THE UNIVERSITY *of York*

Discussion Papers in Economics

No. 2000/55

The Geometry of Similar Tests for Structural Change

by

Giovanni Forchini

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

The Geometry of Similar Tests for Structural Change

G. Forchini*
University of York

First version: February, 1999
This version: February, 2000

Abstract

This paper analyses similar tests for structural change for the normal linear regression model. By focusing on the geometric properties of the critical regions rather than on the test statistics themselves, we are able to derive optimal tests and to better understand the properties of existing test procedures such as the CUSUM, the CUSUM of squares and the sup F tests. Moreover, by analysing the geometry of the critical regions we identify an *intrinsic difficulty* of testing for structural change which affects the power of all tests. A measure of this intrinsic difficulty is suggested.

Key Words: Similar Tests, Structural Change

JEL Classification: C12, C31, C52

Address for Correspondence
Department of Economics and Related Studies
University of York
Heslington
York YO1 5DD
UK
e-mail: gf7@york.ac.uk

*I would like to thank Grant Hillier for patient help, suggestions and useful discussions, and the Editor, Peter C.B. Phillips, and two anonymous referees for comments which led to an improved version of the paper. I also thank Patrick Marsh, Jan Podivinsky, Richard Smith and the participants of a York economics seminar and of the 1997 EC^2 conference in Amsterdam for helpful comments. Grants from ESRC (N. R00429424115) and the University of Southampton are gratefully acknowledged.

1 Introduction

Testing for structural change in the linear regression model has received considerable attention in the statistical and econometric literature in the last thirty years. This intense activity has led to the development of many tests for structural change which 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 change point is known because of institutional changes, wars, etc. (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 for this problem can be found in Hillier (1987).

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 non-standard 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, the CUSUM of squares of Brown, Durbin and Evans (1975), the fluctuation test of Ploberger, Krämer and Kontrus (1987), the sup F test, and the tests recently suggested by Andrews, Lee and Ploberger (1996)) are similar, i.e. their sizes do not depend on nuisance parameters.

In this paper we analyse similar tests for structural change in the linear model with Gaussian errors, in finite samples. The emphasis is on the geometrical aspects of the testing situation, because the difficulty of testing for structural change hinges on the geometry of the problem.

Since, no uniformly most powerful (UMP) similar test exists, even for simple cases, when it is not known where the structural breaks occur, we suggest a weaker optimality criterion leading to a tests analogous to those suggested by Andrews, Lee and Ploberger (1996). However, we will argue that the geometry of the testing problem generates an "intrinsic difficulty", which affects the power of *all* testing procedures. Thus, following Hillier (1995), we emphasize the importance of reporting a measure of this intrinsic difficulty, i.e. an index of the potential loss of power with respect to the situation where both number and location of the change points is known (this problem is further discussed in Section 7).

The paper stresses the idea that it is reasonable to test for structural change when the intrinsic difficulty of testing is small (i.e. when we know approximately where the structural breaks occur). However, in this case there is no need to use tests such as the sup F test or the optimal tests proposed by Andrews, Lee and Ploberger (1996) or those derived in Section 4. An F test against any fixed changepoint could do just as well. Tables 2, 3 and 4 in Andrews, Lee and Ploberger (1996) show that the midsample F test performs as well as the other tests when the change point is more or less in the middle of the sample (i.e. when it is approximately known where the structural change occurs).

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 reviews the geometrical aspect of the standard F test, and Section 5 derives optimal tests for structural change (the proofs of all statements are in the Appendix). Section 5 reviews other tests for structural change and shows how they are affected by the geometry of the testing problem In Section 7 we propose a measure of the intrinsic difficulty in testing for structural change. The conclusions end the paper.

2 The model and characterization of similar tests

In this Section, we characterize similar tests for a linear regression model with $t+1$ subsamples, containing respectively τ_1, τ_2, ..., τ_{t+1} observations:

$$y = \begin{pmatrix} Z_{\tau_1} & 0 & \dots & 0 & 0 \\ 0 & Z_{\tau_2} & \dots & 0 & 0 \\ \dots & \dots & \ddots & \dots & \dots \\ 0 & 0 & \dots & Z_{\tau_t} & 0 \\ 0 & 0 & \dots & 0 & Z_{\tau_{t+1}} \end{pmatrix} \begin{pmatrix} \eta \\ \eta + \gamma_1 \\ \vdots \\ \eta + \gamma_{t-1} \\ \eta + \gamma_t \end{pmatrix} + V\zeta + u \qquad (1)$$

where y is a $T \times 1$ vector of dependent variables , V is a $T \times n$ matrix of independent variables, η and γ_1,...,γ_t are $k \times 1$ vectors of parameters, ζ is a $n \times 1$ vector of parameters, and u is $N\left(0,\sigma^2 I_T\right)$. The matrices $Z_{\tau_1}, Z_{\tau_2}, ..., Z_{\tau_t}, Z_{\tau_{t+1}}$ have the form

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

where $\bar{\tau}_i = \sum_{j=0}^{i} \tau_j$ and z_i is the i-th $(1 \times k)$ vector of observations.

In this paper we identify the change point by an index τ which represents a partition of T into $t+1$ parts, $\tau = (\tau_1, \tau_2, ..., \tau_{t+1})$, $\tau_i > 0$ for all i, $\sum_{i=1}^{t+1} \tau_i = T$. The subset of partitions of T of interest (i.e. the set of all possible changepoints in the model) is denoted by Υ.

Let $Z = (z_1', ..., z_T')'$, so that (1) can be written as

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

where $X = (Z, V)$ is $T \times p$, $\beta = (\eta', \zeta')'$ is $p \times 1$, $\gamma = (\gamma_1', \gamma_2', ..., \gamma_t')'$ is $K \times 1$ and $Z(\tau)$ is the $T \times K$ matrix obtained by deleting the first k columns in diag $\left(Z_{\tau_1}, Z_{\tau_2}, ..., Z_{\tau_t}, Z_{\tau_{t+1}}\right)$ 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. Moreover, we do not allow for changes over time for σ^2, either, to avoid Behrens-Fisher type problems.

Note that 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 y, and the following analysis will be conditional on them. By so doing, we cover the case of both stationary and nonstationary regressors. Note also that the set-up considered is fairly general: the change points can be approximately located (for example 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 $H_0 : y \sim N\left(X\beta, \sigma^2 I_T\right)$, and in this case the statistics $\hat{\beta} = (X'X)^{-1} X'y$ and $s_1^2 = y'P_X y$ $\left(P_X = I_T - X(X'X)^{-1} X'\right)$ are jointly sufficient for the nuisance parameters β and σ^2. Since they are boundedly complete, any similar region of size α is a fraction α of the surface $\left(\hat{\beta}, s_1^2\right)$ =constant (Hillier (1987)).

