Econometric Theory, **14**, 1998, 663–669. Printed in the United States of America.

A NOTE ON SPURIOUS BREAK

JUSHAN BAI
Massachusetts Institute of Technology

When the disturbances of a regression model follow an $I(1)$ process there is a tendency to estimate a break point in the middle of the sample, even though a break point does not actually exist. In this note, we provide a mathematical proof for this phenomenon.

1. INTRODUCTION

Recently, Nunes, Kuan, and Newbold (1995) (henceforth NKN) pointed out that when the disturbances of a regression model follow an $I(1)$ process there is a tendency to estimate a break point in the middle of the sample, even though a break point does not actually exist. This phenomenon is called a "spurious break" by the authors and was discovered by a simulation experiment. In this note, we provide a mathematical proof for this phenomenon.

It is of interest to ask the following question, and it is, in fact, often asked. Given a regression model with no break point, and supposing that a break point is entertained in estimation, how does the estimated break point behave? Let $\hat{k}$ denote the estimated break point and T the sample size. Define $\hat{\lambda}_T = \hat{k}/T$. Thus, $\hat{\lambda}_T$ denotes the estimated break fraction. For $I(0)$ disturbances, it can be shown that the estimated break fraction converges to the boundary (i.e., either 0 or 1). Because boundary values imply no break in the sample, the estimated break point conforms to the true model of no break. However, a different phenomenon emerges for $I(1)$ disturbances. NKN found that the estimated break point will stay in the middle of the sample, suggesting the existence of a break point. The authors call this phenomenon spurious break in an analogy to spurious regression. Although the problem of spurious regression is much better understood and is well documented (see Granger and Newbold, 1974; Phillips, 1986; Durlauf and Phillips, 1988), the problem of spurious break is less well studied. This note takes up the issue. In particular, we shall deliver a mathematical proof that $\hat{\lambda}_T$ does not converge to 0 or 1, corroborating and confirming the simulation findings of NKN. It should be emphasized that our analysis assumes the absence of a break in the data-generating process.

The result suggests that caution should be exercised in estimating a break point when disturbances are $I(1)$. Diagnostic testing should be performed prior to es-

I thank three anonymous referees and Bruce Hansen (co-editor) for their comments and suggestions that led to this much improved version. Partial financial support is acknowledged from the National Science Foundation under grant SBR-9414083. Address correspondence to: Jushan Bai, Department of Economics, E52-274B, Massachusetts Institute of Technology, Cambridge, MA 02139, USA; e-mail: jbai@mit.edu.

 0266-4666/98 $9.50

timating a break point, and test statistics that are robust to $I(1)$ errors should be used. Vogelsang (1994a, 1994b) proposed a number of test statistics useful for this purpose.

2. NOTATION AND ASSUMPTION

Consider the model

$$y_t = \begin{cases} x_t'\beta_1 + \varepsilon_t & t = 1,2,\ldots,k \\ x_t'\beta_2 + \varepsilon_t & t = k+1,\ldots,T. \end{cases}$$

Let $\hat{\beta}_1(k)$ be the least squares estimator of β_1 based on the first k observations and $\hat{\beta}_2(k)$ be the least squares estimator of β_2 based on the last $T-k$ observations, i.e.,

$$\hat{\beta}_1(k) = \left(\sum_{t=1}^{k} x_t x_t'\right)^{-1}\left(\sum_{t=1}^{k} x_t y_t\right),$$
$$\hat{\beta}_2(k) = \left(\sum_{t=k+1}^{T} x_t x_t'\right)^{-1}\left(\sum_{t=k+1}^{T} x_t y_t\right).$$

Define the sum of squared residuals for the full sample as

$$S_T(k) = \sum_{t=1}^{k}(y_t - x_t'\hat{\beta}_1(k))^2 + \sum_{t=k+1}^{T}(y_t - x_t'\hat{\beta}_2(k))^2$$

and define the break point estimator as

$$\hat{k} = \underset{1\le k\le T}{\operatorname{argmin}}\, S_T(k) = \underset{1\le k\le T}{\operatorname{argmin}}\left\{S_T(k) - \sum_{t=1}^{T}\varepsilon_t^2\right\}.$$

