ISSN 1440-771X

MONASH University

Australia

Department of Econometrics
and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

Weighted Average Power Similar Tests for Structural Change for the Gaussian Linear Regression Model

G. Forchini

August 2005

Working Paper 20/05

Weighted Average Power Similar Tests for Structural Change for the Gaussian Linear Regression Model

G. Forchini *†

Monash University

August 2005

Abstract

The average exponential tests for structural change of Andrews and Ploberger (*Econometrica*, 62, 1994) and Andrews, Lee and Ploberger (*Journal of Econometrics* 70, 1996) and modifications thereof maximize a weighted average power which incorporates specific weighting functions in order to make the resulting test statistics simple. Generalizations of these tests involve the numerical evaluation of (potentially) complicated integrals. In this paper we suggest a uniform Laplace approximation to evaluate weighted average power test statistics for which a simple closed form does not exist. We also show that a modification of the avg-F test is optimal under a very large class of weighting functions and can be written as a ratio of quadratic forms. Finally, we discuss how the computational burden of averaging over all possible change-points can be addressed.

Key Words: Linear Regression Model, Similar Tests, Invariant Tests, Structural Change, Weighted Average Power Tests, Laplace Approximation, Uniform Laplace Approximation

JEL Classification: C12, C21, C22

*Address for correspondence: Giovanni Forchini, Department of Econometrics and Business Statistics, Monash University, Clayton, Victoria 3800, Australia. E-mail address: Giovanni.Forchini@BusEco.monash.edu.au

†I thank Peter Phillips, Patrick Marsh and and Simone Grose for helpful comments and discussions.

1 Introduction

Andrews and Ploberger (1994) and Andrews, Lee, and Ploberger (1996) suggest finite sample similar tests for structural change at unknown change-points in the Gaussian linear regression model which maximize a weighted average power (WAP). They obtain a class of optimal tests for the case where the disturbance variance is known. For the case where the error variance is unknown, they propose replacing the unknown variance by an estimate, and show that the resulting tests are still similar. Andrews and Ploberger (1994) also prove that these tests are asymptotically optimal.

Forchini (2002) extends the results of Andrews and Ploberger (1994) and Andrews, Lee, and Ploberger (1996), and derives similar WAP tests for structural change at unknown change-points which allow for an unknown variance. These tests are optimal for any sample size and are equivalent to those of Andrews and Ploberger (1994) and Andrews, Lee, and Ploberger (1996) in large samples.

Unfortunately, existing WAP tests for structural change at unknown change-points have two drawbacks. (i) Firstly, they incorporate specific choices of weighting functions which have been selected in such a way that the resulting test statistics have relatively simple functional forms. The use of different weighting functions to accommodate the relative importance of different departures from the null hypothesis is not viable because of the need to evaluate complicated integrals numerically. (ii) Secondly, these tests require the evaluation of several F-tests (or other equivalent tests) for all possible change-points. Since WAP tests have non-standard distributions, calculating their critical values may be very computer intensive especially when the sample size is large.

This paper contributes to the literature by investigating the construction of WAP tests for general weighting functions. Firstly, we find that the WAP test for local departures from the null denoted by LR_0 by Forchini (2002) is optimal for a large class of weighting functions, and can be written as a ratio of quadratic forms in the vector of residuals calculated under the null hypothesis. These properties make the test very attractive in practical applications. The avg-F test of Andrews and Ploberger (1994) and Andrews, Lee, and Ploberger (1996) is also optimal in large samples for a larger class of weighting functions than the one originally used in its derivation, however, its computation is more involved than that of the LR_0 test statistic.

Secondly, we study WAP tests for general weighting functions for alternatives hypotheses which are not necessarily local to the null hypothesis. We find that the use of (uniform) Laplace approximations (e.g. Bleistein and Handelsman (1986)) provides easily computable expressions for WAP test statistics. These approximations can be easily implemented, and there is plenty of evidence in the literature that they are very accurate. We briefly discuss the problem of averaging over all possible change-points, and suggest ways of reducing the computational burden that it involves.

The rest of the paper is organized as follows. Section 2 presents the model, the notation, and reviews existing results about WAP tests. Section 3 gives the main results. All proofs are in the Appendix. Section 4 briefly discusses the problem of averaging over all possible-change points, and Section 5 concludes.

2 The model and WAP Tests for Structural Change

We consider a Gaussian linear regression model with $t+1$ sub-samples, containing respectively τ_1, τ_2, ..., τ_{t+1} ($\sum_{i=1}^{t+1} \tau_i = T$) observations:

$$y = X\beta + Z(\tau)\gamma + u \tag{1}$$

where y is a $T \times 1$ vector of dependent variables, $X = (Z, W)$ is a $T \times p$ matrix of independent and fixed regressors. The sub-matrix Z is partitioned as $Z = \left(Z'_{\tau_1}, Z'_{\tau_2}, ..., Z'_{\tau_t}, Z'_{\tau_{t+1}}\right)'$ where block Z_{τ_i} contains τ_i observations ($i = 1, 2, ..., t+1$) on k variables, and the $T \times K$ matrix $Z(\tau)$ is obtained by deleting the first k columns, and all the columns which can be obtained as a linear combination of the remaining ones (to keep the notation as simple as possible we do not index K by τ) in the block diagonal matrix diag $\left(Z_{\tau_1}, Z_{\tau_2}, ..., Z_{\tau_t}, Z_{\tau_{t+1}}\right)$. Using this notation we identify change-points by an index τ which represents a partition of T into $t+1$ integer parts, $\tau = (\tau_1, \tau_2, ..., \tau_{t+1})$, $\tau_i > 0$ for all i, $\sum_{i=1}^{t+1} \tau_i = T$. The subset of all partitions of T of interest (i.e. the set of all possible change-points in the model) is denoted by Υ. For further discussion of this notation see Forchini (2002).