The alternative hypothesis is $H_1 : y \sim N\left(X\beta + Z(\tau)\gamma, \sigma^2 I_T\right)$. 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 $\mathbb{R}^T$, $\bar{V}^\tau = \left\{Z(\tau)\gamma : \gamma \in \mathbb{R}^K\right\}$, indexed by τ, where m is the number of elements in Υ. Note that the number of change points determines the dimension of the subspaces $\bar{V}^\tau$, while 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 $H_1 : y \sim N(X\beta + W\gamma, \sigma^2 I_T)$, and W is a completely known $(T \times K)$ matrix. In this case the columns of W span a fixed K-dimensional subspace of $\mathbb{R}^T$, as γ varies over $\mathbb{R}^K$, $\bar{V} = \{W\gamma : \gamma \in \mathbb{R}^K\}$. In the structural change case this is no longer so.

Note also that 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 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 $Z^*(\tau)$.

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

Theorem 1 *(Hillier (1991), Theorem 1) 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 $CC' = P_X$, $C'C = I_{T-p}$ 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 $S_{T-p-1} = \{v \in \mathbb{R}^{T-p} : v'v = 1\}$.*

The matrix C can be chosen so that w_1 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 following pages. The CUSUM, the CUSUM of squares, the fluctuation test, the sup F test and the tests of Andrews and Ploberger (1993) and Andrews, Lee and Ploberger (1996) (even in the case where the error variance is unknown) are similar because they are all functions of v.

Before discussing the structural break case, we will give a geometrical interpretation to the familiar F-test in the problem of testing $H_0 : y \sim N(X\beta, \sigma^2 I_T)$ against $H_1 : y \sim N(X\beta + W\gamma, \sigma^2 I_T)$. This will allow us to introduce some tools and some ideas which will be developed in later Sections.

3 The geometry of the standard F-test

When constructing tests econometricians tend to focus on the properties of tests statistics rather than critical regions. This Section shows that by emphasising the importance of critical regions, geometrical considerations come into play even for the well known F test. This is also useful to better understand the traditional technique used to overcome the problem of the nonexistence of uniformly most powerful tests.

In the the simple set-up where $H_0 : y \sim N(X\beta, \sigma^2 I_T)$ is tested against the alternative $H_1 : y \sim N(X\beta + W\gamma_1, \sigma^2 I_T)$, where W is a fixed matrix and γ_1 is known, the most powerful (similar) critical region ω exists, and has the form of a "cap" on S_{T-p-1} centred around the (known) straight line going through the origin and the point $C'W\gamma_1/\sigma$, with an area equal to a fraction α of the surface area of the unit sphere, where α is the size of the test and C is a $T \times T - p$ matrix such that $CC' = P_X$, $C'C = I_{T-p}$ and $C'X = 0$. Moreover, the power of the test depends on the distance of $C'W\gamma_1/\sigma$ from the origin even though the critical region does not.

When the null hypothesis $H_0 : y \sim N(X\beta, \sigma^2 I_T)$ is tested against the alternative $H_1 : y \sim N(X\beta + W\gamma, \sigma^2 I_T)$, where γ is not known, the optimal critical region ω depends on the unknown $C'W\gamma/\sigma$. Since γ can be any vector in $\mathbb{R}^K$, $C'W\gamma/\sigma$ can be any point in the fixed k-dimensional hyperplane generated by the columns of $C'W$ as γ/σ vary over $\mathbb{R}^K$. The nonexistence of UMP tests is thus related to fact that $S_{T-p-1} \cap \{C'W\gamma/\sigma : \gamma/\sigma \in \mathbb{R}^K\}$ consists of more than one point.

A procedure which can be traced back to Wald (1943) and which is often used to handle the existence of nuisance parameters consists in weakening the optimality criterion, i.e. in choosing to critical region to maximize an "average" power. Hillier (1987) suggests to maximize the average power on the surface of constant λ, by averaging over all possible directions of $C'W\gamma/\sigma$. The most powerful similar critical region (in terms of average power) is given by the points on S_{T-p-1} for which the angle with the subspaces $\{C'W\gamma/\sigma : \gamma/\sigma \in \mathbb{R}^K\}$ is small (Hillier (1987), Section 3). That is, ω is a "strip" on S_{T-p-1} close to $S_{T-p-1} \cap \{C'W\gamma/\sigma : \gamma/\sigma \in \mathbb{R}^K\}$.

Note that ω is the union of all the critical regions (of a suitable size α^*) for a given γ as γ varies in $\mathbb{R}^K$. Therefore the most powerful critical region (in terms of average power) contains subsets of S_{T-p-1} for which the (unconditional) power is very small: these are the sets in ω for which the angle between v and $C'W\gamma/\sigma$ is small. The average power depends on the dimension of the subspace $\left\{C'W\gamma/\sigma : \gamma/\sigma \in \mathbb{R}^K\right\}$, i.e. on K, because it increases the average distance of points in ω from the subspace $C'W\gamma/\sigma$. The power of ω depends on λ, but the critical region itself does not.

Finally, note that the angle between v and subspace generated by the columns of $C'W$ is related to the F-statistic, f, for testing $H_0 : \gamma = 0$ against $H_1 : \gamma \neq 0$. However, by focusing on the test statistics the geometrical intuition is missed out.

4 Optimal tests for structural change

In this Section we analyse the construction of optimal similar tests for structural change. We will find that the power of the optimal test is affected by the relative position of the subspaces

$$V^\tau = \left\{C'Z(\tau)\gamma/\sigma : \gamma/\sigma \in \mathbb{R}^K\right\} = C'\bar{V}^\tau.$$

We consider optimal similar critical regions for tests of $H_0 : y \sim N\left(X\beta, \sigma^2 I_T\right)$ against $H_1 : y \sim N\left(X\beta + Z(\tau)\gamma, \sigma^2 I_T\right)$, where γ is unknown. Since when γ varies in $\mathbb{R}^K$, the columns of $Z(\tau)$ 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. To do this we assign a weight $p(\tau)$ to each subspace V^τ, and maximize the average power over the surfaces of constant $\lambda_\tau = \frac{1}{\sigma^2}\gamma Z(\tau)' P_X(\tau)\gamma$ in each V^τ, and among the subspaces V^τ. However this is not enough to yield a test which is UMP in terms of average power.

Theorem 2 *The critical region ω, which maximizes the average power $\bar{P}_\omega = \sum_{\tau\in\Upsilon} p(\tau) P_\omega^\tau$ where P_ω^τ is the power of ω for each fixed τ, is*

$$\omega = \left\{ v \in S_{T-p-1} : \sum_{\tau\in\Upsilon} p(\tau) \exp\left\{-\frac{1}{2}\lambda_\tau\right\} {}_1F_1\left(\frac{T-p}{2}; \frac{K}{2}; \frac{\lambda_\tau}{2}(\cos\theta_\tau)^2\right) > c \right\} \quad (3)$$

where $0 < \theta_\tau < \frac{\pi}{2}$ is the angle between v and V^τ.The constant c is determined by $\left[\Gamma\left(\frac{T-p}{2}\right)/2\pi^{(T-p)/2}\right]\int_\omega (dv) = \alpha$, where (dv) denotes the unnormalized Haar measure on S_{T-p-1}. Thus no UMP test in terms of average power exists.

