Change point estimation in regressions with $I(d)$ variables

Chih-Chiang Hsu*

Department of Economics, National Central University, Chung-Li 32054, Taiwan

Received 18 February 2000; accepted 15 July 2000

Abstract

In this paper we study the least-squares change-point estimator in regressions with stationary and invertible $I(d)$ regressors and disturbances. We find that the least-squares estimator remains consistent when there is a one-time break, but it may identify a spurious change when there is none. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Change point estimation; Fractional Brownian motion; Spurious change

JEL classification: C22

1. Introduction

The change-point problem has long been a leading research topic in the statistics and econometrics literature. For a mean shift in weakly dependent $I(0)$ processes, Bai (1994) established the consistency and rate of convergence results for the least-squares change-point estimator under fairly general conditions. These results were subsequently extended by Kuan and Hsu (1998) for fractionally integrated $I(d)$, $-0.5 < d < 0.5$, processes with a mean change. Similar results were also obtained for stationary regression models, time trend models, and cointegration models; see e.g., Nunes et al. (1995) and Bai (1997, 1998). On the other hand, Nunes et al. (1995) and Kuan and Hsu (1998) observed that, for the data that are nonstationary $I(1)$ series or have long memory ($I(d)$ with $0 < d < 0.5$), the least-squares estimator may suggest a spurious change point when there is none. In this paper, we extend the results of Kuan and Hsu (1998) to regression models with stationary and invertible $I(d)$ regressors and disturbances. It is shown that the least-squares estimator remains consistent when there is a one-time change. When there is no change and the disturbances have long memory, we demonstrate that existing structural-change tests will reject the null with probability
approaching one and hence may incorrectly suggest that a change has occurred. Moreover, the least-squares estimator may also identify a spurious change point in the middle of the sample. Interestingly, these phenomena are irrelevant to the dependence structure of regressors. In contrast with the spurious regression results of Tsay and Chung (2000), our finding indicates that a spurious change may still occur even when there is no spurious regression. This paper proceeds as follows. Least-squares consistency is established in Section 2. The spurious-change results of test and estimation are discussed in Section 3. Section 4 gives some concluding remarks. All proofs are given in Appendix A.

2. Consistency

We consider the data generating process (DGP) of \( y_t \) which has a one-time change at unknown point \( k_0 \):

\[
y_t = x_t' \beta_1 + x_t' \lambda_1 \mathbf{1}_{(t > k_0)} + \varepsilon_t, \quad t = 1, \ldots, T,
\]

where \( x_t = [1, x_t]' \), \( \beta = [a, b]' \), \( i = 1, 2 \), \( \lambda = \beta_2 - \beta_1 = [\lambda_1, \lambda_2]' \) denotes the magnitude of parameter changes, and \( \mathbf{1} \) is the indicator function. Also let \( k_0 = [T \tau_0] \), the integer part of \( T \tau_0 \), so that \( \tau_0 \) represents the relative position of \( k_0 \) in the sample. Given (1), we impose the following conditions.

[A1]. \( \tau_0 \in [\tau, \overline{\tau}] \), where \( \tau < \overline{\tau} \) and \( [\tau, \overline{\tau}] \) is a proper subset of \( [0, 1] \).

[A2]. Both \( x_t \) and \( \varepsilon_t \) are stationary and invertible ARFIMA (autoregressive, fractionally integrated, moving average) processes:

\[
\Phi_x(L)(1 - L)^d x_t = \Theta_x(L)v_t, \quad \Phi_x(L)(1 - L)^d \varepsilon_t = \Theta_x(L)u_t,
\]

where \( \Phi(L) \) and \( \Theta(L) \) are polynomials in the lag operator \( L \) such that their roots lie outside the unit circle, and \(-0.5 < d_x, d_\varepsilon < 0.5 \).

