04/RT/11

Páipéar Taighde Teicniúil
Research Technical Paper

Structural Breaks
An Instrumental Variable Approach

Denis Conniffe and Robert Kelly

Banc Ceannais na hÉireann
Central Bank of Ireland
Eurosystem

Structural Breaks
An Instrumental Variable Approach*

Denis Conniffe[†] and Robert Kelly[‡]
University College Dublin and Central Bank of Ireland

Abstract

A structural change test and corresponding change estimator of an instrumental variable nature are proposed. The strengths of the approach lie in its ease of application and the strong test power. It does not suffer from critical value adjustments required by the CUSUM type tests and is unique in that it tests and measures the size of the break in one operation. The power of this test is compared to others in the literature, both algebraically and through simulations, with favourable results.

JEL classification: C01, C12, C15.

Keywords: Structural Break Test, Structural Break Estimation, Instrumental Variable.

*The views expressed in this paper are those of the authors and do not necessarily reflect those of the Central Bank of Ireland or the ESCB.Any errors are the sole responsibility of the authors.

[†] This work was completed before Denis Conniffe's untimely passing in Janurary 2011.

[‡] Contact: Central Bank of Ireland - Financial Stability Department, North Wall Quay, Spencer Dock, Dublin 1, Ireland. E-mail: robert.kelly@centralbank.ie

Non-Technical Summary

At the most basic level, undetected structural breaks yield biased and inefficient OLS estimates. Therefore, the issue of detecting and estimating structural breaks has been of high importance in the literature for many decades. An approach based on an instrumental variable is proposed with three major strengths over previous tests; (i) ease of application, (ii) strong test power and (iii) ability to detect and estimate gradual breaks.

As a starting point, the instrumental variable can be assumed to be a time trend across the whole sample - a significant coefficient provides both a test for change and unbiased estimate of size. Usually, one knows some information relating to the location of the break point. We may not know if it had a lead effect due to expectations or a lag due to required time adjustment, but it is possible to specify a most probable region for the break point with locations further away from that region being less likely. This information can be incorporated into the IV variable and improve the power of our test.

Estimation of change, as distinct from testing for it, has been relatively neglected in the literature. The usual approach is to fit a dummy variable to the predicted point of change. If this point is not precisely correct or the change is gradual, all estimates of break size are biased. The IV estimate is obtainable irrespective of assumptions about the number or pattern of changes although its bias and efficiency depend on the true change structure.

1 Introduction

The topic of testing for structural change and the associated one of measuring that change if it occurred has been of interest for many years. Often, the invariance of a regression equation is the issue in question. Does a model $y_i = w_i^{'}b + u_i$ hold over the whole sample of n time periods or should it be,

$$\begin{aligned} y_i &= w_i^{'}b + u_i & \text{i} &= 1, 2, 3, ..., r \\ y_i &= w_i^{'}(b + \Delta) + u_i & \text{i} &= r + 1, r + 2, ..., n \end{aligned} \tag{1}$$

where r, the break point, is usually unknown a priori.[1] Indeed some of the literature has considered multiple break points giving,

$$\begin{aligned} y_i &= w_i^{'}b + u_i & \text{i} &= 1, 2, 3, ..., r_1 \\ y_i &= w_i^{'}(b + \Delta_1) + u_i & \text{i} &= r_1 + 1, r_1 + 2, ..., r_2 \\ y_i &= w_i^{'}(b + \Delta_2) + u_i & \text{i} &= r_2 + 1, r_2 + 2, ..., r_3 \end{aligned} \tag{2}$$

etc. with all r_i unknown. While a large amount of research has appeared in the statistical and econometric literature, the problem is far from satisfactorily resolved. In particular, the accuracy of the distributional approximations on which currently available methodologies rely are very dubious unless sample size is huge and consequently test sizes and reported powers can be incorrect.

In this paper we propose an instrumental variable type estimator for the size of change, which also provides a test statistic for the existence of a change. Our method is conceptually simple and performs well in comparison to currently recommended methods. We do make some assumptions to reduce the complexity of the problem. We assume that interest focuses

[1] In reality we will often have some, but uncertain, information about the location of a break point. For example, Ireland joining the European Union no doubt altered coefficients in many Irish economic models. But whether these occurred with a lead due to expectations or a lag due to required time adjustment or indeed occurred in stages over a period may be unclear. We could perhaps specify a most probable region for the break point with locations further away from that region being less likely. We can probably totally exclude times greatly before or after joining.

on detecting the occurrence of change and estimating the size of change. So, at least initially, we are not concerned with the precise location of the change point or even whether change occurred in stages at consecutive time points.[2] So we will be testing for the occurrence of at least one break point and estimating total change. In future research we will discuss how these assumptions might be relaxed, although we consider they are often quite plausible.

2 Previous Approaches

Clearly, if we could postulate one particular time period for a single change, the appropriate test should be based on how much two models, fitted before and after that point, seem to differ from each other. Employing the classical F test for this assessment is the common approach and is usually called a Chow test. The change in coefficients would then be estimable from regressions fitted before and after the break point.

The obvious difficulty is that the break point will usually be unknown and the approach of doing F tests at every possible break point invalidates the significance level (size) of each test. Quandt (1958) initially thought he had avoided the difficulty by estimating the break point by maximum likelihood (ML) and performing and F test at that point. But it is easily shown that the ML estimate of the break point actually corresponds to the point at which the largest F value occurs, so that the procedures are actually identical and the test size wrong.[3] Quandt (1960) reverted to suggesting always test for a break point at $r = n/2$, on the idea that if there was a real break anywhere coefficient estimates from before and after $r = n/2$ should differ at least somewhat. Clearly the test will be powerful only if there is a single true break point near $r = n/2$.[4] Perhaps even more importantly, the estimated

[2] In this we are no different from previous authors who choose not to specify the precise alternative model to the null of no structural change.

