Optimal tests for nested model selection with underlying parameter instability

Barbara Rossi[†]

Abstract

This paper develops optimal tests for model selection between two nested models in the presence of underlying parameter instability. These are joint tests for both parameter instability and a null hypothesis on a subset of the parameters. They modify the existing tests for parameter instability to allow the parameter vector to be unknown. These tests statistics are useful if one is interested in testing a null hypothesis on some parameters but is worried about the possibility that the parameters may be time varying. The paper provides the asymptotic distributions of this class of test statistics and their critical values for some interesting cases.

JEL: C52, C53. Keywords: parameter instability, model selection.

[†] Barbara Rossi, Department of Economics, Duke University, P.O. Box 90097, Durham, NC 27708 USA. E-mail: brossi@econ.duke.edu; phone: (919) 660 1801; fax: (919) 684 8974. First draft: November 2000. This draft: February, 2004

Acknowledgments: I would like to thank M. Watson for the idea of this paper and for numerous discussions, suggestions, comments and teaching. I am grateful to T. Clark, G. Chow, R. Gallant, F. Sowell, N. Swanson, E. Tamer, A. Tarozzi and seminar participants at the University of Virginia, ECARES Universite' Libre de Brussels, the 2001 Triangle Econometrics Conference, the 2002 NBER Summer Institute, the 2003 Summer Meetings of the Econometric Society, and the 2003 EC^2 Conference for comments. Financial support from the IFS Summer Research, Princeton University, is gratefully acknowledged. The most recent draft of this paper can be downloaded from: http://www.econ.duke.edu/~brossi/. All mistakes are mine.

1. Introduction

This paper develops optimal tests for model selection between two nested models in the presence of underlying parameter instability in the data. The model selection procedure considered in this paper is hypothesis testing; in fact, when the competing models are nested, the problem of testing which model is best among the two is to test the significance of additional variables that are present only under the largest model. The tests proposed in this paper thus *jointly* test for *both* parameter instability *and* a null hypothesis on a subset of the parameters.

The main contribution of this paper is to address *simultaneously* the two problems of testing parameter instability and model selection among nested models. It is argued that tests for model selection fail to detect parameter instability and that tests for parameter instability are not designed to choose between nested models. If the goal is to jointly test parameter stability and select a model, then it is possible to identify a class of optimal tests. The optimal tests modify existing tests for parameter instability in order to allow them to reject the incorrect model. This is achieved by *imposing*, rather than *estimating*, the parameters of interest under the null, thus making the statistic *not* invariant to shifts in these parameters.

The tests presented in this paper are useful in situations in which one is interested not only in whether the explanatory variables proposed by some economic model are statistically significant in explaining the observed data, but also in whether this relationship is stable over time. For example, these tests would be useful if one is interested in testing whether inflation or exchange rates are random walks but is also worried about the possibility that parameters may be varying over time; see Clark and McCracken (2002) and Rossi (2003).

The strand of research closest to this paper is that concerning tests for parameter instability, in particular the works by Chow (1960), Quandt (1960), Ploberger et al. (1990 and 1992), Andrews (1993), Andrews and Ploberger (1994), Sowell (1996), Ghysels and Hall (1998) and Elliott and Müller (2002). However, these tests are designed to detect parameter instability *only,* whereas this paper is also concerned about testing hypotheses on the parameter vector and, hence, treats it as *unknown.*

An alternative way to deal with model selection issues in the presence of parameter instability is to do a two-stage procedure: first test whether there is parameter instability, then test which model, among the competing ones, is the best description of the data. In some special cases analyzed in this paper, that is for the special weighting distributions over the local alternatives analyzed in section 3, the test statistics in the two stages are asymptotically independent. In this case, it is easy to fix the size in each stage of the procedure so that the two-stage procedure will have an overall correct size asymptotically. However, this result is not true for general weighting distributions. In addition, two-stage tests have advantages and disadvantages. The advantage is that if we reject we know which part of the alternative we reject; the disadvantage is that the test will not have the optimal weighted average power for alternatives that are equally likely.

The paper is organized as follows. Section 2 derives the optimal tests for testing the joint hypothesis of parameter stability and model selection and provides their asymptotic distribution. Section 3 discusses special tests and reports their asymptotic critical values, and Section 4 compares their asymptotic local powers. Section 5 concludes. Proofs of the Results are in Appendix 1, whereas Appendix 2 contains the tables of asymptotic critical values.

2. Model selection in the presence of underlying parameter instability

2.1 Heuristics

In order to gain some intuition about the results in this paper, consider a simple example where the Data Generating Process (DGP) is the following and the time of the break is known:

$$y_t = \beta_t + \epsilon_t \; ; \; \beta_t = \begin{cases} \beta_1 \text{ for } t = 1, 2, ..\tau \\ \beta_2 \text{ for } t = \tau + 1, ...T \end{cases} ; \qquad \epsilon_t \sim iid\ N(0, \sigma_\epsilon^2) \tag{1}$$

If the researcher is interested in testing whether the parameter β_t is constant over time and equal to a specific value β_0, a possible test statistic would be:

$$Chow_T^* = \frac{\sum_{t=1}^{T} (y_t - \beta_0)^2 - \left(\sum_{t=1}^{\tau} \left(y_t - \widehat{\beta}_1\right)^2 + \sum_{t=\tau+1}^{T} \left(y_t - \widehat{\beta}_2\right)^2 \right)}{\frac{1}{T} \left(\sum_{t=1}^{\tau} \left(y_t - \widehat{\beta}_1\right)^2 + \sum_{t=\tau+1}^{T} \left(y_t - \widehat{\beta}_2\right)^2 \right)} \tag{2}$$

where $\widehat{\beta}_1 = \frac{1}{\tau}\sum_{t=1}^{\tau} y_t$ and $\widehat{\beta}_2 = \frac{1}{T-\tau} \sum_{t=\tau+1}^{T} y_t$ are the sample averages of y_t in the two sub-samples. By adding and subtracting the full sample average $\widehat{\beta} = \frac{1}{T}\sum_{t=1}^{T} y_t$ inside the square of the first addend on the numerator, (2) can be rewritten as:

$$Chow_T^* = \frac{T\left(\widehat{\beta} - \beta_0\right)^2}{\widehat{\sigma}_\epsilon^2} + \frac{\sum_{t=1}^{T}\left(y_t - \widehat{\beta}\right)^2 - \left(\sum_{t=1}^{\tau}\left(y_t - \widehat{\beta}_1\right)^2 + \sum_{t=\tau+1}^{T}\left(y_t - \widehat{\beta}_2\right)^2\right)}{\frac{1}{T}\left(\sum_{t=1}^{\tau}\left(y_t - \widehat{\beta}_1\right)^2 + \sum_{t=\tau+1}^{T}\left(y_t - \widehat{\beta}_2\right)^2\right)} \tag{3}$$

where $\widehat{\sigma}_\epsilon^2 \equiv \frac{1}{T}\left(\sum_{t=1}^{\tau}\left(y_t - \widehat{\beta}_1\right)^2 + \sum_{t=\tau+1}^{T}\left(y_t - \widehat{\beta}_2\right)^2\right) \underset{p}{\rightarrow} \sigma_\epsilon^2$. Thus, the test is decomposed in two components: the one on the left is a test on β and the one on the right is the standard Chow test for structural break. Hence, the test achieves power in detecting deviations from β_0 by adding to the traditional test for structural break a component that is variant to constant shifts in the mean. The asymptotic distribution of the test can easily be found in this case because the two components are independent.[1] Let $B_1(.)$ denote a scalar Brownian Motion and $BB_1(.)$ denote a scalar Brownian Bridge and $\pi = [\tau/T]$. Note that the first component on the right hand side in (3) is asymptotically the square of a standardized normal ($B_1(1)^2$) whereas the distribution of the second component is known from Andrews (1993) to be $BB_1(\pi)^2/\pi(1-\pi)$. As a result, the asymptotic distribution of this modified Chow test will be:

$$Chow_T^* \Rightarrow B_1(1)^2 + \frac{BB_1(\pi)^2}{\pi(1-\pi)} \tag{4}$$

Thus, the first component, which makes the test powerful in detecting constant shifts in the mean, adds a chi-square component to the limiting distribution of a standard Chow test for parameter instability. This example provides an easy and intuitive explanation of the asymptotic distribution of the tests considered in this paper.

2.2 Framework

This section describes the class of models considered in this paper and the assumptions under which the results are valid. The parametric model applies to a stationary and ergodic time series process:

---

[1] In fact, the standard Chow test can be rewritten as a Wald test: $\frac{\tau(T-\tau)}{T}\frac{\left(\widehat{\beta}_1-\widehat{\beta}_2\right)^2}{\sigma_\epsilon^2} + op(1)$ and $\widehat{\beta}$ and $\left(\widehat{\beta}_1 - \widehat{\beta}_2\right)$ are independent. To see why, note that $cov(\widehat{\beta}_1-\widehat{\beta}_2, \widehat{\beta}) = cov(\widehat{\beta}_1-\widehat{\beta}_2, \pi\widehat{\beta}_1+(1-\pi)\widehat{\beta}_2) = \pi var(\widehat{\beta}_1)-(1-\pi)var(\widehat{\beta}_2) = \pi\frac{1}{\pi} - (1-\pi)\frac{1}{1-\pi} \;= 0$

Assumption 1: *For each T, the sequence* $\{x_{t,T}\}_{t=1}^{T}$ *consists of the first T elements of an r-dimensional stationary and ergodic process. The parameter space* Θ *is a compact subset of* R^k*. For notational simplicity,* x_t *will be used to denote* $x_{t,T}$*.*

The class of local alternatives allows both for structural changes and for nonlinear hypotheses on the parameters.

Assumption 2: *The local alternatives are specified as:*

$$\theta_{t,T} = \theta^* + \frac{1}{\sqrt{T}} g\left(\gamma, \pi, \frac{t}{T}\right) \tag{5}$$

$$a\left(\theta^*\right) = \frac{1}{\sqrt{T}} \theta_A \tag{6}$$

where: $g(\gamma, \pi, s)$*, for* $s \in [0,1]$*, is a k-dimensional step function,* $\gamma \in R^i$*,* $\pi \in (0,1)^j$ *denotes the times of the structural changes as fractions of the sample size (j being the number of such breaks);* $a\left(\theta^*\right) = 0$ *is a possibly nonlinear restriction that identifies the true parameter value under the null hypothesis when there is no structural change, and* θ_A *denotes its local alternative.*

Hence, the parameter θ is unknown and possibly time-varying. The class of estimators considered here are extremum estimators that minimize the objective function $Q_T(\theta)$, which depends on both the data and the sample size. The focus will be on the restricted estimator $\widetilde{\theta}$:

$$\widetilde{\theta} = \arg\min_{\theta \in \Theta} \widehat{Q}_T(\theta) \text{ s.t. } a\left(\theta\right) = 0, \text{ where } \widehat{Q}_T(\theta) \equiv F_T(\theta)' W_T F_T(\theta) \tag{7}$$

where $F_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} f(x_t, \theta)$ is the sample analogue of $E\left(f(x_t, \theta)\right)$, the moment condition that is equal to zero at the true parameter value and $E\left(.\right)$ is the expected value function. The moment condition is such that $f : R^r \times R^k \rightarrow R^m$ and W_T is a (sequence of) positive semi-definite matrices.

The next assumptions are sufficient to ensure consistency of the estimator, its identification and asymptotic normality. Furthermore, the class of estimators is restricted to efficient GMM estimators, and Assumption 6 provides a sufficient condition for efficiency.

Assumption 3 (Identification): $\lim_{T \rightarrow \infty} E[f(x, \theta)] = 0$ *only if* $\theta = \theta^*$*.*

Assumption 4 (Consistency and Asymptotic Normality): *(i)* $\theta^* \in interior(\Theta)$*; (ii)* $f(x, \theta)$ *is continuously partially differentiable in a neighborhood* Υ *of* θ^**,* $\forall \theta \in \Theta$*; (iii) The functions*

$f(x,\theta)$ and $\nabla_\theta f(x,\theta)$ are measurable functions of x $\forall\theta \in \Theta$ and $E[||f(x,\theta^*)||^2]$ is finite; (iv) $E[f(x_t,\theta^*)] = 0, E[f(x_t,\theta^*)'f(x_t,\theta^*)] < \infty$ and $\sup_{\theta\in\Theta} ||f(x,\theta)|| < \infty$ $\forall t = 1,...T$ and $T = 1,2,...$ Each element of $f(x_t,\theta_{t,T})$ is uniformly square integrable $\forall t = 1,...T$ and $T = 1,2,...$; (v) for $M = \underset{T\to\infty}{plim}\nabla_\theta f(x,\theta^*) \in R^{m\times k}$ of full column rank, $M'W_T M$ is non-singular; (vi) $\{x_t\}$ is strong mixing with strong mixing coefficients $\left\{\alpha(n)^{1-2/\beta}\right\} < \infty$ with $\beta > 2$ and the individual elements of $f(x_t,\theta_{t,T})$ have finite absolute moments $E[|f^{(i)}(x_t,\theta_{t,T})|^\beta]$ for $i = 1,..m$.

