Econometric Theory, **18**, 2002, 853–867. Printed in the United States of America.
DOI: 10.1017.S0266466602184027

OPTIMAL SIMILAR TESTS FOR STRUCTURAL CHANGE FOR THE LINEAR REGRESSION MODEL

G. FORCHINI
University of York

This paper analyzes similar tests for structural change for the normal linear regression model in finite samples. Using the approach of Wald (1943, *American Mathematical Society Transactions* 54, 426–482), Hillier (1987, *Econometric Theory* 3, 1–44), Andrews and Ploberger (1994, *Econometrica* 62, 1382–1414), and Andrews, Lee, and Ploberger (1996, *Journal of Econometrics* 70, 9–36), we characterize a class of optimal similar tests for the existence of (possibly multiple) changepoints at unknown times. We extend the analysis of Andrews et al. (1996) by deriving weighted optimal similar tests for the case where the error variance is not known. We also show that when the sample size is large, the tests of Andrews et al. constructed by replacing the error variance with an estimate are equivalent to the optimal test derived in this paper. Power comparisons are provided by a small simulation study.

1. INTRODUCTION

Testing for structural change in the linear regression model has received considerable attention in the statistical and econometric literature in the last 30 years. This intense activity has led to the development of many tests for structural change that are widely used by practitioners.

Two different testing situations have been studied under the heading of "testing for structural change." The first one arises when the changepoint is assumed to be known, e.g., because of institutional changes, wars, etc. (see Chow, 1960; for a survey of results, see Pesaran, Smith, and Yeo, 1984). In this setup, testing for structural change is equivalent to testing for linear restrictions on the parameters (provided the variance does not change over time), and a characterization of similar tests (i.e., tests for which their sizes do not depend on nuisance parameters) for this problem can be found in Hillier (1987).

I thank Grant Hillier for patient help, suggestions, and useful discussions and the editor, Peter Phillips, and two anonymous referees for comments that led to an improved version of the paper. I also thank Karim Abadir, Les Godfrey, Patrick Marsh, Jan Podivinsky, Richard Smith, and the participants of the 1997 EC^2 conference in Amsterdam for helpful comments. Grants from ESRC (R00429424115) and the University of Southampton are gratefully acknowledged. Address correspondence to: G. Forchini, Department of Economics and Related Studies, University of York, Heslington, York Y01 5DD, UK.

© 2002 Cambridge University Press 0266-4666/02 $9.50

https://doi.org/10.1017/S0266466602184027 Published online by Cambridge University Press

The second testing situation, the focus of this paper, occurs when there is no information available regarding the time of the possible structural changes. This gives rise to a nonstandard testing setup where classical tests are not directly applicable because of the existence of nuisance parameters under the alternative but not under the null hypothesis of no structural change. Even in this case, several of the most popular tests suggested in the literature (the CUSUM and the CUSUM of squares of Brown, Durbin, and Evans, 1975; the fluctuation test of Ploberger, Krämer, and Kontrus, 1989; the sup F test of Quandt, 1960; and Andrews, 1993; and the tests recently suggested by Andrews and Ploberger, 1994; and Andrews, Lee, and Ploberger, 1996) are similar.

For the case of unknown changepoints, Andrews et al. (1996) suggest tests for structural change in the Gaussian linear regression model that maximize a weighted average power. They derive a class of optimal tests for the case where the disturbance variance is known. For the more realistic case where the error variance is not known, they propose replacing the unknown variance by an estimate and show that the resulting tests are still similar (although they cannot prove the finite sample optimality).

In this paper we analyze similar tests for structural change in the Gaussian linear regression model in finite samples. The assumption of normality allows us to simplify the analysis, even though the results obtained can be indicative for more general models. We first use results in Hillier (1987) to characterize the class of similar tests, and we then derive "optimal" tests for structural change for the case of unknown error variance. Because it is well known that no uniformly most powerful similar test exists even for simple cases, we follow the suggestion of Wald (1943), Hillier (1987), Andrews and Ploberger (1994), and Andrews et al. (1996) and weaken the optimality criteria by maximizing a weighted power. The tests for structural change at unknown changepoints that we suggest improve upon those of Andrews et al. (1996), because they are optimal (in terms of weighted power) in the case where the error variance is unknown for every sample size. By showing that the tests suggested by Andrews et al. (1996) for the case of an unknown disturbance variance are the same as the optimal tests when the sample size is large, we also provide a proof of the asymptotic optimality of such tests in terms of our weighting function.

The geometrical aspects of the testing setup will be emphasized throughout the paper. We believe that these are important to better understand the effect of the uncertainty about the changepoint upon the inferential procedure and to indicate what kind of weighting functions is needed to derive optimal tests in terms of weighted power.

The remaining part of the paper is organized as follows. The model is described in Section 2, where similar tests are also characterized. Section 3 derives optimal tests for structural change (the proofs of all statements are in the Appendix). The results of a Monte Carlo simulation are reported in Section 4. Section 5 concludes.

https://doi.org/10.1017/S0266466602184027 Published online by Cambridge University Press