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. If c is close to zero then small deviations from the null are given a large weight. If c is large more weight is given to large parameters.

Theorem 3 *The critical region ω which maximizes the unconditional power averaged over (i) the partition τ with weights $p(\tau)$, (ii) the direction of μ_τ with uniform weights on S_{K-1}, and (iii) the λ_τ with weights proportional to a the density function of a chi-square random variable with K degrees of freedom, is given by*

$$LR_c > c_1$$

where

$$\begin{aligned} LR_c &= (1+c)^{\frac{T-p-K}{2}} \sum_{\tau\in\Upsilon} p(\tau) \left(1 + c\sin^2\theta_\tau\right)^{-\frac{T-p}{2}}, & (4)\\ &= (1+c)^{\frac{T-p-K}{2}} \sum_{\tau\in\Upsilon} p(\tau) \left(\frac{1+\frac{K}{T-p-K} f_\tau}{1+c+\frac{K}{T-p-K} f_\tau} \right)^{\frac{T-p}{2}}. & (5) \end{aligned}$$

where $0 < \theta_\tau < \frac{\pi}{2}$ 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

$$\frac{\Gamma\left(\frac{T-p}{2}\right)}{2\pi^{\frac{T-p}{2}}} \int_{\left\{v\in S_{T-p-1}:LR_c>c_1\right\}} (dv) = \alpha,$$

where (dv) denotes the unormalized Haar measure on S_{T-p-1}.

The statistic LR_c is called the “likelihood ratio” by Andrews and Ploberger (1994) and Andrews, Lee and Ploberger (1996).

Remarks

(i) The derivation of the LR_c test is based on a particular weighting function for λ_τ, and can be very sensitive to its choice. Moreover, since it requires averaging over three weighting functions, its power may be very small. Only if the subspaces V^τ

are "very close" to each other (i.e. if the intersection of the optimal critical regions for each fixed τ is "large"), we can expect the LR_c test to be powerful. In this case however, an F test against any fixed τ could do just as well. If, however, the subspaces V^{τ} are very far from each other (i.e. optimal critical regions for fixed τ do not have many parts in common) the power of the LR_c test will be low. Thus, reporting an index of distance between the subspaces is important as a measure of "goodness" of the test.

(ii) Andrews and Ploberger (1994) and Andrews, Lee and Ploberger (1996) suggest to average the power also over all values of $(\beta', \gamma')'$ with weights proportional to the density of a normal distribution over V^{τ}, and obtain a class of similar tests for the case where σ^2 is known. Our approach is the same as taking a weight for $C'Z(\tau)\gamma/\sigma$ proportional to a $N(0, cI_K)$ density, where c is an arbitrary positive constant which scales the variance matrix. Averaging the unconditional power with respect to $C'Z(\tau)\gamma/\sigma \sim N(0, cI_K)$ is equivalent to averaging the power of a critical region ω with respect with both the direction of $C'Z(\tau)\gamma/\sigma$ (i.e. μ_{τ} uniform on S_{K-1}) and its length (i.e. λ_{τ} with $\lambda_{\tau}/c \sim \chi^2_K$). Note that the weighting function we use is different from that of Andrews and Ploberger (1994) and Andrews, Lee and Ploberger (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 $Z(\tau)\gamma$ for arbitrary β, so there is no loss of information in constructing inference about $Z(\tau)\gamma$ from w_1 only (which, also, does not depend on β). Moreover, our results are valid for unknown σ^2, too (note that the average exponential Wald test suggested by Andrews, Lee and Ploberger (1996) for the case of unknown σ^2 is also similar since it is a function of v only through the f_{τ}).

(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 non-informative prior.

A normalized version of the LR_c test for this case is

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

If c is small, the distribution of $C'Z(\tau)\gamma/\sigma$ tends to be more concentrated around 0. Letting c tend to zero, a normalized LR_c test is

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

as a consequence LR_0 should have power in detecting small deviations from the null. It can be easily shown that LR_0 maximize the average slope of the power function given the uniform weighting function for μ_τ and $\{p(\tau)\}_{\tau\in\Upsilon}$.

(iv) Monte Carlo simulations reported in Andrews, Lee and Ploberger (1996) suggest that the choice of the (arbitrary) constant c does not affect the outcome of the test. This statement seems reasonable in our case, too, because $(1+c)^{-(T-p)/2}$ also scales LR_c under H_0. Moreover, the factor $\frac{c}{1+c}$ changes very slowly as c increases.

(v) The distribution of LR_c under H_0 is difficult to derive, and the critical values will depend on the geometrical relations among the spaces V^τ. However, it can be simulated for the given regressors.

5 Other tests for structural change

In the preceeding Section an optimal test for structural changed has been derived. However, it was noticed that the power of the optimal test is strongly affected by the relative position of the subaspaces V^τ. Since the subspaces V^τ do not depend on the particular test chosen, they will affect all test procedures. This Section shows how some existing tests are affected.

5.1 The CUSUM and the CUSUM of squares tests

The CUSUM and CUSUM of squares test statistics of Brown, Durbin and Evans (1975) can be written in terms of the vector v as follows:

(i) The CUSUM test consists in rejecting the null of no structural change if $|a_r' v| > c_{r,\alpha}$, for at least one $r = p+1, ..., T$, where $c_{r,\alpha}$ is a constant depending on r and on the size, α, of the test (see Brown, Durbin and Evans (1975) Section 2.3), and a_r is a $T-p$ dimensional vector for which thefirst $r-p$ components are equal to $(T-p)^{-\frac{1}{2}}$ and the remaining $(T-r)$ are zero. If θ_r denotes the angle between v and a_r, the CUSUM test will reject the null of no structural change if $|\cos\theta_r| > c_{\alpha,r}$ for at least one r. Therefore, the critical region of the CUSUM test consists in the union of "caps" (of different dimensions) near the projection of a_r and $-a_r$ on S_{T-p-1}.

(ii) The CUSUM of squares test rejects the null hypothesis of no structural change if the angle θ_r between v and the space spanned by the columns of

$$B_r = \begin{pmatrix} I_{r-p} & 0 \\ 0 & 0 \end{pmatrix},$$

satisfies,

$$(\cos\theta_r)^2 > c_\alpha' + \frac{r-p}{T-p}$$

or

$$(\cos\theta_r)^2 < \frac{r-p}{T-p} - c_\alpha',$$

for some $r = p+1, ..., T$, where c_α' is a constant depending on the size α of the test.

The critical region or the CUSUM of squares test is the union of "strips" near the intersection of the subspace spanned by the columns of B_r and S_{T-p-1} and "strips" near the intersection of the subspace spanned by the column of $I_T - B_r$ and S_{T-p-1}.