[3] Quandt (1958) also showed likelihood-ratio test methodology could be applied, but the validity conditions fail when the null hypothesis of no change is true.

[4] Presumably Quandt would have chosen a point other than the mid-point if there was some (even inconclusive) evidence for one.

coefficient change could be badly biased.

Performing multiple tests requires considerable modification to the critical test value criteria to compensate for the number of tests been performed, so as to keep the true size level equal to the nominal one. The Bonferroni bound approach can assist if there are only a few possible change points (very small sample size), but that is rarely the situation.[5] Various test criteria, including largest F, have been proposed and almost all amount to functions of the maximum partial sum (CUSUM) of regression residuals or their squares. The exact finite stable distributions of these criteria seem intractable so approximations have been adopted. Most appeal to the highly developed corpus of results about boundary crossing probabilities with continuous time stochastic processes. Brown, Durbin & Evans (1975) looked at partial sums of recursive (i.e. independent) residuals and approximated the distribution through conversion to a continuous time process (Brownian motion) and derived a test from known probability formulae for that process crossing a boundary. Similar approaches to cumulative sums of ordinary residuals which sum to zero lead to Brownian Bridge processes and several authors have developed the corresponding tests. These include Ploberger & Kramer (1992), Andrews (1993), Bai (1994) and more recent papers. But these Wiener process style approximations are never accurate unless sample size is very large. When it is not, the test sizes are often much greater than the nominal levels leading to incorrect rejection of the null much too often and a corresponding distortion of the power of the tests when the change is real. Some papers have extended the framework to discuss multiple possible change points including Andrews, Lee and Ploberger (1996), Lui, Wu & Zidek (1997) and Bai & Perron (1998). However, continuous time approximations then require large samples between each possible pair of break points exacerbating finite sample inaccuracies.

Other approaches have appeared in the literature aimed at better approximation of cumulative sum distributions at finite sample size. These include James, James & Siegmund (1987) and Conniffe & Spencer (2000) and their approximations proved far more accurate

[5]Efron (1997) has outlined a possible approach.

than Wiener approximation approaches for the simple model considered. However, there may be a formidable difficulty in extending the approaches to more general regression models.

The exponents of Wiener process approximations are not unaware of inaccuracy. Indeed Ploberger & Kramer (1992) remarked that simulation showed their test overestimated size appreciably for $n = 120$.[6] Bai & Perron (2003) suggested corrections to their (1998) test for multiple change points to adjust for size errors. These corrections are not straightforward, requiring case specific adjustments depending on sample size and the values of explanatory variables. Altissimo & Corradi (2003) also found excess test size for a moderate number of observations and suggested a correction. In general, the literature has focused on testing for a change and locating the change point rather than estimating the actual change. This is perhaps because once a change point is estimated, it is easy to fit models to each segment. But as will be demonstrated, the resulting estimates are not necessarily best. So while the literature is impressive in many respects, there remains substantial room for improvement.

3 An Instrumental Variable Approach

To illustrate the idea, take the simplest case,[7]

$$y_t = \mu + e_t \qquad \text{t} = 1, 2, 3, ..., r$$
$$y_t = \mu + \delta + e_t \qquad \text{t} = r + 1, r + 2, ..., n$$

which could be written,

$$y_t = \mu + \delta x_t + e_t \qquad \text{t} = 1, 2, 3, ..., n$$

where $x_t = 0$ when $t \leq r$ and $x_t = 1$ when $t > r$. In reality we do not know r precisely, so we might 'guess' a value $r*$ and fit the model

$$y_t = \mu + \delta x_t^* + e_t \qquad \text{t} = 1, 2, 3, ..., n$$

[6] Conniffe & Spencer (2000) showed it overestimated by far greater amounts at lower n.

[7] Although simple, it is the illustrative case used frequently in the literature. For example, see Altissimo & Corradi (2003)

where $x_t^* = 0$ when $t \leq r^*$ and $x_t^* = 1$ when $t > r^*$. A test of change could be based on the significance of the difference from zero of $\tilde{\delta} = Sx^*y/Sx^{*2}$. Of course, this would just be Quandt's(1960) approach if $r^* = n/2$. Now suppose, z_t is a non-decreasing variable, correlated to x_t^*, but invariant to the discrepancy between x_t and x_t^*. We could seek such an appropriate instrumental variable just as we do in other econometric applications, but as we will show shortly, it is always possible to construct one. We take the instrumental variable type estimator, $\hat{\delta} = Szy/Szx^*$, as the change estimator and base our test for a change point on the significance of its difference from zero.

Just making the trivial assumption that $E(e_i) = 0$, it is easy to show,

$$E(\hat{\delta}) = \delta \frac{(n-r)}{n-r^*} \frac{\bar{z}_r - \bar{z}}{\bar{z}_{r^*} - \bar{z}}$$

where $\bar{z}_r$, $\bar{z}_r$ and $\bar{z}$ are the mean values of z_t for $t > r$, $t > r^*$ and for all n, respectively. If $r < r^*$, that is if the true change point occurs before our guessed point, $(n-r)/(n-r^*) > 1$. But since z_t is non-decreasing $\bar{z}_{r^*}/\bar{z}_r$ so $(\bar{z}_r - \bar{z})/(\bar{z}_{r^*} - \bar{z}) < 1$ and the product may remain close to unity. The estimator $\hat{\delta}$ will usually be biased unless $r = r^*$, but may be much less so than alternatives. An example of a z vector is $(0, 0, 0, ..., 0, 1/s, 2/s, 3/s, ..., 1, 1, 1, ..., 1)'$ where s is the time range over which change could have taken place. If we thought any point was a viable contender for change-point, though as previously noted, we think it will usually be possible to rule out some early and some late time points,[8] this variable would become $(1, 2, 3, ..., n)'/n$, a time trend variable.