2. THE MODEL AND CHARACTERIZATION OF SIMILAR TESTS

This section characterizes similar tests for a Gaussian linear regression model with $t+1$ subsamples, containing, respectively, $\tau_1, \tau_2, \ldots, \tau_{t+1}$ ($\sum_{i=1}^{t+1} \tau_i = T$) observations:

$$
\begin{aligned}
y_{\tau_1} &= Z_{\tau_1}\eta + W_{\tau_1}\zeta + u_{\tau_1},\\
y_{\tau_2} &= Z_{\tau_2}(\eta + \gamma_1) + W_{\tau_2}\zeta + u_{\tau_2},\\
&\cdots\\
y_{\tau_{t+1}} &= Z_{\tau_{t+1}}(\eta + \gamma_t) + W_{\tau_{t+1}}\zeta + u_{\tau_{t+1}},
\end{aligned}
\tag{1}
$$

where $y_{\tau_i} = (y_{\bar{\tau}_{i-1}+1}, y_{\bar{\tau}_{i-1}+2}, \ldots, y_{\bar{\tau}_i})'$, $\bar{\tau}_i = \sum_{j=0}^{i} \tau_j$, is a $\tau_i \times 1$ vector of dependent variables, W_{τ_i} is a $\tau_i \times n$ matrix of independent variables, η and $\gamma_1, \ldots, \gamma_t$ are $k \times 1$ vectors of parameters, ζ is an $n \times 1$ vector of parameters, and u_{τ_i} is a $\tau_i \times 1$ vector of disturbances. The matrices $Z_{\tau_1}, Z_{\tau_2}, \ldots, Z_{\tau_t}, Z_{\tau_{t+1}}$ have the form

$$
Z_{\tau_1} = \begin{pmatrix} z_1 \\ \cdots \\ z_{\tau_1} \end{pmatrix}, Z_{\tau_2} = \begin{pmatrix} z_{\tau_1+1} \\ \cdots \\ z_{\tau_1+\tau_2} \end{pmatrix}, \ldots, Z_{\tau_i} = \begin{pmatrix} z_{\bar{\tau}_{i-1}+1} \\ \cdots \\ z_{\bar{\tau}_i} \end{pmatrix}, \ldots, Z_{\tau_{t+1}} = \begin{pmatrix} z_{\bar{\tau}_t+1} \\ \cdots \\ z_T \end{pmatrix},
$$

where z_i is the ith $(1 \times k)$ vector of observations on k explanatory variables.

In this paper we identify changepoints by an index τ that represents a partition of T into $t+1$ parts, $\tau = (\tau_1, \tau_2, \ldots, \tau_{t+1})$, $\tau_i > 0$ 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 changepoints in the model) is denoted by Υ.

Let $y = (y'_{\tau_1}, y'_{\tau_2}, \ldots, y'_{\tau_{t+1}})'$, $Z = (z'_1, \ldots, z'_T)'$, $W = (W'_{\tau_1}, W'_{\tau_2}, \ldots, W'_{\tau_{t+1}})'$, and $u = (u'_{\tau_1}, u'_{\tau_2}, \ldots, u'_{\tau_{t+1}})'$, so that (1) can be written as

$$
y = X\beta + Z(\tau)\gamma + u, \tag{2}
$$

where $X = (Z, W)$ is $T \times p$, $\beta = (\eta', \zeta')'$ is $p \times 1$, $\gamma = (\gamma'_1, \gamma'_2, \ldots, \gamma'_t)'$ is $K \times 1$, $u \sim N(0, \sigma^2 I_T)$, and $Z(\tau)$ is the $T \times K$ matrix obtained by deleting the first k columns in the block diagonal matrix $(Z_{\tau_1}, Z_{\tau_2}, \ldots, Z_{\tau_t}, Z_{\tau_{t+1}})$ in (1). Moreover, $p = n + k$, $K = kt$. We assume that $T - p - K = T - n - (1+t)k \geq 0$; otherwise the model is not identified even if τ is known. We also do not allow for changes over time for the error variance (σ^2), so that we can avoid Behrens–Fisher type problems. No assumption on the elements of the matrices X and $Z(\tau)$ has been made. These can be considered as either fixed or random. In the latter case they are assumed to be ancillary to the parameters $(\beta, \gamma, \sigma^2, \tau)$, and the following analysis will be conditional on them.

The setup considered is very general: the changepoints can be approximately located (e.g., they can occur about a particular date, or year) or only their number can be specified, or they could be totally unknown. We want to test $H_0: y \sim N(X\beta, \sigma^2 I_T)$, and in this case the statistics $\hat{\beta} = (X'X)^{-1}X'y$ and $s_1^2 = y'M_X y$ ($M_X = I_T - X(X'X)^{-1}X'$) are jointly sufficient for the nuisance parameters β and σ^2. Because they are boundedly complete, any similar region of size α is a

https://doi.org/10.1017/S0266466602184027 Published online by Cambridge University Press

fraction α of the surface $(\hat{\beta}, s_1^2) = \text{constant}$ (Hillier, 1987). It may happen that $\tau_i < k$ for some i, so that the vector $Z(\tau)\gamma$ can be zero even if γ is different from zero. Thus, the null hypothesis under consideration is different from $H_1: \gamma = 0$, because this might not be testable (Breusch, 1986). If $\tau_i < k$ for some i, $Z(\tau)\gamma$ can be written as a linear combination of the linearly independent columns of $Z(\tau)$, $Z^*(\tau)$, say, and under the alternative $Z^*(\tau)\gamma^* = 0$. We assume that $Z(\tau)\gamma$ is replaced by $Z^*(\tau)$ and γ by γ^* if $\tau_i < k$ for some i, but we will continue to indicate the resulting full rank matrix by $Z(\tau)$ and the parameter vector by γ. The number K is thus to be understood as the number of columns of the reduced matrix $Z^*(\tau)$.

The alternative hypothesis is $H_1: y \sim N(X\beta + Z(\tau)\gamma, \sigma^2 I_T)$. The term $Z(\tau)\gamma$ depends on t, τ, and γ, and as γ varies over $\mathbb{R}^K$, it spans m (K-dimensional) subspaces of $\mathbb{R}^T$, $\bar{V}^\tau = \{Z(\tau)\gamma : \gamma \in \mathbb{R}^K\}$, indexed by τ, where m is the number of elements in Υ. Note that the number of changepoints determines the dimension of the subspaces $\bar{V}^\tau$, whereas the position of the potential breaks influences the number of such subspaces. This setup is quite different from the classical testing situation where the alternative is specified as $H_1: y \sim N(X\beta + Q\gamma, \sigma^2 I_T)$ and Q is a completely known $(T \times K)$ matrix. In this case the columns of Q span a fixed K-dimensional subspace of $\mathbb{R}^T$, as γ varies over $\mathbb{R}^K$, $\bar{V} = \{Q\gamma : \gamma \in \mathbb{R}^K\}$. In the structural change case this is no longer so.

