

	Model A			Model B			Model C		
α	c_{ZA}	$\hat{F}(c_{ZA})$	c_M	c_{ZA}	$\hat{F}(c_{ZA})$	c_M	c_{ZA}	$\hat{F}(c_{ZA})$	c_M
1	−5.34	0.95	−5.33	−4.93	1.12	−4.96	−5.57	1.07	−5.59
2.5	−5.02	2.90	−5.07	−4.67	2.50	−4.67	−5.30	2.71	−5.32
5	−4.80	5.65	−4.84	−4.42	4.98	−4.42	−5.08	5.33	−5.10
10	−4.58	10.24	−4.59	−4.11	10.72	−4.14	−4.82	10.79	−4.85
50	−3.75	51.88	−3.78	−3.23	51.62	−3.26	−3.98	52.30	−4.02
90	−2.99	91.39	−3.04	−2.48	90.22	−2.49	−3.25	90.83	−3.28
95	−2.77	95.90	−2.83	−2.31	94.78	−2.30	−3.06	95.33	−3.08
97.5	−2.56	98.15	−2.64	−2.17	97.18	−2.15	−2.91	97.46	−2.91
99	−2.32	99.31	−2.41	−1.97	99.01	−1.97	−2.72	98.93	−2.71

Table 4 compares our results with the ones obtained by Zivot and Andrews (1992) in their limited Monte Carlo experiment.[5] Each line refers to a given quantile α of the distribution (in percent); for each case, three columns are reported: c_{ZA}, the critical value reported in Tables 2–4 in the original article by Zivot and Andrews (1992); $\hat{F}(c_{ZA})$, the actual left-hand tail probability corresponding to c_{ZA}, computed via Eq. (5) and the corrected critical value c_M. Note that the critical values are reported in the original article with only two decimals, which limits the accuracy of the comparison to some extent.

The accordance between the original estimates and ours is generally rather good, especially so for model B, although it should be noted that the largest discrepancies appear to affect the tails of the distribution, with the left-hand tail arguably being the one of greatest practical importance. Those discrepancies lead to slightly less conservative appraisals: for example, the test statistic reported in the original article[6] for the "real wages" series (bottom of Table 6) is −4.7441, for which model C was used on a sample with 62 observations. In the original paper this number is compared to the asymptotic 10% critical value of −4.85 and therefore the statistic was seen to lie quite comfortably outside the rejection region, whereas its p-value is equal to 10.06% according to the approximation proposed here, which takes finite-sample effects into account.

4. Conclusions

This paper provides easy to use formulae to compute approximate p-values for the three varieties of the Zivot–Andrews test, thereby improving on the original rough tables of critical values. Although the differences with the customarily used critical values are relatively minor, it is advisable to use the approximation provided here, especially in the light of the fact that the largest differences appear between the fifth and tenth percentiles of the distribution, the most important area in practice. In that area, our p-values are slightly smaller (with differences not exceeding the first or second digit when the p-value is expressed as a percentage).

Moreover, the method proposed here takes into account that the asymptotic distribution for models A and C can provide a poor approximation in small samples, such as those that can be encountered in practical macroeconomic applications.

[5] In the original article the asymptotic distribution was approximated via a numerical simulation with $T = 1000$, so the same choice was kept here.

[6] Data used in the original article, as well as Perron (1989), come from the famous Nelson and Plosser (1982) dataset.