Assumption 5 (Constraints): *$a(\theta)$ is continuously partially differentiable in a neighborhood Υ of θ^*, $\forall\theta \in \Theta$; $A \equiv \nabla_\theta a(\theta^*) \in R^{r\times k}$ has rank $r \leq k$.*

Assumption 6 (Efficiency in the class of GMM estimators): *The asymptotic variance of the GMM estimator is efficient in the class of GMM estimators: $\left\{W_T^{-1}\right\}_{T=1}^{\infty} \underset{p}{\to} \Sigma \equiv \lim_{T\to\infty} E\left[TF_T(\theta^*)F_T(\theta^*)'\right] \in R^{m\times m}$.*

When the alternative hypothesis of interest is *either* (5) *or* (6) then optimal tests are available. In the former case, an optimal test when the break date is known is the Chow (1960) test and, when the break date is unknown, a class of tests with optimal weighted average power is that of Andrews and Ploberger (1994).[2] In case the alternative is (6) *only*, the Likelihood Ratio test (and the asymptotically equivalent Wald and Lagrange Multiplier tests) is asymptotically locally most powerful among all invariant tests and, hence, it is optimal (see Engle, 1984).

However, when *both* hypothesis are of interest then considering separately tests for parameter instability and Likelihood Ratio tests is not sufficient anymore. This paper identifies a class of tests that are optimal, in the sense of having the highest asymptotic local power function for some specified alternatives. This class of tests is discussed in the next sub-section.

2.3 Optimal tests

We are interested in constructing a Lagrange Multiplier test statistic for testing jointly alternatives (5) and (6). The test builds on partial sums of the form:

$$F_{sT}(\widetilde{\theta}) \equiv \frac{1}{T}\sum_{t=1}^{[sT]} f(x_t,\widetilde{\theta}), \qquad s \in [0,1] \tag{8}$$

[2] Although Andrews' Sup-LR test (see Andrews, 1993) is not a member of that class, Andrews and Ploberger (1993) show its asymptotic admissibility against alternatives that are sufficiently distant form the null hypothesis.

where the partial sums are evaluated at the restricted estimator vector, $\widetilde{\theta}$. When assumptions 1-6 are satisfied, the asymptotic distribution of the partial sums of sample moments under the null and the alternative hypotheses is stated in Results 1 and 2. For notational convenience, let $\overline{M} \equiv \Sigma^{-1/2}M \in R^{m\times k}$ and partition it as: $\overline{M} = \left(\overline{M}_\beta \ , \ \overline{M}_\delta \right)$. Also, let "$\Rightarrow$" denote weak convergence to the relevant stochastic process, and "$\underset{p}{\rightarrow}$" denote convergence in probability.

Result 1. Distribution under the alternative hypothesis: *If assumptions 1-6 are satisfied, then:*

$$\begin{aligned}\sqrt{T}W_T^{1/2}F_{sT}(\widetilde{\theta}) \quad &\Rightarrow \quad Z\left(s\right) \equiv B_m\left(s\right) - s\overline{H}B_m\left(1\right) - s\overline{M}D'\theta_A \\ &\qquad - \overline{M}\int_0^s g\left(\gamma,\pi,r\right)dr + s\overline{M}B^{-1/2}HB^{1/2}\int_0^1 g\left(\gamma,\pi,r\right)dr\end{aligned} \tag{9}$$

where $B_m(\cdot)$ *is an m-dimensional standard Brownian Motion,* $D' \equiv B^{-1}A'\left(AB^{-1}A'\right)^{-1}$, $B \equiv \overline{M}'\overline{M}$, $H \equiv I_k - B^{-1/2}A'\left(AB^{-1}A'\right)^{-1}AB^{-1/2}$, I_k *is a k-dimensional identity matrix,* $\overline{H} \equiv \overline{M}B^{-1/2}HB^{-1/2}\overline{M}'$, *and both* H *and* $\overline{H}$ *are idempotent with rank equal to* $(k-r)$.

See Appendix 1 for proofs. Result 2 shows the asymptotic distribution of the standardized moment condition under the null hypothesis that there is no parameter instability in any of the coefficients and that a subset of parameters satisfies some restriction condition:

Assumption 7 (Null hypothesis): *Under the null hypothesis:* $\theta_{t,T} = \theta^*$ *for all* t, T.

Result 2. Distribution under the null hypothesis: *If assumptions 1, 3-7 are satisfied then:*

$$CW_T^{1/2}\sqrt{T}F_{sT}(\widetilde{\theta}) \Rightarrow \begin{pmatrix} BB_{k-r}(s) \\ B_r(s) \\ B_{m-k}(s) \end{pmatrix} \tag{10}$$

for an orthonormal matrix C *such that* $\overline{H} = C'\Lambda C$, $CC' = I_m$ *and* $\Lambda = \begin{pmatrix} I_{k-r} & 0 \\ 0 & 0 \end{pmatrix}$. $BB_{k-r}(s)$ *is a (k-r)-dimensional Brownian Bridge and* $[B_r(s)', B_{m-k}(s)']'$ *is an (m-k+r)-dimensional vector Brownian Motion. The Brownian Motions and the Brownian Bridges are independent.*

Note that, under the null hypothesis, the asymptotic distribution of the standardized partial sum of moment conditions is composed by both Brownian Bridges and Brownian Motions. The $(k-r)$-dimensional Brownian Bridge component derives from the parameters that are not specified under the null. In fact, this component is a partial sum of mean zero moment conditions, where the zero mean is obtained by estimating the drift.[3]

The alternative hypothesis will add drift components to the moment conditions, as Result 1 shows. In particular, the drift components originate both from deviations from the parameter stability hypothesis and from deviations from the specified null hypothesis on the value of the parameters. For the local alternatives considered in this paper, the normalized partial sum of the sample moments evaluated under the null hypothesis converges to a stochastic process denoted by $Z(s)$. Under the local alternative, $Z(s)$ satisfies the following stochastic differential equation:

$$dCZ(s) = \begin{pmatrix} dBB_{k-r}(s) \\ dB_{m-k+r}(s) \end{pmatrix} + Cv(s)ds \tag{11}$$

where $v(s) \equiv -\overline{M}D'\theta_A - \overline{M}g(\gamma,\pi,s) + \overline{M}B^{-1/2}HB^{1/2}\left(\int_0^1 g(\gamma,\pi,r)dr\right)$. Under the null hypothesis, the same expression holds with $v(s) = 0$. To get some insight, rearrange (9):[4]

$$\begin{aligned} Z(s) \;&=\; \left(I_m - \overline{H}\right)\left(B_m(s) - \overline{M}\left(\int_0^s g(\gamma,\pi,v)dv\right) - s\overline{M}D'\theta_A\right) + \\ &\quad + \overline{H}\left(BB_m(s) - \overline{M}\int_0^s \left(g(\gamma,\pi,v) - \left(\int_0^1 g(\gamma,\pi,r)dr\right)\right)dv\right) \end{aligned} \tag{12}$$

[3] For example, when the process is univariate and such that: $y_t = \beta_0 + \epsilon_t$, $\epsilon_t \sim iid\ N(0,1)$, $\beta_0 = 0$ (as in the introductory example at the beginning of the paper) then the partial sum of moments is: $\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]}\left(y_t - \widehat{\beta}\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]} y_t - \frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]}\frac{1}{T}\sum_{t=1}^{T} y_t = \frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]} y_t - s\frac{1}{\sqrt{T}}\sum_{t=1}^{T} y_t$ and the origin of the Brownian Bridge is evident. If there were no restrictions under the null hypothesis, then the asymptotic distribution of $CW_T^{1/2}\sqrt{T}F_{sT}(\beta_0,\widetilde{\delta})$ would be $(BB_q(s)'\ B_{m-q}(s)')'$, which is Sowell (1996) result. When there are restrictions on a subset of p parameters under the null hypothesis, these will show up as p-Brownian Motions, in addition to the previous components. These are Brownian Motions because they are the limiting distribution of a partial sum of mean zero moment conditions, where the zero mean is obtained by *imposing*, rather than *estimating*, the drift. In the previous example, in this case the partial sum of moments is: $\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]}(y_t - \beta_0) = \frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]} y_t$ and the origin of the Brownian Motion is clear. The $B_{m-q}(s)$ component corresponds to the over-identified moment restrictions.

[4] The result follows because: $\overline{M}B^{-1/2}HB^{1/2} = \overline{M}B^{-1/2}HB^{-1/2}\overline{M}'\overline{M} = \overline{HM}$, $\overline{HM}D' = 0$ and $\left(I - \overline{H}\right)\overline{M}D' = \overline{M}D'$, which can be verified by direct calculations.

so that:

$$\begin{aligned} dCZ(s) &= (I-\Lambda)C\left[dB_m(s)-\overline{M}g(\gamma,\pi,s)-\overline{M}D'\theta_A\right]+ \\ &+\Lambda C\left[dBB_m(s)-\overline{M}\left(g(\gamma,\pi,s)-\left(\int_0^1 g(\gamma,\pi,r)dr\right)\right)\right] \\ &= \begin{pmatrix} dBB_{k-r}(s) \\ dB_{m-k+r}(s) \end{pmatrix} - \begin{pmatrix} C^{(1)}\overline{M}\left(g(\gamma,\pi,s)-\left(\int_0^1 g(\gamma,\pi,r)dr\right)\right) \\ C^{(2)}\overline{M}\left(g(\gamma,\pi,s)+D'\theta_A\right) \end{pmatrix} \end{aligned}$$

where $C^{(1)}$ and $C^{(2)}$ are, respectively, the first $(k-r)$ and the last $(m-k+r)$ rows of C. Thus, the null hypothesis puts restrictions on *both* the Brownian Motions and the Brownian Bridge components. In fact, it is a joint hypothesis on parameter instability (affecting the Brownian Bridge component) and on the parameters (affecting the Brownian Motion component). This differs from Sowell (1996) case (see the discussion below his eq. (3), pag. 1091), where the alternative *only* places restrictions over Brownian Bridges. However, Sowell (1996) derived optimal tests in terms of the Radon-Nikodym derivative of the measure implied by the null hypothesis for *both* the Brownian Motion and the Brownian Bridge components (see the proof of his Theorem 3) so we can apply a similar argument. Thus, the test with the greatest weighted average power, according to some weighting functions $R(\eta,\pi)$ (on η for every π) and $J(\pi)$ (on π) rejects the joint null hypothesis of no structural break and $a(\theta^*)=0$ if:

$$\int\int \zeta(\eta,\pi)dR(\eta,\pi)dJ(\pi) \geq k_\alpha \tag{13}$$

$$\text{where } \zeta(\eta,\pi)=\exp\left\{\int_0^1 v(s)'dZ(s)-\frac{1}{2}\int_0^1 v(s)'v(s)ds\right\} \tag{14}$$

$\eta \equiv [\theta_A', \gamma']' \in R^{2p\times 1}$, and k_α is defined so that the test has size α.