A characterization of similar tests of $H_0: y \sim N(X\beta, \sigma^2 I_T)$ is given by the following theorem of Hillier (1987).

THEOREM 1. *(Hillier, 1987). The class of all similar tests for* $H_0: y \sim N(X\beta, \sigma^2 I_T)$ *against any alternative whatever is characterized by the vector* $v = C'y/(s_1^2)^{1/2}$, *where* C *is a* $T \times T - p$ *matrix such that* $CC' = M_X$, $C'C = I_{T-p}$, *and* $C'X = 0$. *That is, a critical region for testing* H_0 *has size independent of* β *and* σ^2 *if and only if it is defined in terms of* v *alone. Moreover, under* H_0, *the vector* v *is uniformly distributed over the unit* $(T - p)$*-sphere* $S_{T-p} = \{v \in \mathbb{R}^{T-p} : v'v = 1\}$.

The matrix C can be chosen so that $C'y$ is the vector of recursive residuals for the model $y = X\beta + u$ (but we do not need to impose this condition on the tests considered in the subsequent discussion). For the null model considered, the CUSUM, the CUSUM of squares of Brown et al. (1975), the fluctuation test of Ploberger et al. (1989), the sup F test, and the tests of Andrews et al. (1996) are similar because they are all functions of v.

3. OPTIMAL TESTS FOR STRUCTURAL CHANGE

This section studies the optimal similar critical regions for tests of $H_0: y \sim N(X\beta, \sigma^2 I_T)$ against $H_1: y \sim N(X\beta + Z(\tau)\gamma, \sigma^2 I_T)$, where both γ and τ are unknown.

It is well known that even in the case where τ is known there is no uniformly most powerful test because the power of the test depends on the unknown (nuisance) parameter γ. A procedure, which can be traced back to Wald (1943) and is often used to handle the existence of nuisance parameters, consists in weak-

https://doi.org/10.1017/S0266466602184027 Published online by Cambridge University Press

ening the optimality criterion, i.e., in choosing the critical region to maximize an "average" power. For the case of known τ, Hillier (1987) suggests maximizing the average power on the surface of constant $\lambda_\tau = \gamma' Z(\tau)' M_X Z(\tau)\gamma/\sigma^2$ by averaging over all possible directions of $C'Z(\tau)\gamma/\sigma$. The most powerful similar critical region (in terms of average power) is given by the points on S_{T-p} for which the angle with the subspaces V^τ is small (Hillier, 1987, Sect. 3). That is, ω is a "strip" on S_{T-p} close to $S_{T-p} \cap V^\tau$. Note that the angle θ_τ between v and the subspace generated by the columns of $C'Z(\tau)$ is related to the F-statistic, f_τ, for testing $H_0: \gamma = 0$ against $H_1: \gamma \neq 0$, τ fixed, by the relation $(\cos\theta_\tau)^2 = [K/(T-p-K)]\, f_\tau (1 + [K/(T-p-K)]\, f_\tau)^{-1}$.

When τ is itself a nuisance parameter under H_1, (as γ varies in $\mathbb{R}^K$) the columns of $C'Z(\tau)$ generate a hyperplane V^τ indexed by τ. When weakening the optimality criterion to maximize an "average" power, we are confronted with the problem of averaging *over and among* different spaces V^τ. An obvious generalization of the traditional (fixed τ) procedure consists in assigning a weight $p(\tau)$ to each subspace V^τ and in maximizing the average power over the surfaces of constant λ_τ for each V^τ. However, this is not enough to yield a test that is uniformly most powerful in terms of average power, as can be deduced from the proof of Theorem 2, which follows.

Averaging among subspaces requires much more information than just averaging over the surface of constant λ_τ for fixed τ. It requires taking into account λ_τ itself, i.e., the distance from the origin of the alternative for each fixed τ. Thus a further weight on λ_τ is needed. We will assume that the weight for λ_τ/c is proportional to the density of a chi-square distribution with K degrees of freedom. The constant c determines the relative weight given to different alternatives. When c is close to zero then small deviations from the null are given a large weight. On the other hand, when c is large relatively more weight is given to large parameters.

THEOREM 2. *The critical region ω that maximizes the unconditional power averaged over (i) the partition τ with weights $p(\tau)$, (ii) the direction of $C'Z(\tau)\gamma/\sigma$ with uniform weights on S_K, and (iii) $\lambda_\tau = \gamma' Z(\tau)' M_X Z(\tau)\gamma/\sigma^2$ with weights proportional to the density function of a chi-square random variable with K degrees of freedom is given by*

$$LR_c > c_1,$$

where

$$LR_c = (1+c)^{-(K/2)} \sum_{\tau \in \Upsilon} p(\tau) \left(1 + \frac{c}{1+c}\cos^2\theta_\tau\right)^{-[(T-p)/2]}$$

$$= (1+c)^{(T-p-K)/2} \sum_{\tau \in \Upsilon} p(\tau) \left(\frac{1 + \dfrac{K}{T-p-K} f_\tau}{1 + c + \dfrac{K}{T-p-K} f_\tau} \right)^{(T-p)/2}, \qquad (3)$$

https://doi.org/10.1017/S0266466602184027 Published online by Cambridge University Press

where $0 < \theta_\tau < \pi/2$ *is the angle between* v *and* V^τ *and* f_τ *is the* F *test statistic for testing* $H_0: y \sim N(X\beta, \sigma^2 I_T)$ *against* $H_1: y \sim N(X\beta + Z(\tau)\gamma, \sigma^2 I_T)$ *for a fixed* τ*. The critical value,* c_1*, is determined by the condition*

$$\frac{\Gamma\left(\frac{T-p}{2}\right)}{2\pi^{(T-p)/2}} \int_{\{v \in S_{T-p}: LR_c > c_1\}} (dv) = \alpha,$$

where (dv) *denotes the unnormalized Haar measure on* S_{T-p}*.*

This result is proved in the Appendix. The statistic LR_c is called the "likelihood ratio" following Andrews et al. (1996).