[A3]. The processes \( \{v_t\} \) and \( \{u_t\} \) are two mutually independent sequences of i.i.d. random variables with zero mean, finite variances \( \sigma_v^2 \) and \( \sigma_u^2 \), and \( \mathbb{E}|v_t|^d < \infty \) with \( \delta_v \equiv \max\{4, -8d_x/(1 + 2d_x)\} \), and \( \mathbb{E}|u_t|^\delta_u < \infty \) with \( \delta_u \equiv \max\{4, -8d_\varepsilon/(1 + 2d_\varepsilon)\} \).

Conditions [A2] and [A3] ensure that \( x_t \) and \( \varepsilon_t \) obey the functional central limit theorem (FCLT):

\[
\frac{1}{T^{0.5 + d_x}} \sum_{t=1}^{[T \tau]} x_t \Rightarrow \kappa \mathbb{B}^{d_x}(\tau), \quad \frac{1}{T^{0.5 + d_\varepsilon}} \sum_{t=1}^{[T \tau]} \varepsilon_t \Rightarrow \kappa \mathbb{B}^{d_\varepsilon}(\tau),
\]

where \( \Rightarrow \) denotes weak convergence, \( \kappa \) is a positive constant, and \( \mathbb{B}^{d_x} \) is the fractional Brownian motion. In what follows we also write \( \to \) as convergence in probability and \( \Rightarrow \) as convergence in distribution.

For each hypothetical change point \( k \), the least-squares change-point estimator solves \( \hat{k} = \arg\min_{1 \leq k \leq T} \text{RSS}(k) \), where
and $\hat{\beta}_1(k)$ and $\hat{\beta}_2(k)$ are the pre- and post-change least-squares estimators of $\beta_1$ and $\beta_2$. The estimated break fraction, $\hat{\tau}$, is defined as

$$\hat{\tau} = \inf \{ \tau^*: \tau^* = \arg\min_{\tau \in [\tau_0]} \text{RSS}([T\tau]) \},$$

and we may write $\hat{\tau} = \hat{k}/T$. It is easy to verify that

$$\text{RSS}([T\tau]) = \sum_{i=1}^{T} (e_i + x_i' A_1_{(i \geq [T\tau_0])})^2 - \bar{H}_f([T\tau]),$$

where

$$\bar{H}_f([T\tau]) = \bar{E}_f(0,\tau)M_T^{-1}(0,\tau)\bar{E}_f(0,\tau) + \bar{E}_f(\tau,1)M_T^{-1}(\tau,1)\bar{E}_f(\tau,1),$$

with

$$\bar{E}_f(\tau_1,\tau_2) = \sum_{i=[T\tau_1]+1}^{[T\tau_2]} x_i (e_i + x_i' A_1_{(i \geq [T\tau_0])}) \quad \text{and} \quad M_f(\tau_1,\tau_2) = \sum_{i=[T\tau_1]+1}^{[T\tau_2]} x_i x_i'.$$

Hence, the least-squares estimator of (3) is the same as that obtained from maximizing (4). Bai (1997) establishes the consistency result for $\hat{\tau}$ in regression models with weakly dependent data; Kuan and Hsu (1998) also prove consistency for the location model with ARFIMA data. In this section, we generalize these consistency results to regression models with ARFIMA regressors and disturbances.

**Theorem 2.1.** Given DGP (1), suppose that Conditions [A1]–[A3] hold. Then $\hat{\tau} \xrightarrow{p} \tau_0$ as $T \to \infty$, where $\hat{\tau}$ is given by (3).