3. Special tests[5]

The leading case of the class of alternatives for structural break is that of alternatives that are linear in the parameters, that is: $g(\gamma,\pi,s) = \widetilde{G}(\pi,s)\gamma$. In the case of a single structural break, $\widetilde{G}(\pi,s) = 1(s\geq\pi)G$, where $1(s\geq\pi)$ is the indicator function, equal to one if $s\geq\pi$ and

[5] I consider only two-sided alternatives here; one may generalize the argument to one-sided alternatives.

zero otherwise, and G is a $(k \times p)$ matrix identifying the p-dimensional vector of time-varying parameters, say $G = [I_p \;\; 0_{q\times p}]$. Let's define:

$$A(\pi) = \begin{pmatrix} -D\overline{M}' & 0 \\ -(1-\pi)\, G'A'D\overline{M}' & G'\overline{M}' \end{pmatrix} \begin{pmatrix} Z(1) \\ Z(\pi) - \pi Z(1) \end{pmatrix} \tag{15}$$

$$V(\pi) = \begin{pmatrix} DBD' & (1-\pi)\, DB^{1/2}(I-H)\, B^{1/2}G \\ (1-\pi)\, G'B^{1/2}(I-H)\, B^{1/2}D' & (1-\pi)\, G'B^{1/2}[I-(1-\pi)\, H]B^{1/2}G \end{pmatrix} \tag{16}$$

The optimal test statistic described by (13) becomes $\int\int \exp\left\{\eta' A(\pi) - \frac{1}{2}\eta' V(\pi)\eta\right\} dR(\eta,\pi) dJ(\pi)$. As in Sowell (1996), different choices of the weighting function $R(\eta,\pi)$ lead to different test statistics. The weighting function considered here is an $(r+p)$-dimensional multivariate normal distribution with zero mean and covariance $U(\pi)$. When the time of the break is not known and we are interested in the test statistic that gives equal weight to alternatives that are equally difficult to detect when π is known, so that $U(\pi)^{-1} = \frac{1}{c}V(\pi)$, then the test statistic in (13) becomes $\int_{\Pi} \left(\exp\left\{\frac{1}{2}\frac{c}{1+c}\Phi^*(\pi)\right\}\right) dJ(\pi)$, where Π is the support of $J(\pi)$ and $\Phi_T^*(\pi) = A(\pi)'V(\pi)^{-1}A(\pi)$. The latter is a Wald test for the fixed and known π scenario. The test statistic can be estimated as:

$$\begin{aligned} TS_{c,T}^{AP*} &= \int_{\Pi} \left(\exp\left\{\frac{1}{2}\frac{c}{1+c}\Phi_T^*(\pi)\right\}\right) dJ(\pi) \\ \Phi_T^*(\pi) &= LM_1 + LM_2(\pi) \end{aligned} \tag{17}$$

where: $LM_1 \equiv \frac{1}{T}\left(W_T^{1/2}\sum_{t=1}^{T} f_t(\widetilde{\theta})\right)' \widehat{\Omega}_1 \left(W_T^{1/2}\sum_{t=1}^{T} f_t(\widetilde{\theta})\right)$

$$\begin{aligned} LM_2(\pi) &\equiv \frac{1}{\pi(1-\pi)}\frac{1}{T}\left(\sum_{t=1}^{[T\pi]} f_t(\widetilde{\theta}) - \pi\sum_{t=1}^{T} f_t(\widetilde{\theta})\right)' \widehat{\Sigma}^{-1/2\prime}\widehat{\Omega}_2\widehat{\Sigma}^{-1/2}\left(\sum_{t=1}^{[T\pi]} f_t(\widetilde{\theta}) - \pi\sum_{t=1}^{T} f_t(\widetilde{\theta})\right) \\ \widehat{\Omega}_1 &= \widehat{C}_1'\left(\widehat{C}_1\widehat{C}_1'\right)^{-1}\widehat{C}_1, \quad \widehat{C}_1 \equiv \left(\widehat{A}\widehat{B}^{\,-1}\widehat{A}^{\,\prime}\right)\widehat{A}\widehat{B}^{\,-1}\widehat{\overline{M}}', \quad \widehat{\Omega}_2 = \widehat{C}_2'\left(\widehat{C}_2\widehat{C}_2'\right)^{-1}\widehat{C}_2, \qquad \widehat{C}_2 \equiv G'\widehat{\overline{M}}' \\ \widehat{B} &= \widehat{\overline{M}}'\widehat{\overline{M}}, \quad \widehat{A} = \frac{1}{T}\sum_{t=1}^{T}\nabla_\theta a(\widetilde{\theta}), \quad \widehat{\overline{M}} = \widehat{\Sigma}^{-1/2}\frac{1}{T}\sum_{t=1}^{T}\nabla_\theta f_t(\widetilde{\theta}) \\ \widehat{\Sigma} &= \frac{1}{T}\sum_{t=1}^{T}\left(f_t(\widetilde{\theta}) - \frac{1}{T}\sum_{t=1}^{T} f_t(\widetilde{\theta})\right)\left(f_t(\widetilde{\theta}) - \frac{1}{T}\sum_{t=1}^{T} f_t(\widetilde{\theta})\right)' \end{aligned}$$

if $f_t(.)$ are iid, otherwise $\widehat{\Sigma}$ is estimated with a Newey-West HAC estimator. The limiting distribution of this test statistic under the null hypothesis is described in the following Proposition:

Proposition 1. *Let Assumptions 1-6 hold. The test statistic for testing* $a(\theta^*) = 0$ *against* (5) *and* (6) *with the greatest average power according to the weighting function* $R(\eta, \pi) \sim N\left(0, cV(\pi)^{-1}\right)$, *for* $V(\pi)$ *defined in (16) is (17). Its asymptotic distribution under the null hypothesis is:*

$$TS_{c,T}^{AP*} \Rightarrow \int_{\Pi} \left(\exp\left\{ \frac{1}{2} \frac{c}{1+c} \Phi^*(\pi) \right\} \right) dJ(\pi) \tag{18}$$

$$\Phi^*(\pi) \equiv \left(\frac{BB_p(\pi)' BB_p(\pi)}{\pi (1-\pi)} + B_r(1)' B_r(1) \right) \tag{19}$$

Proposition 1 shows that $TS_{c,T}^{AP*}$ is a weighted average of Wald tests. As noted above, the difference between the asymptotic distribution of the tests defined in this paper and that of the test for structural break only is that the latter do not have the $B_r(1)' B_r(1)$ component. This component arises from testing restrictions on θ over the whole sample. In fact, it corresponds to a centered chi-square with r degrees of freedom, the usual limiting distribution of the Wald test statistic for testing hypotheses on a parameter vector. Appendix 1 shows that both the tests for structural break and the classical tests obtain as special cases of (25).

From now until the end of this section, we specialize the above findings to situations in which the researcher is interested in testing hypotheses on a subset of the parameters. This is discussed in the following Corollary.

Corollary 1. Null hypotheses on subsets of parameters. *Let the parameter vector* $\theta \in R^k$ *be partitioned as* $\theta = [\beta', \delta']'$, *where* $\beta \in R^p$ *and* $\delta \in R^q$. *Let Assumptions 1, 3-6 hold. Let Assumption 2 be replaced by Assumption 2':* $\beta_{t,T} = \beta^* + \frac{1}{\sqrt{T}} g_\beta\left(\gamma, \pi, \frac{t}{T}\right)$ *and* $\beta^* = \beta_0 + \frac{1}{\sqrt{T}} \beta_A$. *It follows that:*

$$\begin{aligned} \sqrt{T} W_T^{1/2} F_{sT}(\beta_0, \widetilde{\delta}) \quad &\Rightarrow \quad B_m(s) - s\overline{P}_\delta B_m(1) - s\left(I_m - \overline{P}_\delta\right) \overline{M}_\beta \beta_A \\ &\quad -\overline{M}_\beta \int_0^s g_\beta(\gamma, \pi, r)\, dr + s\overline{P}_\delta \overline{M}_\beta \int_0^1 g_\beta(\gamma, \pi, r)\, dr \end{aligned} \tag{20}$$

where $\overline{P}_\delta \equiv \overline{M}_\delta \left(\overline{M}_\delta{}' \overline{M}_\delta\right)^{-1} \overline{M}_\delta{}' \in R^{m \times m}$. *Also,* $v(s)$ *in (11) becomes:* $v(s) \equiv -\overline{M}_\beta g_\beta(\gamma, \pi, s)$ $+\overline{P}_\delta \overline{M}_\beta \left(\int_0^1 g_\beta(\gamma, \pi, v) dv \right) - \left(I_m - \overline{P}_\delta\right) \overline{M}_\beta \beta_A$.

We will finally consider special cases of $TS_{c,T}^{AP*}$ that have been considered in the literature for tests for structural break only. Each of these special cases have greatest weigthed average power against particular forms of parameter instability. We will analyze the form that the optimal test proposed in this paper assumes for these particular forms of parameter instability.

INSERT TABLE 1

Andrews and Ploberger test

Let $\beta_t = \beta_1(\pi)$ for $t = 1, 2...[T\pi]$ and $\beta_t = \beta_2(\pi)$ for $t = [T\pi] + 1, ...T$, where $[\cdot]$ denotes the greatest integer function. Also, to simplify notation, let $f_t(\beta_k) \equiv f_t(x_t, \beta_k, \delta)$, $f_t\left(\widehat{\beta}_k\right) \equiv f_t\left(x_t, \widehat{\beta}_k, \widehat{\delta}\right)$, $f_t(\beta_0) \equiv f_t\left(x_t, \beta_0, \widetilde{\delta}\right)$, $k = 1, 2$. Let $\theta(\pi) \equiv (\beta_1, \beta_2, \delta)$, $\widehat{\theta}(\pi) \equiv (\widehat{\beta}_1, \widehat{\beta}_2, \widehat{\delta})$ be the unrestricted GMM estimator under the hypothesis that there is a break at the fraction $[T\pi]$ of the sample, and $\widetilde{\theta}(\pi)$ be the constrained estimator. Thus, the Wald test for a fixed and known π can be estimated as either:[6]

$$\text{Wald:} \quad \Phi_T^*(\pi) = T\left(R\widehat{\theta}(\pi) - r\right)'\left(RV(\widehat{\theta}(\pi))R'\right)^{-1}\left(R\widehat{\theta}(\pi) - r\right) \tag{21}$$

$$\text{Distance Metric form:} \quad \Phi_T^*(\pi) = \overline{Q}(\widetilde{\theta}(\pi)) - \overline{Q}(\widehat{\theta}(\pi)) \tag{22}$$

$$\text{Lagrange Multiplier:} \quad \Phi_T^*(\pi) = LM_1 + LM_2(\pi) \tag{23}$$

where notation is in Table 1. The table assumes that $f_t(.)$ consists of mean zero uncorrelated random variables. When $f_t(.)$ consists of mean zero but serially correlated random variables then consistent estimation of $\widehat{\Sigma}_1$, $\widehat{\Sigma}_2$, $\widehat{\Sigma}$ and $\widehat{\Gamma}$ requires a HAC estimator, e.g. Newey and West (1987). Note that (23) is particularly easy to calculate. It is simply the sum of the two LM tests to test (5) and (6) separately. Then, Proposition 2 follows.

Proposition 2. *Let Assumptions 1, 2', 3-6 hold. The test statistic for testing $\beta = \beta^*$ against $\beta_{t,T} = \beta^* + \frac{1}{\sqrt{T}}\beta_A + \frac{1}{\sqrt{T}}\gamma 1(s \geq \pi)$ with the greatest average power according to the weighting*

[6] For completeness, let us mention that the Lagrange Multiplier statistic can also be obtained as: $T \cdot \nabla_\beta \overline{Q}(\beta_0, \beta_0, \widetilde{\delta})'\widehat{\Omega}^{-1}\nabla_\beta \overline{Q}(\beta_0, \beta_0, \widetilde{\delta})$ where and $\widehat{\Omega}$ is a consistent estimator of $E\left(T \cdot \nabla_\beta \overline{Q}(\widetilde{\theta})\nabla_\beta \overline{Q}(\widetilde{\theta})'\right) \in R^{2p \times 2p}$. However, the LM formula provided in the main text is easier to calculate.

function $R(\eta,\pi) \sim N\left(0, cV(\pi)^{-1}\right)$, *for* $V(\pi)$ *defined in (16) (17) with either (21), (22) or (23). Its asymptotic distribution under the null hypothesis is:*

$$TS_c^{AP*} \Rightarrow \int_{\Pi} \left(\exp\left\{ \frac{1}{2}\frac{c}{1+c}\Phi^*(\pi) \right\} \right) dJ(\pi) \tag{24}$$

$$\Phi^*(\pi) \equiv \left(\frac{BB_p(\pi)' BB_p(\pi)}{\pi(1-\pi)} + B_p(1)' B_p(1) \right) \tag{25}$$

As, special cases, we have:

$$\text{(a) } c \to \infty : \quad Exp - W_T^* \equiv \underset{c\to\infty}{\text{plim}} \; TS_{c,T}^{AP*} \Rightarrow \int_{\Pi} \left(\exp\left\{ \frac{1}{2}\Phi^*(\pi) \right\} \right) dJ(\pi) \tag{26}$$

$$\text{(b) } c \to 0 : \quad Mean - W_T^* \equiv \underset{c\to 0}{\text{plim}} \; 2\left(\frac{TS_{c,T}^{AP*} - 1}{c} \right) \Rightarrow \int_{\Pi} \Phi^*(\pi) dJ(\pi) \tag{27}$$

The special cases that correspond to extreme values of the parameter c are similar to those in Andrews and Ploberger. When $c \to \infty$ ($c \to 0$), more weight is assigned to alternatives about parameter instability further from (closer to) the null hypothesis.