Remarks.

(i) The derivation of the LR_c test is based on a particular weighting function for λ_τ, and the actual power of the test could be strongly dependent on the choice of such weighting function. The critical values for the LR_c test are difficult to derive because they depend on the complex geometrical relations among the spaces V^τ. However, they can be simulated for the given regressors because of the similarity property. For the derivation of the asymptotic distributions of the test statistics derived previously it might be useful to impose additional assumptions on the regressors in (1) (such as stationarity, ergodicity, and weak exogeneity) as in Andrews (1993) and Andrews and Ploberger (1994). Recently, Hansen (2000) has derived the asymptotic distribution of the sup F tests and the tests of Andrews et al. (1996) in the presence of exogenous stochastic trend, polynomial trend, or structural shifts in the regressors. The results can be used to derive the asymptotic critical values for the LR_c test in view of the results of Corollary 1, which follows.

(ii) Andrews et al. (1996) suggest averaging the power over all values of $(\beta', \gamma')'$ with weights proportional to the density of a normal distribution over V^τ, and they obtain a class of similar tests for the case where σ^2 is known, given by a weighted average of the form

$$\text{exp-F}_c = (1+c)^{-(K/2)} \sum_{\tau \in Y} p(\tau) \exp\left\{\frac{cKf_\tau}{2(1+c)}\right\}.$$

Our approach is the same as taking a weight for $C'Z(\tau)\gamma/\sigma$ proportional to a $N(0, \sqrt{c} I_K)$ density, where c is an arbitrary positive constant that scales the variance matrix. Averaging the unconditional power with respect to $C'Z(\tau)\gamma/\sigma \sim N(0, \sqrt{c} I_K)$ is equivalent to averaging the power of a critical region ω with respect to both the direction of $C'Z(\tau)\gamma/\sigma$ and its length (i.e., λ_τ with $\lambda_\tau/\sqrt{c} \sim \chi^2_K$). Note that the weighting function we use is different from that of Andrews et al. (1996), because ours does not depend on the nuisance parameter β. Our approach can be justified by the fact that the statistic $\hat{\beta}$ is not informative about $Z(\tau)\gamma$, so there is no loss of information in constructing inference about $Z(\tau)\gamma$ from $C'y$ only (which also does not depend on β). Moreover in contrast to Andrews et al. (1996), we are able to show that the tests suggested are optimal

https://doi.org/10.1017/S0266466602184027 Published online by Cambridge University Press

when σ^2 is unknown. The relationship between the LR_c and the exp-F_c tests as functions of the individual F test statistics (f_τ) is given by Corollary 1.

COROLLARY 1. *Let LR_c indicate the LR_c statistic in (3). Then, as T goes to infinity, for a fixed p, the exp-F_c and the LR_c test statistics coincide:*

$$\text{exp-F}_c - LR_c = o_P(1).$$

Thus the exp-F_c tests of Andrews et al. (1996) are optimal in terms of the average power of Theorem 2 when the sample size is large and the $\sup_{\tau\in\Upsilon} f_\tau$ is bounded in probability. Thus the asymptotic critical values of LR_c tests are the same as the asymptotic critical values of the exp-F_c tests. Note also that because Andrews and Ploberger (1994) show that the exp-F_c tests are asymptotically optimal under more general conditions than those considered in this paper, these asymptotic optimality properties can be extended to the LR_c tests using Corollary 1. Moreover, the proof of Corollary 1 is based on the convergence of $(1 + 1/n)^{xn}$ to $\exp\{x\}$, which is very slow. Therefore the convergence of LR_c to exp-F_c is also very slow. I thank one of the referees, who noted that the validity of this result did not need one of the assumptions originally made on Υ.

(iii) As far as the choice of c is concerned, there does not seem to be any optimal value. However, as $c \to \infty$, $C'Z(\tau)\gamma/\sigma$ tends to be uniformly distributed over $\mathbb{R}^K$, and this, using a Bayesian interpretation, might be taken as a noninformative prior. A normalized version of the LR_c test for this case is

$$\begin{aligned} LR_\infty &= \lim_{c\to\infty} (1+c)^{K/2} LR_c \\ &= \sum_{\tau\in\Upsilon} p(\tau)(1-\cos^2\theta_\tau)^{-[(T-p)/2]} \\ &= \sum_{\tau\in\Upsilon} p(\tau)\left(1+\frac{K}{T-p-K} f_\tau\right)^{(T-p)/2}. \end{aligned} \tag{4}$$

Note that the Andrews et al. (1996) test statistic in this case is exp-F_∞ = $\ln(\sum_{\tau\in\Upsilon} p(\tau)\exp\{Kf_\tau/2\})$. On the other hand, letting c tend to zero (so that possible departures of $C'Z(\tau)\gamma/\sigma$ tend to be small), a normalized LR_c test is

$$\begin{aligned} LR_0 &= \frac{2}{T-p}\lim_{c\to 0}\frac{LR_c - (1+c)^{-(k/2)}}{c} \\ &= \sum_{\tau\in\Upsilon} p(\tau)\cos^2\theta_\tau, \\ &= \sum_{\tau\in\Upsilon} p(\tau)\frac{\frac{K}{T-p-K} f_\tau}{1+\frac{K}{T-p-K} f_\tau}. \end{aligned} \tag{5}$$

As a consequence LR_0 should have power in detecting small deviations from the null. The corresponding statistic in Andrews et al. (1996) is avg-F = $\sum_{\tau\in\Upsilon} p(\tau) f_\tau$.

https://doi.org/10.1017/S0266466602184027 Published online by Cambridge University Press

4. MONTE CARLO RESULTS

This section presents Monte Carlo results comparing the optimal tests derived previously with the sup F test and the exp-F_c tests of Andrews et al. (1996). We consider three models, and we concentrate on the case where there can be only two possible subsamples.