The following assumptions are supposed to hold:

Assumptions:

(1) $u \sim N\left(0, \sigma^2 I_T\right)$

(2) $T - p - K \geq 0$

(3) X and $Z(\tau)$ are fixed for all $\tau \in \Upsilon$

(4) $Z(\tau)' M_X Z(\tau) \gamma / (T-p) = Q_\tau + o(1)$ for all $\tau \in \Upsilon$, where Q_τ is a finite positive definite matrix, and $M_X = I_T - X(X'X)^{-1} X'$

(5) $K = O(T-p)$

Assumptions (1), (2), (3) and (4) are standard in this literature. Assumption (5) reflects the fact that the number of possible change-points may increase as the sample size increases.

By writing the model as in equation (1) one can easily show that both the class of tests invariant under the transformations $y \to ay + X\vartheta$ (with $a > 0$, $\vartheta \in \mathbb{R}^p$) and the class of similar tests for $H_0 : y \sim N\left(X\beta, \sigma^2 I_T\right)$ against any alternative whatever are characterized by the vector $v = C'y/\left(y' M_X y\right)^{1/2}$, where C is a $T \times T - p$ matrix such that $CC' = M_X$, $C'C = I_{T-p}$ and $C'X = 0$ (cf. King and Hillier (1985) and Hillier (1987)).

The power of the critical region ω is (e.g. equation (A.3) of Forchini (2002))

$$\begin{aligned} P_\omega &= \frac{1}{2\pi^{(T-p)/2}} \exp\left\{-\frac{1}{2}(T-p)\lambda_\tau\right\} \\ &\quad \int_\omega \sum_{j=0}^{\infty} \frac{\Gamma\left(\frac{T-p+j}{2}\right) 2^{j/2}}{j!} \left((T-p)^{1/2} \lambda_\tau^{1/2} \phi_\tau' \Lambda_\tau' v\right)^j (dv), \end{aligned}$$

where

$$\Lambda_\tau = C' Z(\tau) \left(Z(\tau)' M_X Z(\tau)\right)^{-1/2}$$
$$\phi_\tau = (T-p)^{-1/2} \left(Z(\tau)' M_X Z(\tau)\right)^{1/2} (\gamma/\sigma) / \lambda_\tau^{1/2}$$
$$\lambda_\tau = \phi_\tau' \phi_\tau = \gamma' Z(\tau)' M_X Z(\tau) \gamma / \left[(T-p)\sigma^2\right].$$

No uniformly most powerful test exists in this set-up, so one usually considers WAP tests (e.g. Wald (1943) and Cox and Hinkley (1974)). Andrews and Ploberger (1994) and Andrews, Lee, and Ploberger (1996) suggest averaging over the partitions $\tau \in \Upsilon$ with weights $p(\tau)$, and over all values of $(\beta', \gamma')'$ with a weighting function proportional to the density of a normal distribution. They show that if the error variance σ^2 is known, a WAP test has the form

$$\text{exp-F}_c = \sum_{\tau \in \Upsilon} \frac{p(\tau) \exp\left\{\frac{cK_\tau f_\tau}{2(1+c)}\right\}}{(1+c)^{K/2}} \tag{2}$$

where f_τ is the F test statistic for testing the null hypothesis $H_0 : \gamma = 0$ against the alternative $H_1 : \gamma \neq 0$ for a fixed change-point τ.

Forchini (2002) extends the results of Andrews and Ploberger (1994) and Andrews, Lee, and Ploberger (1996) by deriving a WAP test for structural change for the case where σ^2 is unknown. This is done by averaging the power over the partitions $\tau \in \Upsilon$ with weights $p(\tau)$ (as suggested by Andrews and Ploberger (1994) and Andrews, Lee, and Ploberger (1996)) and over all possible directions of $C' Z(\tau) \gamma/\sigma$ with uniform weight (as advised by Wald (1943) and Hillier (1987)). However, since all this is not enough to obtain uniformly most powerful tests in terms of WAP, a further averaging over $\lambda_\tau > 0$

with weight $g(\lambda_\tau)$ is needed. In this case the WAP of the critical region ω is

$$\bar{P}_\omega = \frac{2\Gamma(b)}{\Gamma(q)\Gamma(b-q)} \sum_{\tau\in\Upsilon} p(\tau) \int_{\lambda_\tau>0} \int_\omega (\cos\theta_\tau)^{K-1} \\ (\sin\theta_\tau)^{T-p-K-1} \exp\{-bh(\lambda_\tau;\theta_\tau)\} g(\lambda_\tau) d\theta_\tau d\lambda_\tau,$$

where

$$h(\lambda;\theta) = \lambda - b^{-1} \ln\left\{{}_1F_1\left(b;q;b\lambda\cos^2\theta\right)\right\} \tag{3}$$

$\cos^2\theta_\tau = [q/(b-q)] f_\tau (1 + [q/(b-q)] f_\tau)^{-1}$, and $b = (T-p)/2$, $q = K/2$. Here and in the rest of the paper we make use of the standard notation for hypergeometric functions (e.g. Slater (1960)).

