Econometric Theory, **19**, 2003, 1128–1143. Printed in the United States of America.
DOI: 10.1017/S0266466603196107

CRITICAL VALUES AND *P* VALUES OF BESSEL PROCESS DISTRIBUTIONS: COMPUTATION AND APPLICATION TO STRUCTURAL BREAK TESTS

ARTURO ESTRELLA
Federal Reserve Bank of New York

The *p* values of structural break tests, when the break date or dates are unknown, must be calculated in terms of the probability distributions of functions of Bessel processes. The literature so far has maintained that direct computation of these *p* values and of the corresponding critical values is too difficult and has relied on approximations based on simulations, asymptotic expansions, or curve fitting. This paper presents a fast simple method of calculating exact *p* values and critical values and uses the method to evaluate the accuracy of the various approximations.

1. INTRODUCTION

The last decade has seen a surge in the development of statistical tests for breaks in time series models, particularly when the exact break point is unknown. For example, Andrews (1993) provides an extensive study of the properties of unknown break point tests based on generalized method of moments (GMM) estimation. These tests correspond to the classical Wald, likelihood ratio, and Lagrange multiplier tests. Ghysels, Guay, and Hall (1997) have developed an alternative test, also applicable to GMM estimates, that is based on the accuracy of predictions of the orthogonality conditions used in estimation. Bai and Perron (1998) and Bai (1999) present tests for multiple unknown break points.

In contrast to tests of known break points, which have asymptotic chi square distributions, the asymptotic distribution of unknown break point tests is related to less standard Bessel process distributions. These Bessel distributions present several challenges for the applied researcher. For instance, there are no closed-form expressions for the distribution function, which makes it difficult to compute *p* values. Moreover, the distribution depends on two parameters, which makes it somewhat inconvenient to tabulate critical values.

The author is grateful for comments and suggestions from Don Andrews, Clint Cummins, Jeff Fuhrer, Ken Garbade, Jim Mahoney, Tony Rodrigues, Josh Rosenberg, Sebastian Schich, participants in a workshop at the Federal Reserve Bank of New York, and the referees. The views expressed in this paper are those of the author and do not necessarily represent those of the Federal Reserve Bank of New York or the Federal Reserve System. Address correspondence to: Arturo Estrella, Federal Reserve Bank of New York, 33 Liberty Street, New York, NY 10045, USA; e-mail: arturo.estrella@ny.frb.org.

 © 2003 Cambridge University Press 0266-4666/03 $12.00

https://doi.org/10.1017/S0266466603196107 Published online by Cambridge University Press

Thus, the literature has generally deemed it too difficult to calculate exact p values and critical values and has relied almost exclusively on approximations to the distribution function or to its inverse function. These approximations are based on simulations (as in Andrews, 1993), asymptotic formulas (based on the limit of the distribution function as the test value approaches infinity, as in DeLong, 1981, and others), and curve fitting (as in Hansen, 1997).

The main objectives of this paper are to show that it is actually quite feasible to calculate the exact values of the asymptotic distribution function and to examine the accuracy of the approximations that have been employed in the past. The computational method suggested is easily programmed using existing statistical packages, be it as a subroutine or as a feature of the package itself.

Of course, a compelling case for the calculation of exact values would make the use of approximations less relevant going forward. It is nevertheless important to gauge the accuracy of the approximations, because the literature already contains many applications and it is useful to know whether existing results would be affected by the substitution of approximations for exact values.

The following section describes the Bessel process distribution and the role it plays in tests for unknown break points. Section 3 then considers the calculation of exact p values and critical values and presents exact results analogous to those of Andrews (1993), which are based on simulations. Sections 4 and 5 deal with the asymptotic and Hansen approximations, respectively, and with the accuracy of those approximations. Section 6 concludes. Proofs of the propositions are found in the Appendix.

2. THE USE OF BESSEL PROCESS DISTRIBUTIONS IN BREAK TESTING

2.1. Tests for a Single Unknown Break Point

Suppose that the model to be tested is indexed by a vector of parameters θ_t, where t is an index of time. The subscript indicates that the value of the parameter vector may change from period to period. The null hypothesis is that the parameters are constant throughout the sample, that is,

$$H_0 : \theta_t = \theta_0, \qquad t = 1,2,\ldots,T. \tag{1}$$

The alternative is that the parameter vector changes at a proportion π of the sample, that is, at time $T\pi$:

$$H_a : \theta_t = \begin{cases} \theta_1(\pi), & t = 1,\ldots,T\pi \\ \theta_2(\pi), & t = T\pi + 1,\ldots,T. \end{cases} \tag{2}$$

To perform this test, Andrews (1993) discusses GMM analogues of the three classical tests: Wald, likelihood ratio, and Lagrange multiplier. The exact form of the statistics may be found in that paper. Like their classical counterparts,

https://doi.org/10.1017/S0266466603196107 Published online by Cambridge University Press

these statistics are distributed asymptotically as chi squared when the break point parameter π is known. If π is unknown, but is suspected to lie in a proper subset of the unit interval, the break point may be estimated by locating the value of π that corresponds to the strongest indication of a break, that is, to the largest value of one of the three statistics. Andrews and Ploberger (1984) also consider statistics based on the average or exponential average value of the break statistic over a subset of the unit interval. In this paper, we focus on the tests based on the supremum of the statistic, which are widely used and have a direct intuitive interpretation.

Thus, let $\Pi = [\pi_1, \pi_2]$ be a closed interval contained in (0,1) and let S be the set of observations from $t = [\pi_1 T]$ to $t = [\pi_2 T]$, where $[\cdot]$ denotes rounding to the nearest integer, that is, $S = \{t | [\pi_1 T] \leq t \leq [\pi_2 T]\}$. Consider the supremum of the Wald (W), likelihood ratio (LR), or Lagrange multiplier (LM) statistic from among the values computed at every point in S and let ζ denote one of $\sup_S W$, $\sup_S LR$, or $\sup_S LM$. Andrews (1993) shows that, under certain regularity conditions, the distribution of ζ is a function of multivariate Brownian motion, which may be constructed as follows.