Taking the simplest model $y_t = \mu + e_t$ when $t \leq r$ and $y_t = \mu + \delta + e_t$ when $t > r$, we can illustrate bias by comparing $\hat{\delta}$, based on the time trend, with the estimator, $\tilde{\delta} = Sx^*y/Sx^{*2}$, the 'OLS' estimator, with, for convenience, r^* chosen as $n/2$. Testing if $\hat{\delta}$ differs from zero would, of course, be Quandt's test for this model. Of course comparisons with the modern CUSUM (Wiener process style) tests will be more interesting, but because of the complexity involved, we defer this to the next section.

[8]We realise that this implies a very simple test for change could be based on early verus late points (which could actually be formulated as an IV test) but this would omit much of the data information.

Now suppose change actually occurred at $t = n/2 - L$, that is with a lead of L, it is easy to show

$$E(\tilde{\delta}) = E\left(\frac{Sx^*y}{Sx^{*2}}\right) = \delta\left(1 - \frac{2L}{n}\right) \tag{3}$$

while

$$E(\hat{\delta}) = E\left(\frac{Sty}{Stx^*}\right) = \delta\left(1 - \frac{4L^2}{n^2}\right) \tag{4}$$

Obviously replacing $\tilde{\delta}$ by $\hat{\delta}$ has reduced the bias from $O(n^{-1})$ to $O(n^{-2})$ for a fixed L.[9] Of course if r is $n/2$, so that $L = 0$, both (3) and (4) are unbiased and $\hat{\delta}$ will have larger variance than $\tilde{\delta}$.

In fact $Var(\hat{\delta}) = (16/3)(n^2 - 1/n^3)\sigma^2$, while $Var(\tilde{\delta}) = 4\sigma^2/n$ so that, approximately, $Var(\hat{\delta}) = 1.3(Var(\tilde{\delta}))$. So mean square-error, (MSE) = ((Bias)^2+Variance) is probably the fairest criterion on which to base comparisons. MSE($\tilde{\delta}$)-MSE($\hat{\delta}$) is,

$$\frac{4\delta^2L^2}{n^2}\left(1 - \frac{4L^2}{n^2}\right) - 0.3\frac{\sigma^2}{n}$$

Clearly if $L = 0$ this is negative, but as L increases it becomes positive if δ^2 is appreciable. Switching the consideration to testing rather than estimation, the power of a test based on $\hat{\delta}$ will be inferior to the Quandt test if the true break point is at $n/2$. We can easily quantify by how much through comparing the non-centrality parameters (NCP) of the tests, which determine the power of F or t tests. For the Quandt and IV tests the NCP values are

[9] IV estimators are not unbiased, but usually described as consistent. But how potential length of lead or lag increases with n needs consideration. It may not be 'fixed' but could conceivably increase with n. However, as noted previously, we do not think it could increase indefinitely with n. Assuming L is proportional to $\sqrt{n}$ may be generous enough and then the bias of (4) is $O(n^{-1})$ so that it is $\sqrt{n}$ consistent, in the usual asymptotic sense, while (3) is not.

$$\frac{n\delta^2}{4\sigma^2} \quad \text{and} \quad \frac{3n^3\delta^2}{16\sigma^2(n^2-1)}$$

respectively. For large n this is approximately 1/4 v 3/16. But this is the maximum penalty for using the IV approach. On the other hand for lag length L the corresponding NCP values are

$$\frac{n\delta^2}{4\sigma^2}\left(1-\frac{2L}{n}\right)^2 \quad \text{and} \quad \frac{3\delta^2}{\sigma^2}\left(\frac{n^2}{4}-L^2\right)^2 / \left(n(n^2-1)\right)$$

respectively. Now the IV test is more powerful unless L/n is very small. The IV test approach can clearly also have power in testing for an overall change which occurs in stages. Obviously, IV would not have power against detecting a jump at one point, followed later by an equal fall - a zero net change. But that might not be considered a true change in the mean (or coefficients) at all, but an increase in volatility.

For this special case of z a time trend, $\hat{\delta} = Sty/Stx^*$, the test of $\delta = 0$ is easily shown to be the same as for a time trend in the model $y = \mu + bt + e_t$, a test which was proposed and investigated by Farley, Hinich & McGuire (1975)[10], although they did not address the problem of estimating change should the test prove significant. They compared their test with some others, in particular with Quandt's (1960) 'always in the middle test' and found it to be quite competitive. Of course, assuming z to be a pure time trend assumes there is no such trend in the model to begin with.

The extension to allow for explanatory variables is obvious. For example, with one explanatory variable, both μ and β could change in $y_t = \mu + \beta w_t + e_t$. If we knew r we could estimate and test for changes by fitting the model,

[10]The test was first derived by Farley & Hinich (1970), but not in a form immediately recognisable as a time trend test.

$$y_t = \mu + \delta x_t + \beta w_t + \theta x_t w_t + e_t$$

with r unknown we can replace it by our 'guesstimate' r^*, generate x_t^* and in the terminology of two stage least squares replace it by its OLS predicator from z_t, say $\hat{x}_t$ and regress y_t on the 'instruments' $\hat{x}_t$, w_t and $\hat{x}_t w_t$. Note that two variables involving x_t^*, does not mean a single z variable is inadequate.