The second equality follows because $\sum_{t=1}^{T}\varepsilon_t^2$ does not depend on k. Finally, let

$$\hat{\lambda}_T = \min\left\{(\lambda : \lambda = \underset{u\in[\underline{\lambda},\bar{\lambda}]}{\operatorname{argmin}}\left[S_T([Tu]) - \sum_{t=1}^{T}\varepsilon_t^2\right]\right\},$$

where $0 < \underline{\lambda} < \bar{\lambda} < 1$. The behavior of λ_T can be determined by exploring the limiting process of $S_T(k) - \sum_{t=1}^{T}\varepsilon_t^2$. In the absence of a break, i.e., $\beta_1 = \beta_2$, it is easy to show that

$$S_T(k) - \sum_{t=1}^{T}\varepsilon_t^2 = M_T(k),$$

where

$$M_T(k) = \left(\sum_{t=1}^{k}\varepsilon_t x_t\right)'\left(\sum_{t=1}^{k} x_t x_t'\right)^{-1}\left(\sum_{t=1}^{k} x_t\varepsilon_t\right) + \left(\sum_{t=k+1}^{T}\varepsilon_t x_t\right)'\left(\sum_{t=k+1}^{T} x_t x_t'\right)^{-1}\left(\sum_{t=k+1}^{T} x_t\varepsilon_t\right). \quad \textbf{(1)}$$

To obtain the limiting process for $M_T(k)$, we need the following assumptions, which are similar to those of NKN.

Assumption A1. For an $I(1)$ error process ε_t, we assume that $T^{-2}\sum_{t=1}^{[T\lambda]}\varepsilon_t^2 \Rightarrow c\int_0^\lambda W^2(u)\,du$ with $W(u)$ being a standard Wiener process and $c > 0$ a constant.

Assumption A2. There exists a diagonal matrix D_T such that

$$D_T^{-1/2}\left(\sum_{t=1}^{[T\lambda]} x_t x_t'\right)D_T^{-1/2} \Rightarrow Q(\lambda).$$

The matrix $Q(\lambda)$ is assumed to be positive definite for all $\lambda > 0$, and $Q(v) - Q(u)$ is positive definite for all $v - u > 0$. In addition, $Q(0) = 0$.

Assumption A3. For some $\alpha \geq 0$,

$$T^{-\alpha/2}D_T^{-1/2}\sum_{t=1}^{[T\lambda]} x_t\varepsilon_t \Rightarrow G(\lambda), \tag{2}$$

where $G(\lambda)$ is a stochastic process having continuous sample path with $G(0) = 0$.[1]

The assumptions here are quite general, encompassing many models used in practice. Various special cases of $G(\lambda)$ and $Q(\lambda)$ are given in NKN. Typically, $G(\lambda)$ is a functional of a Gaussian process.

3. MAIN RESULT

In this section, we characterize the limiting behavior of $M_T(k)$. When ε_t is $I(0)$, $M_T(k)$ has a proper limit for $k \in [aT, bT]$ with $a, b \in (0,1)$ and $a < b$. This will not be true when ε_t is $I(1)$. A normalization is required. Define

$$M_T^*(k) = T^{-\alpha}M_T(k).$$

We also note that $\hat{k} = \operatorname{argmax}_k M_T(k) = \operatorname{argmax}_k M_T^*(k)$ because $T^{-\alpha}$ does not depend on k. Under Assumptions A1–A3, it can be shown that (see NKN)

$$\hat{\lambda}_T \xrightarrow{d} \operatorname*{argmax}_{\lambda\in[\underline{\lambda},\bar{\lambda}]} M^*(\lambda), \tag{3}$$

where

$$M^*(\lambda) = G(\lambda)'Q(\lambda)^{-1}G(\lambda) + [G(1) - G(\lambda)]'[Q(1) - Q(\lambda)]^{-1}[G(1) - G(\lambda)]. \tag{4}$$