Let $B_p(\cdot)$ be a p-dimensional vector of independent Brownian motions. Then $\|B_p(\pi)\| = (B_p(\pi)' B_p(\pi))^{1/2}$ is the Bessel process of order p and $BB_p(\pi) = B_p(\pi) - \pi B_p(1)$, known as a Brownian bridge. If $B_p(\cdot)$ is defined on [0,1] and restricted to Π, then

$$Q_p(\pi) = \frac{(B_p(\pi) - \pi B_p(1))'(B_p(\pi) - \pi B_p(1))}{\pi(1-\pi)} = \sum_{i=1}^{p} \frac{(BB_p(\pi)_i)^2}{V(BB_p(\pi)_i)} \tag{3}$$

is the square of the standardized tied-down Bessel process of order p.

When one of the GMM break tests is performed with a known value of π, the asymptotic distribution of the test statistic corresponds to $Q_p(\pi)$, which in that case has a χ^2_p distribution, as demonstrated by Andrews and Fair (1988). However, when π is unknown and is estimated by locating the supremum of one of the test statistics, Andrews (1993) shows that the relevant distribution is

$$P(\zeta > c) = P\left(\sup_{\pi \in \Pi} Q_p(\pi) > c\right) = P\left(\sup_{1<s<\lambda} \frac{\|B_p(s)\|}{s^{1/2}} > c^{1/2}\right), \tag{4}$$

where $\lambda = \pi_2(1-\pi_1)/\pi_1(1-\pi_2)$. Note that the probability depends on π_1 and π_2, not individually, but only through λ. Note also that when $\pi_1 = 1 - \pi_2 = \pi_0$, then $\lambda = ((1-\pi_0)/\pi_0)^2$. The interpretation of equation (4) is that the distribution of the supremum of the tied-down Bessel process is equivalent to the distribution of the supremum of the standard Bessel process exceeding a square-root boundary, which has been studied by DeLong (1981).

In addition to Andrews (1993), other papers have used Bessel process distributions to test for single unknown breaks in models. Among these are Siegmund (1986) for a single parameter, James, James, and Siegmund (1987)

https://doi.org/10.1017/S0266466603196107 Published online by Cambridge University Press

for multiple parameters, and Akman and Raftery (1986) for a single Poisson parameter.

Ghysels et al. (1997) introduce a test for multiple parameters, which is based on estimating the model by GMM over a portion of the sample and predicting the value of the GMM orthogonality conditions over the remaining portion. When the model is just-identified, their statistic measures the distance of the orthogonality conditions from zero, which would hold if the model fit perfectly over the remaining portion of the sample. In this case, the asymptotic distribution of the statistic is the same as in Andrews (1993).

2.2. Tests for Multiple Breakpoints

Some tests for multiple break points also use the Bessel process distribution from equation (4). For instance, Bai (1999) constructs a LR test for $m + 1$ breaks under the null of m breaks. Let $0 < \varphi_1 < \cdots < \varphi_m < 1$ be the proportions of the sample representing the m break points estimated under the null, with $\varphi_i - \varphi_{i-1} > \pi_0$ for some small value π_0. Define also $\eta_i = \pi_0/(\varphi_i - \varphi_{i-1})$, $i = 1, \ldots, m + 1$, where we take $\varphi_0 = 0$ and $\varphi_{m+1} = 1$.

The LR test is obtained by estimating $m + 1$ break points and comparing the likelihood to the m break likelihood. In this case, the distribution of the LR is asymptotically

$$P(LR > c) = 1 - \prod_{i=1}^{m+1} \left(1 - P\left(\sup_{\pi \in [\eta_i, 1-\eta_i]} Q_p(\pi) > c\right)\right). \tag{5}$$

Thus, p values for LR may be computed by repeated application of the formula for p values in the single-break case.

Bai and Perron (1998) also use the Bessel process distribution in tests for multiple break points. They develop two types of tests, one for no breaks vs. a fixed number k of breaks and the other a sequential test for m vs. $m + 1$ breaks. In the former, the asymptotic distribution is the Bessel process when $k = 1$, whereas in the sequential test, the asymptotic distribution is given by repeated application of the Bessel process, as in Bai (1999).

2.3. Other Statistical Applications

Bessel process distributions are also useful in statistical applications other than structural break tests. For instance, Anderson and Darling (1952) construct a variant of the Kolmogorov–Smirnov test for goodness of fit, whose distribution is given by the Bessel process. Siegmund (1985) is a comprehensive survey of sequential analysis tests, several of which are asymptotically distributed as the Bessel process. For instance, a repeated significance test for nonzero drift in Brownian motion $W(t)$ is based on the first time that $|W(t)| \geq b\sqrt{t}$ within a time interval $m_0 \leq t \leq m$. Its distribution is essentially the same

https://doi.org/10.1017/S0266466603196107 Published online by Cambridge University Press

as (4). Siegmund (1986) contains a more detailed summary of these and other applications.

3. EXACT *P* VALUES AND CRITICAL VALUES

3.1. DeLong's Exact Formula

Equation (4) in the previous section presents two equivalent formulations for the asymptotic distribution of the supremum-based break tests, one in terms of the standardized tied-down Bessel process and one in terms of the Bessel process itself. In an important and influential paper, DeLong (1981) gives an exact expression for the Bessel process formulation, which involves the crossing of a square-root boundary. Virtually all the numerical applications of the Bessel process to break testing because 1981 correspond to approximations of DeLong's formula.

The reason for the use of approximations is the generally perceived intractability of the expression of DeLong (1981). In this section, we examine the expression and provide some results that simplify its calculation. With these results, we argue that the calculation is in fact fairly quick and simple and that there is little reason to resort to approximations.