The critical region which maximizes WAP has the form

$$S_{g,p} = \sum_{\tau\in\Upsilon} p(\tau) \Im(\theta_\tau) > k_\alpha \tag{4}$$

for a suitable constant k_α, such that the size of the test is α, where

$$\Im(\theta) = \int_{\lambda>0} \exp\{-bh(\lambda;\theta)\} g(\lambda) d\lambda. \tag{5}$$

A closed form for the WAP test can be obtained by choosing $g(\lambda_\tau)$ proportional to a certain power of λ_τ. For example, if one chooses $g(\lambda_\tau)$ in such a way that $(T-p)\lambda_\tau/\sqrt{c} \sim \chi^2_K$, the resulting test statistic is

$$\mathrm{LR}_c = \sum_{\tau\in\Upsilon} \frac{p(\tau)}{(1+c)^{K/2}\left(1 + \frac{c}{1+c}\cos^2\theta_\tau\right)^{(T-p)/2}}. \tag{6}$$

Forchini (2002) (Corollary 1) shows that (2) and (6) are approximately the same as T increases for fixed p. The statistic LR_c seems cumbersome because it depends on $\cos^2\theta_\tau$ which does not seem to have an econometric interpretation. However, the following result holds.

Proposition 1 *The quantity* $\cos^2\theta_\tau$ *is the coefficient of determination of the OLS regression of* $M_X y$ *on* $M_X Z(\tau)$.

For an arbitrary weighting function g, the integral (5) over λ_τ in (4) cannot be evaluated explicitly. In the next section we will generalize the WAP tests to cover such situations.

3 Main results

Our first result deals with a WAP test statistic for local departures from the null hypothesis, obtained as $\mathrm{LR}_0 = b^{-1} \lim_{c\to 0} \left[\left(\mathrm{LR}_c - (1+c)^{-q}\right)/c\right]$. Theorem 1 shows that LR_0 has the same functional form for a large class of weighting functions.

Theorem 1 *Let $f(\lambda)$ to be a piecewise continuous function such that $\int_{-\infty}^{\infty} |f(\lambda)| \, d\lambda < \infty$ and $\int_{-\infty}^{\infty} f(\lambda) \, dx = 1$, and define $g_a(\lambda) = a^{-1} f\left(a^{-2}\lambda^2\right)$ then the WAP test statistic $S_{\pi_0,p} = \lim_{a\to 0} S_{\pi_a,p}$ is equal, after a suitable normalization, to*

$$LR_0 = b^{-1} S_{\pi_0,p} + 1 = \sum_{\tau\in\Upsilon} p(\tau) \cos^2\theta_\tau.$$

Moreover,

$$LR_0 = \frac{\hat{u}' A_\Upsilon \hat{u}}{\hat{u}'\hat{u}}$$

where $\hat{u} = M_X y$ is the vector of residuals of the OLS regression of y on X, and

$$A_\Upsilon = \sum_{\tau\in\Upsilon} p(\tau) Z(\tau) \left(Z(\tau)' M_X Z(\tau)\right)^{-1} Z'(\tau).$$

Therefore for all weighting functions $g_a(\lambda)$ satisfying the conditions of Theorem 1, the WAP test for local departure is an average of the coefficients of determination of the auxiliary OLS regressions of $M_X y$ on $M_X Z(\tau)$, $\tau \in \Upsilon$. Moreover, in order to calculate the LR_0, one just needs to run one OLS regression (of y on X) and to evaluate a quadratic form, since the $T \times T$ matrix A_Υ must be computed once only. This is a very appealing property because it is a WAP test for which the computation burden is low. One may notice that the calculation of the critical values for the LR_0 can be efficiently done numerically using Imhof (1961)'s procedure.

For large T we have that

Corollary 1 *The avg-F test based on the statistic avg-F=* $\sum_{\tau\in\Upsilon} p(\tau) f_\tau$ *is optimal in large samples for the class of weighting functions* $g_a(\lambda)$ *specified in Theorem 1.*

Note that the avg-F test cannot be written as ratio of quadratic forms in $\hat{u}$, because the denominator of the F-test statistic for fixed τ is the estimate of σ^2 based on the unrestricted model (1).

Apart from this special case, the optimal test depends on the specific weighting function $g(\lambda)$. If such function is more complicated than a mixture of polynomials and simple exponentials, $\Im(\theta)$ in (5) does not have a closed form. Therefore, it is reasonable to approximate the integral $\Im(\theta)$, given its structure, using a Laplace expansion. It can be easily checked that $h'(\lambda;\theta) = 0$ is equivalent to

$$\frac{{}_1F_1\left(b;q;b\lambda\cos^2\theta\right)}{{}_1F_1\left(b+1;q+1;b\lambda\cos^2\theta\right)} = \frac{b}{q}\cos^2\theta. \tag{7}$$

The expression on the left-hand-side is a strictly increasing function of λ and has a minimum at $\lambda = 0$. So the minimum of $h(\lambda;\theta)$, λ_0, occurs on the boundary ($\lambda_0 = 0$) if $\cos^2\theta \leq q/b$, and at an interior point ($\lambda_0 > 0$) if $\cos^2\theta > q/b$. Thus, one has to consider three cases:

1. if $\cos^2\theta < q/b$, then

$$\Im(\theta) \sim \Im_1(\theta) = \frac{g(0)}{bh'(0;\theta)} = \frac{g(0)}{b(1-(b/q)\cos^2\theta)} \tag{8}$$

   since $h(0;\theta) = 0$, $h'(0;\theta) = 1-(b/k)\cos^2\theta$ (e.g. Section 4.3 of De Bruijn (1961));

2. if $\cos^2\theta > q/b$, then a standard Laplace expansion (e.g. De Bruijn (1961)) gives

$$\Im(\theta) \sim \Im_2(\theta) = \frac{(2\pi)^{1/2}\exp\{-bh(\lambda_0;\theta)\}\, g(\lambda_0)}{[h''(\lambda_0;\theta)\, b]^{1/2}} \tag{9}$$

   where λ_0 solves (7); and,

3. if $\cos^2\theta = q/b$ then

$$\Im(\theta) \sim \Im_2(\theta)/2. \tag{10}$$

The expansions above are not uniform in θ, and (10) cannot be obtained as a limiting case of (8) or (9) as $\cos^2\theta \to q/b$. As as consequence, these approximations to $\Im(\theta)$ can be extremely poor when $\cos^2\theta$ is nearly equal to q/b. Thus, we need to find an asymptotic expansion which holds uniformly with respect to θ.