Andrews (1993) Sup-LR test

A test statistic commonly considered in the literature of structural breaks is the Quandt Likelihood Ratio (QLR) test statistic (or Sup-LR test), which is the supremum (over all possible break dates) of the Chow statistic designed for these alternatives for a fixed break date. Andrews (1993) derived its asymptotic distribution. The modified QLR test statistic for the alternatives specified in this paper can be obtained by letting $\frac{c}{1+c} \to \infty$ in (18), which gives:

$$QLR_T^* = \sup_{\pi} \Phi_T^*(\pi) \tag{28}$$

The limiting distribution of (28) under the null hypothesis is given in the following Proposition:

Proposition 3. *Let Assumptions 1, 2', 3-6 hold. The test statistic for testing* $\beta = \beta^*$ *against* $\beta_{t,T} = \beta^* + \frac{1}{\sqrt{T}}\beta_A + \frac{1}{\sqrt{T}}\gamma 1(s \geq \pi)$ *with the greatest average power according to the weighting function* $R(\eta,\pi) \sim N\left(0, cV(\pi)^{-1}\right)$, *for* $V(\pi)$ *defined in (16) and c such that* $\frac{c}{1+c} \to \infty$ *is (28), whose asymptotic distribution under the null hypothesis is:*

$$QLR_T^* \Rightarrow \sup_{\pi} \Phi^*(\pi) \tag{29}$$

Nyblom (1989) test

Another test for parameter instability is that considered by Nyblom (1989) and Nyblom and Mäkeläinen (1983). These authors derive the locally most powerful invariant (to translations and scale transformations) test for constancy of the parameter process against the alternative that the parameters follow a random walk process:[7]

$$\beta_t = \beta_{t-1} + e_t, \quad e_t \sim N\left(0, \frac{1}{T^2}\sigma_e^2 \overline{\Gamma}\right) \tag{30}$$

The modified Nyblom test statistic for testing whether β_t is equal to β_0 is:

$$Nyblom_T^* = \int_0^1 \left(T \cdot \nabla_\beta Q_{1,[\pi T]}(\beta_0, \widetilde{\delta})' \widehat{\Omega}_N^{-1} \nabla_\beta Q_{1,[\pi T]}(\beta_0, \widetilde{\delta})\right) J(\pi)\, d\pi \tag{31}$$

where $\widehat{\Omega}_N = \overline{M}'_\beta \left(I_m - \overline{P}'_\delta\right) \overline{M}_\beta \in R^{p \times p}$ and the gradient of the objective function is defined as $\nabla_\beta Q_{1,[\pi T]}(\beta_0, \widetilde{\delta}) \equiv \frac{1}{T} \sum_{t=1}^{[T\pi]} \nabla_\beta F_T(\beta_0, \widetilde{\delta})'\ \Sigma^{-1}\ f_t\left(x_t, \beta_0, \widetilde{\delta}\right)$.

Note that (31) is a generalization of the locally best invariant test statistic proposed by Nabeya and Tanaka (1988) for the case in which β is known and equal to β_0. The test proposed in this paper is more general than Nabeya and Tanaka, as estimation is not restricted to the ordinary least square case, and β can be a vector. Appendix 1 shows that the asymptotic distribution of the modified Nyblom statistic under the null hypothesis is:

Proposition 4. *Let Assumptions 1, 2', 3-6 hold. The test statistic for testing $\beta = \beta_0$ against $\beta_{t,T} = \beta_0 + \frac{1}{\sqrt{T}}\beta_A$, $\beta_{t,T} = \beta_{T,t-1} + e_t$, $e_t \sim N\left(0, \frac{1}{T^2}\sigma_e^2\overline{\Gamma}\right)$ with the greatest average power according to the weighting function $R(\eta, \pi) \sim N\left(0, cV(\pi)^{-1}\right)$, for $V(\pi)$ defined in (16) is (31). Its asymptotic distribution under the null hypothesis is:*

$$Nyblom_T^* \Rightarrow \int_0^1 B_p(\pi)' B_p(\pi) J(\pi)\, d\pi \tag{32}$$

Tables 2 to 5 in Appendix 2 report critical values for the optimal tests for $J(\pi)$ uniformly distributed on [0.15, 0.85].[8] The significance levels considered in the tables are 10%, 5%, 2.5% and

[7] The notation is the same as in Nyblom and Mäkeläinen (1983), $\overline{\overline{\Gamma}}$ is a known matrix and σ_e^2 is a scalar. See also King (1980), King and Hillier (1985) and Stock and Watson (1998).

[8] Trimming values are required. See Andrews (1993).

1%. The critical values are obtained by simulating the asymptotic distributions described in this section. The number of Monte Carlo replications is 5,000. Notice that all the values in Tables 2 to 4 are higher than those for the corresponding tests for structural break only, the reason being that the optimal tests add the non-negative component $B_p(1)' B_p(1)$ (see equation (25)).

4. Asymptotic local power analysis

The local power properties of the optimal tests derived above can be compared with those of tests for parameter instability only and those of tests for $a(\theta^*) = 0$ only. The comparison can be made both theoretically and by Monte Carlo simulations.

Let's first consider the theoretical local power properties of the various tests. To facilitate a comparison with the tests existing in the literature, we focus on the tests discussed in the second part of Section 3, and, for brevity, we analyze only (26).[9] Let $\widetilde{\theta} = (\beta_0, \widetilde{\delta})$. From (23) and (26), and using the notation in Table 1, we have that:

$$log\left(Exp - W_T^*\right) = \frac{1}{2} LM_1 + log \int_{\Pi} \left(\exp\left\{ \frac{1}{2} LM_2(\pi) \right\} \right) dJ(\pi)$$

Appendix 1 shows that:

(a) $\Omega_1^{1/2} \frac{1}{\sqrt{T}} W_T^{1/2} \sum_{t=1}^{T} f_t(\widetilde{\theta}) \Rightarrow Z_p^{(1)}$

$Z_p^{(1)} \equiv B_p(1) - \Omega_1^{1/2} \left(I - \overline{P}_\delta\right) \overline{M}_\beta \beta_A - \Omega_1^{1/2} \left(I - \overline{P}_\delta\right) \overline{M}_\beta \int_0^1 g_\beta(s)\, ds$

(b) $\Omega_2^{1/2} \left(\frac{1}{\sqrt{T}} W_T^{1/2} \left[\sum_{t=1}^{[T\pi]} f_t(\widetilde{\theta}) - \pi \sum_{t=1}^{T} f_t(\widetilde{\theta}) \right] \right) \Rightarrow Z_p^{(2)}(\pi)$

$Z_p^{(2)}(\pi) \equiv BB_p(\pi) - (1-\pi)\, \Omega_2^{1/2} \overline{M}_\beta \int_0^\pi g_\beta(\gamma, \pi, r)\, dr + \pi \Omega_2^{1/2} \overline{M}_\beta \int_\pi^1 g_\beta(\gamma, \pi, r)\, dr.$[10]

(c) Under the null hypothesis, LM_1 and $LM_2(\pi)$ are asymptotically independent; this follows from the fact that $B_p(1)$ and $BB_p(\pi)$ are independent. Thus, if one performs two tests, LM_1 and $\int_{\Pi} \left(\exp\left\{ \frac{1}{2} LM_2(\pi) \right\} \right) dJ(\pi)$, each at size $1 - \sqrt{1-\alpha}$, then the joint test will have size α. However, this two-stage test, by construction, will not have the highest weighted average power according to the weight function in Proposition 1.

[9] A similar analysis applies to the optimal Mean Wald, QLR and Nyblom tests.

[10] See also Appendix 1 for more details. Note that: $-\overline{M}_\beta \int_0^\pi g_\beta(\gamma, \pi, r)\, dr + \pi \overline{M}_\beta \int_0^1 g_\beta(\gamma, \pi, r)\, dr =$
$-(1-\pi)\, \overline{M}_\beta \int_0^\pi g_\beta(\gamma, \pi, r)\, dr + \pi \overline{M}_\beta \int_\pi^1 g_\beta(\gamma, \pi, r)\, dr$

(d) $LM_1 \Rightarrow Z_p^{(1)\prime} Z_p^{(1)}$ and thus it may have no power to detect (5). In fact, $Z_p^{(1)}$ does not depend on $\int_0^{\pi} g_\beta(s)\,ds$. Thus, for example, the power versus alternatives where the break function is such that $\int_0^1 g_\beta(s)\,ds = 0$ will be equal to the size.

(e) $LM_2 \Rightarrow Z_p^{(2)}(\pi)' Z_p^{(2)}(\pi)$ and thus it has no power to detect (6). In fact, $Z_p^{(2)}$ does not depend on β_A. Thus, for example, the power versus alternatives in which there is no break but $\beta_A \neq 0$ will be equal to the size;

(f) The test for parameter instability only can be obtained by substituting $a(\theta) = 0$, $A = 0$, $\overline{H} = I$ in the proof of Result 1, so that asymptotically it behaves like $Z_p^{(2)}(\pi)$, which is the same as in Andrews (1993), and conclusions similar to those in (d) hold. In fact, upon inspection, it is clear that LM_1 is the standard LM test for testing $\beta = \beta_0$, whereas $LM_2(\pi)$ has the same asymptotic distribution the LM test for parameter instability.

To verify these insights, we perform some Monte Carlo simulations. A variety of DGPs is considered, paying particular attention to situations where the standard tests fail to detect the alternative hypothesis. For simplicity, only a univariate model is considered:

$$y_t = \beta_{t,T} + \epsilon_t \qquad \epsilon_t \sim N(0,1), \quad T = 100, \quad \beta_0 = 0 \tag{33}$$

The likelihood ratio LR_1 tests whether the parameter equals β_0 whereas parameter instability tests check whether $\beta_{t,T}$ is constant; optimal tests jointly test the two hypotheses. The parameter instability tests (TVP) considered here are the Andrews and Ploberger Exponential Wald tests ($Exp\!-\!W_T$), the Nyblom test ($Nyblom_T$) and the Quandt Likelihood Ratio (QLR_T). The optimal tests are $Exp - W_T^*$, QLR_T^* and $Nyblom_T^*$ defined in section 3. The nominal size is 5%.

We consider the following DGPs:[11]

Design 1. $\beta_{t,T} = \beta_0 + \beta_A \frac{1}{\sqrt{T}}\ \forall t$

The upper, left panel in Figure 1 shows the asymptotic local power of the tests as a function of β_A. It shows that when the parameter is not time-varying, the likelihood ratio LR_1 is the most powerful test, according to the Neyman and Pearson lemma. The test designed to detect

[11] Note that different Monte Carlo experiments could be designed, in which all models are possibly (dynamically) mis-specified, as in Corradi and Swanson (2001). This setup would not be the one for which the optimal tests proposed in this paper are designed, so it is not investigated.

structural break, $Exp - W_T$, has a flat power function around the size of the test whereas the $Exp - W_T^*$ test is almost as powerful as the LR_1 test.

Design 2. $\beta_{t,T} = \beta_0 + \frac{1}{\sqrt{T}}\beta_A \; 1\left(t > [\frac{T}{2}]\right)$

Design 2 involves a single break in the data. This particular alternative is both a deviation of the parameter vector from the null hypothesis and a structural break, so all the tests (the most powerful likelihood ratio test, LR^*,[12] the TVP and the optimal tests) should detect it. This is in fact what the upper right panel in Figure 1 shows.