To state DeLong's expression, let $M(a, b, z)$ be the confluent hypergeometric function

$$M(a,b,z) = \sum_{k=0}^{\infty} \frac{(a)_k z^k}{(b)_k k!}, \tag{6}$$

where $(a)_k \equiv a(a+1)\ldots(a+k-1)$ is the Pochhammer symbol, with $(a)_0 \equiv 1$. The infinite series converges as long as b is not a negative integer. Let $M_1(a, b, z)$ be the partial derivative of this function with respect to its first argument. DeLong (1981) shows that

$$P(\zeta > c) = P\left(\sup_{1<s<\lambda} \|B_p(s)\| > c^{1/2} s^{1/2} \right) = 1 - \sum_{j=1}^{\infty} \alpha_j \lambda^{-\beta_j}, \tag{7}$$

where β_j is the jth root of $M(\beta + p/2, p/2, -c/2)$ as a function of $\beta > 0$,

$$\alpha_i = -\frac{\exp(-c/2)(c/2)^{p/2} M(\beta_i + p/2, p/2 + 1, -c/2)}{\Gamma(p/2+1)\beta_i M_1(\beta_i + p/2, p/2, -c/2)},$$

and $\Gamma(\cdot)$ is the gamma function.

Computationally, the only real challenge in this formula is the identification of the roots β_j. Once these values are known, the infinite sums in (6) and (7) converge quickly, and their computation to any desired degree of accuracy is straightforward. Let us therefore focus on the calculation of the roots β_j. The main problem is that there is an infinity of β_j, and there is no known closed-

https://doi.org/10.1017/S0266466603196107 Published online by Cambridge University Press

form expression for these values. However, given a relatively close first approximation $\hat{\beta}_j$, Newton's method may be used to compute β_j accurately. Thus, how can $\hat{\beta}_j$ be obtained?

DeLong (1981) provides an asymptotic approximation (as j approaches infinity) that is very accurate for large values of j, and fairly accurate even for small j, unless the number of parameters p is large. Thus,

$$\hat{\beta}_j = \frac{\pi^2}{2c}\left(j - \frac{3}{4} + \frac{p}{4}\right) - \frac{p}{4} + O(1). \tag{8}$$

Newton's method could be applied using this value as an initial guess ($\beta_j^{(0)} = \hat{\beta}_j$). If $\beta_j^{(n)}$ is the nth guess, the $n+1$ approximation is given by

$$\beta_j^{(n+1)} = \beta_j^{(n)} - \frac{M(\beta_j^{(n)} + p/2, p/2+1, -c/2)}{M_1(\beta_j^{(n)} + p/2, p/2, -c/2)}. \tag{9}$$

The derivative M_1 may be computed analytically or numerically, and $\beta_j^{(n)}$ generally converges to β_j quickly.

The problem with this method is that $\hat{\beta}_j$ is not always sufficiently accurate, particularly if the value of p is large. In such cases, Newton's method may converge to the wrong β_j. A more reliable method, though computationally somewhat more costly, is to perform a grid search to identify the approximate location of each β_j and then to apply Newton's method using each approximate location as an initial approximation. Because of its reliability, an optimally constructed grid search is the preferable method.

Thus, consider a grid for β based on an increment of size $\delta > 0$ and look successively at $M(n\delta + p/2, p/2, -c/2)$, $n = 0,1,\ldots$ until a change of sign is detected. If the sign changes for β between $n\delta$ and $(n+1)\delta$, Newton's method may be used locally to compute the zero.[1] The initial guess is based on a linear interpolation of the function M between $n\delta$ and $(n+1)\delta$. The value of δ must be chosen to be small enough so that the fineness of the $n\delta$ grid is able to detect all zeros. However, the smaller the δ, the longer the program takes to execute. We present a few results that suggest that $\delta = 1$ is optimal. Smaller values are more costly and offer no additional information, whereas larger values may not identify some of the roots.

PROPOSITION 1. *Let* $m_n = M(n+b, b, z)$ *for* $n = 0,1,\ldots$. *Then* $m_0 = \exp(z)$, $m_1 = ((b+z)/b)m_0$, *and* $m_{n+1} = (2 + (z-b)/(n+b))m_n + (b/(n+b) - 1)m_{n-1}$ *for* $n \geq 1$.

PROPOSITION 2. *Let* β_j *and* β_{j+1} *be two consecutive roots of* $M(\beta + p/2, p/2, -c/2)$ *as a function of* $\beta > 0$. *Then* $|\beta_{j+1} - \beta_j| = \beta_{j+1} - \beta_j > 1$.

PROPOSITION 3. $\dfrac{M(\beta_j + p/2, p/2+1, -c/2)}{M_1(\beta_j + p/2, p/2, -c/2)} < 0$, *so that* $\alpha_j \lambda^{-\beta_j} > 0$ *for all* j.

https://doi.org/10.1017/S0266466603196107 Published online by Cambridge University Press

Proposition 2 implies that we can take $\delta = 1$ with no danger of missing any of the roots β_j. Any interval one unit long contains at most one root, and this root is more than one unit away from the nearest roots. Moreover, to speed up the grid search, we may use the recursive formula in Proposition 1 and obviate the need to evaluate M at every point $n = 0,1,\ldots$. Once a change is detected, specifically when $m_{n+1}m_n \leq 0$, we can find a root in the interval $(m_n, m_{n+1}]$ using Newton's method. The initial guess is determined by linear interpolation in this interval.

This process may be used to continue identifying roots until the number of terms in the infinite sum in (7) is sufficient to obtain the desired level of accuracy. Proposition 3 implies that all the terms in (7) are positive and one may stop when the last term computed is smaller than some positive tolerance level.

3.2. Numerical Results

To check the accuracy of the computational method described previously, the method was used to replicate the relevant tables in DeLong (1981). DeLong's Table I (first four panels) contains probabilities that correspond to the application of (7) to 1,500 combinations of the parameters p, λ, and c. The method was applied with tolerance levels that provide accuracy to at least the four decimal places given by DeLong. The results matched DeLong's exactly in more than 98% of cases.