Theorem 2 *Let ν_θ be 1 if $\cos^2\theta < q/b$ and -1 otherwise, λ_0 be the minimum of $h(\lambda;\theta)$ in the region where ${}_1F_1\left(b;q;b\lambda\cos^2\theta\right)$ is positive, and $\Phi(x)$ denote the cumulative distribution function of a standard normal distribution. Suppose that $g(\lambda)$ has no singularity in $[0,+\infty)$. Then, for large b,*

$$\Im(\theta) \sim \Im_A(\theta) = \Im_2(\theta)\left(1-\Phi\left(\nu_\theta\sqrt{-2bh(\lambda_0;\theta)}\right)\right) + \Im_1(\theta) - \frac{\nu_\theta g(\lambda_0)}{b\sqrt{-2h(\lambda_0;\theta)\,h''(\lambda_0;\theta)}}, \tag{11}$$

uniformly in θ, where $\Im_1(\theta)$ and $\Im_2(\theta)$ are defined in equations (8) and (9) respectively.

In order to achieve uniformity, the asymptotic expansion of $\Im(\theta)$ in Theorem 2 is slightly more complicated than the standard ones presented earlier on in equations (8), (9) and (10). It is a weighted average of $\Im_1(\theta)$ and $\Im_2(\theta)$ plus a correction term. Since it requires the evaluation of $h(\lambda;\theta)$ and $h''(\lambda;\theta)$ at the saddlepoint λ_0 (even though λ_0 may not be in $[0,+\infty)$) and of $h'(0;\theta)$ only, it can be easily computed. The restriction that $g(\lambda)$ does not have singularities can be relaxed by using the techniques of Chapter 9 of Bleistein and Handelsman (1986).

In order to implement the approximate WAP test using equation (11) we need to calculate numerically the saddlepoint λ_0. The following result gives an asymptotic expansion for λ_0 which can be inserted directly in (11) or can be used to obtain a starting point for a numerical calculation of λ_0.

Theorem 3 *Let $a = b/q - 1 = O(1)$, then, for large b, the saddlepoint for $h(\lambda;\theta)$ in (3) is approximately*

$$\lambda_0 \sim \tilde{\lambda}_0 = -\frac{1-a-(1+a)\cos(2\theta)}{2(1+a)\sin^2\theta}.$$

We will see in Section 3.1 that the approximation is good when $\cos^2\theta \geq q/b$, but it may be poor when $\cos^2\theta < q/b$.

Finally, one may note that the test statistic $\sum_{\tau\in\Upsilon} p(\tau)\,\Im_A(\theta_\tau)$ is a complicated function of $\cos^2\theta_\tau$ and in general has a non-standard asymptotic distribution. However, since under the null hypothesis its distribution is free of nuisance parameters, the techniques of Monte Carlo tests can be used to calculate p-values efficiently (see for instance Dufour and Khalaf (2001)).

3.1 Numerical Results

We now present some numerical results aiming at evaluating the performance of the approximations suggested in Theorems 2 and 3. We start with Theorem 3 since the approximation depends only on $h(\lambda;\theta)$.

Table 1 gives examples of exact (i.e. numerical), λ_0^N, and approximate, $\tilde{\lambda}_0$, solutions to equation (7) for various values of b, q and $\cos^2\theta$. It shows that the approximation is fairly accurate (even if q and b are as small as 1 and 10 respectively) when λ_0 is positive, but can be poor for negative values of λ_0.

[TABLE 1 APPROXIMATELY HERE]

We now gives some numerical evidence concerning the approximation in Theorem 2. Table 2 gives the exact and approximate values of $\Im(\theta)$ when $g(\lambda) = 1$ and $g(\lambda) = \sqrt{2/\pi}\exp\left\{-\lambda^2/2\right\}$ for $b = 19$ and $q = 2$. Notice that for $g(\lambda) = 1$ the integral $\Im(\theta)$ can be evaluated exactly as

$$\Im(\theta) = \int_{\lambda>0} \exp\{-bh(\lambda;\theta)\}\,d\lambda = b^{-1}\,{}_2F_1\left(1,b;q;\cos^2\theta\right) \tag{12}$$

where ${}_2F_1\left(1,b;q;\cos^2\theta\right)$ denotes Gauss hypergeometric function (e.g. Slater (1960)). When $g(\lambda) = \sqrt{2/\pi}\exp\left\{-\lambda^2/2\right\}$, we evaluate the integral numerically.

The approximation is very accurate for both weighting functions despite the small value of b considered and despite $\cos^2\theta$ being close to $q/b = 2/19 \simeq 0.105$.

[TABLE 2 APPROXIMATELY HERE]

4 Further Remarks

In Section 3 we have discussed how the construction of WAP tests can be extended to more general weighting functions g. One of the problems in the application of WAP tests is the averaging over the partitions $\tau \in \Upsilon$ because it requires the computation of several F-tests. This problem is worsened by a large sample size because, as this increases, the number of possible change-points also increases. In this Section we discuss possible ways of overcoming this situation.

We have already noticed in Theorem 1 that the statistic LR_0 is a ratio of quadratic forms of the OLS residuals of the regression on y on X, and that the matrix A_Υ in the numerator needs to be evaluated only once. This property makes the LR_0 test very appealing.