Design 3. $\beta_{t,T} = \begin{cases} \beta_0 + \beta_A \frac{1}{\sqrt{T}} & \text{for } t = 1 \text{ up to } t = [\frac{T}{2}] \\ \beta_0 - \beta_A \frac{1}{\sqrt{T}} & \text{for } t = [\frac{T}{2}] + 1 \text{ up to } T \end{cases}$

The lower panel on the right in Figure 1 shows that, in this design, the shift in the parameter vector is not detected by a simple likelihood ratio (LR_1) because the statistic on which it is based (the average of the observations) is invariant to it; in fact, notwithstanding the structural break, the average over the whole sample is asymptotically equal to β_0. While the TVP test is the most powerful, the optimal test is powerful too.

Design 4. $\beta_t = \beta_0 + \beta_{t-1} + u_t$, where $u_t \sim N\left(0, \frac{\sigma_u^2}{T^2}\right)$ is independent from ϵ_t and $\sigma_u^2 \geq 0$

The asymptotic local power functions for this design are depicted in the bottom right panel in Figure 1 as functions of the parameter $\sigma_u^2 \geq 0$. When $\sigma_u^2 = 0$ then β_t is constant whereas when $\sigma_u^2 \neq 0$ then β_t is a random walk with no drift. The test designed for this hypothesis is the Nyblom test; LR_1 test is also powerful. The reason is that LR_1 is detecting deviations from the null hypothesis by comparing the sample average with the null hypothesis and the sample average is not a consistent estimate of the true parameter value. Note that the optimal Nyblom test is powerful too.

The results of the simulations suggest the following conclusions. First, *the tests that maintain some power across all the designs considered here are the optimal tests.* For all the other tests there is at least one design (a particular direction away from the null hypothesis) in which the power is flat around the size of the test. Hence, they are not "robust" across designs, whereas the optimal tests are. Second, let's consider a two-stage testing procedure, where the first stage tests whether there is a structural break (by using either QLR_T or $Exp - W_T$) and the second

[12] LR* is the likelihood ratio test for testing $\beta_2 = \beta_0$ conditional on knowing that $\beta_1 = \beta_0$ (see the example at the beginning of section 2).

stage, conditionally on the first stage, tests hypotheses on the parameters (by using the LR_1 test). Let the tests be labeled "*Seq.QLR*" and "*Seq.Exp-W*" respectively. In the special cases considered in Section 3 (obtained with particular weighting matrices), the two stages of the test are asymptotically independent. By choosing a size equal to $1-\sqrt{1-0.95}$, the joint significance level will be the desired nominal level, 0.95. Figure 2 shows that there is no clear ranking between the sequential tests and the optimal tests. The power ranking will depend on the direction of the alternative hypothesis. However, by construction, the optimal tests will have the greatest average local power. Two stage independent tests have advantages and disadvantages. The advantages are that if we reject we know which part of the alternative we reject and that the first stage test could be used if the researcher is unsure about which elements of the parameter vector are subject to instability. The disadvantage is that they will not have the optimal weighted average power for alternatives that are equally likely; in other words, if we want tests that are invariant to non-singular linear transformations of the hypothesis, we cannot construct the test as formed by two independent components, as two-stage tests are not invariant to these transformations.

INSERT FIGURES 1 AND 2

5. Conclusions

This paper shows that there exists a class of locally most powerful tests for testing the joint hypothesis of model selection between two nested models and parameter stability. This paper introduces this class of tests, states the assumptions under which they are valid and works out their asymptotic distributions. It also derives some special cases, that apply for specific forms of parameter instability. These tests are easy to calculate and this paper reports their (asymptotic) critical values.

References

Andrews, Donald W. K., "Tests for Parameter Instability and Structural Change with Unknown Change Point", *Econometrica*, 61(4), July 1993, pp. 821-856.

Andrews, Donald W. K., Inpyo Lee, and Werner Ploberger, "Optimal Change-point Tests for Normal Linear Regression", *Journal of Econometrics*, 70, 1996, pp. 9-38.

Andrews, Donald W. K., and Werner Ploberger, "Optimal Tests When a Nuisance Parameter is Present only under the Alternative", *Econometrica*, 62(6), November 1994, pp. 1383-1414.

Andrews, Donald W. K., and Werner Ploberger, "Admissibility of the Likelihood Ratio Test When a Nuisance Parameter is Present only under the Alternative", *Cowles Foundation Discussion Paper*, July 1993.

Andrews, Donald W. K., "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation", *Econometrica*, 59, 1991, pp. 817-858.

Chow, Gregory C., "Tests of Equality Between Sets of Coefficients in Two Linear Regressions", *Econometrica*, 28, 1960, pp. 591-605.

Clark, Todd E., and M.W. McCracken, "Forecast-Based Model Selection in the Presence of Structural Breaks", *mimeo*, August 2002.

Corradi, Valentina and Norman Swanson, "Bootstrap Conditional Distribution Tests Under Dynamic Misspecification", *mimeo*, 2001.

Elliott, Graham and Ulrich K. Müller, "Optimally Testing General Breaking Processes in Linear Time Series Models", *mimeo*, November 2002.

Engle, Robert F., "Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics", in: Z. Griliches and M. Intriligator, *Handbook of Econometrics, Volume 2*, Amsterdam: North Holland, 1984.

Ghysels, Eric, and Alastair Hall, "Predictive Tests for Structural Change with Unknown Breakpoint", *Journal of Econometrics* 82(2), February 1998, pp. 209-33.

King, M. L., "Robust Tests for Spherical Symmetry and their Application to Least Square Regression", *The Annals of Statistics*, 80(6), 1980, pp. 1265-1271.

King, M. L., and Grant H. Hillier, "Locally Best Invariant Tests of the Error Covariance Matrix of the Linear Regression Model", *Journal of the Royal Statistical Society, series B*, 47(1), 1985, pp. 98-102.

Nabeya, Seiji, and Katsuto Tanaka, "Asymptotic theory of a test for the constancy of regression coefficients against the random walk alternative", *The Annals of Statistics 16*(1), 1988, pp. 218-235.

Newey, Whitney, and Daniel McFadden, "Large Sample Estimation and Hypothesis Testing", in: Engle, R. and D. L. McFadden, *Handbook of Econometrics, Vol. IV*, Amsterdam: Elsevier-North Holland, 1994.

Newey, Whitney, and Kenneth West, "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica* 55, 1987, pp. 703-708.

Nyblom, Jukka, "Testing for the Constancy of Parameters Over Time", *Journal of the American Statistical Association*, 84(405), March 1989, pp. 223-230.

Nyblom, Jukka, and Timo Mäkeläinen, "Comparisons of Tests for the Presence of Random Walk Coefficients in a Simple Linear Model", *Journal of the American Statistical Association*, 78(384), December 1983, pp. 856-864.

Ploberger, Werner, and Walter Krämer, "The Local Power of the Cusum and Cusum of Squares Tests", *Econometric Theory*, 6, 1990, pp. 335-347.

Ploberger, Werner, and Walter Krämer, "The CUSUM test with OLS Residuals", *Econometrica*, 60, 1992, pp. 271-286.

Quandt, R. E., "Tests of the Hypothesis that a Linear Regression System Obeys Two Separate Regimes", *Journal of the American Statistical Association*, 55, 1960, pp. 324-330.

Rossi, Barbara, "Are Exchange Rates Really Random Walks? Some Evidence Robust to Parameter Instability", *mimeo*, December 2003.

Sowell, Fallaw, "Optimal Tests for Parameter Instability in the Generalized Method of Moments Framework", *Econometrica*, 64(5), September 1996, pp. 1085-1107.

Stock, James H., and Mark W. Watson, "Median Unbiased Estimation of Coefficient Variance in a Time-Varying Parameter Model", *Journal of the American Statistical Association*, 93(441), March 1998, pp. 349-358.

Figures

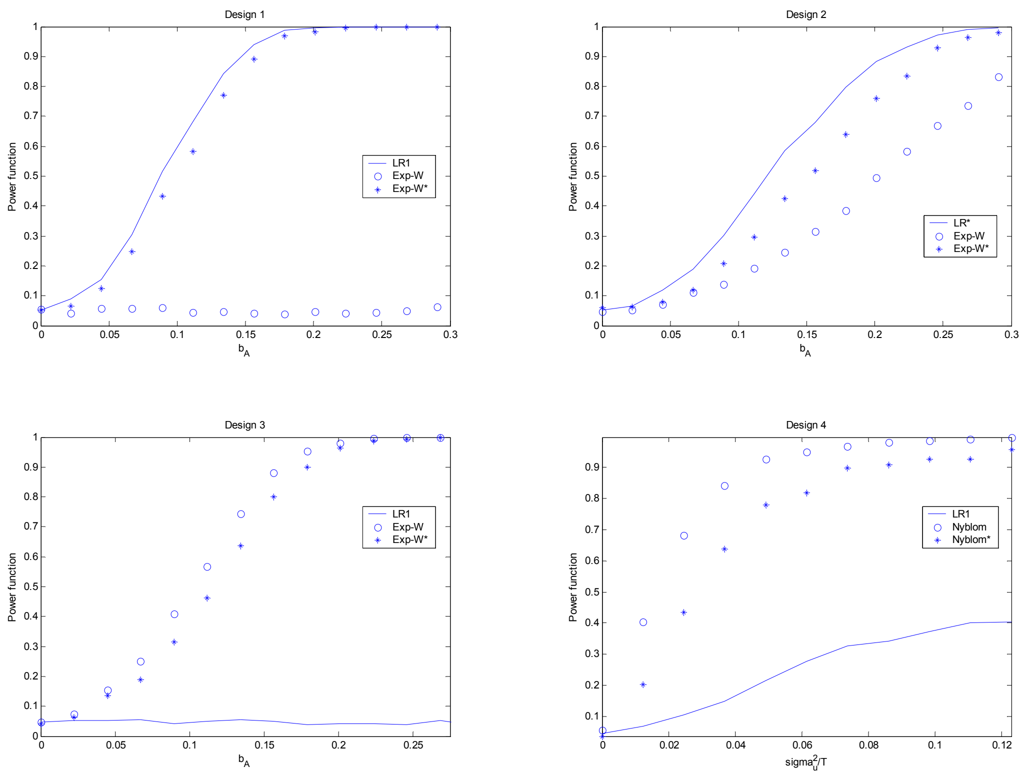

Figure 1. Asymptotic power functions of 5% tests for designs 1 to 4.

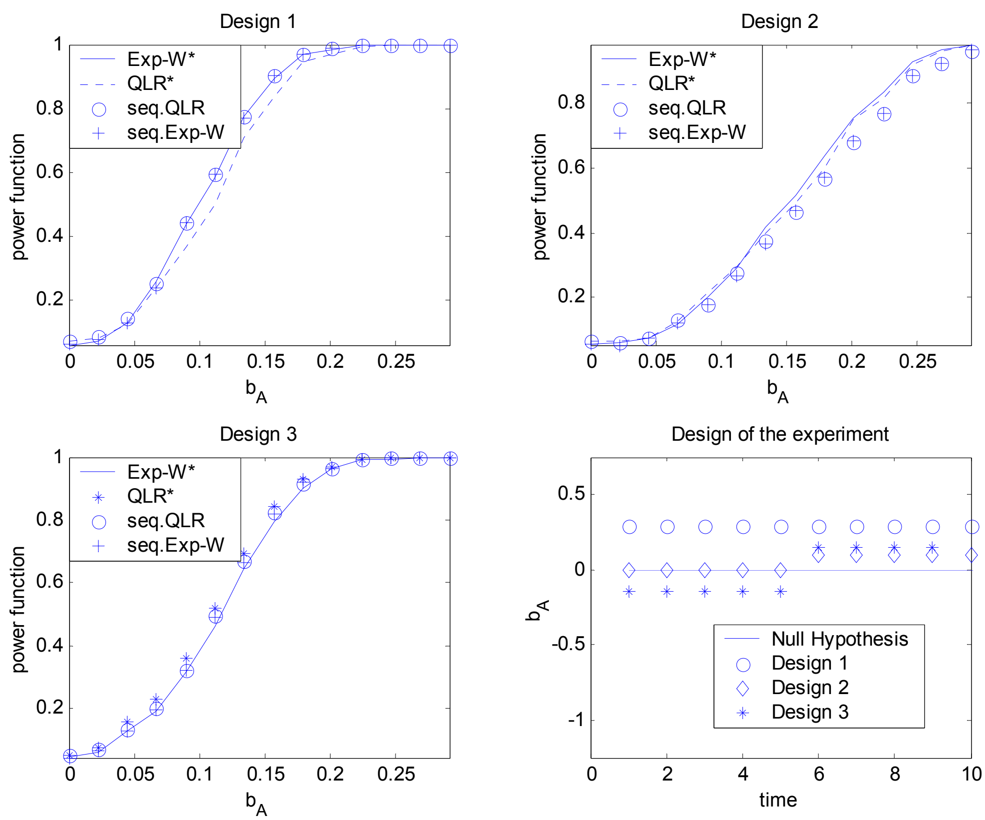

Figure 2. Comparison of asymptotic power functions of 5% selected optimal tests with the naive sequential test (across different designs).