In cases where the results differ, DeLong's figure was always slightly lower, suggesting perhaps the inclusion of fewer terms from expression (7) (because all are positive by Proposition 3). Most differences were very minor, however, only 0.0001. A few larger differences were found only in cases in which $p = 1$ and $c = 1$. These differences were all less than 0.0042 and in most instances were much smaller. As a robustness check in light of these puzzling discrepancies, the method of Section 3.1 was applied with $p = 1$ and $c = 1 \pm \varepsilon$, for small values of ε. The same results as with $c = 1$ were obtained.

DeLong's Table II presents critical values c, to two decimal places, corresponding to various combinations (375 altogether) of the parameters p and λ. This table was replicated by inverting the numerical probability function of Section 3.1, using another application of Newton's method. The results of DeLong's Table II were matched exactly.

Finally, the method was used with the same tolerance level to recalculate Table 1 of critical values in Andrews (1993), which is widely cited as a reference in both the theoretical and applied literatures. Exact results are presented here in Table 1. Note that, following Andrews (1993), these results are tabulated according to π_0, the proportion of data points dropped at either end of the sample, rather than λ, as given in equation (4). Andrews calculates his figures by stochastic simulation of the Bessel process over the unit interval. These simulation results are subject to discretization bias and to sampling error. Discretization bias results because continuous Brownian motion is approximated using

https://doi.org/10.1017/S0266466603196107 Published online by Cambridge University Press

a large but finite number of normal distributions over small disjoint intervals. This bias is notoriously difficult to eliminate, even when using very small intervals.

Andrews (1993) splits the unit interval into 3,600 segments, and yet his critical values tend to be a bit low, most likely as a result of discretization. In most cases, however, the approximations are reasonably close. Exceptions are the values for $p = 8$, which are particularly low for no obvious reason. Andrews (2003) provides a revised table based on a larger number of simulation iterations. The entries for $p = 8$ are no longer outliers, and the sampling error is reduced. However, some discretization bias remains.

Discretization does not necessarily lead to less accurate results, because the timing of observations in a finite sample maps to a discrete set of points in the unit interval, rather than the infinity of points corresponding to the asymptotic theory. However, use of the exact asymptotic distribution seems preferable in most applications. One important reason is that simple discretization assumes that increments are independent and normally distributed, a result that derives from the application of central limit theorems (as in Andrews, 1993, Theorem 1) in the asymptotic case. Assuming independence and normality directly in finite samples is possible but is much more restrictive than the typical assumptions of the asymptotic GMM approach.

4. ASYMPTOTIC APPROXIMATION

4.1. Approximate Probability for Large c

Because DeLong's formula has been considered intractable, various researchers have proposed using an asymptotic approximation to the formula, which is accurate when c is large. In fact, DeLong (1981) himself derives such an approximation, though special cases had been known earlier. The approximation is based on an expansion of the probability from equation (4) in terms of powers of c. Specifically,

$$P\left(\sup_{1<s<\lambda} \|B_p(s)\| > c^{1/2}s^{1/2}\right) = \frac{1}{\Gamma(p/2)}\left(\frac{1}{2}c\right)^{p/2} e^{-1/2c}\left(\log(\lambda)\left(1-\frac{p}{c}\right)+\frac{2}{c}+O(c^{-2})\right). \quad \textbf{(10)}$$

The asymptotic approximation is obtained by ignoring the term $O(c^{-2})$ in (10), resulting in a closed-form expression as a function of the arguments p, λ, and c. Dirkse (1975) studies essentially the same distribution but with the Bessel process constructed from univariate Brownian motion ($p = 1$). He does not derive an expression for the exact probability but proceeds directly to the calculation of the asymptotic approximation, which matches (10) exactly (with $p = 1$).[2]

https://doi.org/10.1017/S0266466603196107 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466603196107 Published online by Cambridge University Press

TABLE 1. Exact asymptotic critical values: c_α such that $P(\sup_{\pi\in\Pi} Q_p(\pi) > c_\alpha) = \alpha$, $\alpha = 0.1, 0.05, 0.01$

		$p=1$			$p=2$			$p=3$			$p=4$			$p=5$		
π_0	λ	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
0.50	1.00	2.71	3.84	6.63	4.61	5.99	9.21	6.25	7.81	11.34	7.78	9.49	13.28	9.24	11.07	15.09
0.49	1.08	3.46	4.73	7.77	5.58	7.09	10.53	7.38	9.06	12.81	9.03	10.85	14.85	10.59	12.54	16.76
0.48	1.17	3.78	5.09	8.22	5.98	7.53	11.05	7.83	9.55	13.37	9.53	11.39	15.46	11.14	13.12	17.40
0.47	1.27	4.03	5.37	8.55	6.28	7.86	11.43	8.18	9.92	13.79	9.91	11.79	15.90	11.55	13.55	17.87
0.45	1.49	4.41	5.80	9.06	6.76	8.37	12.01	8.72	10.49	14.42	10.50	12.41	16.57	12.18	14.21	18.58
0.40	2.25	5.11	6.57	9.94	7.60	9.27	13.00	9.67	11.49	15.48	11.54	13.48	17.71	13.29	15.35	19.77
0.35	3.45	5.64	7.14	10.57	8.23	9.93	13.70	10.37	12.21	16.24	12.29	14.26	18.50	14.10	16.17	20.61
0.30	5.44	6.09	7.61	11.09	8.76	10.47	14.27	10.95	12.80	16.85	12.92	14.89	19.14	14.76	16.84	21.28
0.25	9.00	6.50	8.04	11.55	9.23	10.96	14.77	11.47	13.33	17.38	13.48	15.45	19.70	15.36	17.42	21.86
0.20	16.00	6.90	8.45	11.97	9.69	11.41	15.23	11.96	13.82	17.87	14.01	15.97	20.22	15.91	17.98	22.40
0.15	32.11	7.30	8.86	12.40	10.14	11.87	15.69	12.46	14.31	18.36	14.54	16.50	20.73	16.47	18.52	22.93
0.10	81.00	7.74	9.31	12.86	10.64	12.37	16.19	13.00	14.85	18.88	15.11	17.06	21.28	17.07	19.12	23.50
0.05	361.00	8.31	9.90	13.45	11.28	13.01	16.82	13.69	15.53	19.55	15.84	17.78	21.97	17.84	19.87	24.21