Model 1. The first model considered is an example of a stationary model:

$$y_i = \beta_0 + (-1)^i \beta_1 + N(0,1), \tag{6}$$

where $T = 120$ and the possible structural change affects the intercept

$$y_i = \begin{cases} \beta_0 + (-1)^i \beta_1 + N(0,1) & \text{for } i = 1,\ldots,\tau \\ \beta_0 + a_0 + (-1)^i \beta_1 + N(0,1) & \text{for } i = \tau+1,\ldots,T, \end{cases}$$

where a_0 is a constant.

Model 2. The second model is an example of a time-trend model:

$$y_i = \beta_0 + \left(i - \frac{T+1}{2}\right)\beta_1 + N(0,1), \tag{7}$$

for $T = 120$, and under the alternative there is a structural change in the intercept

$$y_i = \begin{cases} \beta_0 + \left(i - \dfrac{T+1}{2}\right)\beta_1 + N(0,1) & \text{for } i = 1,\ldots,\tau \\ \beta_0 + a_0 + \left(i - \dfrac{T+1}{2}\right)\beta_1 + N(0,1) & \text{for } i = \tau+1,\ldots,T, \end{cases}$$

where a_0 is again a constant

Model 3. Finally the third model is

$$y_i = \beta_0 + \beta_1 GNP_i + \beta_2 r_i + N(0,\sigma^2),$$

$i = 1,2,\ldots,T(= 39)$, where the values of the regressors are the annual gross national product (GNP) (GNP_i) seasonally adjusted in 1992 constant dollars, and the interest rate (r_i) from Tables B-2 and B-73 of the 1999 *Economic Report of the President* (Council of Economic Advisers 1999). (These data extend the data set given by Greene (1991) in Example 7.9 to the period 1959–1997).

The following alternative model is considered:

$$y_i = \begin{cases} \beta_0 + \beta_1 GNP_i + \beta_2 r_i + N(0,\sigma^2) & \text{for } i = 1,\ldots,\tau \\ \beta_0 + a_0 + \beta_1 GNP_i + \beta_2 r_i + N(0,\sigma^2) & \text{for } i = \tau+1,\ldots,T, \end{cases}$$

where $T = 39$ and a_0 is a constant.

https://doi.org/10.1017/S0266466602184027 Published online by Cambridge University Press

Model 1 and Model 2 have also been studied by Andrews et al. (1996), and they have been chosen to facilitate the comparison. For these models we will assume, without loss of generality, that $(\beta_0, \beta_2) = (0,0)$ and $\sigma^2 = 1$. For Model 3 we also choose $(\beta_0, \beta_1, \beta_2) = (0,0,0)$ and $\sigma^2 = 1$.

The tests compared are based on the following test statistics: sup$p$-F $= \sup_{\tau=p+1,\ldots,T-p} f_\tau$, avg-F, exp-$F_1$, exp-$F_\infty$, LR_0, LR_1, LR_∞, and the F test for the correct changepoint, with the appropriate values of p, K, and T, and $p(\tau) = 1/(T - 2p)$ and $\tau = (t, T - t)$ with $t = p + 1, \ldots, T - p$. Note that we assume that no knowledge is available about which parameter is affected by the structural change.

The 5% critical values of the tests have been calculated by simulating the distribution of the tests using 100,000 replications. Because the tests are similar we can set $(\beta_0, \beta_1, \beta_2) = (0,0,0)$ and $(\beta_0, \beta_1) = (0,0)$ and $\sigma^2 = 1$ to obtain

TABLE 1. Power for Model 1

Alternative Distribution		Test Statistic								
τ	a_0	sup15-F	sup2-F	avg-F	exp-F_1	exp-F_∞	LR_0	LR_1	LR_∞	F
15	0.4	.09	.09	.11	.10	.10	.11	.10	.10	.23
	0.6	.21	.21	.19	.21	.21	.19	.21	.21	.49
	0.8	.42	.42	.32	.40	.41	.31	.39	.41	.75
	1.0	.68	.66	.50	.63	.65	.49	.62	.65	.91
30	0.4	.19	.17	.21	.21	.18	.21	.21	.19	.39
	0.6	.46	.42	.46	.47	.45	.45	.47	.46	.74
	0.8	.78	.74	.74	.78	.77	.73	.78	.77	.94
	1.0	.95	.94	.92	.95	.95	.92	.95	.95	.99
45	0.4	.25	.21	.30	.28	.25	.30	.29	.25	.47
	0.6	.60	.54	.63	.63	.60	.63	.64	.60	.83
	0.8	.88	.85	.88	.89	.87	.88	.89	.88	.97
	1.0	.98	.98	.98	.99	.99	.98	.99	.99	1
60	0.4	.26	.22	.33	.31	.27	.33	.31	.27	.49
	0.6	.60	.55	.67	.65	.61	.67	.66	.61	.84
	0.8	.91	.88	.92	.92	.91	.92	.92	.91	.98
	1.0	.99	.98	.99	.99	.99	.99	.99	.99	1
75	0.4	.25	.22	.31	.28	.25	.31	.29	.26	.46
	0.6	.58	.53	.63	.62	.58	.63	.63	.59	.83
	0.8	.88	.85	.89	.89	.88	.88	.89	.88	.98
	1.0	.98	.98	.98	.99	.98	.98	.99	.98	1
90	0.4	.19	.17	.21	.21	.19	.21	.21	.19	.37
	0.6	.46	.41	.46	.47	.45	.46	.47	.45	.71
	0.8	.77	.71	.73	.76	.75	.73	.76	.75	.93
	1.0	.95	.93	.92	.94	.94	.91	.94	.94	.99
105	0.4	.11	.11	.11	.11	.11	.11	.11	.11	.24
	0.6	.21	.20	.17	.19	.20	.17	.19	.20	.47
	0.8	.43	.41	.32	.39	.41	.32	.38	.40	.74
	1.0	.67	.65	.50	.62	.64	.50	.62	.64	.91

https://doi.org/10.1017/S0266466602184027 Published online by Cambridge University Press