Appendix 1. Proofs

Proof of Result 1

To simplify notation, let $f_t(x_t, \widetilde{\theta})$ be denoted as $f_t(\widetilde{\theta})$ and $\theta_{t,T}$ be denoted by θ_t. The restricted estimator $\widetilde{\theta}$ satisfies the following FOCs for minimizing the Lagrangean $Q(\theta) + a(\theta)' \lambda$, where λ is the $(r \times 1)$ vector of Lagrange multipliers:

$$\begin{aligned} 0 &= \nabla_\theta Q(\widetilde{\theta}) + \nabla_\theta a(\widetilde{\theta})'\widetilde{\lambda} \\ 0 &= a(\widetilde{\theta}) \end{aligned} \quad (34)$$

Take a mean value expansion of $f_t(\widetilde{\theta})$ around θ^* :

$$f_t(\widetilde{\theta}) = f_t(\theta^*) + \nabla_\theta f_t(\overline{\theta}) \cdot (\widetilde{\theta} - \theta^*) \quad (35)$$

where $\overline{\theta}$ is a intermediate point (in Euclidean distance) between $\widetilde{\theta}$ and θ^*, and by consistency of $\widetilde{\theta}$, $\overline{\theta} \underset{p}{\rightarrow} \theta^*$. Summing (35) from $t = 1$ to $[sT]$ gives $F_{sT}(\widetilde{\theta}) = F_{sT}(\theta^*) + \nabla_\theta F_{sT}(\overline{\theta}) \cdot (\widetilde{\theta} - \theta^*)$, which, evaluated at $s = 1$ and pre-multiplied by $\nabla_\theta F_T(\widetilde{\theta})' W_T$ gives:

$$\nabla_\theta Q(\widetilde{\theta}) = \nabla_\theta F_T(\widetilde{\theta})' W_T F_T(\theta^*) + \nabla_\theta F_T(\widetilde{\theta})' W_T \nabla_\theta F_T(\overline{\theta}) \cdot (\widetilde{\theta} - \theta^*) \quad (36)$$

Another mean value expansion of $a(\widetilde{\theta})$ around θ^* gives:

$$a(\widetilde{\theta}) = a(\theta^*) + A\left(\overline{\theta}\right)(\widetilde{\theta} - \theta^*) \quad (37)$$

Thus, combining (34), (36) and (37), and $A\left(\overline{\theta}\right) \underset{p}{\rightarrow} A$:

$$\begin{pmatrix} -\nabla_\theta F_T(\widetilde{\theta})' W_T F_T(\theta^*)\sqrt{T} \\ -a(\theta^*)\sqrt{T} \end{pmatrix} = \begin{pmatrix} \nabla_\theta F_T(\widetilde{\theta})' W_T \nabla_\theta F_T(\overline{\theta}) & A' \\ A & 0 \end{pmatrix} \begin{pmatrix} (\widetilde{\theta} - \theta^*)\sqrt{T} \\ \lambda\sqrt{T} \end{pmatrix} + o_p \quad (38)$$

Define: $B \equiv \overline{M}'\overline{M}$; $A \equiv A(\theta^*)$; $P \equiv B^{-1/2} A' \left(AB^{-1}A'\right)^{-1} AB^{-1/2}$, $H \equiv I - P$. Solving (38) for $(\widetilde{\theta} - \theta^*)$ gives:

$$\sqrt{T}(\widetilde{\theta} - \theta^*) = -B^{-1/2} H B^{-1/2} \nabla_\theta F_T(\widetilde{\theta})' W_T F_T(\theta^*)\sqrt{T} - B^{-1} A' \left(AB^{-1}A'\right)^{-1} a(\theta^*)\sqrt{T} + o_p \quad (39)$$

By substituting (39) in (35), summing from $t = 1$ to $[sT]$ and pre-multiplying by $\sqrt{T} W_T^{1/2}$, we have:

$$\begin{aligned} \sqrt{T} W_T^{1/2} F_{sT}(\widetilde{\theta}) &= \sqrt{T} W_T^{1/2} F_{sT}(\theta^*) - \frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} W_T^{1/2} \nabla_\theta f_t(\overline{\theta}) B^{-1/2} H B^{-1/2} \nabla_\theta F_T(\widetilde{\theta})' W_T F_T(\theta^*) \\ &\quad - \frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} W_T^{1/2} \nabla_\theta f_t(\overline{\theta}) B^{-1} A' \left(AB^{-1}A'\right)^{-1} a(\theta^*) + o_p \end{aligned} \quad (40)$$

Next, a mean value expansion of $F_{sT}(\theta^*)$ around θ_t implies:

$$F_{sT}(\theta^*) = F_{sT}(\theta_t) + \frac{1}{T}\sum_{t=1}^{[sT]}\nabla_\theta f_t(\overline{\theta}_t)\left(\theta^* - \theta_t\right) \tag{41}$$

where $\overline{\theta}_t$ is an intermediate point between θ_t and θ^*. Substituting (41) in (40), we have:

$$\begin{aligned}
\sqrt{T}W_T^{1/2}F_{sT}(\widetilde{\theta}) &= \sqrt{T}W_T^{1/2}F_{sT}(\theta_t) + \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}W_T^{1/2}\nabla_\theta f_t(\overline{\theta}_t)\left(\theta^* - \theta_t\right) \\
&\quad -\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}W_T^{1/2}\nabla_\theta f_t(\overline{\theta})B^{-1/2}HB^{-1/2}\nabla_\theta F_T(\widetilde{\theta})'W_T F_T(\theta_t) + \\
&\quad -\frac{1}{T}\sum_{t=1}^{[sT]}W_T^{1/2}\nabla_\theta f_t(\overline{\theta})B^{-1/2}HB^{-1/2}\nabla_\theta F_T(\widetilde{\theta})'W_T\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\nabla_\theta f_t(\overline{\theta}_t)\left(\theta^* - \theta_t\right) \\
&\quad -\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}W_T^{1/2}\nabla_\theta f_t(\overline{\theta})B^{-1}A'\left(AB^{-1}A'\right)^{-1}a\left(\theta^*\right) + o_p
\end{aligned} \tag{42}$$

Letting $T \to \infty$, we have:

$$\sqrt{T}W_T^{1/2}F_{sT}(\theta_t) \Rightarrow B_m(s)$$

$$\frac{1}{T}\sum_{t=1}^{[sT]}W_T^{1/2}\nabla_\theta f_t(\overline{\theta}_t)\sqrt{T}\left(\theta^* - \theta_t\right) \underset{p}{\to} -\overline{M}\int_0^s g\left(\gamma,\pi,r\right)dr, \text{ (included } s=1\text{)}$$

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}W_T^{1/2}\nabla_\theta f_t(\overline{\theta})B^{-1/2}HB^{-1/2}\nabla_\theta F_T(\widetilde{\theta})'W_T F_T(\theta_t) \Rightarrow s\overline{M}B^{-1/2}HB^{-1/2}\overline{M}'B_m(1)$$

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}W_T^{1/2}\nabla_\theta f_t(\overline{\theta})B^{-1/2}MB^{-1/2}\nabla_\theta F_T(\widetilde{\theta})'W_T\frac{1}{T}\sum_{t=1}^{T}\nabla_\theta f_t(\overline{\theta}_t)\left(\theta^* - \theta_t\right) \\
&\underset{p}{\to} -s\overline{M}B^{-1/2}HB^{-1/2}\overline{M}'\overline{M}\int_0^1 g\left(\gamma,\pi,r\right)dr
\end{aligned}$$

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}W_T^{1/2}\nabla_\theta f_t(\overline{\theta})B^{-1}A'\left(AB^{-1}A'\right)^{-1}a\left(\theta^*\right) \underset{p}{\to} s\overline{M}B^{-1}A'\left(AB^{-1}A'\right)^{-1}\theta_A$$

By substituting the above expressions in (42), we have:

$$\begin{aligned}
\sqrt{T}W_T^{1/2}F_{sT}(\widetilde{\theta}) &\Rightarrow B_m(s) - s\overline{H}B_m(1) - s\overline{M}B^{-1}A'\left(AB^{-1}A'\right)^{-1}\theta_A \\
&\quad - \overline{M}\int_0^s g\left(\gamma,\pi,r\right)dr + s\overline{M}B^{-1/2}HB^{1/2}\int_0^1 g\left(\gamma,\pi,r\right)dr
\end{aligned} \tag{43}$$

where $\overline{H} \equiv \overline{M}B^{-1/2}HB^{-1/2}\overline{M}'$, which proves Result 1.

Proof of Result 2

To prove Result 2, note that under the null hypothesis $\theta_A = 0$ and $g(.) = 0$ so that only the first two components on the right hand side of (43) are relevant. Note also that $\overline{H}$ is a projection matrix with rank $(k-r)$ so that $\overline{H} = C'\Lambda C$, where $\Lambda = \begin{pmatrix} I_{k-r} & 0 \\ 0 & 0 \end{pmatrix}$ and C is an orthonormal matrix such that $CC' = I_m$. Hence, Result 2 follows.

Proof of Corollary 1

Let $B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$ and $B^{-1} = \begin{pmatrix} B_{11}^{-} & B_{12}^{-} \\ B_{21}^{-} & B_{22}^{-} \end{pmatrix}$. Also, let the restrictions be linear restrictions on subsets of the parameters, so that $A = [I_{p\times p} \vdots 0_{p\times q}]$. Let $\overline{M}_\theta = [\overline{M}_\beta \vdots \overline{M}_\delta]$, $\overline{P}_\delta \equiv \overline{M}_\delta \left(\overline{M}_\delta' \overline{M}_\delta\right) \overline{M}_\delta'$. Corollary 1 follows from (43) by using the following results (a)-(e).[13]

(a) $\overline{H} = \overline{P}_\delta$

(b) $\overline{M} B^{-1/2} H B^{1/2} = \overline{P}_\delta \overline{M}$

(c) $\overline{M} B^{-1} A' \left(A B^{-1} A'\right)^{-1} = \left(I - \overline{P}_\delta\right) \overline{M}_\beta$

(d) $\widetilde{\beta} = \beta_0$

(e) $g(.) = [g_\beta(.) \;\; 0_{q\times p}]$

Proof of (15) and (16)

Let Assumption 2 hold and let the class of alternatives be linear in the parameters: $g(\gamma, \pi, s) = \widetilde{G}(\pi, s)\gamma$. Thus $v(s)$, defined below (11), becomes:

$$v(s) = \begin{pmatrix} -\overline{M}D' & -\overline{M}\widetilde{G}(\pi, r) + \overline{M} B^{-1/2} H B^{1/2} \left(\int_0^1 \widetilde{G}(\pi, r)' dr\right) \end{pmatrix} \begin{pmatrix} \theta_A \\ \gamma \end{pmatrix}$$

Let $\eta \equiv (\theta_A', \gamma')'$ and define $a(s)$ to be such that $v(s)' = \eta' a(s)$, that is:

$$a(s) = \begin{pmatrix} -D\overline{M}' \\ -\widetilde{G}(\pi, r)'\overline{M}' + \left(\int_0^1 \widetilde{G}(\pi, r)' dr\right)' B^{1/2} H B^{-1/2} \overline{M}' \end{pmatrix}$$

$A(\pi)$ is defined as $\int_0^1 a(s) dZ(s)$ and $V(\pi)$ is defined as $\int_0^1 a(s) a(s)' ds$. When there is only one break,

[13] (a)-(d) follow from direct calculation. Details are provided in an appendix available upon request.