In the more general case the sum over partitions cannot be avoided. However, equation (4) shows that the WAP test $S_{g,p}$ is the expected value of $\Im(\theta_\tau)$, $\tau \in \Upsilon$. As such it can be estimated by taking a sample of n observations $\Im(\theta_{\tau^i})$ (occurring with probability $p(\tau)$), $\tau^1, \tau^2, ..., \tau^n$, say, and computing the sample mean,

$$\hat{S}_{g,p} = n^{-1} \sum_{i=1}^{n} \Im(\theta_{\tau^i}).$$

The expected value of $\hat{S}_{g,p}$ equals $S_{g,p}$, and its variance is a decreasing function of n. Therefore, by choosing n sufficiently large we can obtain a precise estimate of the $S_{g,p}$. The computational burden can be reduced by choosing n smaller than the number of partitions in Υ so that the number of F-tests to calculate is on average smaller than n.

As an example of efficiency of this procedure consider 120 i.i.d observations y_i obtained as $y_i = \beta_0 + (-1)^i \beta_1 + N(0,1)$ with $\beta_0 = \beta_1 = 0$. We allow for one break at $5 \leq t \leq 116$. The critical values for several WAP tests based on 10000 replications are reported in Table 3. The second column of Table 3 contains the critical values calculated in the standard way, the third and forth contain the critical values for the same test statistic when the sum over all possible change point is approximated as indicated above. In this case we take $n = 112$ and $n = 50$ giving an average number of different F-test statistics in each iteration approximately equal to, respectively, 69 and 17. The approximation seems to perform well.

[TABLE 3 APPROXIMATELY HERE]

As an alternative to the above procedure, one could try to find optimality criteria that would deliver a simple test statistic which does not require the evaluation of several F test statistics. Nyblom (1989) suggests a locally most powerful test for parameter constancy in model (1) with $\gamma = 0$ by assuming that β is a martingale process. A recent development (cf. Carrasco (2004)) is based on an *average model.* That is one could average y in equation (1) over all possible change-points and obtain

$$y = X\beta + \bar{Z}\gamma + u$$

where

$$\bar{Z} = \sum_{\tau \in \Upsilon} p(\tau) Z(\tau),$$

and test whether γ is zero or not using an F test. This procedure is not based on any classical statistical criteria, and it may be difficult to justify both model averaging and the optimality of the resulting test. However, its critical values are easily calculated, and this is certainly an appealing property.

5 Conclusions

This paper has studied WAP tests for structural change in a Gaussian linear regression model. We have shown that the LR_0 test is optimal for a large class of weighting functions and is also easy to compute because it requires the evaluation of a quadratic form in the vector of residuals only. This properties make the test very attractive since it is the simpler test in the class of WAP tests considered, and has also good power properties.

We have also shown that WAP tests can be constructed for very general weighting functions by means of uniform Laplace approximations. These perform very well even for a small sample size. A discussion of ways to reduce the computational burden of averaging over all possible changepoints is also given.

A Proofs

A.1 Proof of Theorem 1

Write

$$\begin{aligned} S_{\pi_\infty,p} &= \lim_{a\to\infty}\sum_{\tau\in\Upsilon} p(\tau)\int_{\lambda_\tau>0}\exp\left\{-\frac{1}{2}(T-p)\lambda_\tau\right\} \\ &\quad {}_1F_1\left(\frac{T-p}{2};\frac{K}{2};\frac{(T-p)\lambda_\tau\cos^2\theta_\tau}{2}\right)g_a'(\lambda_\tau)\,d\lambda_\tau \\ &= \sum_{\tau\in\Upsilon} p(\tau)\int_{\lambda_\tau>0}\exp\left\{-\frac{1}{2}(T-p)\lambda_\tau\right\} \\ &\quad {}_1F_1\left(\frac{T-p}{2};\frac{K}{2};\frac{(T-p)\lambda_\tau\cos^2\theta_\tau}{2}\right)\lim_{a\to 0}g_a'(\lambda_\tau)\,d\lambda_\tau. \end{aligned}$$

The first part of the theorem follows from the fact that $g_a'(\lambda_\tau) = dg_a(\theta)/d\theta|_{\theta=\lambda_\tau}$ converges to the derivative of a delta function $\delta(\lambda_\tau)$. The second part of the theorem follows from the definition of $\cos^2\theta_\tau$.

A.2 Proof of Corollary 1

The corollary follows from Theorem 1 and from Corollary 1 of Forchini (2002).

A.3 Proof of Theorem 2

We could not find a reference for this result in the literature. However, since, it can be easily obtained using the methods described in Chapter 9 of Bleistein and Handelsman (1986), we give here an outline of the proof only.

Consider the integral of equation (5). Since the minimum of $h(\lambda;\theta)$ can be anywhere in $[0,+\infty)$, it can be on the boundary, and this is the source of the problems.

Define a new variable of integration so that

$$h(\lambda;\theta)=\phi(t;\gamma)=\frac{t^2}{2}+\gamma t$$

so that $\lambda=0$ is mapped to $t=0$, $\lambda=+\infty$ is mapped to $t=+\infty$. Choose γ so that $\lambda=\lambda_0$ is mapped to $t=-\gamma$, a critical point of $\phi(t;\gamma)$. Thus, we must have

$$h(\lambda_0;\theta)=\frac{(-\gamma)^2}{2}+\gamma(-\gamma)=-\frac{\gamma^2}{2}$$

so that $\gamma^2=-2h(\lambda_0;\lambda_0)$ (note that $h(\lambda_0;\lambda_0)\leq 0$). The correct solution is $\gamma=-\sqrt{-2h(\lambda_0;\lambda_0)}$ if $\lambda_0>0$ and $\gamma=\sqrt{-2h(\lambda_0;\lambda_0)}$ if $\lambda_0=0$. That is $\gamma=\nu_\theta\sqrt{-2h(\lambda_0;\lambda_0)}$.