Note that the CUSUM and CUSUM of squares identify respectively $T-p$ directions (the a_r) or subspaces (B_r). These are chosen independently of the subspaces V^τ, so that the critical regions of both tests could be quite far from such spaces V^τ. From Theorems 2 and 3, we would like the critical region to be close to (some of) these subspaces. This explains why the CUSUM and CUSUM of squares can have no power in detecting structural breaks as is well known in the literature (Garbade

(1977)). Note that their powers depend on the unknown parameter τ, even though their sizes do not. The failure of these tests stresses the importance of taking into account the subspaces V^{τ}.

The fluctuation test of Ploberger, Kramer and Kontrus (1987) has the same structure of the CUSUM and the CUSUM of squares tests as it can be written in terms of v and identifies $T-p$ directions on S_{T-p-1}. Although in the flucutation test such directions depend on the matrix of regressors (in contrat to the CUSUM and CUSUM of squares tests), it is not clear how they might be related to the subspaces V_{τ} and to the direction corresponding to the "true" alternative.

5.2 The $\sup F$ test

The $\sup F$ test represents an attempt to take into account the existence of the subspaces V^{τ}. This is based on Roy's union-intersection principle (Roy (1953)) which can be stated as follows: take as critical region of size α, ω, for testing $H_0 : y \sim N\left(X\beta, \sigma^2 I_T\right)$ against $H_1 : y \sim N\left(X\beta + Z\left(\tau\right)\gamma, \sigma^2 I_T\right)$, the union of the most powerful critical regions ω_{τ}, $\omega = \bigcup_{\tau \in \Upsilon} \omega_{\tau}$, of a properly chosen size α^*, for a fixed $\tau \in \Upsilon$. Since each ω_{τ} can be characterized in terms of the F test, Roy's union-intersection principle together with Lemma 2 in the Appendix produces the $\sup F$ test. The critical region of the $\sup F$ test is the union of "strips" on S_{T-p-1} along the intersections of S_{T-p-1} and the subspaces V^{τ}.

Remarks.

i) The $\sup F$ reduces to the classical F-test when the change points are known. In spite of this the $\sup F$ test *does not have* any optimal property.

ii) For a fixed α^*, the size of the $\sup F$ test depends in a complicated way upon the angles among the subspaces V^{τ}. For example, if ρ is the angle between V^{τ_1} and V^{τ_2}, the size of the test is

$$\alpha\left(\rho\right) = \begin{cases} 2\alpha^* & \text{if } \rho \geq \alpha^* \pi \\ \alpha^* + \rho & \text{If } 0 \leq \rho < \alpha^* \pi \end{cases} .$$

Clearly the size of the $\sup F$ test is in the interval $[\alpha^*, 1]$, however there is no easy way of measuring α. The use of Bonferroni-type inequalities could require the use of very small α^*, if we do not want ω to be the whole S_{T-p-1}. Moreover, asymptotic critical values for the $\sup F$ test (see for example Andrews (1993)), have been tabulated ignoring the geometry of the testing problem, i.e. the angles between the subspaces V^τ, and this can also result in serious size distortions in finite samples. However, the distance between the subspaces V^τ can be measured (see Section 7), and provides an indirect way of measuring α: if it is small, then the size of the $\sup F$ test will be approximately α^*, i.e. the size of the F test for any fixed τ.

iii) The area of the region ω (and thus the geometry of the V^τ) measures the variability of the optimal (in terms of average power) critical regions as τ varies under the alternative. Thus, the spatial relation between the spaces V^τ affects the power properties of the $\sup F$ test as well as its size. The F test for a fixed $\tau = \tau_1$ will not be very powerful (in terms of average power) if the true τ is τ_2 and V^{τ_1} is far from V^{τ_2}. As a consequence, the average power over τ_1 and τ_2 could be very low. On the other hand, if the angle between V^{τ_1} and V^{τ_2} is small, the average power over τ_1 and τ_2 will be almost the same as the power of the F test against any fixed τ, that is, the averaging process does not significantly reduce the power. Confirmation of this can be found in the simulation study of Table 1 of Andrews, Lee and Ploberger (1996), where the power of the $\sup F$ appears to be significantly less than the power of the F test for known changepoint, when the changepoint is near the beginning or the end of the sample (so that the subspace V^τ corresponding to the true τ is "far" from most of the other subspaces).

iv) In the previous Sections we have shown that the position of the subspace V^τ in $\mathbb{R}^{T-p}$ affects both the $\sup F$ and the LR_c tests. If such subspaces are close, the outcomes of both tests will be similar, and their interpretation will be easy. Note however, that in this case both critical regions will be approximately the critical region of an F test for a fixed τ. If this is the case, using a (simple) F test for any fixed τ will yield a test approximately as powerful as the (more complicated) $\sup F$ and LR_c tests. Tables 2, 3 and 4 of Andrews, Lee and Ploberger (1996) support this statement, as the midsample F test performs as well as the other tests they consider, if the structural change occurs approximately in the middle of the sample. If the subspaces V^τ are well spread out over $\mathbb{R}^{T-p}$, the $\sup F$ and the LR_c test have very different critical regions, and the outcomes of such tests could be totally different (see Figure 1). Therefore, there are problems in choosing a test. The $\sup F$ test does not have optimal properties, but its critical region covers the optimal critical region for the true unknown τ. On the other hand the LR_c test has some optimal properties,

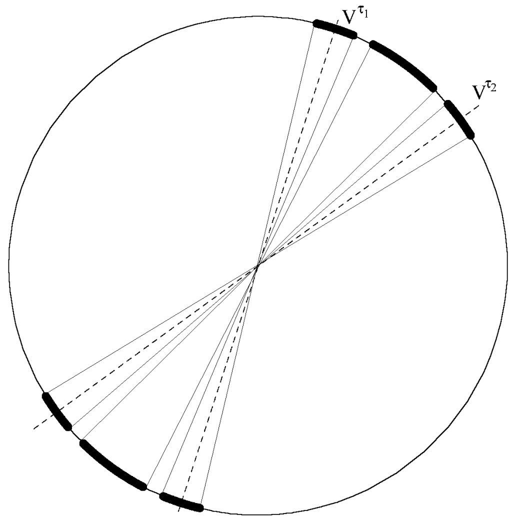

Figure 1: Critical regions on S_1 for the $\sup F$ and the LR_c tests of $H_0 : y \sim N(X\beta, \sigma^2 I_T)$ against $H_1 : y \sim N(X\beta + Z(\tau)\gamma, \sigma^2 I_T)$, $\tau = \tau_1, \tau_2$ ($T - p = 2$, $K = 6$, $c = 1$, $\alpha = 0.1$).

but its critical region could be quite far from the optimal critical region for the true unknown τ.

6 Local properties of tests for structural change

In Section 4 optimal tests for structural changed has been derived and Section 5 has compared these with existing tests which are used by practitioners. This Section studies the local properties of similar tests for structural change with reference to local optimality and unbiasedness.