4 Comparisons with CUMSUM Approach: Size and Power of Tests

It is more interesting to compare the IV test and estimator with the more modern CUSUM (Wiener process based) approaches. Initially, we take the IV variable to be a time trend. We commence with comparisons of tests in terms of size and power. Because of the complexity of the finite sample distribution of the maximum partial sum it seems impossible to obtain algebraic expressions permitting comparisons. So a simulation approach will be adopted here. We commence by employing the time trend as z.

4.1 Actual and Nominal Test Sizes

Firstly, we examine the size of the tests, that is the probability of detecting a break when none exists. The first column of Table 1 gives the proportion of times the significance level ($\alpha = 0.05$) is exceeded by the p-value of the Ploberger & Kramer (1992) test, in a Monte Carlo simulation with 10,000 replications for each sample size. Sample sizes range from 10 to 100 in increments of 10, all drawn from the standard normal distribution. Columns two and three report the results for the Bai and Perron (1998) and IV tests respectively. In the case of Bai & Perron (1998), a test of one break is compared to a null of no breaks.

The Ploberger & Kramer (1992) results show their test yields too high a proportion of values in the tails, with p-values double that of significance levels at low levels of n. Even with $n = 100$, the rejection region is overestimated by 40 per cent. Although the Bai & Perron (1998) test performs better, particularly as sample size increases, size is consistently poorer than the instrumental variable approach.

Bai & Perron (2003) investigate the practical issues surrounding the empirical application of their (1998) test, in particular the behavior of estimators and test statistics in finite samples. This involves simulating the limiting distribution of the estimators and test statistics, yielding an adjustment coefficient, ϵ, that depends on sample size. The break test critical values, ϵ are all lower than Bai & Perron (1998) and thus providing poorer size than the results presented in the Bai & Perron case in Table 1.

4.2 Power of the Tests

For a given size, the objective of a test is to maximise statistical power (or minimise Type II error). So we compare the statistical power of the IV and CUSUM approaches taking the Ploberger & Kramer (1992) test as representative of the CUSUM based tests. Ploberger & Kramer, like ourselves, did not develop their test in the context of a specific alternative model. They visualised the alternative model as change possibly occurring in steps at a number of change points or even in a continuous manner over time. However, they illustrate the power of their test by simulation for a single change point and so we will exercise this case. Firstly, the Ploberger & Kramer test must be size corrected to avoid an upward bias in statistical power. Two samples, of equal size are drawn from the normal distribution with a difference in means equal to the break size. The break size is calculated as $X/\sqrt{n}$, where n is the sample size and X takes on values 0 to 12 in increments of 0.02.[11] This is consistent with the view that there is an inverse relationship between the power to detect

[11] Due to the extremely poor power of CUSUM tests for small samples, X takes the values 0 to 24 in increments of 0.02, when the sample size is 10. This is necessary for the power of the CUSUM tests to converge to 1.

smaller breaks and sample size. The break point location is taken as the sample mid-point. The power of the tests is calculated in a Monte Carlo simulation with 10,000 replications for each break size.

Figure 1 shows a comparison of power for various sample sizes.[12] The power of the IV approach, again using a time trend, is consistent over sample size, easily outperforming the power of CUSUM tests. For very small samples (e.g. 10), when the power of the IV test is approaching 1, the CUSUM test is only reporting power of 0.20. In fact, a sample size of approximately 5,000 is required before the power of both approaches is comparable.

4.3 Changing Break Point Location

The power of both the Ploberger & Kramer and the IV test can depend on the location of the true change point as is perhaps obvious. So simulations were repeated at a series of break points and the results are summerised in Figures 2 & 3 for sample sizes of 20 and 80 respectively. In general, the symmetry between break locations equidistant from start or finish is evident. Figure 2 shows for small samples, the IV test displays better power with the superiority mentained over a much wider distribution of break size and location (The poor power of CUSUM test is evident from the relatively narrower peak and smaller area of green). For a larger sample soze of 80, Figure 3 shows neither test possesses much power at break points close to the start of finish. Although nethier test performs well at the start and end of smaples, the proportiate advantages of the IV test remain. For example, the IV approach displays 33 per cent greater statistical power for break point loacted at 8 or 72. Overall, the superiority of the IV test as found when the break location was a mid-sample is not only maintained but increased as the location moves away from mid-sample. It is also worth noting, these simulations display the worse case for the IV approach as the dummy, x_i^* can be located anywhere in the sample, drastically improving power.

[12] Although results for four sample sizes are shown, simulations were conducted for sample size up to 10,000 with these graphs depicting the general trend. This is purely to limit the amount of simulation results to be reported in figures.

4.4 Prior Information about Break Point Location

The above simulations are based on the special case of the IV approach that there is no known prior information about the location of the break point. This corresponds to the weakest power for the IV estimator. Therefore, we conduct a simulation comparison which allows for prior location information. However, it could be argued that the comparisons of IV and CUSUM made in the previous section cannot fairly be taken over an IV test with location information. If the CUMSUM test users possessed the information implicit in this choice IV, they would correspondingly limit the range of the partial sums of residuals.

Therefore, a second simulation is performed to compared the power of this test to the previous IV and CUSUM approaches. We take a sample of 100 derived from a standard normal distribution, with a prior knowledge of a break within a band 10 per cent either side of the mid-point of the sample is assumed, when the true break occurs at $n/2$. The IV approach uses an instrumental variable as defined in section 3, with increments only between observations 40 and 60. The residuals for the CUSUM are estimated over the entire sample but the range of partial sums tested are limited to the central 20 per cent of the sample.