		$p=6$			$p=7$			$p=8$			$p=9$			$p=10$		
π_0	λ	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
0.50	1.00	10.64	12.59	16.81	12.02	14.07	18.48	13.36	15.51	20.09	14.68	16.92	21.67	15.99	18.31	23.21
0.49	1.08	12.10	14.15	18.58	13.56	15.71	20.32	14.98	17.23	22.01	16.38	18.72	23.66	17.75	20.18	25.27
0.48	1.17	12.68	14.77	19.25	14.17	16.36	21.03	15.63	17.91	22.74	17.05	19.42	24.42	18.46	20.91	26.05
0.47	1.27	13.12	15.22	19.74	14.63	16.84	21.54	16.11	18.41	23.28	17.56	19.94	24.97	18.98	21.45	26.62
0.45	1.49	13.79	15.92	20.48	15.34	17.57	22.31	16.85	19.17	24.08	18.33	20.74	25.80	19.78	22.27	27.47
0.40	2.25	14.96	17.12	21.73	16.57	18.83	23.60	18.14	20.48	25.41	19.67	22.09	27.17	21.17	23.67	28.89
0.35	3.45	15.81	17.98	22.60	17.47	19.72	24.50	19.07	21.41	26.34	20.64	23.06	28.13	22.18	24.67	29.87
0.30	5.44	16.52	18.68	23.29	18.20	20.45	25.22	19.84	22.17	27.09	21.43	23.84	28.89	22.99	25.47	30.66
0.25	9.00	17.14	19.30	23.90	18.85	21.09	25.85	20.51	22.83	27.73	22.13	24.52	29.55	23.71	26.18	31.33
0.20	16.00	17.72	19.87	24.45	19.46	21.69	26.42	21.14	23.45	28.32	22.78	25.16	30.16	24.39	26.83	31.96
0.15	32.11	18.30	20.44	25.00	20.07	22.28	26.99	21.77	24.06	28.90	23.43	25.79	30.76	25.06	27.48	32.57
0.10	81.00	18.94	21.06	25.59	20.72	22.92	27.59	22.45	24.72	29.53	24.14	26.47	31.40	25.78	28.18	33.22
0.05	361.00	19.74	21.84	26.33	21.55	23.73	28.36	23.31	25.55	30.31	25.02	27.32	32.20	26.69	29.05	34.04

https://doi.org/10.1017/S0266466603196107 Published online by Cambridge University Press

		$p = 11$			$p = 12$			$p = 13$			$p = 14$			$p = 15$		
π_0	λ	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
0.50	1.00	17.28	19.68	24.72	18.55	21.03	26.22	19.81	22.36	27.69	21.06	23.68	29.14	22.31	25.00	30.58
0.49	1.08	19.11	21.61	26.85	20.45	23.03	28.41	21.77	24.42	29.94	23.09	25.80	31.45	24.39	27.17	32.94
0.48	1.17	19.84	22.37	27.65	21.20	23.80	29.23	22.55	25.22	30.78	23.88	26.63	32.31	25.21	28.01	33.82
0.47	1.27	20.38	22.92	28.24	21.76	24.38	29.83	23.13	25.82	31.40	24.48	27.23	32.94	25.82	28.64	34.47
0.45	1.49	21.21	23.77	29.12	22.62	25.25	30.73	24.01	26.71	32.32	25.39	28.16	33.88	26.75	29.58	35.43
0.40	2.25	22.65	25.22	30.57	24.10	26.74	32.22	25.53	28.24	33.84	26.95	29.72	35.44	28.35	31.18	37.02
0.35	3.45	23.68	26.24	31.58	25.17	27.79	33.25	26.63	29.32	34.90	28.07	30.82	36.52	29.50	32.31	38.12
0.30	5.44	24.53	27.07	32.38	26.03	28.64	34.07	27.52	30.19	35.74	28.98	31.72	37.37	30.43	33.22	38.99
0.25	9.00	25.27	27.80	33.07	26.79	29.39	34.78	28.30	30.95	36.46	29.78	32.49	38.11	31.25	34.02	39.74
0.20	16.00	25.96	28.47	33.71	27.51	30.07	35.43	29.03	31.66	37.12	30.53	33.21	38.79	32.01	34.75	40.43
0.15	32.11	26.65	29.13	34.34	28.21	30.76	36.07	29.75	32.35	37.77	31.27	33.93	39.45	32.77	35.48	41.10
0.10	81.00	27.39	29.85	35.01	28.97	31.49	36.75	30.53	33.10	38.47	32.07	34.69	40.16	33.58	36.26	41.82
0.05	361.00	28.32	30.74	35.84	29.93	32.40	37.61	31.51	34.04	39.34	33.06	35.64	41.04	34.60	37.23	42.72