When ε_t is $I(0)$, a similar limiting process will be obtained. This implies that, in the absence of a break point, the estimated break point $\hat{\lambda}_T$ is a random variable with support in $[\underline{\lambda}, \bar{\lambda}]$. This is true regardless of whether the error is $I(0)$ or $I(1)$. Thus if the attention is focused on the compact interval $[\underline{\lambda}, \bar{\lambda}]$, not much difference between the two cases can be discerned. It is important to explore the be-

havior of $M^*(\lambda)$ for λ near the boundary. For $I(0)$ error process ε_t, NKN proved that $M(\lambda) \to \infty$ as $\lambda \to 0$ or 1, thus $\hat{\lambda}_T \to \{0,1\}$, if $\underline{\lambda} \to 0$ and $\bar{\lambda} \to 1$ (also see Andrews, 1993). In their Remark 1 (p. 742), NKN pointed out that they were unable to characterize the limiting behavior of $M^*(\lambda)$ for λ near 0 or 1. Through simulation, they found that $M^*(\lambda)$ does not diverge to infinity as λ decreases to zero or increases to 1.

In the following, we shall prove that $M^*(\lambda)$ is a well-defined process on $[0,1]$ and is uniformly bounded in probability over $[0,1]$.

In Assumption A3, we shall assume that $\alpha \geq 2$. This is true whenever x_t contains a nonzero mean regressor (e.g., a constant or a trend). When all components of x_t are $I(0)$ and have zero means, it is possible that $\alpha = 1$ (for an example, see NKN). (Actually, an $I(1)$ dependent variable with $I(0)$ regressors is unlikely to be a useful model.) In any case, our proof does not apply to the situation for which $0 < \alpha < 2$ and thus is not considered in this paper.

THEOREM 1. *Assume that Assumptions* A1–A3 *hold. For the* α *defined in* (2), *assume* $\alpha \geq 2$. *We have*

$$\sup_{\lambda \in (0,1)} M^*(\lambda) = O_p(1). \tag{5}$$

Proof of Theorem 1. For an arbitrary vector z and an arbitrary projection matrix P, we have $z'Pz \leq z'z$. Apply this inequality to $M_T^*(k)$ to obtain

$$M_T^*(k) \leq T^{-\alpha} \sum_{t=1}^{T} \varepsilon_t^2 \leq T^{-2} \sum_{t=1}^{T} \varepsilon_t^2 \quad \text{for all } k \in [1,T], \tag{6}$$

from $\alpha \geq 2$. Because $T^{-2} \sum_{t=1}^{T} \varepsilon_t^2$ does not depend on k, and it has a limit by Assumption A1, $M_T^*(k)$ is uniformly bounded in probability. Thus its limit, $M^*(\lambda)$, is uniformly bounded in probability for $\lambda \in (0,1)$. ■

Therefore, $M^*(\lambda)$ is stochastically bounded even when $\lambda \to \{0,1\}$. This is in contrast to the case of $I(0)$ errors for which the corresponding process grows without bound as λ tends to the boundary. To rule out the possibility that $\hat{\lambda}_T \to \{0,1\}$, we need to further examine the behavior of $M^*(\lambda)$ for λ near 0 and 1. Strictly speaking, $M^*(\lambda)$ is not defined yet at $\lambda = 0$ and $\lambda = 1$. As the limit of $M^*(\lambda)$ when $\lambda \to 0$, $M^*(0)$ should be defined as

$$M^*(0) = G(1)'Q(1)^{-1}G(1), \tag{7}$$

which is obtained from (4) by taking $G(\lambda) = 0$, $Q(\lambda) = 0$, and $G(\lambda)'Q(\lambda)^{-1}G(\lambda) = 0$ for $\lambda = 0$. Note that the term $G(\lambda)'Q(\lambda)^{-1}G(\lambda)$ is the limit of the first term of (1) on the right-hand side divided by $T^{-\alpha}$. Now