Appendix. Implementation

What follows is an example implementation of the *p*-value calculation algorithm in Python 3.

```
import numpy as np
from scipy.stats import norm

def ZA_pval(x, model, T):
  """
  Inputs:
  x    : test statistic
  model : 0 for model A, 1 for model B, 2 for model C
  T    : sample size (irrelevant for model B)

  Output:
  pvalue for the Zivot-Andrews test
  """
  # parameters for the 3 mixture densities (from Table 1)
  mixparam = np.array([
    [ -3.5996599318743 , -3.2953169619380 , -3.8208292227089 ],
    [ -3.7137959349271 , -3.8443307759503 , -4.2468694259444 ],
    [ -4.0896680493990 , -2.6643311985097 , -4.2596709690664 ],
    [ -2.0437103996984 ,  0.0000000000000 ,  2.2647562978488 ],
    [  1.8126809565698 ,  0.0000000000000 , -3.5956125697606 ],
    [  3.1183341773607 ,  0.0000000000000 ,  2.5215376489005 ],
    [  0.5869570034084 ,  0.5049155767709 ,  0.4634122749765 ],
    [  0.4722198037167 ,  0.6542940408129 ,  0.7594679205523 ],
    [  0.6571031977369 ,  0.3818231624152 ,  0.5962311904965 ],
    [  2.2515395169676 ,  0.0000000000000 ,  0.3643460273986 ],
    [  0.1150624184873 ,  0.0000000000000 , -0.5742100745147 ],
    [  0.0719931996014 ,  0.0000000000000 , -0.0501705651492 ],
    [  0.1687967777211 ,  0.5764454420524 ,  0.3635845655491 ],
    [  0.4126878946409 ,  0.2233088988575 ,  0.1090155191074 ],
    [  0.4185153276381 ,  0.2002456590901 ,  0.5273999153436 ]
  ])

  # Extract parameter block relevant for the chosen case

  b0 = mixparam[0:3, model]
  b1 = mixparam[3:6, model]
  a0 = mixparam[6:9, model]
  a1 = mixparam[9:12, model]
  lam = mixparam[12:15, model]

  # Means and standard deviations of the three mixture components
  m = b0.copy()
  s = a0.copy()

  if model != 1:
    # adjust for sample size
    m += b1 / np.sqrt(T)
    s += a1 / np.sqrt(T)

  # Normal CDF of the standardized mixture components
  P = norm.cdf((x - m[None, :]) / s[None, :])

  # Compute mixture and return as a scalar double
  mix = P @ lam
  return float(mix[0])
```

Data availability

No data was used for the research described in the article.

References

Blackman, D., Vigna, S., 2021. Scrambled linear pseudorandom number generators. ACM Trans. Math. Softw. (TOMS) 47 (4), 1–32.
Boswijk, H.P., Doornik, J.A., 2005. Distribution approximations for cointegration tests with stationary exogenous regressors. J. Appl. Econometrics 20 (6), 797–810.
Doornik, J.A., 1998. Approximations to the asymptotic distributions of cointegration tests. J. Econ. Surv. 12 (5), 573–593.
Gregory, A.W., Hansen, B.E., 1996. Residual-based tests for cointegration in models with regime shifts. J. Econometrics 70 (1), 99–126.
Lumsdaine, R.L., Papell, D.H., 1997. Multiple trend breaks and the unit-root hypothesis. Rev. Econ. Stat. 79 (2), 212–218.
Marsaglia, G., Tsang, W.W., 2000. The Ziggurat method for generating random variables. J. Stat. Softw. 5, 3–30.
Nelson, C.R., Plosser, C.R., 1982. Trends and random walks in macroeconomic time series: some evidence and implications. J. Monet. Econ. 10 (2), 139–162.
Perron, P., 1989. The great crash, the oil price shock, and the unit root hypothesis. Econometrica 1361–1401.
Ridders, C., 1979. A new algorithm for computing a single root of a real continuous function. IEEE Trans. Circuits Syst. 26 (11), 979–980.
Sephton, P.S., 2017. Finite sample critical values of the generalized KPSS stationarity test. Comput. Econ. 50 (1), 161–172.
Sephton, P.S., 2022. Finite sample lag adjusted critical values of the ADF-GLS test. Comput. Econ. 59 (1), 177–183.
Zivot, E., Andrews, D.W.K., 1992. Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. J. Bus. Econom. Statist. 20 (1), 25–44.