COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale University
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1016

OPTIMAL CHANGEPOINT TESTS FOR NORMAL LINEAR REGRESSION

by

Donald W.K. Andrews, Inpyo Lee and Werner Ploberger

April 1992

OPTIMAL CHANGEPOINT TESTS FOR NORMAL LINEAR REGRESSION

Donald W. K. Andrews
Cowles Foundation for Research in Economics
Yale University

Inpyo Lee
Cowles Foundation for Research in Economics
Yale University

Werner Ploberger[1]
Cowles Foundation for Research in Economics
Yale University
and
Technische Universität, Wien

April 1992

SUMMARY

This paper determines a class of finite sample optimal tests for the existence of a changepoint at an unknown time in a normal linear multiple regression model with known variance. Optimal tests for multiple changepoints are also derived. Power comparisons of several tests are provided based on simulations.

JEL Classification No. 211.

Keywords: Changepoint test, linear regression, multiple changepoints, optimal test, structural change test.

1. INTRODUCTION

This paper derives a class of finite sample optimal tests for one or more changepoints at unknown times in a multiple linear regression model. The regression model has fixed regressors and normal errors with known variance. Under the null hypothesis, the regression parameter vector is constant across all time periods. Under the alternative, a sub–vector of the regression parameter changes at m or fewer unknown points in time for some integer $m \geq 1$. The optimal tests are given by exponential averages of the Wald test statistics that apply when the changepoints are known. The commonly used likelihood ratio test is not in the class of optimal tests.

The testing situation considered in this paper is one in which a nuisance parameter (the vector of changepoints) appears under the alternative hypothesis but not under the null hypothesis. (For asymptotic analyses of such problems, see Davies (1977, 1987), Andrews and Ploberger (1991), Hansen (1991), and King and Shively (1991).) It is well known that standard optimality results for tests, both finite sample and asymptotic, do not apply in testing situations of the type described above. For example, Wald's (1942) finite sample average power optimality results for tests of linear restrictions in normal linear regression or analysis of variance models do not apply. In this paper, we extend the Wald (1942)–type optimality results to cover the case of tests with unknown changepoints. We consider a weighted average power criterion function. We obtain a class of test statistics that are indexed by a scalar measure c of the magnitude of the parameter changes against which the test's power is directed. For the special case of a known changepoint vector, our test statistics reduce to the standard optimal test statistic for this case.

The changepoint tests introduced here can be used to test the null hypothesis of parameter constancy against the alternative of multiple parameter shifts at unknown times or, more generally, against parameter changes of a less specific nature. Alternatives of the latter sort are often of interest in econometrics. Test consistency results of Andrews (1989) suggest that the tests will have asymptotic power against a very broad range of alternatives. In the second part of this paper, we assess the power of the optimal tests against alternatives of both types. In particular, we consider one–time parameter changes and martingale parameter changes. Of prime interest are (1) the sensitivity of the power of the optimal tests to the scalar measure c mentioned above and (2) the relative power of the optimal tests to other tests in the literature such as the likelihood ratio test, the midpoint F test, the CUSUM test of Brown, Durbin, and Evans (1975), and a test introduced by Nyblom (1989) for martingale parameter changes. In brief, we find that the power of the optimal tests is not very sensitive to changes in c. We suggest the use of $c = \infty$. We also find that the optimal tests perform quite well in finite samples vis–a–vis the other tests considered, both for the alternatives for which they are designed and for the other alternatives considered.

Previous work on changepoint problems of the sort considered here includes, among others, Chernoff and Zacks (1964), Hinckley (1969), Gardner (1969), Farley and Hinich (1970), James, James, and Siegmund (1987), Kim and Siegmund (1989), Andrews (1989), and Jandhyala and MacNeill (1991).

The remainder of this paper is organized as follows. Section 2 defines the optimal test statistics and establishes their optimality. Section 3 discusses the case where the error variance σ^2 is unknown. Section 4 presents the Monte Carlo power comparisons.

2. OPTIMAL TESTS FOR REGRESSION WITH NORMAL ERRORS

In this section, we establish finite sample optimality results for a class of tests of structural change for the linear regression model with normal errors.

ASSUMPTION 1. *The model is given by*

$$Y_t = \begin{cases} X_t'\delta_1 + Z_t'\delta_2 + U_t & \text{for } t = 1, \ldots T\pi_1 \\ X_t'(\delta_1 + \beta_1) + Z_t'\delta_2 + U_t & \text{for } t = T\pi_1+1, \ldots, T\pi_2 \\ \vdots & \\ X_t'(\delta_1 + \beta_m) + Z_t'\delta_2 + U_t & \text{for } t = T\pi_m+1, \ldots, T \end{cases}$$

for some $\pi = (\pi_1, ..., \pi_m)' \subset \Pi$, *where* $U_t \sim \text{iid } N(0, \sigma^2)$, σ^2 *is known and hence* wlog $\sigma^2 = 1$, $X_t, \delta_1, \beta_1, ..., \beta_m \in R^v$, $Z_t, \delta_2 \in R^w$, $\{(X_t, Z_t) : t = 1, ..., T\}$ *are non-random,* $\Pi \subset \{(\pi_1, ..., \pi_m)' : 0 < \pi_1 < \pi_2 < \cdots < \pi_m < 1,\ T\pi_j$ *is an integer, and* $\Sigma_{T\pi_j+1}^{T\pi_{j+1}} X_t X_t'$ *is full rank* v $\forall j = 0, 1, ..., m$, *where* $\pi_0 = 0$ *and* $\pi_{m+1} = 1\}$, *and* $\Sigma_1^T Z_t Z_t'$ *is full rank* w.

The unknown parameters of the model are given by

$$\theta = (\beta_1', ..., \beta_m', \delta_1', \delta_2')' \in R^s, \text{ where } s = (m+1)v + w\,. \tag{2.1}$$

The null and alternative hypotheses of interest are:

$$\begin{aligned} &H_0 : \beta_1 = \beta_2 = \cdots = \beta_m = 0 \text{ and} \\ &H_1 : \beta_j \neq 0 \text{ for some } j = 1, ..., m \text{ and } \pi \in \Pi \text{ for some specified set } \Pi. \end{aligned} \tag{2.2}$$

When the term $Z_t'\delta_2$ appears in the regression function, only the parameter vector for the X_t regressor is subject to change under the alternative H_1. This case is referred to as a test of *partial structural change.* When no term $Z_t'\delta_2$ appears in the regression function under H_0 or H_1, the parameter vector for the entire regressor vector is subject to change under H_1. This case is called a test of *pure structural change.* Note that the error variance $\sigma^2 = 1$ is taken to be constant across the observations in either case.

To determine an optimal test of H_0 versus H_1, we use a weighted average power criterion. We specify a weight function $J(\cdot)$ over the discrete set Π of changepoint vectors. For each changepoint vector $\pi \in \Pi$, we specify a weight function $Q_\pi(\cdot)$ over the different parameter vectors θ that lie in the alternative hypothesis H_1. A test of level α is said to be optimal for weight functions $(J(\cdot), Q_\pi(\cdot))$ if its weighted average power exceeds that of any other test of level α. This criterion is quite similar to that used by Wald (1942) to obtain classical optimality results for tests of linear restrictions in the linear regression model. In fact, for any fixed value of π, the weight function we consider has the same contours as that considered by Wald. Further, the optimal tests introduced here reduce to the standard Wald test when the changepoint vector is specified, i.e., $\Pi = \{\pi^*\}$.

Let θ_0 denote some parameter vector in the null hypothesis:

$$\theta_0 = (0', \delta_{10}', \delta_{20}')' \in R^S \text{ for some } \delta_{10} \in R^V \text{ and } \delta_{20} \in R^W . \tag{2.3}$$

Any parameter vector $\theta \in R^S$ can be written as the sum of the null parameter vector θ_0 and some perturbation vector $h \in R^S$. That is,

$$\theta = \theta_0 + h . \tag{2.4}$$

Given θ_0 and change point π, we specify a weight function $Q_\pi(\cdot)$ over perturbation vectors h. The weight function we use is a singular multivariate normal distribution whose support lies in the orthogonal complement (with respect to a particular inner product) of the linear subspace defined by the null hypothesis.