TABLE 2. Power for Model 2

Alternative Distribution		Test Statistic								
τ	a_0	sup15-F	sup2-F	avg-F	exp-F_1	exp-F_∞	LR_0	LR_1	LR_∞	F
15	0.4	.10	.10	.12	.12	.11	.12	.12	.11	.16
	0.6	.19	.17	.24	.23	.21	.24	.23	.21	.33
	0.8	.31	.30	.40	.40	.36	.40	.40	.37	.56
	1.0	.50	.47	.58	.58	.54	.57	.58	.55	.76
30	0.4	.12	.11	.14	.14	.13	.14	.14	.13	.19
	0.6	.21	.18	.27	.25	.22	.27	.25	.23	.38
	0.8	.36	.32	.43	.42	.38	.43	.42	.38	.62
	1.0	.56	.52	.61	.61	.57	.61	.61	.58	.81
45	0.4	.08	.08	.09	.09	.09	.09	.09	.09	.17
	0.6	.15	.13	.16	.16	.15	.16	.16	.15	.31
	0.8	.27	.23	.25	.27	.26	.25	.27	.26	.52
	1.0	.44	.39	.38	.42	.42	.38	.42	.42	.73
60	0.4	.07	.07	.08	.08	.07	.08	.08	.07	.15
	0.6	.13	.11	.10	.12	.12	.10	.12	.12	.31
	0.8	.21	.19	.15	.18	.19	.14	.18	.19	.49
	1.0	.37	.33	.22	.29	.33	.21	.29	.33	.69
75	0.4	.08	.07	.09	.09	.08	.09	.09	.08	.17
	0.6	.15	.13	.16	.16	.15	.16	.16	.15	.34
	0.8	.26	.23	.25	.27	.26	.25	.27	.26	.54
	1.0	.44	.39	.39	.43	.42	.39	.42	.42	.74
90	0.4	.10	.09	.14	.13	.11	.14	.13	.11	.18
	0.6	.20	.18	.26	.25	.22	.26	.25	.22	.38
	0.8	.35	.30	.43	.40	.36	.43	.41	.37	.60
	1.0	.56	.51	.62	.62	.58	.63	.62	.59	.81
105	0.4	.10	.09	.12	.11	.10	.12	.11	.10	.17
	0.6	.18	.16	.23	.22	.20	.23	.22	.20	.35
	0.8	.33	.30	.40	.40	.37	.40	.40	.37	.55
	1.0	.51	.48	.60	.60	.56	.60	.60	.57	.76

the critical values. The powers reported in Tables 1–3, for different values of τ, are based on 5,000 replications.

From Tables 1 and 2 it appears that the LR_c tests and the exponential tests of Andrews et al. (1996) have identical powers for a medium large sample of 120 observation for both Model 1 and Model 2. For a smaller sample size, such as that of Model 3, the differences between the powers of the tests are still very small (see Table 3). No test appears to be uniformly better than the others.

In comparison with the sup F test, the LR tests and those of Andrews et al. (1996) do not perform significantly better unless the structural change occurs outside the range of possible change points implicitly considered in the construction of the sup F test.

https://doi.org/10.1017/S0266466602184027 Published online by Cambridge University Press

TABLE 3. Power of tests for structural change for Model 3

Alternative Distribution		Test Statistic								
τ	a_0	sup15-F	sup3-F	avg-F	exp-F_1	exp-F_∞	LR_0	LR_1	LR_∞	F
10	0.5	.05	.05	.05	.05	.05	.06	.05	.05	.07
	1	.07	.08	.09	.09	.08	.09	.09	.08	.17
	1.5	.10	.15	.14	.16	.16	.14	.15	.16	.35
	2	.14	.30	.22	.29	.30	.21	.26	.30	.61
20	0.5	.07	.06	.07	.07	.06	.07	.07	.06	.08
	1	.18	.13	.17	.16	.14	.17	.17	.15	.23
	1.5	.37	.26	.33	.31	.28	.32	.33	.29	.48
	2	.62	.48	.55	.54	.51	.54	.56	.52	.76
30	0.5	.06	.06	.06	.06	.06	.06	.06	.06	.07
	1	.07	.08	.09	.09	.08	.09	.09	.08	.16
	1.5	.11	.16	.15	.15	.16	.15	.15	.15	.34
	2	.16	.27	.23	.27	.27	.23	.25	.27	.58

5. CONCLUSIONS

This paper has derived optimal similar tests for structural change in the linear model under the hypothesis of normality of errors. This was done by first characterizing the class of similar tests using a result from Hillier (1987) and then deriving an average optimal test following the principle suggested by Wald (1943), Hillier (1987), Andrews and Ploberger (1994), and Andrews et al. (1996).

The tests for structural change suggested are analogous to those of Andrews et al. (1996). The weighting function employed is slightly different from that of Andrews et al. (1996). However, we are able to derive optimal tests for the case of unknown error variance and also are able to prove that the exponential F tests of Andrews et al. (1996) (for the case where the error variance is unknown) are optimal according to our criteria when the sample size is large.

A simulation study shows that the tests suggested have power very similar to those of Andrews et al. (1996) for the case where the disturbance variance is unknown. This suggests that the latter tests might be close to optimal even when the sample size is relatively small.