We report some Monte Carlo results in Fig. 1. The data $y_i$ are generated according to (1) with $\beta_1 = [1,0.5]'$, $\beta_2 = [2,1]'$, and $\tau_0 = 0.5$. Four combinations of $d_\xi$ and $d_\zeta$ are considered: $(d_\xi,d_\zeta) = (0.4,0.4), (-0.4,0.4), (0.4,-0.4), (-0.4,-0.4)$. In each experiment, $T = 100$, and the number of replications is 50000. Fig. 1 shows that the empirical distributions of $\hat{k}$ are more concentrated around the true change point, although their precisions vary with $d_\xi$ and $d_\zeta$. These simulations confirm the result of Theorem 2.1. However, it seems that the value of $d_\zeta$ is more crucial in determining the performance of $\hat{k}$. In view of Fig. 1 (a) and (c) (or (b) and (d)), we can see that for $d_\zeta = 0.4$ ($d_\zeta = -0.4$), $\hat{k}$ performs better when $d_\xi$ is negative. On the other hand, comparing Fig. 1 (a) and (b), we observe that given $d_\zeta = 0.4$, $\hat{k}$ performs similarly for $d_\xi = 0.4$ and $d_\xi = -0.4$; in fact, positive $d_\xi$ yields a slightly better result. The same conclusion holds for Fig. 1 (c) and (d).

**3. Spurious change**

We have shown that the least-squares estimator is capable of locating the change point when it exists. In this section, however, we demonstrate that it is difficult to know if there is indeed a
structural change in $I(d)$ data. In practice, researchers usually first test the existence of a change and then decide if they should proceed to estimate the change point. Unfortunately, as noted by Kuan and Hsu (1998), many well-known tests for structural change have significant size distortions when data have long memory. In contrast with Tang and MacNeill (1993), these size distortions do not result from ignoring nuisance parameters of weak dependence. In fact, the null of parameter constancy will be rejected with probability approaching one when data have strong positive correlations. In regressions with $I(d)$ variables, consider the Wald statistic for the null hypothesis $\beta_1 = \beta_2$ against the alternative (1). For each possible break $k$, the Wald statistic is

$$W_k = (\hat{\beta}_1(k) - \hat{\beta}_2(k))' (\hat{V}(1,k) + \hat{V}(k,T))^{-1} (\hat{\beta}_1(k) - \hat{\beta}_2(k)),$$

where $\hat{V}(\tau_1,\tau_2) = M_T^{-1}(\tau_1,\tau_2) \text{RSS}/T$. Under the null hypothesis, $\lambda = 0$, and hence

$$\hat{E}_T(\tau_1,\tau_2) = E_T(\tau_1,\tau_2) = \sum_{t=[T\tau_1]}^{[T\tau_2]} x_t e_t, \quad \hat{H}_T([T\tau]) = H_T([T\tau])$$

$$= E_T'(0,\tau)M_T^{-1}(0,\tau)E_T(0,\tau) + E_T'(\tau,1)M_T^{-1}(\tau,1)E_T(\tau,1).$$

It follows that
\[ W_r([T\tau]) = \frac{H_r([T\tau]) - H_r}{\left( \sum_{i=1}^{T} \varepsilon_i^2 + H_r([T\tau]) \right) / T}, \]

where \( H_r = E_r^t(0,1)M^{-1}_r(0,1)E_r(0,1). \) The next result shows that \( W_r([T\tau]) \) diverges when \( d_\epsilon > 0, \) regardless of the value of \( d. \)

**Theorem 3.1.** Given DGP (1) with \( \lambda = 0, \) suppose that Conditions [A2] and [A3] hold with \( 0 < d_\epsilon < 0.5. \) Then for each \( \tau \in [\hat{\tau}, \hat{\tau}], \)

\[ T^{-2d}W_r([T\tau]) \xrightarrow{D} \frac{\kappa^2_0[B_{d_\epsilon}(\tau) - \tau B_{d_\epsilon}(1)]}{\gamma_\epsilon(0)\tau(1 - \tau)} \]

as \( T \to \infty, \) where \( \gamma_\epsilon(0) \) is the variance of \( \varepsilon. \)