Since

$$\frac{dh(\lambda;\theta)}{dt}=h'(\lambda;\lambda_0)\frac{d\lambda}{dt}=t+\gamma$$

the Jacobian of the transformation $\lambda\to y$ is

$$\frac{d\lambda}{dt}=\frac{t+\gamma}{h'(\lambda;\theta)}.$$

Note that as $t\to-\gamma$ the limit of the ratio must be calculated using l'Hospital rule

$$\lim_{t\to-\gamma}\frac{d\lambda}{dt}=\lim_{t\to-\gamma}\frac{1}{h''(\lambda;\theta)\,\frac{d\lambda}{dt}}$$

so that

$$\lim_{t\to-\gamma}\frac{d\lambda}{dt}=\frac{1}{\sqrt{h''(\lambda_0;\theta)}}.$$

Moreover,

$$\lim_{t\to0}\frac{d\lambda}{dt}=\frac{\gamma}{h'(0;\theta)}$$

if $\gamma \neq 0$ (and $h'(0;\lambda_0) \neq 0$). If $\gamma = 0$, we need to use l'Hospital rule again and obtain

$$\lim_{t\to 0} \frac{d\lambda}{dt} = \frac{1}{\sqrt{h''(0;\theta_0)}}$$

where $\cos^2\theta_0 = q/b$.

We can write

$$\mathfrak{I}(\theta) = \int_{t>0} \exp\left\{-b\left(\frac{t^2}{2} + \gamma t\right)\right\} G(t;\theta)\, dt$$

where

$$G(t;\theta) = g(\lambda)\frac{d\lambda}{dt} = g(\lambda)\frac{t+\gamma}{h'(\lambda;\theta)},$$

and

$$G(-\gamma;\lambda_0) = \lim_{t\to-\gamma} G(t;\lambda_0) = \frac{g(-\gamma)}{\sqrt{h''(\lambda_0;\lambda_0)}}.$$

Write $G(t;\theta)$ as

$$G(t;\theta) = a_0 + a_1 t + t(t+\gamma)H(t;\theta)$$

with

$$\begin{aligned} a_0 &= G(0;\theta) = \begin{cases} \frac{g(0)\gamma}{\sqrt{h'(0;\theta)}} & \text{if } \cos^2\theta < q/b \\ \frac{g(0)}{\sqrt{h''(0;\theta)}} & \text{if } \cos^2\theta \geq q/b \end{cases} \\ a_1 &= \frac{G(-\gamma;\theta) - a_0}{-\gamma} = \frac{G(-\gamma;\theta) - G(0;\theta)}{-\gamma} = \frac{g(0)}{h'(0;\theta)} - \frac{g(\lambda_0)}{\gamma\sqrt{h''(\lambda_0;\theta)}} \end{aligned}$$

for $\lambda > 0$ and

$$\lim_{-\gamma\to 0} \frac{G(-\gamma;\theta) - G(0;\theta)}{-\gamma} = G'(0;\theta) = \left.\frac{d\left(\frac{g(\lambda)}{\sqrt{h''(\lambda;\theta)}}\right)}{dt}\right|_{t=0}.$$

So

$$\begin{aligned} \mathfrak{I}(\theta) &= G(0;\lambda_0)\int_{t>0} \exp\left\{-b\left(\frac{t^2}{2} + \gamma t\right)\right\} dt \\ &\quad + \left(\frac{G(-\gamma;\lambda_0) - G(0;\lambda_0)}{-\gamma}\right)\int_{t>0} \exp\left\{-b\left(\frac{t^2}{2} + \gamma t\right)\right\} t\,dt \\ &\quad + R(b) \end{aligned} \tag{13}$$

One can show, by integration by parts, that the reminder

$$R(b) = \int_{t>0} \exp\left\{-b\left(\frac{t^2}{2}+\gamma t\right)\right\} t(t+\gamma) H(t;\theta)\, dt$$

is asymptotically negligible. Moreover

$$\int_{t>0} \exp\left\{-b\left(\frac{t^2}{2}+\gamma t\right)\right\} dt = \sqrt{\frac{2\pi}{b}} \exp\left\{\frac{b\gamma^2}{2}\right\}\left(1-\Phi\left(\gamma\sqrt{b}\right)\right) \tag{14}$$

where

$$\Phi(x) = (2\pi)^{-1/2} \int_{-\infty}^{x} \exp\left\{-\frac{z^2}{2}\right\} dz$$

is the CDF of the Standard normal distribution. Similarly

$$\int_{t>0} \exp\left\{-b\left(\frac{t^2}{2}+\gamma t\right)\right\} t dt = \sqrt{\frac{2\pi}{b}} (-\gamma) \exp\left\{\frac{b\gamma^2}{2}\right\}\left(1-\Phi\left(\gamma\sqrt{b}\right)\right) + b^{-1}. \tag{15}$$

Inserting (14) and (15) in (13) yields

$$\begin{aligned} \mathfrak{I}(\theta) \quad &\sim \quad G(-\gamma;\lambda_0)\sqrt{\frac{2\pi}{b}} \exp\left\{\frac{b\gamma^2}{2}\right\}\left(1-\Phi\left(\gamma\sqrt{b}\right)\right) \\ &\quad + \left(\frac{G(-\gamma;\lambda_0)-G(0;\lambda_0)}{-\gamma}\right) b^{-1} \end{aligned}$$

so rearranging,

$$\begin{aligned} \mathfrak{I}(\theta) \quad &\sim \quad g(\lambda_0)\sqrt{\frac{2\pi}{bh''(\lambda_0;\lambda_0)}} \exp\left\{\frac{b\gamma^2}{2}\right\}\left(1-\Phi\left(\gamma\sqrt{b}\right)\right) \\ &\quad + \left(\frac{g(0)}{h'(0;\lambda_0)} - \frac{g(\lambda_0)}{\gamma\sqrt{h''(\lambda_0;\lambda_0)}}\right) b^{-1} \end{aligned}$$

and the statement of the theorem follows.