REFERENCES

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

Andrews, D.W.K., I. Lee, & W. Ploberger (1996) Optimal changepoint tests for normal linear regression. *Journal of Econometrics* 70, 9–36.

https://doi.org/10.1017/S0266466602184027 Published online by Cambridge University Press

Andrews, D.W.K., & W. Ploberger (1994) Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* 62, 1383–1414.
Breusch, T.S. (1986) Hypothesis testing in unidentified models. *Review of Economic Studies* 53, 635–651.
Brown, R.L., J. Durbin, & J.M. Evans (1975) Techniques for testing for constancy of regression relationships over time. *Journal of the Royal Statistical Society, Series B* 37, 149–192.
Chikuse, Y. (1991) High dimensional limit theorems and marix decompositions on the Stiefel manifold. *Journal of Multivariate Analysis* 36, 145–162.
Chow, G.C. (1960) Tests of equality between sets of coefficients in two linear regressions. *Econometrica* 28, 591–605.
Council Of Economic Advisers (1999) *Economic report of the president*.
Greene, W.H. (1991) *Econometric Analysis*. New York: Macmillan.
Hansen, B.E. (2000) Testing for structural change in conditional models. *Journal of Econometrics* 97, 93–115.
Hillier, G.H. (1987) Classes of similar regions and their power properties for some econometric testing problems. *Econometric Theory* 3, 1–44.
Pesaran, M.H., R.P. Smith, & J.S. Yeo (1984) Testing for structural stability and predictive failure: A review. *The Manchester School of Economics and Social Studies* 52, 280–295.
Ploberger, W., W. Krämer, & K. Kontrus (1989) A new test for structural stability in the linear regression model. *Journal of Econometrics* 40, 307–318.
Quandt, R.E. (1960) Tests of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American Statistical Association* 53, 324–330.
Wald, A. (1943) Tests of statistical hypotheses concerning several parameters when the number of observations is large. *American Mathematical Society Transactions* 54, 426–482.

APPENDIX

This Appendix contains one Lemma and the proofs of Theorem 2 and Corollary 1.

LEMMA 1. *(Chikuse, 1991). Any vector $v \in S_n$ can be written uniquely as*

$$v = \cos\theta C_1 v_1 + \sin\theta C_2 v_2, \tag{A.1}$$

where $v_1 \in S_k$, $v_2 \in S_{n-k}$, and $0 < \theta < \pi/2$; C_1 and C_2 are $(n \times k)$ and $(n \times n-k)$ matrices satisfying $C_1'C_1 = I_k$, $C_2'C_2 = I_{n-k}$, and $C_1'C_2 = 0$; and *θ is the angle between the vector v and the orthogonal projection of v onto the subspace spanned by the columns of C_1. The Jacobian of the transformation $v \to (\theta, v_1, v_2)$ is*

$$(dv) = (\cos\theta)^{k-1}(\sin\theta)^{n-k-1}(dv_1)(dv_2)\,d\theta, \tag{A.2}$$

where (dv) denotes the unnormalized Haar measure on S_n.

Proof of Theorem 2. The unconditional power of a similar critical region ω of size α is (Hillier, 1987)

https://doi.org/10.1017/S0266466602184027 Published online by Cambridge University Press

$$P_\omega = \frac{1}{2\pi^{(T-p)/2}} \exp\left\{-\frac{1}{2}\lambda_\tau\right\} \int_\omega \sum_{j=0}^{\infty} \frac{\Gamma\left(\frac{T-p+j}{2}\right) 2^{j/2}}{j!} (\phi' \Lambda_\tau' v)^j (dv), \tag{A.3}$$

where

$$\Lambda_\tau = C' Z(\tau)(Z(\tau)' M_X Z(\tau))^{-(1/2)}, \tag{A.4}$$

$$\phi = (Z(\tau)' M_X Z(\tau))^{1/2} \eta, \tag{A.5}$$

$$\lambda_\tau = \phi_\tau' \phi_\tau = \gamma' Z(\tau)' M_X Z(\tau) \gamma / \sigma^2. \tag{A.6}$$

By averaging this on the surface of constant λ_τ, we obtain the average power

$$\frac{\Gamma\left(\frac{T-p}{2}\right)}{2\pi^{(T-p)/2}} \int_\omega \exp\left\{-\frac{1}{2}\lambda_\tau\right\} {}_1F_1\left(\frac{T-p}{2}; \frac{K}{2}; \frac{\lambda_\tau}{2} v' \Lambda_\tau \Lambda_\tau' v\right)(dv), \tag{A.7}$$

and by further averaging it over the set of all possible structural points, Υ, this becomes

$$\frac{\Gamma\left(\frac{T-p}{2}\right)}{2\pi^{(T-p)/2}} \int_\omega \sum_{\tau \in \Upsilon} p(\tau) \exp\left\{-\frac{1}{2}\lambda_\tau\right\} {}_1F_1\left(\frac{T-p}{2}; \frac{K}{2}; \frac{\lambda_\tau}{2} v' \Lambda_\tau \Lambda_\tau' v\right)(dv). \tag{A.8}$$

By using Lemma 1 v can be written uniquely as $v = \cos\theta_\tau \Lambda_\tau v_{1\tau} + \sin\theta_\tau \Lambda_\tau^\perp v_{2\tau}$, where $v_{1\tau} \in S_K$, $v_{2\tau} \in S_{T-p-K}$, $0 < \theta_\tau < \pi/2$, and $\Lambda_\tau^\perp$ is a $T - p \times T - p - K$ matrix orthogonal to Λ_τ such that $(\Lambda_\tau^\perp)' \Lambda_\tau^\perp = I_{T-p-K}$. Because $v_{1\tau}$ and $v_{2\tau}$ are uniformly distributed over S_K and S_{T-p-K}, respectively, the most powerful critical region must contain all such spheres and the test is characterized in terms of $\cos\theta_\tau$, $\tau \in \Upsilon$, only:

$$\frac{2\Gamma\left(\frac{T-p}{2}\right)}{\Gamma\left(\frac{K}{2}\right)\Gamma\left(\frac{T-p-K}{2}\right)} \sum_{\tau \in \Upsilon} p(\tau) \exp\left\{-\frac{1}{2}\lambda_\tau\right\}$$

$$\times \int_\omega (\cos\theta_\tau)^{K-1} (\sin\theta_\tau)^{T-p-K-1} {}_1F_1\left(\frac{T-p}{2}; \frac{K}{2}; \frac{\lambda_\tau}{2} \cos^2\theta_\tau\right) d\theta_\tau. \tag{A.9}$$