		$p = 16$			$p = 17$			$p = 18$			$p = 19$			$p = 20$		
π_0	λ	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
0.50	1.00	23.54	26.30	32.00	24.77	27.59	33.41	25.99	28.87	34.81	27.20	30.14	36.19	28.41	31.41	37.57
0.49	1.08	25.68	28.53	34.41	26.96	29.87	35.87	28.23	31.20	37.32	29.49	32.53	38.75	30.75	33.84	40.17
0.48	1.17	26.52	29.39	35.32	27.82	30.75	36.79	29.11	32.11	38.26	30.39	33.45	39.71	31.67	34.78	41.15
0.47	1.27	27.14	30.03	35.97	28.46	31.41	37.46	29.77	32.77	38.94	31.06	34.13	40.41	32.35	35.47	41.86
0.45	1.49	28.10	30.99	36.96	29.44	32.39	38.47	30.76	33.78	39.96	32.08	35.15	41.44	33.39	36.52	42.91
0.40	2.25	29.74	32.63	38.58	31.11	34.06	40.12	32.48	35.48	41.64	33.83	36.89	43.15	35.18	38.29	44.65
0.35	3.45	30.91	33.78	39.70	32.31	35.24	41.26	33.70	36.68	42.80	35.08	38.11	44.33	36.45	39.53	45.85
0.30	5.44	31.87	34.71	40.58	33.29	36.19	42.16	34.70	37.65	43.72	36.09	39.10	45.26	37.48	40.53	46.79
0.25	9.00	32.70	35.52	41.35	34.14	37.01	42.94	35.57	38.49	44.51	36.98	39.95	46.06	38.38	41.40	47.60
0.20	16.00	33.48	36.27	42.05	34.93	37.78	43.65	36.37	39.27	45.23	37.80	40.74	46.80	39.22	42.20	48.35
0.15	32.11	34.25	37.01	42.73	35.72	38.53	44.35	37.17	40.03	45.94	38.61	41.52	47.52	40.04	42.99	49.08
0.10	81.00	35.08	37.80	43.47	36.56	39.34	45.09	38.03	40.85	46.69	39.49	42.35	48.28	40.93	43.84	49.85
0.05	361.00	36.12	38.80	44.38	37.62	40.34	46.02	39.11	41.88	47.64	40.58	43.39	49.24	42.04	44.90	50.82

4.2. Accuracy of Asymptotic Approximation

To gauge the accuracy of the asymptotic approximation, we perform the following experiment. First, using the exact distribution for a given set of parameter values, we obtain the critical value corresponding to a significance level α, where α is 10, 5, or 1%. Then, we apply the approximation formula to the critical value. The more accurate the approximation, the closer the resulting approximate p value will be to α. The first part of Table 2 provides results for selected values of p ranging from 1 to 20 and for selected values of π_0 ranging from 0.05 to 0.45. These ranges are likely to be representative of values encountered in practice.

As the table indicates, results for $\pi_0 = 0.05$ and 0.15 are quite accurate for all three significance levels, particularly for $\alpha = 0.01$. This pattern is explained by the assumptions underlying the asymptotic approximation, which correspond to low values of both α and π_0. To see this, note that in DeLong's derivation, he makes the assumption that λ is large and, correspondingly, that π_0 is small. Moreover, the result is obtained by allowing the critical value c to be arbitrarily large, which corresponds to an arbitrarily low significance level α. When $\pi_0 = 0.25$, the results deteriorate somewhat, and the errors are more pronounced for larger π_0, particularly when both π_0 and p are large. For instance when $\pi_0 = 0.45$, $p = 20$, and the true significance level is 10%, the formula produces an approximate significance level of 5.75%.

5. HANSEN APPROXIMATION

Hansen (1997) introduces an alternative approach to approximating the p values of the supremum of the Bessel process. The approach involves very extensive preliminary computations, but once those results are obtained and stored in look-up tables, the approximation reduces to the computation of a chi square probability (with noninteger degrees of freedom).

In arriving at this method, Hansen (1997) observes that the shape of the distribution in question is roughly similar to that of the chi square. Thus, for a pair of parameter values (π_0, p), he fits a curve of the form

$$\hat{y}(c) = 1 - \chi^2_\nu(a + bc) \qquad \textbf{(11)}$$

by selecting values of the parameters (a, b, ν) so as to minimize a specific loss function. The "actual" y values are generated from simulations similar to those of Andrews (1993), except that 1,000 points are used in the unit interval, rather than 3,600 as in Andrews. Thus, the discretization bias will tend to be larger. Hansen's loss function is a weighted sum of squared deviations, with larger weights attached to significance levels up to 0.10, which are generally the focus of empirical research.

https://doi.org/10.1017/S0266466603196107 Published online by Cambridge University Press

TABLE 2. Approximate p values when the true significance level is α

π_0	$p = 1$	2	5	10	15	20
Asymptotic approximation formula: $\alpha = 0.10$						
0.05	0.0976	0.1006	0.1039	0.1061	0.1073	0.1080
0.15	0.0917	0.0950	0.0985	0.1008	0.1020	0.1028
0.25	0.0853	0.0884	0.0917	0.0938	0.0948	0.0955
0.35	0.0775	0.0792	0.0811	0.0822	0.0828	0.0832
0.45	0.0705	0.0667	0.0625	0.0598	0.0584	0.0575
Asymptotic approximation formula: $\alpha = 0.05$						
0.05	0.0489	0.0500	0.0512	0.0520	0.0525	0.0528
0.15	0.0467	0.0479	0.0493	0.0502	0.0506	0.0510
0.25	0.0441	0.0453	0.0466	0.0475	0.0479	0.0482
0.35	0.0404	0.0413	0.0422	0.0428	0.0431	0.0433
0.45	0.0358	0.0347	0.0333	0.0324	0.0319	0.0315
Asymptotic approximation formula: $\alpha = 0.01$						
0.05	0.0098	0.0099	0.0101	0.0102	0.0102	0.0103
0.15	0.0096	0.0097	0.0098	0.0100	0.0100	0.0101
0.25	0.0092	0.0093	0.0095	0.0096	0.0097	0.0097
0.35	0.0086	0.0087	0.0089	0.0090	0.0090	0.0091
0.45	0.0075	0.0074	0.0073	0.0072	0.0072	0.0071
Hansen (1997) approximation method: $\alpha = 0.10$						
0.05	0.0887	0.0873	0.0856	0.0842	0.0874	0.0868
0.15	0.0905	0.0895	0.0884	0.0873	0.0896	0.0909
0.25	0.0914	0.0916	0.0890	0.0890	0.0904	0.0914
0.35	0.0909	0.0929	0.0879	0.0880	0.0904	0.0908
0.45	0.0912	0.0919	0.0875	0.0893	0.0921	0.0912
Hansen (1997) approximation method: $\alpha = 0.05$						
0.05	0.0435	0.0426	0.0415	0.0409	0.0427	0.0429
0.15	0.0445	0.0439	0.0435	0.0427	0.0441	0.0454
0.25	0.0450	0.0452	0.0440	0.0439	0.0445	0.0454
0.35	0.0447	0.0464	0.0431	0.0428	0.0442	0.0449
0.45	0.0448	0.0453	0.0426	0.0438	0.0455	0.0452
Hansen (1997) approximation method: $\alpha = 0.01$						
0.05	0.0083	0.0080	0.0076	0.0075	0.0080	0.0083
0.15	0.0086	0.0084	0.0084	0.0081	0.0084	0.0091
0.25	0.0088	0.0088	0.0086	0.0086	0.0086	0.0090
0.35	0.0087	0.0093	0.0083	0.0080	0.0084	0.0088
0.45	0.0087	0.0088	0.0081	0.0084	0.0089	0.0089