More specifically, let V denote the linear subspace of R^S defined by

$$V = \{\theta \in R^S : \theta = (0', \delta_1', \delta_2')' \text{ for some } \delta_1 \in R^V \text{ and } \delta_2 \in R^W\} . \tag{2.5}$$

The null hypothesis can be expressed as $H_0 : \theta \in V$. For given $\pi \in \Pi$, define the inner product

$<h, \ell>_\pi = h'\mathcal{I}(\pi)\ell$ for h, $\ell \in R^s$, where $\mathcal{I}(\pi) = \begin{bmatrix} \mathcal{I}_{1\pi} & \mathcal{I}_{2\pi} \\ \mathcal{I}'_{2\pi} & \mathcal{I}_3 \end{bmatrix}$,

$$\mathcal{I}_{1\pi} = \begin{bmatrix} \Sigma_{T\pi_1+1}^{T\pi_2} X_t X_t' & & & O \\ & \Sigma_{T\pi_2+1}^{T\pi_3} X_t X_t' & \ddots & \\ O & & & \Sigma_{T\pi_m+1}^{T} X_t X_t' \end{bmatrix} \in R^{mv \times mv}, \tag{2.6}$$

$$\mathcal{I}'_{2\pi} = \begin{bmatrix} \Sigma_{T\pi_1+1}^{T\pi_2} X_t X_t' & \Sigma_{T\pi_2+1}^{T\pi_3} X_t X_t' & \cdots & \Sigma_{T\pi_m+1}^{T} X_t X_t' \\ \Sigma_{T\pi_1+1}^{T\pi_2} Z_t X_t' & \Sigma_{T\pi_2+1}^{T\pi_3} Z_t X_t' & \cdots & \Sigma_{T\pi_m+1}^{T} Z_t X_t' \end{bmatrix} \in R^{(v+w)\times mv}, \text{ and}$$

$$\mathcal{I}_3 = \begin{bmatrix} \Sigma_1^T X_t X_t' & \Sigma_1^T X_t Z_t' \\ \Sigma_1^T Z_t X_t' & \Sigma_1^T Z_t Z_t' \end{bmatrix} \in R^{(v+w)\times(v+w)}.$$

Note that $\mathcal{I}(\pi)$ is the information matrix for the parameter θ when the change vector is fixed at π. Denote the orthogonal complement of V under $<\cdot,\cdot>_\pi$ by $V_\pi^\perp$. Since V is a v+w dimensional subspace of R^s, $V_\pi^\perp$ is an mv dimensional subspace of R^s. Let $\{a_{1,\pi}, \ldots, a_{mv,\pi}\}$ be some basis of $V_\pi^\perp$ and define

$$A_\pi = [a_{1,\pi} \cdots a_{mv,\pi}] \in R^{s\times mv}. \tag{2.7}$$

For example, one can take

$$A_\pi = \begin{bmatrix} I_{mv} \\ -\mathcal{I}_3^{-1}\mathcal{I}'_{2\pi} \end{bmatrix}. \tag{2.8}$$

Next, for any choice of A_π as in (2.7), let

$$\begin{aligned} \Sigma_\pi &= A_\pi(A_\pi'\mathcal{I}(\pi)A_\pi)^{-1}A_\pi' \\ &= \begin{bmatrix} (\mathcal{I}_{1\pi}-\mathcal{I}_{2\pi}\mathcal{I}_3^{-1}\mathcal{I}'_{2\pi})^{-1} & -(\mathcal{I}_{1\pi}-\mathcal{I}_{2\pi}\mathcal{I}_{3\pi}^{-1}\mathcal{I}'_{2\pi})^{-1}\mathcal{I}_{2\pi}\mathcal{I}_3^{-1} \\ -\mathcal{I}_3^{-1}\mathcal{I}'_{2\pi}(\mathcal{I}_{1\pi}-\mathcal{I}_{2\pi}\mathcal{I}_3^{-1}\mathcal{I}'_{2\pi})^{-1} & \mathcal{I}_3^{-1}\mathcal{I}'_{2\pi}(\mathcal{I}_{1\pi}-\mathcal{I}_{2\pi}\mathcal{I}_3^{-1}\mathcal{I}'_{2\pi})^{-1}\mathcal{I}_{2\pi}\mathcal{I}_3^{-1} \end{bmatrix}. \end{aligned} \tag{2.9}$$

Also, let $N(0,\Sigma)$ denote a multivariate normal distribution with mean 0 and covariance matrix Σ (possibly singular).

The weight function Q_π is given by

ASSUMPTION 2. $Q_\pi = N(0, c\Sigma_\pi)$ $\forall \pi \in \Pi$ *for some positive constant* c (*that does not depend on* π).

Under Assumption 2, the weight function Q_π on R^s is a singular multivariate normal distribution with covariance matrix of rank $mv < s$. The support of Q_π is $V_\pi^\perp$.

The constant c, which scales the variance matrix of the weight function Q_π, determines whether one gives higher weight to alternatives that are close to the null or distant from the null. Thus, a small value of c corresponds to giving relatively high weight to small parameter changes $(\beta_1, ..., \beta_m)$ and a large value of c corresponds to giving relatively high weight to large parameter changes $(\beta_1, ..., \beta_m)$.

The weighted average power criterion can now be expressed as

$$\sum_{\pi \in \Pi} \int P(\varphi \text{ rejects } H_0 | \theta_0 + h, \pi) dQ_\pi(h) J(\pi) , \tag{2.10}$$

where φ is some level α test and $\{J(\pi) : \pi \in \Pi\}$ is a discrete distribution on Π. For example, one might choose a uniform weight function: $J(\pi) = 1/\text{card}(\Pi)$ $\forall \pi \in \Pi$. An optimal test of level α maximizes the above weighted average power criterion over all tests of level α. (We show below that the optimal test does not depend on the choice of θ_0.)

We determine an optimal test as follows. Let $f_T(y,\theta,\pi)$ denote the density of the T vector of observations $Y = (Y_1, ..., Y_T)'$ evaluated at $y = (y_1, ..., y_T)'$. Note that $f_T(y, \theta_0, \pi)$ does not depend on π for $\theta_0 \in V$ by Assumption 1 and, hence, is denoted $f_T(y, \theta_0)$. Let $\varphi = \varphi(Y)$ denote a test of H_0. That is, $\varphi(Y)$ is a [0,1]–valued function of Y that rejects H_0 with probability γ when $\varphi(Y) = \gamma$. The power of φ against $\theta = \theta_0 + h$ is given by $\int \varphi(y) f_T(y, \theta_0 + h, \pi) dy$. The weighted average power of φ equals

$$\sum_{\pi\in\Pi} \int P(\varphi \text{ rejects } H_0 | \theta_0 + h, \pi) dQ_\pi(h) J(\pi)$$

$$= \sum_{\pi\in\Pi} \int\int \varphi(y) f_T(y, \theta_0 + h, \pi) dy dQ_\pi(h) J(\pi) \tag{2.11}$$

$$= \int \varphi(y) \left[\sum_{\pi\in\Pi} \int f_T(y, \theta_0 + h, \pi) dQ_\pi(h) J(\pi) \right] dy .$$

Equation (2.11) shows that the weighted average power of φ equals the power of φ against the single alternative density specified by

$$g_T(y, \theta_0) = \sum_{\pi\in\Pi} \int f_T(y, \theta_0 + h, \pi) dQ_\pi(h) J(\pi) . \tag{2.12}$$

Hence, a test that maximizes power against the simple alternative $g_T(\cdot, \theta_0)$ also maximizes weighted average power.

The Neyman–Pearson Lemma shows that the best test for testing the simple null $Y \sim f_T(\cdot, \theta_0)$ against the simple alternative $Y \sim g_T(\cdot, \theta_0)$ is based on the likelihood ratio statistic LR defined by

$$LR = g_T(Y, \theta_0)/f_T(Y, \theta_0) = \sum_{\pi\in\Pi} \int f_T(Y, \theta_0 + h, \pi) dQ_\pi(h) J(\pi)/f_T(Y, \theta_0) . \tag{2.13}$$

(Note that this likelihood ratio statistic is not the likelihood ratio commonly considered in the literature. The latter is the ratio $\max_{\pi\in\Pi} \sup_{\theta\in R^s} f_T(Y, \theta, \pi)/(\sup_{\theta\in V} f_T(Y, \theta))$.) We show below that this statistic does not depend on θ_0 and its distribution is the same for all parameter vectors in the null hypothesis. In consequence, (a) an exact similar test of $H_0 : Y \sim f_T(\cdot, \theta_0)$ for some $\theta_0 \in V$ can be constructed based on LR, (b) this test has maximum power among tests of the same level against the alternative $g_T(\cdot, \theta_0)$ for any $\theta_0 \in V$, and (c) this test has maximum weighted average power as defined in (2.10) for any $\theta_0 \in V$ among tests of the same level. In addition, we show below that LR can be written in a simplified form that does not involve an integral with respect to $Q_\pi(\cdot)$. A consequence of this is that the test statistic is straightforward to compute.

In particular, the statistic LR is shown to equal the *exponential Wald statistic* Exp–W_c defined by

$$\text{Exp–W}_c = (1+c)^{-mv/2} \sum_{\pi\in\Pi} \exp\left[\frac{1}{2}\frac{c}{1+c}W(\pi)\right] J(\pi), \text{ where}$$

$$W(\pi) = \hat\beta(\pi)'(H\mathcal{I}(\pi)H')^{-1}\hat\beta(\pi),\ \hat\beta(\pi) = (\hat\beta_1(\pi)', \ldots, \hat\beta_m(\pi)')', \tag{2.14}$$

$$H = [I_{mv} \vdots 0] \in R^{mv\times s},$$

and $(\hat\beta_1(\pi)', \ldots, \hat\beta_m(\pi)', \hat\delta_1(\pi)', \hat\delta_2(\pi)')'$ are the least squares (LS) estimators of $(\beta_1', \ldots, \beta_m', \delta_1', \delta_2')'$ from the regression of Y_t on $(X_t 1(T\pi_1 < t \leq T\pi_2),\ X_t 1(T\pi_2 < t \leq T\pi_3), \ldots, X_t 1(T\pi_m < t \leq T), X_t, Z_t)$. Note that $W(\pi)$ is just the standard Wald test statistic for testing $H_0 : \beta_1 = \beta_2 = \cdots = \beta_m = 0$ against $H_1 : \beta_j \neq 0$ for some $j = 1, \ldots, m$ and the changepoint vector is π.