$$T^{-\alpha} \left(\sum_{t=1}^{k} x_t \varepsilon_t \right)' \left(\sum_{t=1}^{k} x_t x_t \right)^{-1} \left(\sum_{t=1}^{k} x_t \varepsilon_t \right) \leq T^{-\alpha} \sum_{t=1}^{k} \varepsilon_t^2 \leq T^{-2} \sum_{t=1}^{k} \varepsilon_t^2,$$

which converges to zero in probability for any given k or for $k = [T\lambda]$ with $\lambda \to 0$. It follows that $G(\lambda)'Q(\lambda)^{-1}G(\lambda) \to 0$ in probability as $\lambda \to 0$. Thus the definition of (7) is the limit of $M^*(\lambda)$ as $\lambda \to 0$. As a result, $M^*(\lambda)$ becomes continuous at $\lambda = 0$. Similarly, we can define, as the limit of $M^*(\lambda)$ as $\lambda \to 1$, $M^*(1) = M^*(0)$. This extension of $M(\lambda)$ makes it continuous at $\lambda = 1$. We next show that the maximum of $M^*(\lambda)$ is attained neither at 0 nor 1.

THEOREM 2. *Under the conditions of Theorem* 1, *we have*

(i) *With probability* 1,

$$M^*(0) = M^*(1) \le M^*(\lambda), \quad \textit{for every } 0 < \lambda < 1. \tag{8}$$

(ii) *If $G(\lambda)$ has a continuous distribution for each λ, then with probability* 1,

$$M^*(0) = M^*(1) < M^*(\lambda), \quad \textit{for every } 0 < \lambda < 1. \tag{9}$$

Theorem 2(i) implies that as long as $M^*(\lambda)$ is not a constant process, the maximum value of $M^*(\lambda)$ will not be attained at 0 or 1. Let $\lambda^* = \operatorname{argmax}_{\lambda \in [\underline{\lambda}, \bar{\lambda}]} M^*(\lambda)$. Assume that there exists a λ_1 such that $M^*(\lambda_1) > M^*(0)$ (this is true if the process is not a constant). Then $\lambda^* \not\to 0,1$ as $\underline{\lambda} \to 0$ or $\bar{\lambda} \to 1$. This assertion follows from the continuity of $M^*(\lambda)$ at 0 and 1 (there exists a neighborhood N_0 of $\lambda = 0$ such that $M^*(\lambda) < M^*(\lambda_1)$ for all $\lambda \in N_0$; the same is true for $\lambda = 1$). In summary, when $M^*(\cdot)$ is not a constant process, not only does $M^*(\cdot)$ not attain its maximum at 0 or 1, but also the extreme point of $M^*(\cdot)$ on any subset of $[0,1]$ is bounded away from 0 or 1.

From (3), $\hat{\lambda}_T \to \lambda^*$. This implies that the estimated break point will also be bounded away from 0 and 1, provided that $M^*(\lambda)$ is not a constant process. Of course, if $M^*(\cdot)$ is a constant process, any point of $[0,1]$ is an extreme point. It is difficult to construct an example (or model) such that $M^*(\cdot)$ does not depend on λ. For the various concrete examples given in NKN, $M^*(\cdot)$ is not a constant. Theorem 2(ii) gives a sufficient condition to guarantee the nonconstantness of $M^*(\cdot)$. The result of part (ii) is much stronger than needed for the occurrence of spurious breaks.

To prove Theorem 2, we need the following lemma. For a symmetric matrix A, we write $A > 0$ if it is positive definite.