Theorem 3.1 shows that for a positive \( d_\epsilon, W_r([T\tau]) \) diverges in probability. By the continuous mapping theorem, the supremum Wald test of Andrews (1993) and mean Wald and exponential Wald tests of Andrews and Ploberger (1994) also diverge to infinity. Thus, when there is no change, the Wald-type tests would reject the null of parameter constancy with probability approaching one as long as the disturbances have long memory (i.e., \( 0 < d_\epsilon < 0.5 \)), and hence may incorrectly suggest that a change has taken place. As the result of Theorem 3.1 does not depend on \( d, \) this conclusion holds even when the regressor is antipersistent (i.e., \( -0.5 < d_\epsilon < 0 \)). The Lagrange multiplier and likelihood ratio tests also suffer from the same problem. Moreover, when \( \hat{V} \) is replaced with the Newey-West estimator to account for weak dependence, we can show that \( W_r \) still diverges as \( d_\epsilon > 0. \)

Our simulations also confirm the asymptotic analysis of Theorem 3.1. The DGP is generated according to (1) with \( \beta_1 = \beta_2 = [1,2]^\ell \) for different combinations of \( d_\epsilon \) and \( d: d_\epsilon = -0.45, \ldots, 0.45 \) and \( d = 0.1, \ldots, 0.45 \) with the increment 0.05. Thus, there are 162 experiments; for each experiment, \( T = 200, \) the number of replications is 5,000, and the nominal size is 5%. We compute the supremum Wald test of Andrews (1993). In computing the Wald statistic, we estimate \( \hat{V} \) by the Newey-West estimator. The bandwidth is determined by the data-dependent formula of Andrews (1991) with AR(1) specification. The rejection frequencies of the supremum Wald test are summarized in Fig. 2. It can be seen that the type I errors increase from approximately 25% for \( d_\epsilon = 0.1 \) to about 95% for \( d_\epsilon = 0.45, \) regardless of the value of \( d. \) Therefore, as long as \( d_\epsilon \) is positive, the existing tests suffer from serious size distortions.

The previous results suggest that, whether or not there is a change, it is likely that one would be led to estimate the change point after conducting standard structural-change tests. From the proof of Theorem 3.1, we know that

\[ \hat{\tau} = \arg \max_{\tau \in [\hat{\tau}, \hat{\tau}]} T^{-2d}H_r([T\tau]) \]

\[ \xrightarrow{D} \arg \max_{\tau \in [\hat{\tau}, \hat{\tau}]} \frac{\kappa^2_0B_{d_\epsilon}(\tau)}{\tau} + \frac{\kappa^2_0[B_{d_\epsilon}(1) - B_{d_\epsilon}(\tau)]}{(1 - \tau)} \]

\[ = \arg \max_{\tau \in [\hat{\tau}, \hat{\tau}]} G_{d_\epsilon}(\tau), \]
for any $\tau \in [\overline{\tau}, \overline{\tau}]$. By invoking the functional law of iterated logarithm (Taqqu, 1977), we have

$$\limsup_{\tau \to 0} \frac{B_{d_e}(\tau)^2}{\tau} = \limsup_{\tau \to 1} \frac{[B_{d_e}(1) - B_{d_e}(\tau)]^2}{1 - \tau} = 0 \quad \text{a.s.}$$

Thus, the asymptotic behavior of $G_{d_e}(\tau)$ is well-defined on $[0,1]$ with $G_{d_e}(0) = G_{d_e}(1) = \kappa^2B_{d_e}(1)^2$. It follows that $\hat{\tau}$ has a limiting distribution with support equal to $[0,1]$, and that $\hat{\tau} \to \text{argmax}_{\tau \in [0,1]} G_{d_e}(\tau)$. However, the maximum of $G_{d_e}(\tau)$ is not attained at 0 or 1 with probability one, since

$$G_{d_e}(0) - G_{d_e}(\tau) = G_{d_e}(1) - G_{d_e}(\tau) = -\frac{\kappa^2[\tau B_{d_e}(1) - B_{d_e}(\tau)]^2}{\tau(1 - \tau)} < 0,$$

for any $\tau \in (0,1)$. This proves the following theorem.