The unconditional power of a similar critical region ω of size α is (Hillier (1986))

$$P_\omega = \frac{1}{2\pi^{\frac{T-p}{2}}} \exp\left\{-\frac{1}{2}\eta' Z(\tau)' P_X Z(\tau)\eta\right\} \int_\omega \sum_{j=0}^{\infty} \frac{\Gamma\left(\frac{T-p+j}{2}\right) 2^{\frac{j}{2}}}{j!} \left(v'C'Z(\tau)\eta\right)^j (dv), \tag{6}$$

where $\eta = \gamma/\sigma$ and (dv) denotes the unnormalized Haar measure on the $(T-p)$-unit sphere. By expanding P_ω as a Taylor series around $\eta = 0$, we obtain

$$\begin{aligned} P_\omega &= \alpha + \frac{\Gamma\left(\frac{T-p+1}{2}\right)}{2^{\frac{1}{2}}\pi^{\frac{T-p}{2}}} \int_\omega v' C' Z(\tau)\eta\,(dv) - \frac{1}{2}\alpha\eta' Z(\tau)' P_X Z(\tau)\eta \\ &+ \frac{T-p}{2} \frac{\Gamma\left(\frac{T-p}{2}\right)}{2\pi^{\frac{T-p}{2}}} \int_\omega \left(\eta' Z(\tau)' Cv\right)^2 (dv) + o(\eta'\eta) \end{aligned} \tag{7}$$

A level α locally most powerful critical region ω maximizes the second term in the expansion, leading to a critical region of the form $(v'C'Z(\tau)\eta) > c$. Note that this critical region depends on both η and τ, so that it is not uniformly most powerful in a neighbourhood of $\eta = 0$.

By requiring the test to be locally unbiased, the critical region must satisfy

$$\frac{\Gamma\left(\frac{T-p}{2}\right)}{2\pi^{\frac{T-p}{2}}} \int_\omega (v'C'Z(\tau)\eta)\,(dv) = 0 \tag{8}$$

$$\frac{\Gamma\left(\frac{T-p}{2}\right)}{2\pi^{\frac{T-p}{2}}} \int_\omega (dv) = \alpha \tag{9}$$

and

$$-\frac{1}{2}\alpha\eta' Z(\tau)' P_X Z(\tau)\eta + \frac{T-p}{2} \frac{\Gamma\left(\frac{T-p}{2}\right)}{2\pi^{\frac{T-p}{2}}} \int_\omega \eta' Z(\tau)' Cvv'C'Z(\tau)\eta\,(dv) \tag{10}$$

must be large and positive. Note that (8) will be satisfied everytime we choose ω so that $\Gamma\left(\frac{T-p}{2}\right) / \left(2\pi^{\frac{T-p}{2}}\right) \int_\omega \eta' Z(\tau)' Cvv'C'Z(\tau)\eta\,(dv)$ is large. A test satisfying these conditions does not exist uniformly in a neighbourhood of $\eta = 0$ since it depends on η and τ.

A different issue involves the existence of locally unbiased tests. First we assume τ known. Equation (8) can be written as

$$\left[\frac{\Gamma\left(\frac{T-p}{2}\right)}{2\pi^{\frac{T-p}{2}}} \int_\omega v\,(dv)\right]' C'Z(\tau)\eta = 0. \tag{11}$$

That is the average value of v in ω, $v_\omega = \Gamma\left(\frac{T-p}{2}\right) / \left(2\pi^{\frac{T-p}{2}}\right) \int_\omega v\,(dv)$, must be orthogonal to $C'Z(\tau)\eta$. Note that v_ω cannot be the zero vector: suppose it is, then it must be true that

$$\frac{\Gamma\left(\frac{T-p}{2}\right)}{2\pi^{\frac{T-p}{2}}} \int_\omega v' e_i\,(dv) = 0$$

for all unit vectors e_i, $i = 1, 2, ..., T-2$. This shows that ω must have the form of two caps around the point e_i and $-e_i$. Therefore not all components of v_ω can be simultaneously zero. So a locally unbiased test must satisfy (11) η and ,in general, it is not possible to chose v_ω ortogonal to all the subaspaces V^τ. This means we need to choose the critical region so that v is orthogonal to $C'Z(\tau)\eta$ for all possible η. For this choice of critical region, though, (10) might not be positive for all η. If we allow τ also to vary over Υ no unbiased test for structural change exists.

Since there is no locally unbiased test exist, we weaken again our optimality criterion by averaging the power on the surface of constant $\lambda = \sqrt{\eta'\eta}$ to obtain

$$\bar{P}_\omega = \alpha - \frac{\alpha\lambda^2}{2K}\operatorname{tr}\left[Z(\tau)' P_X Z(\tau)\right] + \frac{(T-p)\lambda^2}{2K}\frac{\Gamma\left(\frac{T-p}{2}\right)}{2\pi^{\frac{T-p}{2}}}\int_\omega v'C'Z(\tau)Z(\tau)'Cv\,(dv) + o\left(\lambda^2\right).$$

So, a necessary condition for a test to be locally unbiased is

$$\frac{\Gamma\left(\frac{T-p}{2}\right)}{2\pi^{\frac{T-p}{2}}}\int_\omega v'C'Z(\tau)Z(\tau)'Cv\,(dv) \geq \alpha\frac{\operatorname{tr}\left[Z(\tau)' P_X Z(\tau)\right]}{T-p}.$$

Note that, for a fixed τ, a locally average most powerful test on the surface of constant λ is to reject H_0 if $v'C'Z(\tau)Z(\tau)'Cv > c$, so that a sufficient (but not necessary) condition for unbiasedness is

$$c \geq \frac{\operatorname{tr}\left[Z(\tau)' P_X Z(\tau)\right]}{T-p}.$$

The constant c is a decreasing function of c, so that if α is large the test might not be locally unbiased in terms of average power.

For a fixed $\tau \in \Upsilon$, let $V_\perp^\tau$ denote the subspace of $\mathbb{R}^{T-p}$ orthogonal to V^τ. Then

$$\frac{\Gamma\left(\frac{T-p}{2}\right)}{2\pi^{\frac{T-p}{2}}}\int_{\omega\cap\left(V^\tau\cup V_\perp^\tau\right)} v'C'Z(\tau)Z(\tau)'Cv\,(dv) = \frac{\Gamma\left(\frac{T-p}{2}\right)}{2\pi^{\frac{T-p}{2}}}\int_{\omega\cap V^\tau} v'C'Z(\tau)Z(\tau)'Cv\,(dv)$$

because $v'C'Z(\tau) = 0$ in $V_\perp^\tau$. So a necessary condition for local unbiasedness is that $\omega \cap V^\tau$ is sufficiently large for all τ. When we allow τ to vary in Υ, this condition cannot be satisfied if the subspace V^τ are far from each other. So in general there are no unbiased tests for structural change at unknown changepoints.