One rejects the null hypothesis for large values of Exp–W_c. The resulting test is exactly similar and maximizes weighted average power among tests of equal significance level. These results are summarized in the following theorem. Let $\xi(Y)$ denote a test of level α based on the exponential Wald statistic Exp–W_c.

THEOREM 1: *Suppose Assumptions* 1 *and* 2 *hold. Then,*

(a) $\text{LR} = \text{Exp–W}_c$,

(b) *the distribution of* Exp–W_c *under the null hypothesis is nuisance parameter free, and*

(c) *for any level* α *test* $\varphi(Y)$, $\sum_{\pi\in\Pi}\int \varphi(y)f_T(y, \theta_0 + h, \pi)dydQ_\pi(h)J(\pi)$

$\leq \sum_{\pi\in\Pi}\int \xi(y)f_T(y, \theta_0 + h, \pi)dydQ_\pi(h)J(\pi)\ \ \forall\theta_0 \in V.$

COMMENTS: 1. By Theorem 1(b), the exponential Wald test is an exactly similar test. Its distribution under the null depends on the regressors, however, so it is not possible to provide tables of exact critical values. Exact critical values can be obtained straightforwardly on a case by case basis by simulation. See Section 3 for further discussion of simulation methods. Tables of asymptotic critical values are provided in Andrews and Ploberger (1991) for the case of non–trending regressor variables and $m = 1$.

2. The exponential Wald statistic depends on the constant c that indexes whether power is directed at near or distant alternatives from the null hypothesis. In consequence, one needs to choose a suitable value of c for practical application of the test. Fortunately, the choice of c does not appear to be critical. The simulations reported in Section 4 below show that there is little sensitivity in the power of the exponential Wald test to the choice of c even when c is allowed to range over the entire interval $[0,\infty]$. For reasons outlined in Section 4, we suggest using the test statistic that corresponds to the limiting case as $c \to \infty$ for general purposes:

$$\text{Exp–W}_{\infty} = \lim_{c \to \infty} \log\left[(1+c)^{mv/2}\text{Exp–W}_c\right] = \log \sum_{\pi \in \Pi} \exp\left[\tfrac{1}{2}W(\pi)\right] J(\pi) . \qquad (2.15)$$

The lack of sensitivity of the power of Exp–W_c to the choice of c is not too surprising, because when the nuisance parameter π is known, i.e., $\Pi = \{\pi^*\}$, the power of the test is independent of c. Lastly, we mention that if one desires a finely–tuned choice of c, even though the test is insensitive to c, a method is described in Section 7 of Andrews and Ploberger (1991).

3. The test statistic Exp–W_c is designed for the alternative hypothesis specified in (2.2) that specifies a fixed number of changepoints m. If one wishes to consider alternatives with multiple changepoints of unknown number, then a test statistic can be obtained that is optimal with respect to weighted average power by placing a weight function (prior), say $\{p(m) : m = 1, \ldots, M\}$, over the number of changepoints. The test statistic is $\sum_{m=1}^{M} p(m)\text{Exp–W}_{mc}$, where Exp–$W_{mc}$ is the statistic Exp–W_c defined in (2.14) for the case of m changepoints.

3. UNKNOWN ERROR VARIANCE

The results of Section 2 apply when the variance σ^2 of the errors $\{U_t : t = 1,..., T\}$ is known. In practice, of course, σ^2 is rarely known. In consequence, one usually needs to replace the exponential Wald statistic defined in Section 2 with an analogous statistic that estimates σ^2. The natural way of doing so is to replace $W(\pi)$ by the standard Wald statistic for testing for change occurring at changepoint vector π when σ^2 is unknown. The latter is just mv times the usual F statistic $F(\pi)$:

$$F(\pi) = \hat{\beta}(\pi)' \left[\hat{\sigma}^2(\pi) H \mathcal{I}(\pi) H'\right]^{-1} \hat{\beta}(\pi)/(mv)$$
$$= \frac{[Q^* - Q(\pi)](T-s)}{Q(\pi)mv} \quad \text{, where}$$
$$Q(\pi) = \sum_{j=0}^{m} \Sigma_{T\pi_j+1}^{T\pi_{j+1}}[Y_t - X_t'(\hat{\delta}_1(\pi) + \hat{\beta}_j(\pi))]^2 \text{ for } \pi_0 = 0 \text{ and } \hat{\beta}_0 = 0 \text{ ,} \tag{3.1}$$
$$\hat{\sigma}^2(\pi) = Q(\pi)/(T-s) \text{ ,}$$
$$Q^* = \Sigma_1^T[Y_t - X_t'\tilde{\delta}_1 - Z_t'\tilde{\delta}_2]^2 \text{ ,}$$

$\hat{\beta}(\pi)$ is as in (2.14), and $(\tilde{\delta}_1', \tilde{\delta}_2')'$ is the LS estimator from the regression of Y_t on $(X_t', Z_t')'$.

The analogue of $Exp\text{–}W_c$ for the case of σ^2 unknown is then given by the *exponential* F *statistic* defined by

$$Exp\text{–}F_c = (1+c)^{-mv/2} \sum_{\pi \in \Pi} \exp\left[\frac{mv}{2} \frac{c}{1+c} F(\pi)\right] J(\pi) \text{ .} \tag{3.2}$$

As with $Exp\text{–}W_c$, the statistic $Exp\text{–}F_c$ is nuisance parameter free under the null hypothesis. In consequence, exact tests based on $Exp\text{–}F_c$ can be obtained by simulating the critical values. Since the exact critical values depend on the regressors, it is not feasible to provide tables of them. On the other hand, asymptotic critical values for $Exp\text{–}F_c$ are the same as those for $Exp\text{–}W_c$ and tables are provided in Andrews and Ploberger (1991) for the case of nontrending regressors and $m = 1$.

We found that it was fastest to compute the simulated critical values by using the formula $F(\pi) = ((T-s)/(Tmv))LM(\pi)/(1 - LM(\pi)/T)$, where $LM(\pi)$ is the Lagrange multiplier statistic defined in (4.1) below. The weight matrix of $LM(\pi)$ for each π only needs to be computed once and can be reused for each repetition in the simulation. In addition, the use of $LM(\pi)$ circumvents the need to reestimate the unrestricted model for each value of π. For example, using a Gateway 486 3033 megahertz PC it takes eleven minutes to generate one thousand repetitions for a model with six regressors and one hundred and twenty observations.

The finite sample optimality results of Section 2 do not apply to tests based on Exp–F_c. Nevertheless, for medium to large samples, the difference in power between the tests based on Exp–W_c and Exp–F_c is fairly small (see the Monte Carlo results of Section 4). In consequence, the results of Section 2 imply that the test based on Exp–F_c is *at least* quite close to being optimal in terms of finite sample weighted average power. Furthermore, the test based on Exp–F_c is asymptotically optimal according to the results of Andrews and Ploberger (1991).

Two limiting cases of the exponential F statistic as $c \to 0$ and $c \to \infty$ are given by

$$\begin{aligned} \text{Avg–F} &= \lim_{c \to 0} 2(\text{Exp–F}_c - 1)/(cmv) = \sum_{\pi \in \Pi} F(\pi)J(\pi) \text{ and} \\ \text{Exp–F}_\infty &= \lim_{c \to \infty} \log\left[(1+c)^{mv/2}\text{Exp–F}_c\right] = \log \sum_{\pi \in \Pi} \exp\left[\frac{mv}{2}F(\pi)\right]J(\pi) . \end{aligned} \tag{3.3}$$

(The non–data–dependent renormalizations of Exp–F_c, which are made prior to taking the limits in (3.3), guarantee non–degenerate limiting test statistics. The renormalization has no effect on the actual critical region of each of the tests in the sequence as $c \to 0$ or $c \to \infty$, however, and hence is appropriate.)

The test statistic Avg–F is designed to direct power against alternatives for which the magnitudes of the changes $(\beta_1, \ldots, \beta_m)$ are small. The test statistic Exp–F_∞, on the other hand, is designed to direct power against alternatives for which the magnitudes of the changes (β_1, ..., β_m) are large. The relative performance of these two tests is evaluated for a variety of alternatives in Section 4 via Monte Carlo simulation.

4. MONTE CARLO RESULTS

This section presents Monte Carlo results regarding the finite sample properties of the optimal tests introduced above. Attention is focused throughout on the statistics designed for *one–time* change alternatives for which m = 1. The section has several objectives. First, we analyze the sensitivity of the power of the optimal exponential tests to the choice of c for a variety of alternative distributions. Second, we compare the power of the optimal exponential tests with that of other tests in the literature such as the likelihood ratio or "sup F" test, the CUSUM test of Brown, Durbin, and Evans (1975), the midpoint F test, i.e., F(.5), and a test considered by Nyblom (1989) for the alternative where the parameters form a martingale. We make such comparisons both for a variety of one–time change alternatives and for a variety of martingale parameter alternatives.

Third, we compare the power of the exponential tests for the case where σ^2 is estimated to the case where σ^2 is known. If the difference is small, the optimality of the latter statistics implies near optimality of the former feasible statistics. Fourth, we compare the power of the optimal tests for one–time change alternatives with unknown changepoint to the power of the optimal tests for known changepoint, viz., the F test based on the correct changepoint. Such results show the cost of not knowing the correct changepoint. Fifth, we calculate and compare the nominal size of the tests mentioned above when asymptotic critical values are employed. These results are not of direct interest for linear regression models, because exact critical values can be obtained for these models. For nonlinear models, however, asymptotic critical values are needed for the asymptotically optimal exponential tests introduced in Andrews and Ploberger (1991) and these simulation results should be helpful in assessing the true size of such asymptotic tests.