https://doi.org/10.1017/S0266466603196107 Published online by Cambridge University Press

The method used in the previous section to gauge the accuracy of the asymptotic approximation is applied to the Hansen approximation in the second part of Table 2. In contrast to the results for the asymptotic approximation, the errors in this panel are very consistent across values of the parameters. The errors are not as large as some of those obtained with the asymptotic approximation, but the approximate p values are all lower than the true values, by about 10 to 20% of the p value. Thus, the approximation uniformly overstates significance. It is likely that this pattern is related to the discretization bias in the simulations, which, as noted earlier, can be surprisingly large even with a seemingly large number of points.

6. CONCLUSIONS

This paper has demonstrated that exact asymptotic p values for supremum-based break test statistics are not as intractable as they might appear from the earlier literature. A quick simple method is outlined, which may be executed with existing statistical packages. The method is used to provide a table of exact critical values corresponding to the table in Andrews (1993).

With the availability of this method, the various approximations found in the literature lose some of their appeal. Nevertheless, it is important to examine the accuracy of these approximations to the extent that they have been used in the empirical literature and may have possibly led to misleading conclusions regarding significance of results. Thus, the paper analyzes the accuracy of the approximations and reaches several general conclusions.

First, simulation-based approximations are subject to noticeable discretization bias, even when a large number of points is used. However, for the most part, the simulated critical values provided in Andrews (1993, 2003) are fairly accurate.[3]

Second, the asymptotic approximation is quite accurate for p values up to 10% when $\pi_0 \leq 0.15$. When the proportion of points not tested at the ends is larger than 0.15, the accuracy of the asymptotic approximation can be substantially lower.

Finally, conclusions based on the Hansen (1997) approximation are not likely to be overturned if they are very strong or very weak, because the approximation error is not large and is consistent in magnitude across parameter values. However, results that are reported to be marginally significant may be misclassified by the approximation.

NOTES

1. This approach was implemented with a small arbitrary value of δ in Estrella and Fuhrer (2003).

https://doi.org/10.1017/S0266466603196107 Published online by Cambridge University Press

2. Several papers have proposed approximations that are formally the same as (10) but with $4/c$ replacing $2/c$ within the last brackets. These include Miller and Siegmund (1982), Siegmund (1985, 1986), James et al. (1987), and Bai (1999). Siegmund (1985, equation (11.1)) and (1986, equation (2.29)) also gives the formula with $2/c$, although at least in the 1985 paper the preferred value seems to be $4/c$. A straightforward way to determine the appropriate value of this constant is to evaluate the formula using parameter values that lead to a simplified expression. For example, if $p = 1$ and $\pi_0 = \frac{1}{2}$, the exact distribution is χ_1^2. The asymptotic approximation with $2/c$ (as in (10)) reduces to $A(2) = 2\phi(c^{1/2})/c^{1/2} + o(\phi(c^{1/2})/c^{1/2})$, and it may be shown that $\lim_{c\to\infty} A(2)/(1 - \chi_1^2(c)) = 1$, as expected. If $4/c$ is used, $\lim_{c\to\infty} A(4)/(1 - \chi_1^2(c)) = 2$. Thus, $2/c$ is correct, which confirms the results of DeLong (1981) and Dirkse (1975) and the alternative results of Siegmund (1985, 1986).

3. The only caveat is that caution must be used when the number of parameters is 8 in the 1993 table, because these values contain some large errors.

REFERENCES

Abramowitz, M. & I.A. Stegun (1964) *Handbook of Mathematical Functions*. Washington, D.C.: U.S. Department of Commerce.

Akman, V.E. & A.E. Raftery (1986) Asymptotic inference for a change-point Poisson process. *Annals of Statistics* **14**, 1583–1590.

Anderson, T.W. & D.A. Darling (1952) Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *Annals of Mathematical Statistics* **23**, 193–212.

Andrews, D.W.K. (1993) Tests for parameter instability and structural change with unknown change point. *Econometrica* **61**, 821–856.

Andrews, D.W.K. (2003) Tests for parameter instability and structural change with unknown change point: Corrigendum. *Econometrica* **71**, 395–397.

Andrews, D.W.K. & R.C. Fair (1988) Inference in nonlinear econometric models with structural change. *Review of Economic Studies* **55**, 615–640.

Andrews, D.W.K. & W. Ploberger (1984) Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* **62**, 1383–1414.

Bai, J. (1999) Likelihood ratio tests for multiple structural changes. *Journal of Econometrics* **91**, 299–323.

Bai, J. & P. Perron (1998) Estimating and testing linear models with multiple structural changes. *Econometrica* **66**, 47–78.

DeLong, D.M. (1981) Crossing probabilities for a square root boundary by a Bessel process. *Communications in Statistics—Theory and Methods* **A10**, 2197–2213.

Dirkse, J.P. (1975) An absorption probability for the Ornstein–Uhlenbeck process. *Journal of Applied Probability* **12**, 595–599.