A.4 Proof of Theorem 3

Before proving Theorem 3 we need to find an asymptotic expansion for $h'(\lambda;\theta)$.

Lemma 1 *The following expansion holds*

$$\begin{aligned} q^{-1}\ln\left({}_1F_1\left(-aq;q;-qx\right)\right) \;\sim\; & -\frac{1}{2}\left(1+x\right)+\frac{1}{2}\sqrt{\left(1+x\right)^2+4ax} \\ & -\frac{1}{2}\log\left(1+a\right)-\;a\,\log\left(2\,\left(1+a\right)\right) \\ & +\frac{1}{2}\left(1+2\,a\right)\,\log\left(1+2\,a+x+\sqrt{\left(1+x\right)^2+4ax}\right) \\ & -\frac{1}{2}\log\left(1+x+2\,a\,x+\sqrt{\left(1+x\right)^2+4ax}\right) \end{aligned}$$

and

$$\frac{d}{dx}q^{-1}\ln\left({}_1F_1\left(-aq;q;-qx\right)\right)\sim\frac{-\left(1+x\right)+\sqrt{\left(1+x\right)^2+4ax}}{2x}$$

Proof. The hypergeometric function $y={}_1F_1\left(-aq;q;t\right)$ satisfies the differential equation:

$$ty''\left(t\right)+\left(q-t\right)y'\left(t\right)+aqy\left(t\right)=0 \tag{16}$$

(e.g. equation (1.1.6) of Slater (1960)). By transforming t to $t=-qx$, and defining $w\left(x\right)=y\left(-qx\right)$, equation (16) can be written as

$$-q^{-1}x\frac{w''\left(x\right)}{w\left(x\right)}-\left(1+x\right)\frac{w'\left(x\right)}{w\left(x\right)}+aq=0. \tag{17}$$

By defining $h\left(x\right)=q^{-1}\ln\left(w\left(x\right)\right)$, one can write (17) in terms of $h\left(x\right)$ as

$$\left[a-x\left[h'\left(x\right)\right]^2-\left(1+x\right)h'\left(x\right)\right]q-xh''\left(x\right)=0. \tag{18}$$

The function $h\left(x\right)$ solves equation (18) subject to the condition that it is analytic at $x=0$ and $h\left(0\right)=0$. Thus we can replace the series $h\left(x\right)=\sum_{j=0}^{\infty}q^{-j}P_j\left(x\right)$, where $P_j\left(0\right)=0$ for all $j=0,1,...$, in equation (18) and compare the coefficients of similar powers of q. So equating the coefficient of q to zero gives

$$a-x\left[P_0'\left(x\right)\right]^2-\left(1+x\right)P_0'\left(x\right)=0.$$

There are two solutions to the above differential equation, but only one satisfies $P_0\left(0\right)=0$, corresponding to

$$P_0'\left(x\right)=\frac{-\left(1+x\right)+\sqrt{\left(1+x\right)^2+4ax}}{2x}.$$

Since $h(x)$ is analytic, we have that $h'(x) \sim P_0'(x)$, and this proves the lemma. ■

We can now prove Theorem 3. From (3) one obtains

$$h'(\lambda;\theta) = \sin^2\theta - \cos^2\theta \, \frac{d}{dx} q^{-1} \ln\left\{ {}_1F_1(-aq;q;-qx) \right\} \Big|_{x=(1+a)\lambda\cos^2\theta}.$$

Using Lemma 1, it follows that

$$h'(\lambda) \sim \sin^2\theta - \cos^2\theta \, \frac{-(1+x) + \sqrt{(1+x)^2 + 4ax}}{2x} \Bigg|_{x=(1+a)\lambda\cos^2\theta}$$

The statement of the Theorem follows easily.

References

ANDREWS, D. W. K., I. LEE, AND W. PLOBERGER (1996): "Optimal Changepoint Tests for Normal Linear Regression," *Journal of Econometrics*, 70, 9–36.

ANDREWS, D. W. K., AND W. PLOBERGER (1994): "Optimal Tests When a Nuisance Parameter is Present Only under the Alternative," *Econometrica*, 62, 1383–1414.

BLEISTEIN, N., AND R. A. HANDELSMAN (1986): *Asymptotic Expansions of Integrals.* Dover Publications, Inc., New York.

CARRASCO, M. (2004): "Chi-Square Tests for Parameter Stability," Mimeo, University of Rochester.

COX, D. R., AND D. V. HINKLEY (1974): *Theoretical Statistics.* Chapman and Hall, London.

DE BRUIJN, N. G. (1961): *Asymptotic Methods in Analysis.* North-Holland, Amsterdam.

DUFOUR, J.-M., AND L. KHALAF (2001): "Monte Carlo Test Methods in Econometrics," in *Companion to Theoretical Econometrics*, ed. by B. Baltagi, pp. 494–519. Blackwell, Oxford, U.K.

Forchini, G. (2002): "Optimal Similar Tests for Structural Change for the Linear Regression Model," *Econometric Theory*, 18, 853–867.

Hillier, G. H. (1987): "Classes of Similar Regions and their Power Properties for some Econometric Testing Problems," *Econometric Theory*, 3, 1–44.

Imhof, J. P. (1961): "Computing the Distribution of Quadratic Forms in Normal Variables," *Biometrika*, 48, 419–426.

King, M., and G. H. Hillier (1985): "Locally Best Invariant Tests of the Error Covariance Matrix of the Linear Regression Model," *Journal of the Royal Statistical Society, B*, 47, 98–102.