We now define the test statistics that are considered in the Monte Carlo experiment. Avg–F and Exp–F_∞ are defined as in (3.3) with $J(\pi)$ uniform on {2/120, ..., 118/120}, where the sample size T is 120 and the number of regressors under the null is two. Exp–F_c for c = 1/3, 1, and 3 are the exponential F statistics defined in (3.2) with $J(\pi)$ as above. Sup2–F and Sup15–F equal $\sup_{\pi\in\{2/120,\ldots,118/120\}} F(\pi)$ and $\sup_{\pi\in\{18/120,\ldots,102/120\}} F(\pi)$, respectively, which are the likelihood ratio statistics for $\Pi \subset [.02, .98]$ and $\Pi \subset [.15, .85]$. The latter set Π is smaller than in the definitions of the exponential F tests for reasons of power described in Andrews (1989). Avg–LM, Exp–LM_∞, Exp–LM_c, Sup2–LM, and Sup15–LM are Lagrange multiplier (LM) versions of the tests described above that are defined with $F(\pi)$ replaced by $LM(\pi)/(mv)$, where mv = 2 in the present case and $LM(\pi)$ is the LM test statistic for testing for one–time change occurring at changepoint π (when the error variance σ^2 is unknown). By definition,

$$LM(\pi) = \Sigma_1^{T\pi}(Y_t - X_t'\tilde{\delta})X_t'\left[\left[\Sigma_1^{T\pi}X_tX_t'\right]^{-1} + \left[\Sigma_{T\pi+1}^{T}X_tX_t'\right]^{-1}\right]\Sigma_1^{T\pi}(Y_t - X_t'\tilde{\delta})X_t/\tilde{\sigma}^2, \text{ where}$$
$$\tilde{\delta} = \left[\Sigma_1^{T}X_tX_t'\right]^{-1}\Sigma_1^{T}X_tY_t \text{ and } \tilde{\sigma}^2 = \tfrac{1}{T-2}\Sigma_1^{T}(Y_t - X_t'\tilde{\delta})^2 . \tag{4.1}$$

Nyb–LM denotes a test statistic introduced by Nyblom (1989) for alternatives in which the regression parameters form a martingale (or, more specifically, a random walk in the present case) with constant innovation variance. Nyb–LM equals the Avg–F statistic defined in (3.3) with $F(\pi)$ replaced by $LM(\pi)$ and with $J(\pi)$ proportional to $\pi(1-\pi)$ for $\pi \in \{2/120, ..., 118/120\}$ and 0 elsewhere. Note that Nyb–LM is very close to, and in fact is asymptotically equivalent to, an optimal exponential test statistic for a one–time change alternative with a particular non–uniform choice of $J(\pi)$ and $c \to 0$.

Cusum denotes the Cusum statistic of Brown *et al.* (1975) defined, for example, as in Krämer and Sonnberger (1986, pp. 49–53, 59–61). The Cusum test is the best known parameter instability test in the literature for linear regression models. Another test that also has been suggested for use as a general test of parameter instability is the midpoint Chow F test, denoted here by F(.5).

The exponential F tests discussed above use an estimator of the error variance σ^2 rather than taking it as known. For comparative purposes, we consider Wald versions of each of these tests where the Wald tests take σ^2 as known. The latter tests satisfy the finite sample optimality results of Section 2. Avg–W is defined in the same way as Avg–F except with $W(\pi)$ in place of $F(\pi)$, where $W(\pi)$ is defined in (2.14). Exp–W_c for c = 1/3, 1, and 3 and Exp–W_∞ are defined as in (2.14) and (2.15), respectively, with $J(\pi)$ uniform over {2/120, ..., 118/120}.

For one–time change alternatives with changepoints π_0 = .075, .15, ..., .925, we consider the test statistics that yield the *envelope power function,* viz. $F(\pi_0)$ for π_0 as above. Power results for $F(\pi_0)$ show the power attainable if the true changepoint is known.

Next we describe the two basic models and the variety of alternatives that are considered in the Monte Carlo experiment. The models are:

$$\begin{aligned}&\text{Model S: } Y_t = X_t'\delta_t + U_t,\ X_t = (1, (-1)^t)',\ U_t \sim \text{iid } N(0, \sigma^2),\ t = 1, \ldots, 120,\\ &\text{Model TT: } Y_t = X_t'\delta_t + U_t,\ X_t = (1, t-(T+1)/2)' \text{ for } T = 120,\\ &\qquad U_t \sim \text{iid } N(0, \sigma^2),\ t = 1, \ldots, 120.\end{aligned} \tag{4.2}$$

Model S is a "stationary" model. Model TT is a time trend model. Under the null hypothesis, $\delta_t = \delta_0$ for some $\delta_0 \in R^2$ $\forall t = 1, \ldots, 120$. The following alternative distributions are considered:

$$\text{Alt-1}(\pi_0):\ \delta_t = \begin{cases} \delta_0 & \text{for } t = 1, \ \ldots, \ 120\pi_0 \\ \delta_0 + \beta & \text{for } t = 120\pi_0+1, \ \ldots, \ 120 \end{cases} \quad \text{for some } \delta_0, \beta \in R^2,$$

$$\text{Alt-MG1:}\ \delta_t = \delta_0 + \beta_t,\ \beta_t = \beta_{t-1} + \epsilon_t,\ \epsilon_t \sim \text{iid } N(\underset{\sim}{0}, \tau^2 I_2),\ \beta_0 = \underset{\sim}{0}, \text{ for some } \delta_0 \in R^2, \tag{4.3}$$

$$\text{Alt-MG2:}\ \delta_t = \delta_0 + \beta_t,\ \beta_t = \beta_{t-1} + \epsilon_t,\ \epsilon_t \sim \text{iid } N\left[\underset{\sim}{0}, \frac{\tau^2}{(t/120)(1-t/120)} I_2\right], \beta_0 = \underset{\sim}{0}, \text{ for some } \delta_0 \in R^2,$$

$$\text{Alt-MG}(\pi_0):\ \delta_t = \delta_0 + \beta_t,\ \beta_t = \underset{\sim}{0}\ \ \forall t \le 120\pi_0,\ \beta_t = \beta_{t-1} + \epsilon_t\ \ \forall t = 120\pi_0+1,\ldots,120,\ \epsilon_t \sim \text{iid } N(\underset{\sim}{0}, \tau^2 I_2),\ \beta_0 = \underset{\sim}{0}, \text{ for some } \delta_0 \in R^2,$$

Alt–1(π_0) is a one–time change at $t = 120\pi_0$ alternative. For the case of unknown π_0, this is the alternative for which the exponential tests are designed. Alt–MG1 and Alt–MG2 are commonly considered martingale (MG) parameter alternatives. The statistic Nyb–LM is designed for Alt–MG1. The statistic Avg–LM has some asymptotic optimality properties for Alt–MG2, see Nyblom (1989). Alt–MG(π_0) is a martingale structural change alternative. With this alternative, the linear regression model is properly specified up to time $t = 120\pi_0$ but thereafter it breaks down and its parameters drift according to a martingale specification.

The rejection probabilities of the tests described above do not depend on δ_0 under the null and under the above alternatives. Thus, we set $\delta_0 = (0, 0)'$ without loss of generality. The error variance σ^2 is set equal to one. For Alt–1(π_0), we consider $\pi_0 = .075$, .15, .3, .5, .7, .85, .925. For Alt–MG(π_0), we consider the same values of π_0 except .075. For Model S and Alt–1(π_0), the power of each of the tests except Cusum depends on β only through $\|\beta\|$.[2] In consequence, we take β proportional to $(1, 0)'$ in this case without loss of generality. We take $\|\beta\| = 4.8, 7.2, 9.6, 12.0$. For Model TT and Alt–1(π_0), the power of the tests depends on β through the length and the direction of β. In this case too we consider β proportional to $(1, 0)'$, which corresponds to a shift in intercept. We take $\|\beta\| = 9.6, 12.0, 14.4, 16.8$. For Alt–MG1, Alt–MG2, and Alt–MG(π_0), we consider several values of the martingale innovation standard deviation τ, viz., $\tau = .03, .05, .07, .1, .3$, $\tau = .02, .03, .04, .05, .1$, and $\tau = .05, .1, .5, 1.0$, respectively.

Power results for selected subsets of the test statistics described above are provided in Tables 1–4. Table 1 covers Model S and Alt–1(π_0). Table 2 covers Model TT and Alt–1(π_0). Table 3 covers Model S and Alt–MG1, Alt–MG2, and Alt–MG(π_0). Table 4 covers Model TT and Alt–MG1, Alt–MG2, and Alt–MG(π_0). In all cases, the reported power is for level .05 exact versions of the tests described above with critical values generated by Monte Carlo with 50,000 repetitions.[3] The reported power is based on 1,000 repetitions in each case. True sizes of asymptotic versions of a selected subset of the tests are provided in Table 5 for Model S. The true sizes reported are based on 50,000 repetitions. The asymptotic critical values for these results are taken from Andrews and Ploberger (1991) for Avg–F, Avg–LM, Exp–F_∞, and Exp–LM_∞, from Andrews (1989) for Sup15–F and Sup15–LM, and from Krämer and Sonnberger (1986) for Cusum. For Exp–F_1, Exp–LM_1, Sup2–F, Sup2–LM, Nyb–F, and Nyb–LM, they were generated by Monte Carlo with 50,000 repetitions using the method described in Andrews (1989). Note that these asymptotic critical values are not appropriate for Model TT and, hence, size results for that model are not given.