The power curves, derived from the simulation procedure outlined in section 4.2, for these estimators are presented in Figure 4. Unsurprisingly, the results show no improvement for the CUSUM test, as prior information has no effect on the estimation of partial sums but rather limits the testable region, but, being asymptotically justified, CUSUM test critera do not adjust with sample size. Incorporating prior information yields significant improvement in power for the IV approach. For a power of 0.5, it displays 30 and 22 per cent greater power compared to the CUSUM and original IV tests respectively.

5 Comparing Estimators of Change

Since the Ploberger & Kramer test did not specify a precise alternative model, it is not clear what ought to be estimated without further consideration. However, for consistency with our model we assume one change point is located through the occurrence of the largest partial sum of residuals and change is estimated by the difference in means before and after this location. [13] Whether change should be counted as zero if the test is non-significant, needs consideration. While Ploberger & Kramer did not discuss estimation, Bai & perron (1998) did in a possible multi change context and etimated changes by differences between successive detected regimes. However, the detection of different regimes and hence estimation of the total number of change points was conditional on test significance. So the testing and estimation phases are not fully separable. But limiting estimation to the cases of the simulation where the test is significant would obviously bias the change estimate downwards, so it seems fairer to base the comparisons with the IV estimator on all simulations.[14] Of course, there is another source of bias in the Ploberger & Kramer estimator, having to estimate the break point introduces sampling error which causes a finite sample bias in the change estimate and feeds into the variance of the estimator.

In Figure 5, we compare the biases of the two forms of Ploberger & Kramer estimator with that of the IV estimator. Sample size is 50 and the IV estimator choosen assumes no change before $t = 16$ or after $t = 35$. The break point is taken as the mid-point. Unsurprisingly, the bias of the Ploberger & Kramer estimator setting change to zero for non-significant tests is greatest but even the alternative Ploberger & Kramer estimator shows more bias than the IV estimator. The difference is very large for small changes but decreases at larger changes. The IV estimator appears almost unbiased, partly for the reasons discussed in Section 3 but also because it implicity employs a comparsion of initial and final sets of time points. Figure 6 repeats the simulation but takes the break point at 80 per cent of the mid-point. Previous pattern of biases reoccurs.

[13]Or of course OLS estimates of coefficients where explanatory variables are in the model.

[14]See Figure 5.

Comparisons of esitmators are usually based on mean square-error, (MSE) = ((Bias)^2+Variance), accounting for both the bias and standard error. Figure 7 displays the MSE for the comparison of means based on the CUSUM tests and the estimates from the IV variable constructed as in Figure 5. Sample size, change point region and the true break point also correspond to Figure 5. It is clearly evident that the IV estimator is superior for breaks up to a moderate size. For larger breaks the MSE of both appraoches are comparable.

6 Discussion & Conclusions

At the most basic level, undetected structural breaks yield biased and inefficient OLS estimates. Therefore, the issue of detecting and estimating structural breaks has been of high importance in the literature for many decades. This paper describes an instrumental variable type approach, which leads to an estimate of overall change whatever the number and pattern of changes. Our test for change is simply a test for difference of this estimate from zero. The resulting test avoids some of the shortcomings of previous approaches and allows for ease of application. In particular, this IV test always has correct size which is a problem with the most popular current approaches, based on the maximum sum of of regression residuals (CUSUM).

Comparsions of the power of the IV test with the size corrected Ploberger & Kramer test are favourable to the IV test over a wide range of circumstances. The IV or z variable employed is valid provided it satisfies the conditions specified in Section 2, although of course instrumental variables incorporating more information and more correlated with x^*, show greater precision.

Estimation of change, as distinct from testing for it, has been relatively neglected in the literature. The IV estimate is obtainable irrespective of assumptions about the number or pattern of changes although its bias and efficiency depend on the true change structure. However, the algebric comparsions with the Quandt (1960) estimates and the simulation

comparsions with the CUSUM estimates are relatively favourable to the IV estimate for a single time change. A huge number of situations could concievably be simulated in terms of number of change points, location of them, sizes of changes and sample sizes. However, the findings in this paper are encouraging for the IV estimator at the very least.

References

[1] Altissimo F, Corradi V. 2003. Strong Rules for Detecting the Number of Breaks in a Time Series. *Journal of Econometrics* **117**:207-244

[2] Andrews D.W.K. 1993. Test for Parameter Instability and Structural Change with Unknown Change Point. *Econometrica* **61**: 821-856

[3] Andrews D.W.K, Lee I, Ploberger W. 1996. Optimal Change Point Tests for Normal Linear Regression. *Journal of Econometrics* **70**: 9-38

[4] Bai, J. 1994. Estimation of Structural Change Based on Wald-Type Statistics. *Massachusetts Institute of Technology (MIT), Dept of Economics Working Papers*: **94-06**

[5] Bai J, Perron P. 1998. Estimating and Testing Linear Models with Multiple Structural Changes *Econometrica*: **66**: 47-78

[6] Bai J, Perron P. 2003. Computation & Analysis of Multiple Structural-Change Models. *Journal of Applied Econometrics* **18**: 1-22.

[7] Brown R.L, Durbin J, Evans J., 1975. Techniques for Testing the Constancy of Regression Relationships over Time (with discussion), *Journal of Royal Statistical Society* **37**: 149-192

[8] Chow G. 1960. Tests of Equality between Sets of Coefficients in Two Linear Regressions. *Econometrica***28**: 591-605.

[9] Conniffe D, Spencer J, 2000. Approximating the Distribution of the Maximum Partial Sum of Normal Deviates, *Journal of Statistical Planning & Inference* **88**: 19-27.