In order to weaken the optimality condition even further we need to average the power of different spaces V^τ by $\{p(\tau)\}_{\tau\in\Upsilon}$, leading to

$$\begin{aligned}\bar{\bar{P}}_\omega &= \alpha - \frac{\alpha\lambda^2}{2K}\operatorname{tr}\left[\sum_{\tau\in\Upsilon} p(\tau) Z(\tau)' P_X Z(\tau)\right] + \\ &\quad \frac{(T-p)\lambda^2}{2K}\frac{\Gamma\left(\frac{T-p}{2}\right)}{2\pi^{\frac{T-p}{2}}}\int_\omega v'C'\left[\sum_{\tau\in\Upsilon} p(\tau) Z(\tau) Z(\tau)'\right] Cv\,(dv) + o\left(\lambda^2\right).\end{aligned}$$

In this case a locally average most powerful critical region is given by values of $v'C'\left[\sum_{\tau\in\Upsilon} p(\tau) Z(\tau) Z(\tau)'\right] Cv > c$, and the test will be unbiased provided the following inequality is satisfied

$$\frac{\Gamma\left(\frac{T-p}{2}\right)}{2\pi^{\frac{T-p}{2}}} \int_{\omega} v'C'\left[\sum_{\tau\in\Upsilon} p(\tau) Z(\tau) Z(\tau)'\right] Cv\,(dv) \geq \alpha \frac{\operatorname{tr}\left[\sum_{\tau\in\Upsilon} p(\tau) Z(\tau)' P_X Z(\tau)\right]}{T-p}.$$

A sufficient (but not necessary) conditon for a locally average most powerful test to be locally unbiased is that $c > \operatorname{tr}\left[\sum_{\tau\in\Upsilon} p(\tau) Z(\tau)' P_X Z(\tau)\right] / (T-p)$, which requires the size of the test to be small.

7 A measure of the intrinsic difficulty of testing for structural change

The power properties of any test for structural change is strongly affected by the relative position of the subspaces V^{τ}, $\tau\in\Upsilon$. If these spaces are far enough from each other, tests for structural change have potentially very low power and are not locally unbiased (even in terms of average power). Therefore it is important to measure the "distance" between them, and this can be considered, following Hillier (1995), as a measure of "intrinsic difficulty" of the testing problem, i.e. as a relative measure of the potential loss of power with respect to the situation where both number and location of the change points is known.

If the subspaces V^{τ} are straight lines through the origin, a measure of the intrinsic difficulty can be related to the angles between the V^{τ}, $\tau\in\Upsilon$. More precisely, we can take one of the straight lines as a *reference line*, $V^{\tau'}$, say, and measure the angles between $V^{\tau'}$ and all the other lines V^{τ}, $\tau\in\Upsilon$, $\tau\neq\tau'$. Suppose these angles are $\alpha_{\tau,\tau'}$, $\tau,\tau'\in\Upsilon$, $\tau\neq\tau'$. Then a measure of intrinsic difficulty is $d_{\Upsilon}=\max_{\tau,\tau'\in\Upsilon}\{2\alpha_{\tau,\tau'}/\pi\}$. If d_{Υ} is close to zero there is a small intrinsic difficulty, but if it is near one, the potential loss of power is large.

This idea can be generalized using the critical angles between spaces, i.e. the angles, along orthogonal directions, between two spaces such that the largest angle is the largest possible angle between any two arbitrary points in the two spaces.

Assume that all hyperplanes V^{τ} have the same dimension, and choose a particular

$\tau' \in \Upsilon$, and the associated subspace $V^{\tau'} = \left\{C'Z(\tau)\gamma : \gamma \in \mathbb{R}^K\right\}$. Then, project all $V^{\tau} \cap S_{T-p-1}$, $\tau \in \Upsilon$, on $V^{\tau'}$, by multiplying every element of $V^{\tau} \cap S_{T-p-1}$ by the projection matrix $P_{\tau'} = C'Z(\tau')\left[Z(\tau')' P_X Z(\tau')\right]^{-1} Z(\tau')' C$. It can be shown that the cosines of the critical angles between $V^{\tau'}$ and V^{τ} (i.e. the square of the canonical correlation coefficients between them) are the solutions to the equation (James, (1954))

$$\left| Z(\tau)' P_X Z(\tau') \left[Z(\tau')' P_X Z(\tau')\right]^{-1} Z(\tau')' P_X Z(\tau) - \lambda_\tau Z(\tau)' P_X Z(\tau) \right| = 0. \quad (12)$$

More precisely the cosine of the largest of the critical angles is the square root of the smallest nonzero eigenvalues of (12). Thus, a measure of intrinsic difficulty of the testing situation is $d_\Upsilon = \max_{\tau,\tau'\in\Upsilon}\{2\alpha_{\tau,\tau'}/\pi\}$, where $\alpha_{\tau,\tau'} = \arccos\left(\sqrt{\lambda^*_{\tau,\tau'}}\right)$ and $\lambda^*_{\tau,\tau'}$ is the smallest of the nonzero eigenvalues of the determinental equation (12). Again, a large d_Υ entails a potentially low power for any test for structural change.

If the hyperplanes are of different dimension, we have to choose τ' so that $V^{\tau'}$ is one of the spaces of smaller dimension. The measure of intrinsic difficulty d_Υ can be defined again as above.

Note that hyperplanes orthogonal to another hyperplane, might be orthogonal among themselves as well. The intrinsic difficulty of testing for structural change will certainly be large when no information at all is available regarding the time and the number of possible structural changes, because we can always find two spaces which are orthogonal.

To have an idea of how much the spaces V^{τ}, $\tau \in \Upsilon$, change in a practical application, we calculate the distance between these spaces for Example 7.9 in Greene (1991), for which Greene (1991) finds evidence of parameter instability using the CUSUM of squares test. These data are yearly U.S. data from 1966 to 1985, for the money stock M2 (dependent variable), GNP seasonally adjusted in 1982 constant dollars (x_{GNP}), and the interest rate (x_r). A constant is also included, and following Greene (1991) we assume no prior knowledge about the number and the position of the break points, and the parameters affected.

We will start by considering a very simple situation. Suppose that we know that there is only one possible change point, but its location is unknown. Table 1 gives $2\arccos((\lambda^*_{\tau,\tau'})^{1/2})/\pi$ for all possible partitions of 20 of length 2 excluding the possibility that the structural change occurs in the first and the last three observations.

The variability of d_Υ depends strongly on Υ. If $\Upsilon = \Upsilon_{4,16}$ contains all possible partitions in Table 1, then $d_{\Upsilon_{4,16}} = 0.75$ would indicate a large variability of the best critical regions for fixed τ to changes in τ. If we had some prior information suggesting a structural break at the beginning of the 1970's, we could restrict Υ to

$$\Upsilon_{4,10} = \{(4,16),(5,15),(6,14),(7,3),(8,12),(9,11),(10,10)\}$$

for which $d_{\Upsilon_{4,10}} = 0.37$, which is considerably lower than before. Note that d_Υ depends the number of partitions in it, and also on the values of the elements of the matrix of regressors ($d_{\Upsilon_{4,9}} = 0.37$, $d_{\Upsilon_{5,10}} = 0.13$ and $d_{\Upsilon_{6,11}} = 0.21$, even though the tree sets contain the same number of partitions). The partition $(10,10)$ is the one which is less far from every other partition in term of the index of intrinsic difficulty.