**Theorem 3.2.** Given DGP (1) with $\lambda = 0$, suppose that Conditions [A2] and [A3] hold with $0 < d_e < 0.5$. If $\overline{\tau} = 0$ and $\overline{\tau} = 1$, then $\hat{\tau} \to \text{argmax}_{\tau \in [0,1]} G_{d_e}(\tau)$ and $G_{d_e}(0) = G_{d_e}(1) < G_{d_e}(\tau)$ for any $\tau \in (0,1)$ with probability one.

Ideally, when there is no change, we expect that the probability mass of the change-point estimator concentrates around the two end points of the sample; i.e., $\hat{\tau} \to [0, 1]$. Theorem 3.2 shows that when $d_e > 0$, $\hat{\tau}$ is unlikely to be close to 0 and 1; it is more likely that $\hat{\tau}$ incorrectly identifies a change point in the middle of the sample even when there is none. This is the problem of spurious change, in the sense of Nunes et al. (1995). Nunes et al. (1995) and Bai (1998) showed that the $I(1)$ regressors and errors are responsible for the occurrence of spurious break and spurious regression. Tsay and Chung (2000) also pointed out that the spurious regression may occur when $d_s + d_e > 0.5$. From Theorem 3.2, we can see that the spurious-change problem may still exist even if the regressors and disturbances are stationary and $d_s + d_e < 0.5$. That is, a spurious change may arise when there is no spurious regression.

We examine some simulations which are conducted as that in Section 2 but $\beta_2 = \beta_2 = [1,0.5]'$. Four cases are considered: $(d_s, d_e) = (0.35,0.35), (-0.35,0.35), (0.35, -0.35), (-0.35, -0.35)$. Fig. 3
shows that, for cases (a) and (b), \( \hat{k} \) has the lowest rejection frequencies at the two end points (1 and \( T \)). Hence, the least-square estimator will most likely find a change point in the middle of the sample rather than two end points; this confirms the result of Theorem 3.2. Note that when \( d_i + d_e < 0.5 \) (Fig. 3 (b)), the spurious-change problem still occurs. Moreover, we also consider the cases that \( d_e < 0 \). In contrast with Fig. 3 (a) and (b), (c) and (d) show that the empirical distributions of \( \hat{k} \) now have more probability mass at the two end points. Although we do not give a formal proof, it seems that the spurious-change problem occurs only for a positive \( d_e \). As the results of Section 2, the fractionally differencing parameter of disturbances, \( d_e \), is also crucial for a spurious change.

4. Conclusions

In this paper we study the asymptotic properties of the least-squares estimator of the change point for regressions with \( I(d) \) regressors and disturbances. The consistency of the least-squares change-point estimator is established for stationary \( I(d) \) regressions. It is also found that a spurious change may arise when the disturbances have long memory; whether the regressors have long memory is irrelevant here. Thus, one should be extremely careful in drawing inferences from existing structural-change tests and change-point estimators when data have long memory. A procedure that can properly test and estimate the unknown structural change in a long-memory environment is currently being investigated.
Acknowledgements

I am deeply indebted to C.-M. Kuan for enormous advice and comments.

Appendix A

Proof of Theorem 2.1. If \( d_x \neq 0 \), \( d_z \neq 0 \), and in the light of Tsay and Chung (2000), we know

\[
\sum_{t=1}^{[T_M]} x_t e_t = \begin{cases} 
O_p(T^{d_x+d_z}), & d_x + d_z > 0.5, \\
O_p(T^{0.5} \sqrt{\log T}), & d_x + d_z = 0.5, \\
O_p(T^{0.5}), & \text{otherwise.}
\end{cases}
\]

This result, the FCLT (2), and the Borel–Cantelli lemma imply that

\[
\frac{1}{T} \sum_{t=1}^{[T_M]} x_t e_t \overset{p}{\rightarrow} 0,
\]

and

\[
\frac{1}{T} \sum_{t=\lfloor T_M \rfloor+1}^T x_t e_t \overset{p}{\rightarrow} 0.
\]