We now make several general comments on the tables. First, none of the tables reports power results for LM versions of the exponential and sup tests, i.e., Avg–LM, Exp–LM_c, Exp–LM_∞, Sup2–LM, and Sup15–LM, because these results are very similar to the results for the F versions of the tests. In many cases the power is the same, in some cases it differs by .01, and in only a very few cases does it differ by more than .01.

Second, none of the tables report results for Exp–F_c for $c = 1/3$ or 3, because these results lie between those of Avg–F, Exp–F_1, and Exp–F_∞, respectively, and the latter three exhibit relatively small differences in power except in a few cases.

Third, in Table 1 power for all tests except Cusum is the same for π_0 and $1 - \pi_0$ and, hence, results are only reported for $\pi_0 = .075, .15, .3, .5$. For Cusum, power differs for π_0 and $1 - \pi_0$, so additional columns are provided in the Table for the $1 - \pi_0$ cases. In addition, unlike the other tests, Cusum does not have power that is invariant with respect to the direction of β in Table 1. In consequence, Cusum results are given for two cases: β proportional to (1, 0) and β proportional to (0, 1). The results for two intermediate cases viz., β at angles 36° and 54° with the horizontal axis, are quite similar to those for β proportional to (0, 1) and, for brevity, are not reported.

Fourth, with Model S and Alt–1(π_0) (see Table 1), the power of all tests except Cusum is at a maximum when the changepoint $\pi_0 = .5$ and drops off in a concave fashion as π_0 approaches 0 or 1. In contrast, with Model TT and Alt–1(π_0) (see Table 2), the power of all tests except Cusum and F(.5) is at, or is close to, a minimum at $\pi_0 = .5$ and is maximized at $\pi_0 = .15$ and .85 or $\pi_0 = .3$ and .7. This (possibly) counter–intuitive pattern arises because a regression model with an intercept and time trend is better able to approximate a true regression model that has a constant mean with a single change at $\pi_0 = .5$ than a similar model with a single change at $\pi_0 = .15$, .3, .7, or .85.

Next, we summarize the results presented in the tables. First, we consider the sensitivity of the power of the exponential F tests to c. For Alt–MG1, Alt–MG2, and Alt–MG(π_0), there is little sensitivity to c for Model S or Model TT. For Alt–1(π_0), there are some power differences for different values of c, but they are not huge. More specifically, for Model S and Alt–1(π_0), Exp–F_∞ is best due to its higher power for $\pi_0 = .075$ and .15. For Model TT and Alt–1(π_0), Exp–F_∞ is the best exponential F statistic for $\pi_0 = .5$, but Avg–F and Exp–F_1 are somewhat better for $\pi_0 = .15$ and .85. Overall, we conclude that the exponential F statistics are not very sensitive to the choice of c. For the alternatives considered here, we prefer Exp–F_∞ or Exp–F_1 by a narrow margin over the other exponential F tests.

Second, we compare the power of the exponential F tests with that of other tests in the literature including Sup2–F, Sup15–F, Nyb–LM, Cusum, and F(.5). Across the board Exp–F_∞ is as good as or better than Sup2–F and Sup15–F with the exception of the case of Model S with Alt–1(π_0) and $\pi_0 = .075$ for Sup2–F and of the case of Model TT with Alt–1(π_0) and $\pi_0 = .5$. The differences between the two are small in most cases, but substantial in some. We conclude that Exp–F_∞ is preferable to the likelihood ratio statistics Sup2–F and Sup15–F for the cases considered here.

A comparison of the exponential F tests with Nyb–LM for Alt–MG1 is of interest, because Nyb–LM has some asymptotic optimality properties for these alternatives, see Nyblom (1989). We find that the exponential tests do quite well. Exp–F_∞ slightly outperforms Nyb–LM for some τ values and the reverse occurs for some other τ values. For the other MG alternatives, Exp–F_∞ seems somewhat preferable to Nyb–LM in an overall sense, because the two tests have quite similar power for all cases except those for Alt–MG(π_0) with $\pi_0 = .85$ or .925 for which Exp–F_∞ is clearly superior.

Comparison of the exponential F tests with Cusum is dramatic. For Model S with Alt–1(π_0), the Cusum test has higher power than the exponential F tests when $\pi_0 = .075$ or .15 and $\beta \propto (1, 0)$. In almost all other cases, models, and alternatives, however, the Cusum test performs very poorly both in an absolute sense and relative to the exponential F tests. Thus, we conclude that the exponential F tests have much better overall power properties than the Cusum test.

As expected, the midpoint F test, F(.5), outperforms the exponential F tests for Alt–1(π_0) with $\pi_0 = .5$. In every other case, including all of the MG alternatives, however, the exponential F tests outperform F(.5). The difference is often substantial, especially for cases with π_0 near 0 or 1. Hence, we conclude that the exponential F tests exhibit better all–round power properties than F(.5).

Third, we compare the power of Exp–F_∞ with that of Exp–W_∞. These statistics differ in that the latter takes the error variance σ^2 as known. We find that Exp–W_∞ is more powerful than the feasible statistic Exp–F_∞ against Alt–1(π_0) by between 0.0 and .04 depending on the case considered. (The differences between the exponential F and W tests for other values of c, not reported in the tables, are comparable.) These differences imply that Exp–F_∞ has power that is at least within 0.0–.04 of being optimal for the case of unknown variance σ^2, since Exp–W_∞ possesses the optimality results of Section 2 for the case of known variance. Given that these power differences are fairly small and that any feasible test must sacrifice some power when σ^2 is unknown, we conclude that the exponential F tests must be quite close to being optimal, if they are not strictly optimal, for the case of σ^2 unknown.

Fourth, a comparison of the power of the exponential F tests against Alt–1(π_0) with that of the envelope power (given by F(π_0)) shows that knowledge of the changepoint increases power substantially. Although this is not surprising, its quantification in the tables seems useful. These results indicate that for one–time change alternatives correct prior knowledge that restricts the set Π of changepoints, whether to a single point or just to a smaller set, yields significant power gains.

Fifth, we discuss the size results of Table 5. The use of asymptotic critical values leads to a slight underrejection by the LM versions of the exponential test statistics and a slight overrejection by the F versions of these statistics. The Cusum, Sup2–LM, and Sup15–LM tests exhibit the greatest underrejection of all of the statistics considered, which exacerbates the power problems for the Cusum test. The Avg–F and Nyb–F statistics exhibit the best size properties of all of the statistics considered. In an absolute sense as well, their size properties are excellent. As noted above, these size results are relevant only as an indicator of size results for corresponding tests in nonlinear models, since exact critical values can be obtained for the linear models considered here.

APPENDIX

The proof of Theorem 1(a) uses the following Lemmas. Let

$$\overline{\mathrm{LR}} = \sum_{\pi\in\Pi} \exp\Big[\tfrac{1}{2}\bar\theta(\pi)'\mathcal{I}(\pi)\bar\theta(\pi)\Big]\int \exp\Big[-\tfrac{1}{2}(\bar\theta(\pi)-h)'\mathcal{I}(\pi)(\bar\theta(\pi)-h)\Big]dQ_\pi(h)J(\pi), \text{ where}$$

$$\bar\theta(\pi) = \hat\theta(\pi) - \theta_0 \text{ and } \hat\theta(\pi) = (\hat\beta_1(\pi)', \ldots, \hat\beta_m(\pi)', \hat\delta_1(\pi)', \hat\delta_2(\pi))' \tag{A.1}$$

for $(\hat\beta_1(\pi)', \ldots, \hat\beta_m(\pi)', \hat\delta_1(\pi)', \hat\delta_2(\pi))$ as defined in (2.14).

LEMMA A–1. *Under Assumption* 1, $\mathrm{LR} = \overline{\mathrm{LR}}$.

LEMMA A–2. *For each* $\pi \in \Pi$, *the projection matrix* $P^\perp$ *onto the orthogonal complement* $V_\pi^\perp$ *of* V *with respect to* $<\cdot,\cdot>_\pi$ *is given by*

$$P^\perp\ (= P_\pi^\perp) = A_\pi H = \begin{bmatrix} I_{mv} & 0 \\ -\mathcal{I}_3^{-1}\mathcal{I}_{2\pi}' & 0 \end{bmatrix} \in R^{s\times s},$$

where A_π *and* H *are defined in* (2.8) *and* (2.14) *respectively.*

Proof of Theorem 1. First we establish part (a). By Lemma A–1, it suffices to show that $\overline{\mathrm{LR}} = \text{Exp-}W_c$. To do so, Let $\lambda \sim N(0, c(A'\mathcal{I}A)^{-1})$ and $h = A\lambda$, where $A = A_\pi$ and $\mathcal{I} = \mathcal{I}(\pi)$. Then, $h \sim Q_\pi = N(0, cA(A'\mathcal{I}A)^{-1}A')$ as desired. The density of λ is

$$(2\pi)^{-mv/2}\det{}^{1/2}(A'\mathcal{I}A/c)\exp\Big[-\tfrac{1}{2c}\lambda'A'\mathcal{I}A\lambda\Big] \tag{A.2}$$

with respect to Lebesgue measure on R^{mv}.