and $\widetilde{G}(\pi,s) = 1\,(s \geq \pi)\,G$ then direct calculations show that:

$$A\left(\pi\right) = \begin{pmatrix} -D\overline{M}'Z\left(1\right) \\ -G'\overline{M}'[Z(1)-Z(\pi)] + \left(1-\pi\right)G'B^{1/2}HB^{-1/2}\overline{M}'Z\left(1\right) \end{pmatrix} = \begin{pmatrix} -D\overline{M}' & 0 \\ -\left(1-\pi\right)G'A'D\overline{M}' & G'\overline{M}' \end{pmatrix}\begin{pmatrix} Z(1) \\ Z(\pi)-\pi Z(1) \end{pmatrix} \tag{44}$$

$$V\left(\pi\right) = \begin{pmatrix} DBD' & \left(1-\pi\right)DB^{1/2}\left(I-H\right)B^{1/2}G \\ \left(1-\pi\right)G'B^{1/2}\left(I-H\right)B^{1/2}D' & \left(1-\pi\right)G'B^{1/2}[I-\left(1-\pi\right)H]B^{1/2}G \end{pmatrix} \tag{45}$$

Proof of Propositions 1, 2 and 3

When the weighting function is an $(r+p)$-dimensional multivariate normal distribution with zero mean and covariance $U\left(\pi\right)$ then in this case, and for two-sided alternatives, the optimal tests in (13) simplifies to (by completing the square and integrating out the parameter vector):

$$TS = \int \left(\frac{|U(\pi)^{-1}|^{1/2}}{|V\left(\pi\right)+U(\pi)^{-1}|^{1/2}} \exp\left\{ \frac{1}{2} A\left(\pi\right)' \left(V\left(\pi\right)+U(\pi)^{-1}\right)^{-1} A\left(\pi\right) \right\} \right) dJ(\pi) \tag{46}$$

When $U(\pi)^{-1} = \frac{1}{c}V\left(\pi\right)$ then (up to a constant factor that does not matter):

$$TS = \int \left(\frac{|U(\pi)^{-1}|^{1/2}}{|V\left(\pi\right)+U(\pi)^{-1}|^{1/2}} \exp\left\{ \frac{1}{2} \Phi^*\left(\pi\right) \right\} \right) dJ(\pi),\;\; \Phi^*\left(\pi\right) = A\left(\pi\right)' V\left(\pi\right)^{-1} A\left(\pi\right) \tag{47}$$

By using (45) and standard formulas for the inverse of a partitioned matrix:

$$V\left(\pi\right)^{-1} = \begin{pmatrix} \frac{1}{\pi}AB^{-1}A' & -\frac{1}{\pi}AB^{-1}G \\ -\frac{1}{\pi}G'B^{-1}A' & \frac{1}{\pi(1-\pi)}G'B^{-1}G \end{pmatrix} \tag{48}$$

By combining the above with (44), one finds that:

$$\Phi^*\left(\pi\right) = \left(Z(1)' \vdots \frac{\{Z(\pi)-\pi Z(1)\}'}{\sqrt{\pi\left(1-\pi\right)}} \right) \begin{pmatrix} C_1'(C_1C_1')^{-1}C_1 & 0 \\ 0 & C_2'(C_2C_2')^{-1}C_2 \end{pmatrix} \begin{pmatrix} Z(1) \\ \frac{Z(\pi)-\pi Z(1)}{\sqrt{\pi(1-\pi)}} \end{pmatrix} \tag{49}$$

where $C_1 \equiv \left(AB^{-1}A'\right)^{-1}AB^{-1}\overline{M}'$ has dimension $(r \times m)$ and $C_2 \equiv G'\overline{M}'$ has dimension $(p \times m)$. Notice that $C_1\left(I-\overline{H}\right) = C_1$ so that $C_1 Z\left(1\right) = C_1 B_m\left(1\right)$. Thus, $Z_r(1) \equiv (C_1C_1')^{-1/2}C_1 Z(1)$ is an r-vector of independent standard normals and $\{Z_p(\pi) - \pi Z_p(1)\} \equiv (C_2C_2')^{-1/2}C_2\left\{Z(\pi)-\pi Z(1)\right\}$

is a p-vector of independent Brownian Bridges because $\left\{(C_1C_1')^{-1/2}C_1\right\}\left\{(C_1C_1')^{-1/2}C_1\right\}' = I_p$ (same for C_2). Hence:

$$A(\pi)' V(\pi)^{-1} A(\pi) = Z_r(1)'Z_r(1) + \frac{\{Z_p(\pi) - \pi Z_p(1)\}' \{Z_p(\pi) - \pi Z_p(1)\}}{\pi(1-\pi)} \tag{50}$$

Thus, under the null hypothesis:

$$\Phi^*(\pi) = B_r(1)'B_r(1) + \frac{BB_p(\pi)'BB_p(\pi)}{\pi(1-\pi)} \tag{51}$$

Proposition 1 thus follows from Result 1 and the Continuous Mapping Theorem, and Propositions 2 and 3 follow directly from Proposition 1, Corollary 1 and the results in Andrews and Ploberger (1994).

Asymptotic local power.

Under the alternative hypothesis, and using (12):

$$Z(1) = (I - \overline{H})\left\{B_m(1) - \overline{M}D'\theta_A - \overline{M}\int_0^1 g(s)ds\right\}$$

$$Z(\pi) - \pi Z(1) = BB_m(\pi) - (1-\pi)\overline{M}\int_0^{\pi} g(s)ds + \pi\overline{M}\int_{\pi}^1 g(s)ds$$

and substituting these into (49):

$$\Phi^*(\pi) = Z_r^{(1)\prime}Z_r^{(1)} + Z_p^{(2)}(\pi)'\, Z_p^{(2)}(\pi) \tag{52}$$

where:

$$Z_r^{(1)} \equiv B_r(1) - (C_1C_1')^{-1/2}C_1\overline{M}\left(\int_0^1 g(s)\,ds + D'\theta_A\right) \tag{53}$$

$$Z_p^{(2)}(\pi) \equiv \frac{\{B_p(\pi) - \pi B_p(1)\}}{\sqrt{\pi(1-\pi)}} - (C_2C_2')^{-1/2}C_2\overline{M}\left(\left(\frac{1-\pi}{\pi}\right)^{1/2}\int_0^{\pi} g(s)\,ds - \left(\frac{\pi}{1-\pi}\right)^{1/2}\int_{\pi}^1 g(s)\,ds\right) \tag{54}$$

Note that when $A = [I_p \quad 0_{p\times q}] = G$, $g(.) = [I_p \quad 0_{q\times p}]g_\beta(.)$, then $r = p$, $C_1 \equiv \overline{M}_\beta{}'(I - \overline{P}_\delta)$ (see (c) in the Proof of Corollary 1) and $C_2 \equiv \overline{M}_\beta'$; note also that $\overline{M}_\beta'(I - \overline{P}_\delta)\overline{M}_\beta = \nabla_{\beta\beta}Q - \nabla_{\beta\delta}Q\,(\nabla_{\delta\delta}Q)^{-1}\,\nabla_{\delta\beta}Q$. In addition, note that when π is fixed and $\theta_A = 0$, which is the case examined by Chow (1960) for *testing the existence of structural breaks only*, only $[I\vdots 0]A(\pi)$ and

$[I\vdots 0]V(\pi)[I'\vdots 0']'$ are relevant so that the test becomes $\Phi(\pi) \Rightarrow BB_p(\pi)'BB_p(\pi)/\pi(1-\pi)$, which is Andrews (1993) result (see also Sowell (1993)). Notice also that when $\pi = 1$, which is *the case without structural break*, the result is the classical test statistic for tests on a subset of p parameters: $B_p(1)' B_p(1) \sim \chi^2_{(p)}$ because $BB_p(1) = 0$ and $B_p(1)$ is a p-dimensional multivariate standard normal distribution.

The proof that (21), (22) and (23) are asymptotically equivalent under both the null hypothesis and the local alternatives follows from applying results similar to those in Andrews (1993) and Newey and McFadden (1994).

Proof of Proposition 4

The (modified) Nyblom test statistic for testing both parameter instability and that the parameter vector is equal to some value β_0 was defined as:

$$Nyblom_T^* = \int_0^1 \left(T \cdot \nabla_\beta Q_{1,[\pi T]}(\beta_0, \widetilde{\delta})' \widehat{\Omega}_N^{-1} \nabla_\beta Q_{1,[\pi T]}(\beta_0, \widetilde{\delta})\right) J(\pi)\, d\pi \tag{55}$$

where $\widehat{\Omega}_N \in R^{2p \times 2p}$ is a consistent estimate of the asymptotic variance of $\nabla_\beta Q(\beta_0, \widetilde{\delta})$, the gradient function is defined as:

$$\nabla_\beta Q_{1,[\pi T]}(\beta_0, \widetilde{\delta}) \equiv \frac{1}{T} \sum_{t=1}^{[T\pi]} \nabla_\beta F_T(\beta_0, \widetilde{\delta}, \pi)' \Sigma^{-1} f_t\left(x_t, \beta_0, \widetilde{\delta}\right) \tag{56}$$

Notice that $\frac{1}{T} \sum_{t=1}^{[T\pi]} f_t\left(x_t, \beta_0, \widetilde{\delta}\right)$ is the first component of $\overline{F}_T(\beta_0, \widetilde{\delta}, \pi)$, so that one would expect the asymptotics to be driven by $B(\pi)$. In fact, let $\widetilde{\delta}$ be estimated on observations $1, 2..T$ and take a mean value expansion to obtain:

$$\begin{aligned}
\sqrt{T} \nabla_\beta Q_{1,[\pi T]}(\beta_0, \widetilde{\delta}) &= \nabla_\beta F_T(\theta_0)' \widehat{\Sigma}^{-1} \frac{1}{T} \sum_{t=1}^{[T\pi]} f_t(\theta_0) \sqrt{T} - \frac{1}{T} \sum_{t=1}^{[T\pi]} \nabla_\beta F_T(\theta_0)' \widehat{\Sigma}^{-1} \nabla_\delta f_t(\theta_0) \\
&\quad \left(\nabla_\delta F_T(\theta_0)' \widehat{\Sigma}^{-1} \nabla_\delta F_{1,T\pi}(\theta_0)\right)^{-1} \nabla_\delta F_T(\theta_0)' \widehat{\Sigma}^{-1} F_{1,T\pi}(\theta_0) \sqrt{T} \\
&= \overline{M}_\beta' \left(I_m - \overline{P}_\delta\right) \Sigma^{-1/2} \frac{1}{\sqrt{T}} \sum_{t=1}^{[T\pi]} f_t(\theta_0) + op(1)
\end{aligned}$$

where $F_{1,[T\pi]}(\theta_0) \equiv \frac{1}{T} \sum_{t=1}^{[T\pi]} f_t(x_t, \theta_0)$ has the following asymptotic distribution:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[T\pi]} f_t(\theta_0) \Rightarrow \Sigma^{1/2} B_m(\pi)$$

Hence, (56) is such that:

$$\nabla_\beta Q_{1,[\pi T]}(\beta_0,\widetilde{\delta}) \Rightarrow \overline{M}'_\beta \left(I-\overline{P}_\delta\right) B_m(\pi)$$

$$\begin{aligned} Nyblom^*_T &= \int_0^1 T\cdot\left(\nabla_\beta Q_{1,[T\pi]}(\beta_0,\widetilde{\delta})'\widehat{\Omega}_N^{-1}\nabla_\beta Q_{1,[T\pi]}(\beta_0,\widetilde{\delta})\right) J(\pi)\, d\pi \qquad (57)\\ &\Rightarrow \int_0^1 B_m(\pi)'\left(I_m-\overline{P}_\delta\right)\overline{M}_\beta\left(\overline{M}'_\beta\left(I_m-\overline{P}_\delta\right)\overline{M}_\beta\right)^{-1}\overline{M}'_\beta\left(I_m-\overline{P}_\delta\right)B_m(\pi)\,J(\pi)\,d\pi \\ &= \int_0^1 B_p(\pi)'\,B_p(\pi)\,J(\pi)\,d\pi \end{aligned}$$

Notice that, like the (modified) Andrews and Ploberger case for $c\to 0$, this statistic is a special case of (46); in fact, the $Nyblom^*_T$ and the modified $Mean-Wald^*_T$ statistics simply use two different weighting matrices. Notice that in the structural break case only, the test statistic is constructed on the basis of the first component of $\overline{F}_T(\widehat{\theta},\pi)$ and the estimation of β transforms the Brownian Motion in (57) into a Brownian Bridge, thus originating the Nyblom test statistic: $\int_0^1 B_p(\pi)'\,B_p(\pi)\,J(\pi)\,d\pi$.