Furthermore, by a suitable law of large numbers, it can be seen that

\[
\frac{1}{T} \sum_{t=1}^{[T_M]} x_t x_t' \mathbf{1}_{\{t > \tau_0\}} \overset{p}{\rightarrow} \begin{cases} 
(1 - \tau_0) \Omega_x \lambda, & \tau \leq \tau_0, \\
(1 - \tau) \Omega_x \lambda, & \tau > \tau_0,
\end{cases}
\]

and

\[
\frac{1}{T} \sum_{t=\lfloor T_M \rfloor+1}^T x_t x_t' \mathbf{1}_{\{t > \tau_0\}} \overset{p}{\rightarrow} \begin{cases} 
(1 - \tau_0) \Omega_x \lambda, & \tau \leq \tau_0, \\
(1 - \tau) \Omega_x \lambda, & \tau > \tau_0,
\end{cases}
\]

uniformly in \( \tau \in [\tilde{\tau}, \bar{\tau}] \), where \( \Omega_x = \text{diag}[1, \gamma_x(0)] \) and \( \gamma_x(0) \) is the variance of \( x_t \). Combining these results, we have

\[
\frac{1}{T} \tilde{H}_p([T_M]) \overset{p}{\rightarrow} \begin{cases} 
\frac{(1 - \tau_0)^2}{1 - \tau} (\lambda_1^2 + \gamma_x(0) \lambda_2^2), & \tau \leq \tau_0, \\
(1 - 2\tau_0 + \frac{\tau_0^2}{\tau}) (\lambda_1^2 + \gamma_x(0) \lambda_2^2), & \tau > \tau_0,
\end{cases}
\]

where the limit reaches the global maximum \((1 - \tau_0)(\lambda_1^2 + \gamma_x(0) \lambda_2^2)\) at \( \tau = \tau_0 \). \( \square \)

Proof of Theorem 3.1. Because of the FCLT (2) and (6), we have

\[
T^{-(0.5+d_x)} E_x(0, \tau) \overset{p}{\rightarrow} \left[ \frac{1}{T^{0.5+d_x}} \sum_{t=1}^{[T_M]} e_t \right] \otimes [\kappa(B_{d_x}(\tau), o_{p}(1))^2],
\]

for any \( \tau \in [\tilde{\tau}, \bar{\tau}] \). Similarly, \( T^{-(0.5+d_x)} E_x(\tau, 1) \overset{p}{\rightarrow} \left[ \kappa(B_{d_x}(1) - B_{d_x}(\tau), o_{p}(1))^2 \right] \otimes \Omega_x \). Moreover, by the law of large numbers, we know \( M_x(0, \tau) / T \overset{p}{\rightarrow} \tau \Omega_x \) and \( M_x(\tau, 1) / T \overset{p}{\rightarrow} (1 - \tau) \Omega_x \). Thus,
\[ T^{-2d} H_T([T\tau]) = (T^{-(0.5+d_x)} E_T'(0,\tau))(T M_T^{-1}(0,\tau))(T^{-(0.5+d_x)} E_T(0,\tau)) \]
\[ + (T^{-(0.5+d_x)} E_T'(\tau,1))(T M_T^{-1}(\tau,1))(T^{-(0.5+d_x)} E_T(\tau,1)) \overset{D}{\rightarrow} \kappa^2 \frac{B_{d_x}^2(\tau)^2}{\tau} \]
\[ + \kappa^2 \left[ B_{d_x}(1) - B_{d_x}(\tau) \right]^2 \left( \frac{1}{1 - \tau} \right) , \]

for any \( \tau \in [\tau, \overline{\tau}] \). Similarly, \( T^{-2d} H_T \overset{D}{\rightarrow} \kappa^2 B_{d_x}(1)^2 \). Under the null hypothesis, it can then be shown that \( \left[ \Sigma_{t=1}^{T} e_t^2 + H_T([T\tau]) \right]/T \overset{D}{\rightarrow} \chi_1(0) \). Combining these limits and using the continuous mapping theorem, we obtain the result of (5). \( \square \)

References