For notational simplicity, let $\bar\theta = \bar\theta(\pi)$ and $\mathcal{I} = \mathcal{I}(\pi)$. Then,

$$\overline{\mathrm{LR}} = \sum_{\pi\in\Pi} \zeta(\pi)J(\pi)\ , \text{ where} \tag{A.3}$$

$$\begin{aligned}\zeta(\pi) &= \int \exp\Big[\tfrac{1}{2}\bar\theta'\mathcal{I}\bar\theta - \tfrac{1}{2}(h-\bar\theta)\mathcal{I}(h-\bar\theta)\Big]dQ_\pi(h) \\ &= (2\pi)^{-mv/2}\det{}^{1/2}(A'\mathcal{I}A/c) \\ &\quad \times \int \exp\Big[\tfrac{1}{2}[\bar\theta'\mathcal{I}\bar\theta - (A\lambda-\bar\theta)'\mathcal{I}(A\lambda-\bar\theta) - (A\lambda)'\mathcal{I}A\lambda/c]\Big]d\lambda\ .\end{aligned} \tag{A.4}$$

Let P and $P^\perp$ denote the projection matrices with respect to $<\cdot,\cdot>_\pi$ onto V and $V_\pi^\perp$

respectively. (Note that P and $P^{\perp}$ depend on π since $<\cdot,\cdot>_{\pi}$ and $V_{\pi}^{\perp}$ do.) The term in square brackets in the exponent on the rhs of (A.4), with $A\lambda$ replaced by h for simplicity, now simplifies as follows:

$$\begin{aligned}
&\bar{\theta}'\mathcal{I}\bar{\theta} - (h-\bar{\theta})'\mathcal{I}(h-\bar{\theta}) - h'\mathcal{I}h/c \\
&= \bar{\theta}'\mathcal{I}\bar{\theta} - \left[h - \bar{\theta}\tfrac{c}{1+c}\right]'\mathcal{I}\tfrac{1+c}{c}\left[h - \bar{\theta}\tfrac{c}{1+c}\right] - \tfrac{1}{1+c}\bar{\theta}'\mathcal{I}\bar{\theta} \\
&= \tfrac{c}{1+c}(P\bar{\theta})'\mathcal{I}P\bar{\theta} + \tfrac{c}{1+c}(P^{\perp}\bar{\theta})'\mathcal{I}P^{\perp}\bar{\theta} - \left[h - P^{\perp}\bar{\theta}\tfrac{c}{1+c}\right]'\mathcal{I}\tfrac{1+c}{c}\left[h - P^{\perp}\bar{\theta}\tfrac{c}{1+c}\right] \qquad \text{(A.5)} \\
&\quad - \tfrac{c}{1+c}(P\bar{\theta})'\mathcal{I}P\bar{\theta} \\
&= \tfrac{c}{1+c}(P^{\perp}\bar{\theta})'\mathcal{I}P^{\perp}\bar{\theta} - \left[h - P^{\perp}\bar{\theta}\tfrac{c}{1+c}\right]'\mathcal{I}\tfrac{1+c}{c}\left[h - P^{\perp}\bar{\theta}\tfrac{c}{1+c}\right],
\end{aligned}$$

where the second equality uses the fact that $(P\bar{\theta})'\mathcal{I}h = 0\ \forall h \in V_{\pi}^{\perp}$.

Combining (A.4) and (A.5) gives

$$\begin{aligned}
\zeta(\pi) &= (1+c)^{-mv/2}\exp\left[\tfrac{1}{2}\tfrac{c}{1+c}(P^{\perp}\bar{\theta})'\mathcal{I}P^{\perp}\bar{\theta}\right] \\
&\times \int(2\pi)^{-mv/2}\det{}^{1/2}\left[A'\mathcal{I}A\tfrac{1+c}{c}\right]\exp\left[-\tfrac{1}{2}\left[A\lambda - P^{\perp}\bar{\theta}\tfrac{c}{1+c}\right]'\mathcal{I}\tfrac{1+c}{c}\left[A\lambda - P^{\perp}\bar{\theta}\tfrac{c}{1+c}\right]\right]d\lambda \qquad \text{(A.6)} \\
&= (1+c)^{-mv/2}\exp\left[\tfrac{1}{2}\tfrac{c}{1+c}(P^{\perp}\bar{\theta})'\mathcal{I}P^{\perp}\bar{\theta}\right],
\end{aligned}$$

where the second equality holds because the integral of a normal density equals one.

Using Lemma A–2, $(P^{\perp}\bar{\theta})'\mathcal{I}P^{\perp}\bar{\theta} = (H\bar{\theta})'A'\mathcal{I}AH\bar{\theta}$. Hence, for part (a), it remains to show that $A'\mathcal{I}A = [H\mathcal{I}^{-1}H']^{-1}$. By simple algebra, the left–hand side equals $\mathcal{I}_{1\pi} - \mathcal{I}_{2\pi}\mathcal{I}_3^{-1}\mathcal{I}_{2\pi}'$. The right–hand side equals the inverse of the upper $mv \times mv$ submatrix of $\mathcal{I}(\pi)^{-1}$, which equals $\mathcal{I}_{1\pi} - \mathcal{I}_{2\pi}\mathcal{I}_3^{-1}\mathcal{I}_{2\pi}'$ by the formula for a partitioned inverse. The proof of part (a) is now complete.

To establish part (b), note that under the null hypothesis, we have

$$\hat{\beta}(\pi) = H\hat{\theta}(\pi) = H(X_{\pi}'X_{\pi})^{-1}X_{\pi}'(U + X_{\pi}\theta_0) = H(X_{\pi}'X_{\pi})^{-1}X_{\pi}'U, \qquad \text{(A.7)}$$

since $H\theta_0 = 0$ for $\theta_0 = (0', \delta_{10}', \delta_{20}')'$, where X_{π} is a $T \times s$ matrix with t–th row $(X_t'1(T\pi_1 < t \le T\pi_2), \ldots, X_t'1(T\pi_m < t \le T), X_t', Z_t')'$ and $U = (U_1, \ldots, U_T)'$. Since

Exp–W_c depends on $\{Y_t : t \leq T\}$ only through $\hat{\beta}(\pi)$ and $\hat{\beta}(\pi)$ does not depend on θ_0 under the null, part (b) follows.

To establish part (c), we note that the Neyman–Pearson Lemma (e.g., see Lehmann (1959, Thm. 3.1, p. 65)) implies that a test based on LR is a most powerful test of $H_0(\theta_0) : Y \sim f_T(\cdot, \theta_0)$ versus $H_1(\theta_0) : Y \sim \sum_{\pi \in \Pi} \int f_T(\cdot, \theta_0 + h, \pi) dQ_\pi(h) J(\pi)$. By part (a), Exp–$W_c$ equals LR and by (2.11), the power of a test against $H_1(\theta_0)$ equals its weighted average power defined in (2.10). Hence, for each fixed $\theta_0 \in V$, the test $\xi(Y)$ based on Exp–W_c maximizes weighted average power among tests of the same significance level, say α. By part (b), $\xi(Y)$ has the same significance level α for each hypothesis $H_0(\theta_0)$ for $\theta_0 \in V$. In consequence, $\xi(Y)$ has maximum weighted average power among all tests of level α for the null hypothesis $H_0 : Y \sim f_T(\cdot, \theta_0)$ for some $\theta_0 \in V$. □

Proof of Lemma A–1. Let $\ell(\theta,\pi) = \log f_T(\theta,\pi)$, $D\ell(\theta,\pi) = \frac{\partial}{\partial\theta}\ell(\theta,\pi)$, and $D^2\ell(\theta,\pi) = \frac{\partial^2}{\partial\theta\partial\theta'}\ell(\theta,\pi)$. Simple algebra yields

$$D\ell(\theta_0, \pi) = \begin{bmatrix} \Sigma_{T\pi_1+1}^{T\pi_2}(Y_t - X_t'\delta_{10} - Z_t'\delta_{20})X_t \\ \vdots \\ \Sigma_{T\pi_m+1}^{T}(Y_t - X_t'\delta_{10} - Z_t'\delta_{20})X_t \\ \Sigma_1^T(Y_t - X_t'\delta_{10} - Z_t'\delta_{20})X_t \\ \Sigma_1^T(Y_t - X_t'\delta_{10} - Z_t'\delta_{20})Z_t \end{bmatrix}, -D^2\ell(\theta_0, \pi) = \mathcal{I}(\pi), \tag{A.8}$$

and $\overline{\theta}(\pi) = \mathcal{I}(\pi)^{-1} D\ell(\theta_0, \pi)$.