[10] Efron B. 1997. The Length Heuristic for Simultaneous Hypothesis Tests, *Biometrika* **84**: 143-157.

[11] Farley J, Hinich M, 1970. A Test for a Shifting Slope Coefficient in a Linear Model, *Journal of the American Statistical Association* **65**: 1320-1329.

[12] Farley J, Hinich M, McGuire T, 1975. Some Comparisons of Tests for a Shift in the Slopes of a Multivariate Linear Time Series Model. *Journal of Econometrics* **3**: 297-318.

[13] James B, James K.L, Siegmond D. 1987. Tests for a change point, *Biometrika* **74**: 71-83.

[14] Lui J, Wu S, Zidek J. 1997. On Segmented Multivariate Regression, *Statistica Sinica* **7**: 497-525.

[15] Quandt R. 1958. The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes *Journal of the American Statistical Association* **53**: 873-880.

[16] Quandt R. 1960. Tests of the Hypothesis that a Linear Regression System Obeys Two Separate Regimes, *Journal of the American Statistical Association* **55**: 324-330.

[17] Ploberger W, Kramer W, 1992. The CUSUM Test with OLS Residuals. *Econometrica* **60**: 271-285.

Table 1: Size of Tests based on Monte Carlo Simulation

Sample Size	PK(1992)	BP(1998)	IV Approach
10	0.147	0.122	0.075
20	0.104	0.077	0.055
30	0.089	0.069	0.053
40	0.085	0.060	0.051
50	0.077	0.058	0.050
60	0.075	0.056	0.049
80	0.070	0.054	0.051
100	0.068	0.055	0.050

Notes: *Size of tests based on Monte Carlo Simulation with 10,000 replications. Samples are drawn from the normal distribution. PK = Ploberger & Kramer (1992) and BP = Bia & Perron (1998)*

Figure 1
Comparsion of Statistical Power for Various Sample Sizes

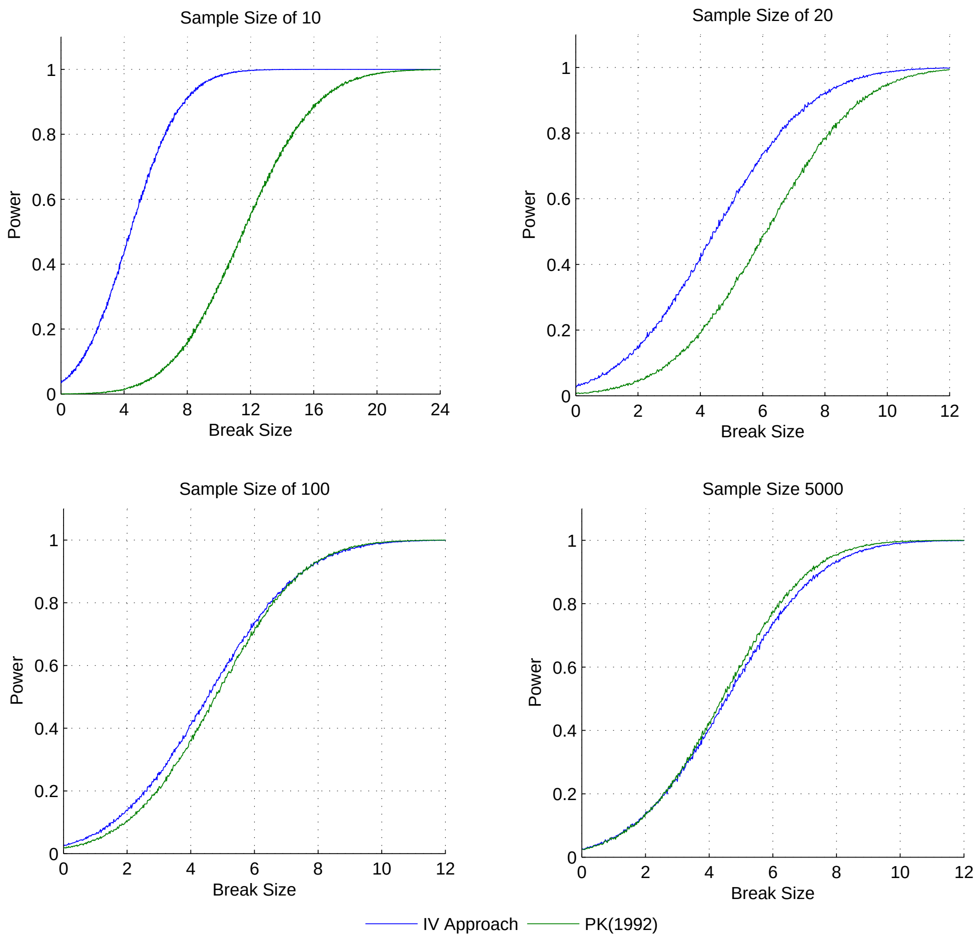

Figure 2
Comparsion of Statistical Power for a Change in Break Lociation (n=20)