To verify how the power of the tests is affected, we have run a Monte Carlo simulation (based on 100000 replications) to compare the relative losses of power of the $\sup F$, LR_0, LR_1, analised in this paper, the avgf, and the expf1 tests of Andrews, Lee and Ploberger (1996) based on the test statistics

$$\begin{aligned}\text{avgf} &= \sum_{\tau\in\Upsilon} p(\tau)\, f_\tau \\ \text{expf1} &= 2^{-\frac{p}{2}} \sum_{\tau\in\Upsilon} p(\tau) \exp\left\{\frac{p f_\tau}{4}\right\}.\end{aligned}$$

F tests for a fixed τ equal or close to the true one are also considered. The model is generated with a structural break at $t = 6$,

$$y_t = \begin{cases} -3169.42 - 14.99223 x_{r,t} + 1.558815 x_{GNP,t} + 175.73 e_t & 1 \le t < 6 \\ -1.5\times 3169.42 - 1.5\times 14.99223 x_{r,t} + 1.5\times 1.558815 x_{GNP,t} + 175.73 e_t & t \ge 6 \end{cases}.$$

where the e_t are independent $N(0,1)$ (and the parameters are those estimated by Greene (1991)). Table 2 shows the power of the $\sup F$, LR_0, LR_1, the avgf, and the expf1 tests, when these are evaluated over two partitions sets $\Upsilon_{4,16}$ and $\Upsilon_{4,10}$. A uniform weight over τ is considered.

A few things are worth noting in Table 2. Although all the parameters are changed by 50% the power of the F test for known $\tau = (6,14)$ ($F_{(6,14)}$) equals .276 only since the sample size is only 20. The power of the $\sup F$ test is about 55% of the power of the $F_{(6,14)}$ test when $\Upsilon = \Upsilon_{4,16}$, and increases to about 70% when $\Upsilon = \Upsilon_{4,10}$. The expf1 test has power slightly larger than the sup F test but significantly less than the powers of the LR_0, $LR_1$and avgf tests. Restricting the set of partitions from $\Upsilon_{4,16}$ to $\Upsilon_{4,10}$ increases the powers of all tests considerably.

F tests for fixed τ close to the true one perform almost as well as the other tests if $\tau = (7,13), (8,12), (9,11)$. For $\tau = (4,16), (5,15)$ the power of F_τ is considerably less than the power of the $F_{(6,14)}$ test because there are very few observations in the first subset.

If nothing at all is known about the possible change points, then all partitions of 20 in at most 20 parts must be considered. In such cases it is enough to take some random partitions and if the angles between the subspaces V^τ are large we have to conclude that every test for structural change has low power. So we randomly generate ten partitions of 20 and compare the largest critical angles between the corresponding subspaces V^τ, $\tau \in \Upsilon$, on $V^{(6,14)}$. The results are summarized in Table 2, where we report $2\arccos((\lambda^*_{\tau,(6,14)})^{1/2})/\pi$ for these ten random partitions. Note that the index of intrinsic difficulty $d_\Upsilon = .99$, which suggests that testing for structural change knowing only that the structural breaks might be associated with these ten partitions, does not lead to very powerful results.

8 Conclusions

This paper has characterized similar tests for structural change in the linear model under the hypothesis of normality of errors with emphasis on their geometrical aspects. We have emphasized the idea that geometrical considerations can help to better understand existing tests and to find out why they can fail.

It has been shown that it is easy to characterize similar tests in terms of the vector $v \in S_{T-p-1}$. The geometry of existing tests (CUSUM, CUSUM of squares and sup F tests) has been analysed, and this has allowed us to understand the setups where such tests can be powerful. Tests optimal in a weak sense have also been suggested which are similar to those of Andrews, Lee and Ploberger (1996). It has been shown that their performances are strongly affected by the geometry of the testing situation.

There is an intrinsic difficulty in testing for structural change, due to the existence of several spaces V^τ, $\tau \in \Upsilon$, and this affects all tests. Unfortunately there is nothing we can do to overcome this problem, apart from acknowledging its existence. In order

	4,16	5,14	6,14	7,13	8,12	9,11	10,10	11,9	12,8	13,7	14,6	15,5
5,15	0.00	...										
6,14	0.00	0.00	...									
7,13	0.02	0.00	0.00	...								
8,12	0.33	0.10	0.00	0.00	...							
9,11	0.37	0.13	0.02	0.00	0.00	...						
10,10	0.37	0.13	0.02	0.01	0.00	0.00	...					
11,9	0.43	0.25	0.21	0.02	0.02	0.00	0.00	...				
12,8	0.50	0.39	0.33	0.06	0.06	0.01	0.00	0.00	...			
13,7	0.66	0.56	0.46	0.30	0.29	0.23	0.05	0.00	0.00	...		
14,6	0.74	0.59	0.51	0.49	0.47	0.31	0.09	0.02	0.00	0.00	...	
15,5	0.75	0.59	0.51	0.49	0.47	0.33	0.12	0.04	0.01	0.00	0.00	...
16,4	0.75	0.59	0.51	0.49	0.47	0.33	0.13	0.04	0.01	0.01	0.00	0.00

Table 1: Measure of intrinsic difficulty for partitions of 20 of length 2.

			$\Upsilon_{4,16}$		$\Upsilon_{4,10}$	
test	Power	Relative Power	Power	Relative Power	Power	Relative Power
sup F			.151	54.71%	.194	70.29%
LR_0			.191	69.20%	.219	79.35%
LR_1			.190	68.84%	.224	81.16%
avgf			.190	68.84%	.224	81.16%
expf1			.155	56.16%	.199	72.10%
$F_{(4,16)}$	.043	15.58%				
$F_{(5,15)}$	.110	39.86%				
$F_{(6,14)}$	.276	100%				
$F_{(7,13)}$	.217	78.62%				
$F_{(8,12)}$	.209	75.72%				
$F_{(9,11)}$	.179	64.86%				

Table 2: Power of tests for structural change.

Partition τ	$\frac{2\arccos\left(\sqrt{\lambda^*_{\tau,\tau'}}\right)}{\pi}$
$(9,3,5,1,2)$	0.8469
$\tau' = (6,14)$	0
$(8,5,1,1,2,3)$	0.2061
$(9,6,5)$	0.8469
$(15,3,2)$	0.9861
$(17,1,2)$	0.9917
$(16,1,2,1)$	0.9901
$(5,1,14)$	0
$(13,1,3,3)$	0.9835
$(9,3,7,1)$	0.8469

Table 3: Measure of intrinsic difficulty for ten random partitions of 20. The reference hyperplane is indexed by the partition (6,14)

to measure its potential impact on the tests we have suggested reporting an index of intrinsic difficulty.