Since $f_T(\theta_0, \pi)$ and $\ell(\theta_0, \pi)$ do not depend on π, we can write

$$LR = \sum_{\pi \in \Pi} \int \exp(\ell(\theta_0 + h, \pi) - \ell(\theta_0, \pi)) dQ_\pi(h) J(\pi). \tag{A.9}$$

Let $h = (h_1', \ldots, h_{m+2}')' \in R^s$, where $h_j \in R^v \; \forall j \leq m+1$ and $h_{m+2} \in R^w$. The integrand of (A.9) simplifies as follows:

$$
\begin{aligned}
&\ell(\theta_0 + h, \pi) - \ell(\theta_0, \pi) \\
&= -\frac{1}{2}\Bigg[\sum_{j=0}^{m} \Sigma_{T\pi_j+1}^{T\pi_{j+1}}[Y_t - X_t'(\delta_{10} + h_{m+1} + h_j) - Z_t'(\delta_{20} + h_{m+2})]^2 \\
&\quad - \Sigma_1^T[Y_t - X_t'\delta_{10} - Z_t'\delta_{20}]^2\Bigg] \\
&= -\frac{1}{2}\Bigg[-2\sum_{j=0}^{m} \Sigma_{T\pi_j+1}^{T\pi_{j+1}}[Y_t - X_t'\delta_{10} - Z_t'\delta_{20}][X_t'(h_{m+1} + h_j) + Z_t'h_{m+2}] \\
&\quad + \sum_{j=0}^{m} \Sigma_{T\pi_j+1}^{T\pi_{j+1}}[X_t'(h_{m+1}+h_j) + Z_t'h_{m+2}]^2\Bigg] \\
&= D\ell(\theta_0, \pi)'h - \frac{1}{2}h'\mathcal{I}(\pi)h.
\end{aligned}
\tag{A.10}
$$

Letting $D\ell$, $\mathcal{I}$, and $\overline{\theta}$ abbreviate $D\ell(\theta_0, \pi)$, $\mathcal{I}(\pi)$, and $\overline{\theta}(\pi)$, respectively, and using $\overline{\theta} = \mathcal{I}^{-1}D\ell$ from (A.8), we obtain

$$
\begin{aligned}
&D\ell'h - \frac{1}{2}h'\mathcal{I}h = \overline{\theta}'\mathcal{I}h - \frac{1}{2}h'\mathcal{I}h = \frac{1}{2}\overline{\theta}'\mathcal{I}\overline{\theta} - \frac{1}{2}(\overline{\theta}'\mathcal{I}\overline{\theta} - 2\overline{\theta}'\mathcal{I}h + h'\mathcal{I}h) \\
&= \frac{1}{2}\overline{\theta}'\mathcal{I}\overline{\theta} - \frac{1}{2}(\overline{\theta}-h)'\mathcal{I}(\overline{\theta}-h).
\end{aligned}
\tag{A.11}
$$

Combining (A.9), (A.10), and (A.11) gives the desired result $LR = \overline{LR}$. □

Proof of Lemma A–2. It suffices to show that (1) $A_\pi H v = 0\ \forall v \in V$ and (2) $A_\pi H m = m$ $\forall m \in V_\pi^\perp$. To show (1), note that $v \in V$ iff $v = (0', v_2')'$ for some $v_2 \in R^{v+w}$. Thus, $A_\pi H v = A_\pi[I_{mv} \vdots 0]\begin{bmatrix} 0 \\ v_2 \end{bmatrix} = 0$. To show (2), note that $m \in V_\pi^\perp$ iff $v'\mathcal{I}(\pi)m = 0\ \forall v \in V$ iff $[0 \vdots I_{v+w}]\mathcal{I}(\pi)m = 0$ iff $[\mathcal{I}_{2\pi}' \vdots \mathcal{I}_3]\begin{bmatrix} m_1 \\ m_2 \end{bmatrix} = 0$, where $m = (m_1', m_2')'$, iff $m_2 = -\mathcal{I}_3^{-1}\mathcal{I}_{2\pi}'m_1$ for $m_1 \in R^{mv}$ and $m_2 \in R^{v+w}$. Thus, for $m \in V_\pi^\perp$, $A_\pi H m = \begin{bmatrix} I_p & 0 \\ -\mathcal{I}_3^{-1}\mathcal{I}_{2\pi}' & 0 \end{bmatrix}\begin{bmatrix} m_1 \\ -\mathcal{I}_3^{-1}\mathcal{I}_{2\pi}'m_1 \end{bmatrix}$ $= \begin{bmatrix} m_1 \\ -\mathcal{I}_3^{-1}\mathcal{I}_{2\pi}'m_1 \end{bmatrix} = m$, as desired. □

FOOTNOTES

[1] Andrews gratefully acknowledges research support from the National Science Foundation through grant number SES–8821021. Ploberger gratefully acknowledges research support from Fonds zur Förderung der wissenschaftlichen Forschung under Schrödingerstipendium Project J–0469–PHY.

[2] This is true only for $\pi_0 = .15, .3, \ldots, .85$, since in these cases the time of change $120\pi_0$ is an even number. For $\pi = .075, .925$, $120\pi_0$ is odd and the power of the tests depends slightly on the direction of β.

[3] For the Cusum test, there is no single critical value since the test rejects if $|\text{Cusum}_t|$ exceeds a line with a given intercept and slope for some t. Exact null rejection rates of .05 were obtained for the Cusum test by taking the slope equal to the value given by the asymptotic version of the test (see Krämer and Sonnberger (1986, pp. 49–53 and 59–61)) and adjusting the intercept of the rejection line appropriately.

REFERENCES

ANDREWS, D. W. K. (1989). Tests for parameter instability and structural change with unknown change point. Cowles Foundation Discussion Paper No. 943 (revised), Yale University.

ANDREWS, D. W. K. & PLOBERGER, W. (1991). Optimal tests when a nuisance parameter is present only under the alternative. Unpublished manuscript, Cowles Foundation, Yale University.

BROWN, R. L., DURBIN, J. & EVANS, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society, Series B* **37**, 149–192.

CHERNOFF, H. & ZACKS, S. (1964). Estimating the current mean of a normal distribution which is subject to changes in time. *Annals of Mathematical Statistics* **35**, 999–1028.

DAVIES, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **64**, 247–254.

DAVIES, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, **74**, 33–43.

FARLEY, J. U. & HINICH, M. J. (1970). A test for a shifting slope coefficient in a linear model. *Journal of the American Statistical Association* **65**, 1320–1329.

GARDNER, L. A., Jr. (1969). On detecting changes in the mean of normal variates. *Annals of Mathematical Statistics* **40**, 116–126.

HANSEN, B. E. (1991). Inference when a nuisance parameter is not identified under the null hypothesis. Working Paper No. 296, Rochester Center for Economic Research, University of Rochester.

HINCKLEY, D. V. (1969). Inference about the intersection in two–phase regression. *Biometrika* **56**, 495–504.

JAMES, B., JAMES, K. L. & SIEGMUND, D. (1987). Tests for a change–point. *Biometrika* **74**, 71–84.

JANDHYALA, V. K. & MACNEILL, I. B. (1991). Tests for parameter changes at unknown times in linear regression models. *Journal of Statistical Planning and Inference* **27**, 291–316.

KIM, H. J. & SIEGMUND, D. (1989). The likelihood ratio test for a change–point in simple linear regression. *Biometrika* **76**, 409–423.

KING, M. L. & SHIVELY, T. S. (1991). Locally optimal testing when a nuisance parameter is present only under the alternative. Unpublished manuscript, Department of Economics, Monash University.

KRÄMER, W. & SONNBERGER, H. (1986). *The Linear Regression Model Under Test.* Heidelberg: Physica–Verlag.

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses.* New York: Wiley.

NYBLOM, J. (1989) Testing for the constancy of parameters over time. *Journal of the American Statistical Association* **84**, 223–230.

WALD, A. (1942). On the power function of the analysis of variance test. *Annals of Mathematical Statistics* **13**, 434–439.

TABLE 1. Test Power for Model S and Alt-1(π_0)

Alternative Distribution		Test Statistic													
π_0 (and $1-\pi_0$)	$\|\beta\|$	Avg-F	Exp-F_1	Exp-F_∞	Sup2-F	Sup15-F	Nyb-LM	Cusum π_0 $\beta \propto (1,0)$	Cusum π_0 $\beta \propto (0,1)$	Cusum $1-\pi_0$ $\beta \propto (1,0)$	Cusum $1-\pi_0$ $\beta \propto (0,1)$	F(.5)	Exp-W_∞ (σ^2 known)	F(π_0)
.075	4.8	.07	.07	.08	.08	.06	.05	.15	.04	.04	.04	.05	.08	.19
	7.2	.10	.13	.15	.16	.09	.08	.31	.04	.04	.03	.06	.15	.38
	9.6	.18	.26	.31	.33	.13	.13	.51	.04	.04	.03	.06	.34	.62
	12.0	.29	.45	.50	.54	.21	.19	.72	.04	.04	.03	.08	.53	.81
.15	4.8	.14	.15	.15	.15	.14	.12	.22	.04	.05	.04	.07	.16	.30
	7.2	.30	.36	.36	.36	.37	.24	.47	.03	.06	.04	.11	.40	.61
	9.6	.54	.65	.66	.64	.65	.44	.73	.03	.07	.03	.18	.70	.87
	12.0	.76	.85	.87	.87	.88	.65	.91	.03	.09	.03	.26	.89	.98
.3	4.8	.32	.32	.30	.26	.31	.32	.24	.04	.07	.04	.19	.33	.49
	7.2	.67	.68	.66	.62	.68	.66	.55	.03	.14	.03	.43	.69	.84
	9.6	.89	.91	.91	.89	.92	.89	.82	.03	.29	.03	.69	.93	.97
	12.0	.99	.99	.99	.98	1.00	.99	.96	.02	.48	.02	.87	1.00	1.00
.5	4.8	.43	.41	.37	.31	.38	.46	.15	.04			.55	.41	.55
	7.2	.79	.79	.77	.72	.79	.81	.40	.03			.90	.81	.90
	9.6	.96	.97	.97	.95	.97	.97	.69	.03			.99	.97	.99
	12.0	1.00	1.00	1.00	1.00	1.00	1.00	.89	.02			1.00	1.00	1.00