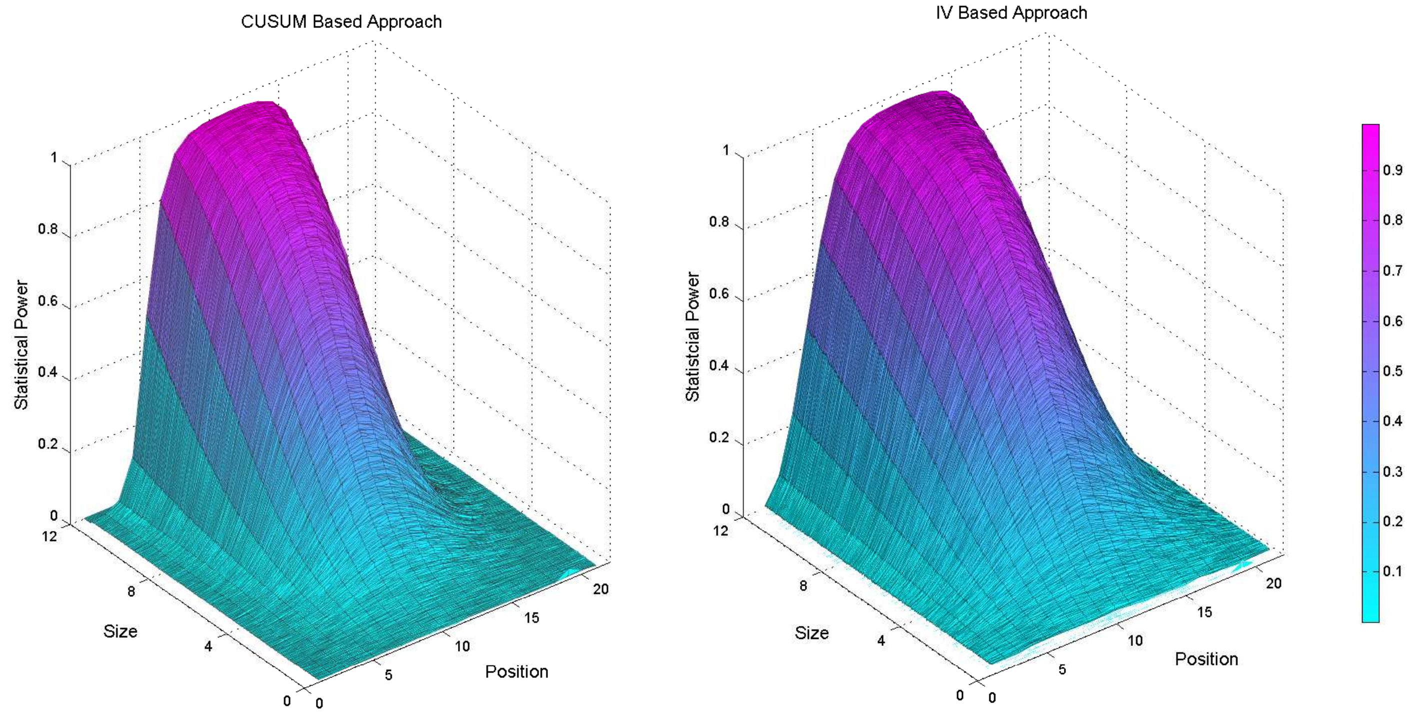

Figure 3
Comparsion of Statistical Power for a Change in Break Point Location (n=80)

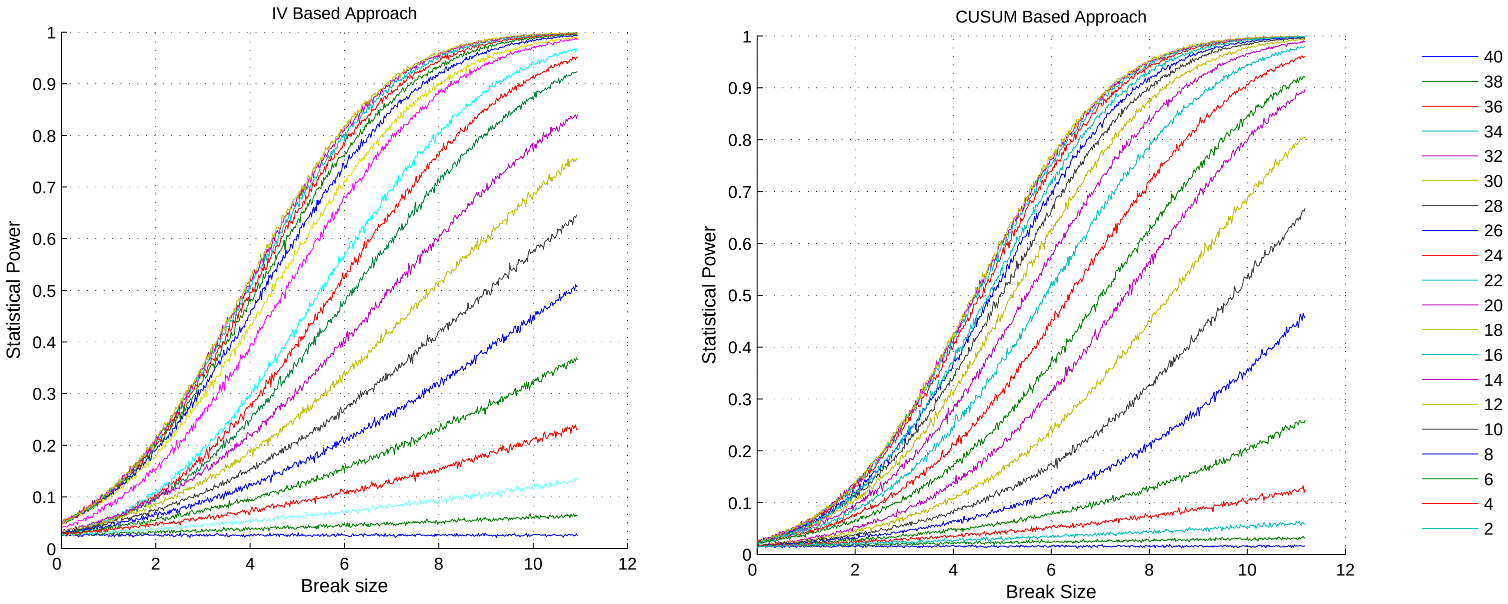

Figure 4
Comparsion of Statistical Power for Modified IV Variable (n=100)