Averaging with respect to λ_τ in (A.9), where the weight for $\lambda_\tau / \sqrt{c}$ is the density function for a random variable having the chi-square distribution with K degrees of freedom, yields

https://doi.org/10.1017/S0266466602184027 Published online by Cambridge University Press

$$\bar{P}_{\omega} = \frac{2\Gamma\left(\frac{T-p}{2}\right)}{\Gamma\left(\frac{K}{2}\right)\Gamma\left(\frac{T-p-K}{2}\right)(1+c)^{K/2}} \times \sum_{\tau\in\Upsilon} p(\tau) \int_{\omega} \frac{(\cos\theta_{\tau})^{K-1}(\sin\theta_{\tau})^{T-p-K-1}}{\left(1-\frac{c}{1+c}\cos^2\theta_{\tau}\right)^{(T-p)/2}} d\theta_{\tau}.$$

The statement of the theorem follows. ■

Proof of Corollary 1. The result is proved by showing that LR_c converges pointwise to exp-F$_c$ as T goes to infinity. If some of the f_{τ} are infinite then both exp-F$_c$ and LR_c are infinite. So we can suppose that the f_{τ} are all finite. The term

$$\exp\left\{\frac{K}{2}\frac{c}{1+c}f_{\tau}\right\} - \left(\frac{1+\frac{K}{T-p-K}f_{\tau}}{1+\frac{K}{(1+c)(T-p-K)}f_{\tau}}\right)^{(T-p)/2}$$

can be rewritten as

$$\exp\{b_{\tau}\} - \left(1+\frac{1}{a_{\tau}+\frac{1}{b_{\tau}}q}\right)^{q},$$

where $b_{\tau} = (K/2)[c/(1+c)]f_{\tau}$, $a_{\tau} = 1 - [(1+c)/cf_{\tau}] = 1 - (K/2b_{\tau})$, and $q = (T-p)/2 > K/2$. By setting $s = a_{\tau} + (1/b_{\tau})q$ we have $q = b_{\tau}(s-a_{\tau}) = b_{\tau}s - a_{\tau}b_{\tau} = b_{\tau}(s-1) + K/2$, and using the triangular inequality we have

$$\left|\exp\{b_{\tau}\} - \left(1+\frac{1}{s}\right)^{b_{\tau}(s-1)+(K/2)}\right| \leq \left|\exp\{b_{\tau}\} - \left(1+\frac{1}{s}\right)^{b_{\tau}s}\right| + \left|\left(1+\frac{1}{s}\right)^{b_{\tau}s}\right|\left|1-\left(1+\frac{1}{s}\right)^{(K/2)-b_{\tau}}\right|.$$

Note that

$$\left|\exp\{b_{\tau}\} - \left(1+\frac{1}{s}\right)^{b_{\tau}s}\right| = \exp\{b_{\tau}\} - \left(1+\frac{1}{s}\right)^{b_{\tau}s}$$

$$\left|\left(1+\frac{1}{s}\right)^{b_{\tau}s}\right| = \left(1+\frac{1}{s}\right)^{b_{\tau}s} \leq \exp\{b_{\tau}\}$$

$$\left|1-\left(1+\frac{1}{s}\right)^{(K/2)-b_{\tau}}\right| \leq \left|1-\left(1+\frac{1}{s}\right)^{K/2}\right| + \left|\left(1+\frac{1}{s}\right)^{K/2}\right|\left|1-\left(1+\frac{1}{s}\right)^{-b_{\tau}}\right|$$

$$= \left[\left(1+\frac{1}{s}\right)^{K/2} - 1\right] + \left(1+\frac{1}{s}\right)^{K/2}\left(1-\left(1+\frac{1}{s}\right)^{-b_{\tau}}\right)$$

https://doi.org/10.1017/S0266466602184027 Published online by Cambridge University Press

and all three terms are increasing in b_τ, so that

$$\left|\exp\{b_\tau\} - \left(1+\frac{1}{s}\right)^{b_\tau(s-1)+(K/2)}\right| \leq \exp\{b^*\} - \left(1+\frac{1}{s}\right)^{b^*s} + \exp\{b^*\}\left[\left(1+\frac{1}{s}\right)^{K/2} - 1\right] + \exp\{b^*\}\left(1+\frac{1}{s}\right)^{K/2}\left(1 - \left(1+\frac{1}{s}\right)^{-b^*}\right),$$

where $b^* = (K/2)[c/(1+c)]\sup_{\tau\in\Upsilon} f_\tau < \infty$. So

$$|\exp - \mathrm{F}_c - LR_c| \leq (1+c)^{-(K/2)} \sum_{\tau\in\Upsilon} p(\tau)\left|\exp\{b_\tau\} - \left(1+\frac{1}{s}\right)^{b_\tau(s-1)+(K/2)}\right|$$

$$\leq (1+c)^{-(K/2)}\left[\exp\{b^*\} - \left(1+\frac{1}{s}\right)^{b^*s} + \exp\{b^*\}\left[\left(1+\frac{1}{s}\right)^{K/2} - 1\right] + \exp\{b^*\}\left(1+\frac{1}{s}\right)^{K/2}\left(1 - \left(1+\frac{1}{s}\right)^{-b^*}\right)\right].$$

As T (and thus s) goes to infinity all the terms in square brackets go to zero. ■

https://doi.org/10.1017/S0266466602184027 Published online by Cambridge University Press