Tables

Table 1

$$\overline{Q}(\beta_1,\beta_2,\delta) \equiv \overline{F}_T(\beta_1,\beta_2,\delta,\pi)'\widehat{\Gamma}\overline{F}_T(\beta_1,\beta_2,\delta,\pi)$$

Wald test

$$\widehat{\theta}(\pi) = \arg\min_{\beta_1,\beta_2,\delta}\overline{Q}(\beta_1,\beta_2,\delta)$$

$$\overline{F}_T(\beta_1,\beta_2,\delta,\pi) = \left(\frac{1}{T}\sum_{t=1}^{[T\pi]} f_t(\beta_1)', \frac{1}{T}\sum_{t=[T\pi]+1}^{T} f_t(\beta_2)'\right)' \in R^{2m\times 1}$$

$$R \equiv \begin{pmatrix} I_p & -I_p & 0_{p\times q} \\ \pi I_p & (1-\pi) I_p & 0_{p\times q} \end{pmatrix}, \qquad r \equiv \begin{pmatrix} 0_{p\times 1} \\ \beta_0 \end{pmatrix}$$

$$V\left(\widehat{\theta}(\pi)\right) = \left[M(\pi)'\widehat{\Gamma}M(\pi)\right]^{-1}$$

$$M(\pi) \equiv \begin{pmatrix} \pi M_\beta & 0 & \pi M_\delta \\ 0 & (1-\pi) M_\beta & (1-\pi) M_\delta \end{pmatrix} \in R^{2m\times 1}$$

$$\widehat{\Gamma} \equiv \begin{pmatrix} \pi\widehat{\Sigma}_1 & 0 \\ 0 & (1-\pi)\widehat{\Sigma}_2 \end{pmatrix}^{-1} \underset{p}{\rightarrow} \Gamma = \begin{pmatrix} \pi\Sigma & 0 \\ 0 & (1-\pi)\Sigma \end{pmatrix}^{-1} \in R^{2m\times 2m}$$

$$\widehat{\Sigma}_1 = \frac{1}{[T\pi]}\sum_{t=1}^{[T\pi]}\left(f_t\left(\widehat{\beta}_1\right) - \frac{1}{[T\pi]}\sum_{t=1}^{[T\pi]} f_t\left(\widehat{\beta}_1\right)\right)\left(f_t\left(\widehat{\beta}_1\right) - \frac{1}{[T\pi]}\sum_{t=1}^{[T\pi]} f_t\left(\widehat{\beta}_1\right)\right)'$$

$$\widehat{\Sigma}_2 = \frac{1}{T-[T\pi]}\sum_{t=[T\pi]+1}^{T}\left(f_t\left(\widehat{\beta}_2\right) - \frac{\sum_{t=[T\pi]+1}^{T} f_t\left(\widehat{\beta}_2\right)}{T-[T\pi]}\right)\left(f_t\left(\widehat{\beta}_1\right) - \frac{\sum_{t=[T\pi]+1}^{T} f_t\left(\widehat{\beta}_1\right)}{T-[T\pi]}\right)'$$

LM test

$$LM_1 \equiv \frac{1}{T}\left(W_T^{1/2}\sum_{t=1}^{T} f_t(\beta_0)\right)'\widehat{\Omega}_1\left(W_T^{1/2}\sum_{t=1}^{T} f_t(\beta_0)\right)$$

$$LM_2(\pi) \equiv \frac{1}{\pi(1-\pi)}\frac{1}{T}\left(\sum_{t=1}^{[T\pi]} f_t(\beta_0) - \pi\sum_{t=1}^{T} f_t(\beta_0)\right)'\widehat{\Sigma}^{-1/2\prime}\widehat{\Omega}_2\widehat{\Sigma}^{-1/2}\left(\sum_{t=1}^{[T\pi]} f_t(\beta_0) - \pi\sum_{t=1}^{T} f_t(\beta_0)\right)$$

$$\widehat{\Omega}_1 = C_1'\left(C_1C_1'\right)^{-1}C_1, \qquad C_1 = \overline{M}_\beta{}'\left(I - \overline{P}_\delta\right)$$

$$\widehat{\Omega}_2 = C_2'\left(C_2C_2'\right)^{-1}C_2, \qquad C_2 = \overline{M}_\beta{}'$$

$$\widehat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}\left(f_t(\beta_0) - \frac{1}{T}\sum_{t=1}^{T} f_t(\beta_0)\right)\left(f_t(\beta_0) - \frac{1}{T}\sum_{t=1}^{T} f_t(\beta_0)\right)'$$

Distance Metric test

$$\widetilde{\theta}(\pi) = \arg\min_{\beta_1,\beta_2,\delta}\overline{Q}(\beta_1,\beta_2,\delta) \quad \text{s.t. } R\theta(\pi) = r$$

Note. Table 1 assumes that $f_t(.)$ consists of mean zero uncorrelated random variables. When $f_t(.)$ consists of mean zero but serially correlated random variables then consistent estimation of $\widehat{\Sigma}_1$, $\widehat{\Sigma}_2$, $\widehat{\Sigma}$ and $\widehat{\Gamma}$ requires a HAC estimator, e.g. Newey and West (1987).

Appendix 2. Asymptotic critical values for the optimal test statistics

Tables 2 to 5 report critical values for the optimal tests and the QLR_T^* test considered in section 3. The significance levels considered in the tables are 10%, 5%, 2.5% and 1%. The critical values are obtained by simulating the asymptotic distributions described section 3. The number of Monte Carlo replications is 5,000.[14]

Table 2 Asymptotic critical values of the Exp-W* statistic

p	0.10	0.05	0.025	0.01	p	0.10	0.05	0.025	0.01
1	2.449	3.1339	3.8168	4.6727	16	22.8516	24.6265	26.2193	27.8914
2	4.2039	5.0154	5.8842	6.8129	17	24.1981	26.0939	27.6806	29.6425
3	5.6563	6.7377	7.7042	8.9198	18	25.5895	27.4142	29.1579	31.2214
4	7.0946	8.1907	9.3191	10.4213	19	26.6029	28.3524	30.1111	31.9303
5	8.744	9.8237	10.9535	12.1939	20	28.1078	30.0753	31.8821	33.854
6	10.0262	11.2035	12.4487	14.0387	25	34.1382	36.1769	37.9084	40.3421
7	11.422	12.6298	13.8575	15.1726	30	39.9551	42.1666	44.3736	46.5745
8	12.872	14.2249	15.3435	16.7515	35	45.9472	48.1833	50.3254	53.1
9	14.1384	15.537	16.9444	18.6276	40	51.8203	54.4758	56.542	59.3382
10	15.4258	16.7608	18.3168	19.7861	50	63.0979	65.6882	68.1909	71.2993
11	16.7578	18.4675	19.5883	21.5468	60	74.5343	77.3691	80.178	83.6545
12	17.9154	19.5819	21.1368	22.6856	70	86.0377	89.291	92.1205	94.9551
13	19.2881	20.945	22.7773	24.986	80	97.3483	100.6474	103.4514	107.5657
14	20.691	22.2846	23.8681	25.5793	90	108.5332	111.5582	114.873	118.682
15	21.6263	23.3855	24.799	26.6754	100	119.7486	123.6853	126.9866	130.7547

[14] The trimming values considered are only 15% and 85% of the available sample and the grid of points is quite sparse, basically each observation is a point in the grid.

Table 3 Asymptotic critical values of the Mean-W* statistic

p	0.10	0.05	0.025	0.01	p	0.10	0.05	0.025	0.01
1	4.2635	5.3643	6.6754	8.151	16	41.3398	44.5649	47.3768	50.5095
2	7.2918	8.7434	10.3015	12.1901	17	43.5754	46.8724	49.7746	54.276
3	10.0139	11.9196	13.5691	15.6527	18	46.0905	49.4028	52.3423	56.2983
4	12.4217	14.3619	16.1385	18.3462	19	48.1377	51.5089	54.585	58.1488
5	15.5389	17.5227	19.3382	21.712	20	50.8136	54.1002	58.0219	62.0964
6	17.7534	19.8775	22.0944	24.7766	25	61.8342	66.2019	69.0632	73.3436
7	20.1055	22.3892	24.4337	27.0243	30	72.7187	76.963	80.8042	86.127
8	22.8585	25.3968	27.3492	30.5082	35	84.3009	88.6818	92.6035	97.1965
9	25.3693	27.8441	30.3481	32.9441	40	95.549	99.9184	104.3606	109.6221
10	27.3176	30.0394	32.3736	35.6678	50	116.9329	121.6882	126.1939	131.5243
11	30.1304	32.9938	35.2951	38.4508	60	138.8518	144.073	148.8159	155.2877
12	32.0777	34.8801	37.3805	41.1871	70	160.5187	166.7235	171.5768	178.3526
13	34.6088	37.6914	40.6089	44.2636	80	182.5128	188.9752	193.4568	200.146
14	37.205	40.4177	43.1792	46.3284	90	204.1154	210.1062	215.1015	222.7963
15	38.9861	42.1837	44.786	47.9691	100	225.5653	232.3089	239.0936	246.1133

Table 4 Asymptotic critical values of the QLR* statistic

p	0.10	0.05	0.025	0.01	p	0.10	0.05	0.025	0.01
1	8.1379	9.8257	11.3324	13.4811	16	51.4352	55.16	58.4212	61.5675
2	12.1958	14.2247	16.0883	17.9423	17	54.1	58.1212	61.5166	65.132
3	15.5623	17.6405	19.9225	22.4823	18	56.91	60.8365	64.4692	68.8502
4	18.6108	21.0546	22.9886	25.9972	19	59.1165	62.7163	66.2358	70.014
5	22.1572	24.5498	26.5899	29.4812	20	62.1529	66.0226	69.7486	73.9967
6	24.8171	27.3766	29.7808	33.2166	25	74.4081	78.6485	82.5688	87.0932
7	27.7544	30.4143	32.8785	35.9484	30	86.2048	90.7147	94.9425	100.0492
8	30.7228	33.7173	36.2602	38.9622	35	98.0434	102.8202	107.0778	112.5657
9	33.5534	36.5523	39.2284	42.9048	40	110.0489	115.2632	119.6745	125.4108
10	36.1735	39.02	41.8908	45.2661	50	132.8999	138.3241	143.1622	150.215
11	38.8891	42.3319	44.8426	48.936	60	155.9591	161.5571	167.068	175.1412
12	41.3338	44.8198	47.7471	51.3894	70	179.0538	185.7491	191.6062	196.8787
13	44.0167	47.5886	51.3857	56.2506	80	201.8076	208.1976	214.2283	222.182
14	46.8766	50.1926	53.3939	56.9349	90	224.1198	230.1001	236.9304	245.06
15	49.0594	52.3466	55.5618	59.0825	100	246.6137	254.605	261.5406	268.5731

Table 5 Asymptotic critical values of the Nyblom* statistic

p	0.10	0.05	0.025	0.01	p	0.10	0.05	0.025	0.01
1	1.1032	1.4037	1.8028	2.2513	16	10.5320	11.4176	12.2337	13.1866
2	1.8762	2.2922	2.6763	3.2492	17	11.1075	12.0112	12.8397	13.9754
3	2.5889	3.1429	3.5900	4.1732	18	11.7596	12.6497	13.4916	14.5962
4	3.1777	3.7502	4.2464	4.8621	19	12.1511	13.1508	13.9863	14.9916
5	3.9724	4.5500	5.0995	5.7410	20	12.9113	13.8067	14.9378	16.0645
6	4.5453	5.1566	5.7730	6.5641	25	15.7323	16.8186	17.7874	18.8564
7	5.1553	5.7525	6.3490	7.0790	30	18.4758	19.5648	20.7284	22.1473
8	5.8693	6.5547	7.0895	8.0249	35	21.3091	22.4827	23.6481	24.8462
9	6.4874	7.1835	7.8843	8.5516	40	24.1748	25.3908	26.6214	28.0032
10	6.9591	7.7067	8.4535	9.2094	50	29.5766	30.8592	32.0037	33.6264
11	7.6836	8.5127	9.1704	9.9631	60	35.1489	36.5507	37.9336	39.4751
12	8.1816	8.9698	9.6757	10.7436	70	40.5455	42.2035	43.6241	45.2102
13	8.8694	9.6591	10.5097	11.3733	80	46.0555	47.7638	49.1463	50.8235
14	9.5038	10.3344	11.1732	12.0471	90	51.5255	53.2490	54.6519	56.2967
15	9.9639	10.8163	11.5417	12.4267	100	56.9164	58.8967	60.6025	62.6136