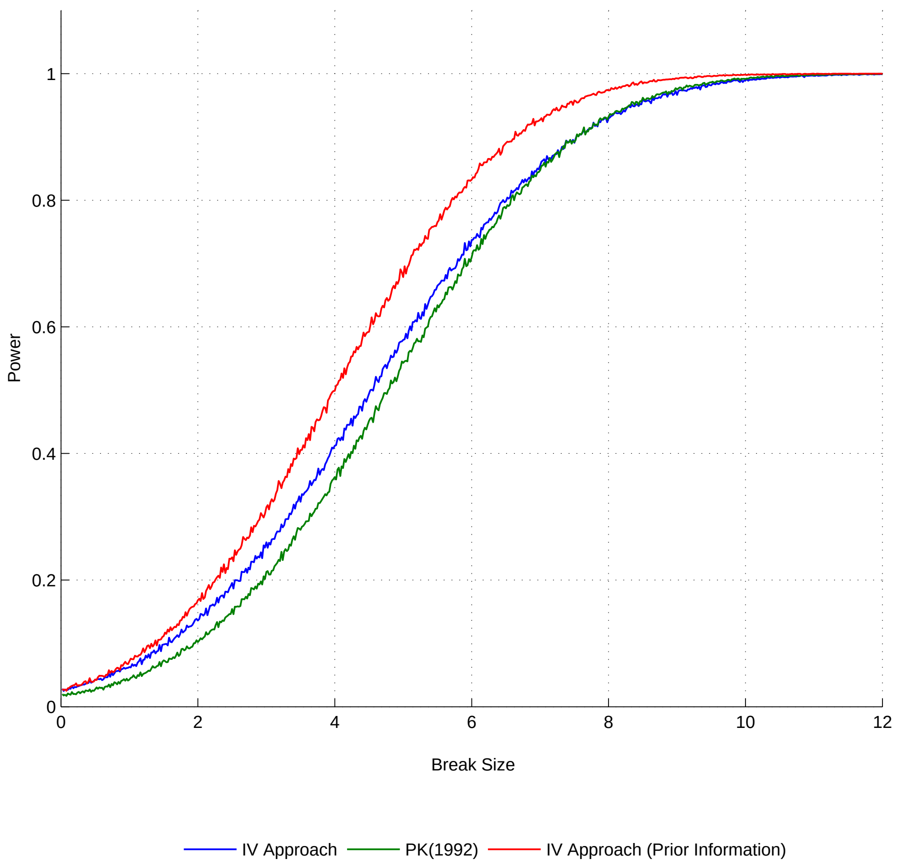

Figure 5
Comparsion of the Bias in the Estimators of Change (n=100)

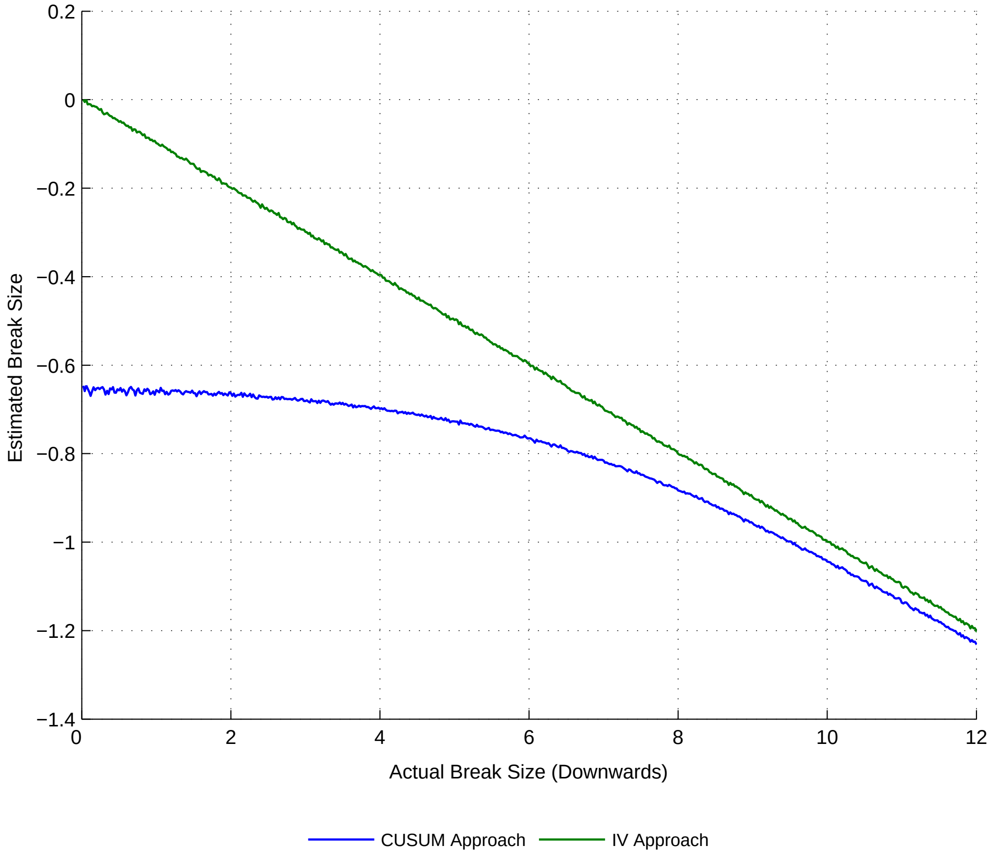

Figure 6
Comparsion of the Estimators of Change Mean Squared Error (n=100)