The main message from our study is not particularly original but it is important: tests for structural change can have very low power, so it is important in practical applications to use all the information available. This message also generalizes to other testing setups involving parameters under the alternative which are not present under the null: sensible testing requires a priori restrictions of the alternative hypothesis. Testing against the "most general" alternative is inappropriate.

9 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. and W. Ploberger (1994), "Optimal Tests when a Nuisance Parameter is Present only under the Alternative", *Econometrica*, 62, 1383-1414.

Andrews, D.W.K., I. Lee, W. Ploberger (1996), "Optimal Changepoint Tests for Normal Linear Regression", *Journal of Econometrics*, 70, 9-36.

Breusch, T.S. (1986), "Hypothesis Testing in Unidentified Models", *Review of Economic Studies*, LIII, 635-651.

Brown, R.L., J. Durbin and 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 Matrix Decompositions on the Stiefel Manifold", *Journal of Multivariate Analysis*, 36, 145-162.

Forchini, G. (1998), "On Diagnostic Tests", Mimeo.

Garbade, K. (1977), "Two Methods for Examining the Stability of Regression Coefficients", *Journal of the American Statistical Association*, 72, 54-63.

Greene, W.H. (1991), *Econometric Analysis*, Macmillan Publishing Company, New York.

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

Hillier, G.H. (1991), "On Multiple Diagnostic Procedures for the Linear Model", *Journal of Econometrics*, 47, 47-66.

Hillier, G.H. (1995), "The Union of Best Critical Regions: Complete Classes, p-Values, and Model Curvature", *Discussion Paper in Economics and Econometrics*, No. 9503, Department of Economics, University of Southampton.

James, A.T. (1954), "Normal Multivariate Analysis and the Orthogonal Group", *The Annals of Mathematical Statistics*, 25, 40-75.

Pesaran, M.H., R.P. Smith and J.S. Yeo (1984), "Testing for Structural Stability and Predictive Failure: A Review", *The Manchester School of Economics and Social Studies*, LII, 280-295.

Ploberger, W., W. Krämer and K. Kontrus (1989), "A New Test for Structural Stability in the Linear Regression Model", *Journal of Econometrics* 40, 307-318.

Roy, S.N. (1953), "On a Heuristic Method of Test Construction and Its Use in Multivariate Analysis", *The Annals of Mathematical Statistics*, 24, 220-238.

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 two Lemmas, and the proofs of Theorems 2 and 3.

Lemma 1 *(Chikuse, 1991) Any vector $v \in S_{n-1}$ can be written uniquely as*

$$v = \cos\theta C_1 v_1 + \sin\theta C_2 v_2 \tag{13}$$

where $v_1 \in S_{k-1}$, $v_2 \in S_{n-k-1}$, and $0 < \theta < \frac{\pi}{2}$; 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 $C_1'C_2 = 0$; 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

$$(dv) = (\cos\theta)^{k-1}(\sin\theta)^{n-k-1}(dv_1)(dv_2)\,d\theta. \tag{14}$$

Lemma 2 *(Hillier, 1995) Let $\omega = \bigcup_{\tau\in\Upsilon}\omega_\tau$ with $\{v : t_\tau(v) > c\}$ where $t_\tau(v)$ is a continuous function of v, and Υ is a set of indexes. Then $\omega = \{v : \sup_{\tau\in\Upsilon} t_\tau(v) > c\}$.*

Proof. If there is an index τ^* such that $t_{\tau^*}(v) > c$, then, $\sup_{\tau\in\Upsilon} t_\tau(v) \geq t_{\tau^*}(v) > c$, so $\omega = \bigcup_{\tau\in\Upsilon}\{v : t_\tau(v) > c\} \subseteq \{v : \sup_{\tau\in\Upsilon} t_\tau(v) > c\}$. On the other hand if $v \in \{v : \sup_{\tau\in\Upsilon} t_\tau(v) > c\}$ then the fact that $t_\tau(v)$ is a continuous function of v guarantees that there is an index τ^* such that $t_{\tau^*}(v) = \sup_{\tau\in\Upsilon} t_\tau(v) > c$, i.e. $v \in \omega_{\tau^*} \subseteq \omega$. So $\{v : \sup_{\tau\in\Upsilon} t_\tau(v) > c\} \subseteq \omega$. ■

Proof of Theorem 2

The unconditional average power over the surface of constant λ_τ of a critical region ω is

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

By averaging it over the partition $\tau \in \Upsilon$ it becomes

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

where $C'Z(\tau)\gamma/\sigma = \lambda_\tau^{1/2}\Lambda_\tau\mu_\tau$, $\lambda_\tau = \frac{1}{\sigma^2}\gamma Z(\tau)' P_X Z(\tau)\gamma > 0$, $\mu_\tau \in S_K$, and $\Lambda_\tau = C'Z(\tau)\left[Z(\tau)' P_X Z(\tau)\right]^{-\frac{1}{2}}$ is a fixed $T-p \times K$ matrix such that $\Lambda'\Lambda = I_K$, and C is a $T \times T-p$ matrix such that $CC' = P_X$, $C'C = I_{T-p}$ and $C'X = 0$.

Using Lemma, 1 v can be written uniquely as $v = \cos\theta_\tau\Lambda_\tau v_{1\tau} + \sin\theta_\tau C_2^\tau v_{2\tau}$, where $v_{1\tau} \in S_{K-1}$, $v_{2\tau} \in S_{T-p-K-1}$, and $0 < \theta_\tau < \frac{\pi}{2}$. Since $v_{1\tau}$ and $v_{2\tau}$ are uniformly distributed over S_{K-1} and $S_{T-p-K-1}$, 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,

$$\frac{2\Gamma\left(\frac{T-p}{2}\right)}{\Gamma\left(\frac{K}{2}\right)\Gamma\left(\frac{T-p-K}{2}\right)} \sum_{\tau\in\Upsilon} p\left(\tau\right)\exp\left\{-\frac{1}{2}\lambda_\tau\right\} \tag{15}$$

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

The statement of the theorem follows. ■

Proof of Theorem 3.

Averaging with respect to λ_τ in (15) in the proof of Theorem 2, where the weight for λ_τ/c is the density function for a random variable having the chi-square distribution with K degrees of freedom, χ^2_K, yields

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

The statement of the theorem follows. ■

The power of any critical region ω is

$$\bar{P}_\omega = \alpha - \frac{1}{2}\alpha\sum_{\tau\in\Upsilon} p\left(\tau\right)\lambda_\tau + \frac{T-p}{2k}\sum_{\tau\in\Upsilon} p\left(\tau\right)\lambda_\tau \frac{\Gamma\left(\frac{T-p}{2}\right)}{2\pi^{\frac{T-p}{2}}}\int_\omega \left(v'\Lambda_\tau\Lambda_\tau' v\right)\left(dv\right) + \sum_{\tau\in\Upsilon} p\left(\tau\right) O\left(\lambda_\tau^2\right)$$