Nyblom, J. (1989): "Testing for the Constancy of Parameters Over Time," *Journal of the American Statistical Association*, 84, 223–230.

Slater, L. J. (1960): *Confluent Hypergeometric Functions.* Cambridge University Press, London.

Wald, A. (1943): "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large," *American Mathematical Society Transactions*, 54, 426–482.

$b = 10$, $q = 1$		
$\cos^2\theta$	$\tilde{\lambda}_0$	λ_0^N
.05	-0.053	-0.182
.10	0.000	0.000
.15	0.059	0.091
.20	0.125	0.169
.25	0.200	0.249
.30	0.286	0.339
.35	0.385	0.441
.40	0.500	0.559
.45	0.636	0.698
.50	0.800	0.864
.55	1.000	1.067
.60	1.250	1.319
.65	1.571	1.644
.70	2.000	2.075
.75	2.600	2.679
.80	3.500	3.583
.85	5.000	5.087
.90	8.000	8.091
.95	17.000	17.095

$b = 50$, $q = 1$		
$\cos^2\theta$	$\tilde{\lambda}_0$	λ_0^N
.05	0.032	0.040
.10	0.089	0.099
.15	0.153	0.163
.20	0.225	0.236
.25	0.307	0.318
.30	0.400	0.412
.35	0.508	0.520
.40	0.633	0.646
.45	0.782	0.795
.50	0.960	0.973
.55	1.178	1.191
.60	1.450	1.464
.65	1.800	1.815
.70	2.267	2.282
.75	2.920	2.936
.80	3.900	3.917
.85	5.533	5.551
.90	8.800	8.818
.95	18.600	18.679

$b = 20$, $q = 10$		
$\cos^2\theta$	$\tilde{\lambda}_0$	λ_0^N
.05	-0.474	-3.015
.10	-0.444	-1.472
.15	-0.412	-0.951
.20	-0.375	-0.683
.25	-0.333	-0.514
.30	-0.286	-0.391
.35	-0.231	-0.290
.40	-1.667	-0.197
.45	-0.091	-0.103
.50	0.000	0.000
.55	0.111	0.119
.60	0.250	0.265
.65	0.429	0.449
.70	0.667	0.691
.75	1.000	1.029
.80	1.500	1.533
.85	2.333	2.370
.90	4.000	4.041
.95	9.000	9.045

$b = 50$, $q = 10$		
$\cos^2\theta$	$\tilde{\lambda}_0$	λ_0^N
.05	-0.158	-0.370
.10	-0.111	-0.159
.15	-0.059	-0.070
.20	0.000	0.000
.25	0.067	0.071
.30	0.143	0.149
.35	0.231	0.239
.40	0.333	0.343
.45	0.455	0.465
.50	0.600	0.611
.55	0.778	0.789
.60	0.100	1.013
.65	1.286	1.299
.70	1.667	1.680
.75	2.200	2.215
.80	3.000	3.016
.85	4.333	4.350
.90	7.000	7.018
.95	15.000	15.019

Table 1: Approximate and exact solutions of $h'(\lambda) = 0$ for various values of b, q and $\cos^2\theta$.

$g(\lambda)$	1			$\frac{\exp\left\{-\frac{\lambda^2}{2}\right\}}{\sqrt{\pi/2}}$		
$\cos^2\theta$	$\mathfrak{I}(\theta)$	$\mathfrak{I}_A(\theta)$	$\mathfrak{I}_A(\theta)/\mathfrak{I}(\theta)$	$\mathfrak{I}(\theta)$	$\mathfrak{I}_A(\theta)$	$\mathfrak{I}_A(\theta)/\mathfrak{I}(\theta)$
0.01	0.058	0.058	0.998	0.046	0.046	1.000
0.02	0.064	0.064	0.995	0.051	0.051	1.000
0.03	0.071	0.071	0.993	0.056	0.056	0.999
0.04	0.079	0.079	0.991	0.063	0.063	0.998
0.05	0.089	0.088	0.990	0.070	0.070	0.998
0.06	0.100	0.099	0.989	0.079	0.079	0.998
0.07	0.112	0.111	0.989	0.089	0.089	0.999
0.08	0.127	0.126	0.989	0.101	0.101	1.000
0.09	0.145	0.144	0.990	0.114	0.115	1.000
0.10	0.166	0.164	0.992	0.130	0.131	1.000
0.11	0.190	0.189	0.993	0.149	0.150	1.010
0.12	0.219	0.218	0.994	0.172	0.174	1.010
0.13	0.253	0.252	0.996	0.198	0.201	1.010
0.14	0.295	0.294	0.997	0.230	0.234	1.020
0.15	0.344	0.343	0.999	0.268	0.273	1.020
0.16	0.403	0.403	1.000	0.313	0.320	1.020
0.17	0.475	0.475	1.000	0.368	0.377	1.030
0.18	0.562	0.563	1.000	0.434	0.446	1.030
0.19	0.668	0.669	1.000	0.513	0.529	1.030
0.20	0.797	0.799	1.000	0.610	0.631	1.030

Table 2: Approximate and exact value of the integral $\mathfrak{I}(\theta)$ for $b = 19$ and $q = 2$.

	Standard	Approx n=112	Approx n=50
avg-F	2.2310	2.2558	2.2553
exp-F_∞	3.7555	3.7491	3.7211
LR_0	0.0366	0.0367	0.0367
LR_∞	36.7754	35.6337	34.8312

Table 3: Critical values for various test statistics based on 10000 replications. The second column contains the critical values for the statistic calculated as an average over all possible partitions. In the third and fourth columns the test statistic is approximated by sampling over the possible change points with uniform weights with n=112 and n=50 respectively.