Estrella, A. & J.C. Fuhrer (2003) Monetary policy shifts and the stability of monetary policy models. *Review of Economics and Statistics* **85**, 94–104.

Ghysels, E., A. Guay, & A. Hall (1997) Predictive tests for structural change with unknown breakpoint. *Journal of Econometrics* **82**, 209–233.

Hansen, B.E. (1997) Approximate asymptotic *P* values for structural-change tests. *Journal of Business and Economic Statistics* **15**, 60–67.

James, B., K.L. James, & D. Siegmund (1987) Tests for a change-point. *Biometrika* **74**, 71–83.

Miller, R. & D. Siegmund (1982) Maximally selected chi square statistics. *Biometrics* **38**, 1011–1016.

Siegmund, D. (1985) *Sequential Analysis: Tests and Confidence Intervals*. New York: Springer-Verlag.

Siegmund, D. (1986) Boundary crossing probabilities and statistical applications. *Annals of Statistics* **14**, 361–404.

https://doi.org/10.1017/S0266466603196107 Published online by Cambridge University Press

APPENDIX: PROOFS OF PROPOSITIONS

Proof of Proposition 1. Abramowitz and Stegun (1964) provide the following results:

$$M(b,b,z) = e^z \text{ (equation 13.6.12)}, \tag{A.1}$$

$$M(b+1,b,z) = e^z M(-1,b,-z) = \frac{b+z}{b}\, e^z \text{ (equation 13.1.27)}, \tag{A.2}$$

$$(b-a)M(a-1,b,z) + (2a-b+z)M(a,b,z) - aM(a+1,b,z) = 0 \text{ (equation 13.4.1)}. \tag{A.3}$$

The second equality in (A.2) follows by direct substitution of the argument values into (6), because $(-1)_j = 0$ for $j > 1$. Expressions for m_0 and m_1 follow immediately from equations (A.1) and (A.2), respectively. Substitution of $a = n + b$ in (A.3) leads to the difference equation in Proposition 1, after solving for m_{n+1}. ■

Proof of Proposition 2. Substitute $b = p/2$ and $z = -c/2$ in the difference equation from Proposition 1, which becomes

$$m_{n+1} = \left(2 - \frac{c+p}{2n+p}\right) m_n - \frac{2n}{2n+p}\, m_{n-1}. \tag{A.4}$$

The strategy is to show that the oscillatory frequency of m_n (hence the frequency of its zeros) increases as a function of c. In the limit as $c \to \infty$, the length of a cycle approaches 2, independently of the value of n, and the zeros are equally spaced at the integers. Thus, for finite c, the difference between any two zeros is greater than 1.

The characteristic equation corresponding to (A.4) is

$$\lambda^2 - \left(2 - \frac{c+p}{2n+p}\right)\lambda + \frac{2n}{2n+p} = 0, \tag{A.5}$$

which has roots given by

$$\lambda_n = \frac{4n - c + p \pm \sqrt{(c-p)^2 - 8nc}}{4n+2p}. \tag{A.6}$$

The local oscillatory frequency of the difference equation is given by $\theta_n = \arg(\lambda_n)$, where arg indicates the argument of the root as a complex number. The subscripts of λ_n and θ_n signify that these measures vary as a function of n, because the coefficients in (A.5) vary with n. If the coefficients were constant, λ_n and θ_n would also be constant, and the solution could be expressed as a simple exponential function of the λ_n.

Now, λ_n is real or complex according to the sign of the expression within the radical in (A.6). Consider the root corresponding to the positive sign in (A.6). The analysis of the other root is analogous. If we focus on the value of c, we have to treat three separate cases. Let

$$c_1 = 4n + p - 2\sqrt{4n^2 + 2np} \tag{A.7}$$

https://doi.org/10.1017/S0266466603196107 Published online by Cambridge University Press

and

$$c_2 = 4n + p + 2\sqrt{4n^2 + 2np}. \tag{A.8}$$

Then, if $0 < c \leq c_1$, λ_n is real and positive; hence $\theta_n = 0$. If $c_1 < c < c_2$, λ_n is complex with positive imaginary part, and its real part is a decreasing function of c (see (A.6)). Thus, θ_n is an increasing function of c. Finally, if $c \geq c_2$, λ_n is real and negative; hence $\theta_n = \pi$. Note that θ_n is continuous at c_1 and c_2 and that these results hold for all n. Thus, for any given n, θ_n is a continuous nondecreasing function of c, which is strictly increasing in the oscillatory cases corresponding to $c_1 < c < c_2$. The distance between two zeros in the oscillatory case corresponds to one-half of the wavelength, π/θ_n, which is clearly decreasing in c.

To complete the proof, we cite DeLong's (1981) result that if β_j is the jth root of $M(\beta + p/2, p/2, -c/2)$ as a function of β, then $\beta_j \to j - 1$ as $c \to \infty$ and hence $\beta_{j+1} - \beta_j \to 1$. Note that this result is consistent with the preceding oscillation analysis, because $c \to \infty$ implies that $\theta_n = \pi$ and that the half-wavelength, which corresponds to the distance between consecutive zeros, is such that $\pi/\theta_n \to 1$. ■

Proof of Proposition 3. Let β_j be the jth zero of $M(\beta + p/2, p/2, -c/2)$ as a function of β. Then by equation 13.4.5 of Abramowitz and Stegun (1964),

$$M(\beta_j + p/2, p/2 + 1, -c/2) = \frac{(\beta_j + p/2)p/2}{(-c/2)\beta_j} M(\beta_j + 1 + p/2, p/2, -c/2). \tag{A.9}$$

Because Proposition 2 implies that there are no zeros between β_j and $\beta_j + 1$, we also have that

$$\text{sign}\{M(\beta_j + 1 + p/2, p/2, -c/2)\} = \text{sign}\{M_1(\beta_j + p/2, p/2, -c/2)\}. \tag{A.10}$$

The result follows immediately. ■

https://doi.org/10.1017/S0266466603196107 Published online by Cambridge University Press