TABLE 2. Test Power for Model TT and Alt-1(π_0)

Alternative Distribution		Test Statistic									
π_0	$\|\beta\|$	Avg-F	Exp-F_1	Exp-F_∞	Sup2-F	Sup15-F	Nyb-LM	Cusum	F(.5)	Exp-W_∞ (σ^2 known)	F(π_0)
.075	9.6	.31	.33	.31	.25	.26	.27	.28	.15	.33	.53
	12.0	.47	.51	.49	.43	.40	.39	.43	.21	.53	.72
	14.4	.64	.70	.68	.62	.56	.56	.58	.29	.71	.87
	16.8	.80	.84	.84	.79	.72	.70	.74	.39	.86	.95
.15	9.6	.52	.52	.46	.39	.41	.51	.22	.35	.49	.67
	12.0	.73	.72	.69	.60	.64	.71	.35	.53	.72	.86
	14.4	.87	.88	.87	.80	.84	.87	.51	.69	.89	.96
	16.8	.96	.96	.96	.94	.94	.95	.67	.82	.96	.99
.30	9.6	.44	.44	.41	.35	.41	.47	.09	.39	.44	.66
	12.0	.64	.66	.65	.58	.65	.68	.13	.57	.67	.86
	14.4	.81	.84	.82	.78	.83	.84	.16	.75	.85	.97
	16.8	.92	.94	.94	.92	.94	.93	.21	.87	.96	.99
.50	9.6	.17	.23	.26	.26	.29	.19	.11	.56	.28	.57
	12.0	.26	.37	.42	.42	.48	.30	.15	.75	.46	.76
	14.4	.39	.58	.63	.64	.69	.44	.20	.90	.65	.91
	16.8	.54	.75	.80	.81	.85	.60	.25	.97	.82	.97
.70	9.6	.47	.47	.45	.38	.45	.52	.11	.41	.47	.69
	12.0	.66	.68	.68	.61	.68	.70	.14	.58	.70	.88
	14.4	.82	.85	.85	.81	.85	.85	.20	.77	.87	.97
	16.8	.93	.95	.95	.93	.96	.95	.28	.88	.96	.99
.85	9.6	.54	.51	.47	.40	.43	.53	.07	.37	.50	.65
	12.0	.74	.74	.69	.60	.64	.74	.08	.55	.72	.85
	14.4	.89	.89	.87	.81	.84	.89	.10	.72	.89	.96
	16.8	.95	.96	.95	.92	.94	.95	.12	.85	.97	.99
.925	9.6	.33	.35	.32	.27	.25	.29	.05	.17	.34	.51
	12.0	.48	.51	.49	.44	.37	.41	.05	.23	.51	.73
	14.4	.66	.69	.67	.61	.54	.57	.05	.31	.70	.88
	16.8	.81	.84	.83	.79	.70	.72	.06	.42	.87	.95

TABLE 3. Test Power for Model S and Alt-MG1, Alt-MG2, and Alt-MG(π_0)

Alternative Distribution		τ	Avg-F	Exp-F_1	Exp-F_∞	Sup2-F	Sup15-F	Nyb-LM	Cusum	F(.5)
Alt-MG1		.07	.26	.26	.25	.22	.24	.26	.08	.22
		.1	.52	.53	.49	.46	.49	.53	.19	.45
		.3	.70	.71	.70	.66	.69	.69	.31	.62
		.4	.86	.87	.86	.83	.86	.84	.44	.75
		.5	1.0	1.0	1.0	1.0	1.0	.99	.74	.92
Alt-MG2		.07	.48	.47	.46	.43	.43	.46	.19	.37
		.1	.70	.72	.70	.66	.66	.68	.34	.57
		.3	.83	.85	.84	.81	.81	.80	.44	.68
		.4	.91	.93	.92	.91	.89	.87	.52	.75
		.5	.99	1.0	1.0	1.0	.98	.98	.72	.87
Alt-MG(π_0)	π_0=.15	.1	.51	.51	.49	.45	.49	.51	.17	.44
		.3	.84	.85	.84	.82	.84	.84	.42	.75
		.5	1.0	1.0	1.0	1.0	1.0	1.0	.75	.95
		1.0	1.0	1.0	1.0	1.0	1.0	1.0	.77	.97
	π_0=.3	.1	.47	.47	.44	.41	.45	.48	.14	.41
		.3	.80	.81	.80	.78	.81	.80	.32	.71
		.5	1.0	1.0	1.0	1.0	1.0	1.0	.68	.96
		1.0	1.0	1.0	1.0	1.0	1.0	1.0	.70	.97
	π_0=.5	.1	.33	.32	.31	.28	.31	.33	.07	.29
		.3	.71	.73	.71	.69	.71	.71	.19	.63
		.5	.99	.99	1.0	1.0	.99	.98	.61	.96
		1.0	1.0	1.0	1.0	1.0	1.0	.99	.66	.97
	π_0=.7	.1	.15	.16	.16	.14	.14	.14	.04	.09
		.3	.40	.42	.42	.41	.42	.36	.06	.22
		.5	.95	.97	.98	.98	.96	.94	.37	.83
		1.0	.99	1.0	1.0	1.0	.99	.97	.47	.90
	π_0=.85	.1	.07	.07	.07	.06	.05	.06	.04	.05
		.3	.13	.15	.15	.16	.12	.10	.04	.07
		.5	.80	.86	.87	.88	.79	.72	.09	.41
		1.0	.93	.96	.97	.98	.92	.88	.16	.67
	π_0=.925	.1	.05	.05	.05	.05	.04	.05	.04	.05
		.3	.06	.06	.06	.07	.05	.05	.04	.05
		.5	.39	.55	.59	.62	.27	.25	.03	.11
		1.0	.76	.86	.89	.90	.63	.56	.03	.20

TABLE 4. Test Power for Model TT and Alt-MG1, Alt-MG2, and Alt-MG(π_0)

Alternative Distribution		τ	Avg-F	Exp-F_1	Exp-F_∞	Sup2-F	Sup15-F	Nyb-LM	Cusum	F(.5)
Alt-MG1		.03	.32	.32	.29	.24	.25	.33	.12	.29
		.05	.47	.49	.47	.41	.45	.48	.20	.44
		.07	.92	.95	.95	.94	.95	.92	.67	.81
		.1	.97	.98	.98	.98	.98	.96	.76	.87
		.3	.99	.99	.99	.99	.98	.98	.80	.90
Alt-MG2		.02	.38	.39	.37	.32	.32	.37	.17	.30
		.03	.64	.66	.64	.61	.61	.61	.37	.51
		.04	.80	.83	.83	.80	.78	.77	.51	.64
		.05	.92	.94	.94	.93	.91	.89	.66	.75
		.1	1.0	1.0	1.0	1.0	1.0	.99	.88	.89
Alt-MG(π_0)	π_0=.15	.05	.43	.43	.41	.37	.40	.44	.13	.39
		.1	.90	.92	.93	.91	.93	.90	.54	.82
		.5	.97	.99	.99	.99	.99	.97	.73	.90
		1.0	1.0	1.0	1.0	1.0	1.0	1.0	.85	.94
	π_0=.3	.05	.37	.39	.37	.32	.35	.38	.11	.35
		.1	.87	.90	.91	.90	.91	.86	.47	.80
		.5	.97	.98	.98	.97	.98	.97	.66	.89
		1.0	.99	1.0	1.0	1.0	1.0	1.0	.79	.94
	π_0=.5	.05	.29	.30	.29	.26	.28	.30	.09	.26
		.1	.82	.86	.87	.85	.87	.81	.37	.73
		.5	.94	.96	.96	.96	.96	.94	.56	.86
		1.0	.99	1.0	1.0	1.0	.99	.99	.75	.94
	π_0=.7	.05	.23	.22	.21	.18	.19	.22	.06	.18
		.1	.67	.72	.72	.69	.69	.65	.20	.54
		.5	.85	.88	.90	.89	.88	.83	.36	.67
		1.0	.95	.98	.99	.98	.98	.93	.58	.80
	π_0=.85	.05	.11	.12	.11	.10	.09	.10	.05	.09
		.1	.42	.45	.45	.42	.38	.40	.06	.30
		.5	.65	.70	.71	.71	.62	.60	.11	.50
		1.0	.85	.91	.92	.93	.84	.80	.25	.69
	π_0=.925	.05	.06	.06	.07	.07	.05	.06	.05	.06
		.1	.20	.21	.22	.20	.16	.16	.05	.10
		.5	.38	.42	.42	.40	.31	.32	.05	.19
		1.0	.63	.69	.72	.72	.56	.57	.04	.43

TABLE 5. True Size of Asymptotic Tests for Model S

Nominal Significance Level	Test Statistic						
	Avg-F	Exp-F_1	Exp-F_∞	Sup2-F	Sup15-F	Nyb-F	Cusum
10%	10.1	10.8	11.7	8.2	9.2	9.9	8.1
5%	5.1	5.6	6.1	4.4	4.6	5.1	3.9
1%	.92	1.23	1.50	.99	1.04	1.02	.64
	Avg-LM	Exp-LM_1	Exp-LM_∞	Sup2-LM	Sup15-LM	Nyb-LM	
10%	9.2	9.2	9.2	5.4	7.0	9.1	
5%	4.4	4.4	4.2	2.4	3.1	4.5	
1%	.70	.74	.73	.36	.44	.74	