LEMMA 1. *For arbitrary positive definite matrices A and B with $A > B$ $(p \times p)$, and arbitrary vectors x and y $(p \times 1)$, we have*

$$x'A^{-1}x - y'B^{-1}y - (x-y)'(A-B)^{-1}(x-y) \le 0. \tag{10}$$

Proof of Lemma 1. Define the matrix

$$H = \begin{pmatrix} (A-B)^{-1} - A^{-1} & -(A-B)^{-1} \\ -(A-B)^{-1} & (A-B)^{-1} + B^{-1} \end{pmatrix}. \tag{11}$$

It suffices to prove H to be positive semidefinite because the left-hand side of (10) is equal to $-z'Hz$ for $z' = (x', y')$. Let $D = (A - B)^{-1} + B^{-1} > 0$. Let C be a matrix with the first p rows $(I,(A - B)^{-1}D^{-1})$ and second p rows $(0, I)$. Using the identity

$$(A - B)^{-1} - A^{-1} = (A - B)^{-1}D^{-1}(A - B)^{-1}$$

we obtain

$$C'HC = \operatorname{diag}(0, D) \geq 0.$$

Thus $C'HC$ is positive semidefinite. This implies that H is positive semidefinite because C has full rank. This proves the lemma. ■

Proof of Theorem 2. The inequality $M^*(0) \leq M^*(\lambda)$ is equivalent to

$$G(1)'Q(1)^{-1}G(1) - G(\lambda)'Q(\lambda)^{-1}G(\lambda) - [G(1) - G(\lambda)]'[Q(1) - Q(\lambda)]^{-1}[G(1) - G(\lambda)] \leq 0.$$

Clearly, part (i) of Theorem 2 follows from Lemma 1 by letting $A = Q(1), B = Q(\lambda)$, $x = G(1)$, and $y = G(\lambda)$. Next, consider (ii). Let $A = Q(1)$ and $B = Q(\lambda)$ and let H be defined in (11). Then $M^*(0) < M^*(\lambda)$ is equivalent to $-\xi'H\xi < 0$, where $\xi = (G(1)', G(\lambda)')'$. Let Γ be an orthogonal matrix such that $\Gamma'H\Gamma = \operatorname{diag}(\lambda_1, \ldots, \lambda_{2p})$ with $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_{2p}$, where λ_i's are the eigenvalues of H. Because $H \geq 0$ and $H \neq 0$, the maximum eigenvalue of H is positive. It follows that

$$-\xi'H\xi = -(\Gamma\xi)'\operatorname{diag}(\lambda_1, \ldots, \lambda_{2p})\Gamma\xi \leq -\eta^2\lambda_1,$$

where η is the first component of $\Gamma\xi$. When $G(\lambda)$ has a continuous distribution, so does ξ. Thus $\Gamma\xi$ is a vector of continuous random variables, implying $-\eta^2\lambda_1 < 0$ with probability 1 because $P(\eta^2 = 0) = 0$. That is, $-\xi'H\xi < 0$ with probability 1. ■

The preceding analysis applies to $I(1)$ regressors with $I(1)$ disturbances. In this case, $Q(\lambda)$ in Assumption A2 is a random positive definite matrix; all the preceding argument applies. The details can be found in Bai (1996). This implies that for spurious regression models, if a break is allowed in estimation, a spurious break will occur.

All these results lend support to the observation that when it comes to $I(1)$ processes, one should be careful about making the hypothesis of a break point. It is well known that a process with a break point may be mistaken as $I(1)$ (Perron 1989); the converse is also true. Our result is simply a rigorous proof of this fact. However, when y_t and x_t are $I(1)$ but are cointegrated, a spurious break will not arise because the underlying disturbances are $I(0)$. Furthermore, should there indeed exist a shift in the cointegrating relationship, the break point can be estimated more precisely than $I(0)$ models (given the same magnitude of shift). Estimating a break point in cointegrating relationship was studied by Bai, Lumsdaine, and Stock (1997) and by an earlier version of this paper.

NOTE

1. The corresponding assumption in NKN (see their (A3′)) is stated in terms of y_t rather than ε_t, which applies to y_t being $I(1)$. The current form allows y_t to depend on deterministic regressors and on an additive $I(1)$ error process.

REFERENCES

Andrews, D.W.K. (1993) Tests for parameter instability and structural change with unknown change point. *Econometrica* 61, 821–856.

Bai, J. (1996) A Note on Spurious Break and Regime Shift in Cointegrating Relationship. Working paper 96-13, Department of Economics, MIT.

Bai, J., R.L. Lumsdaine, & J.H. Stock (1997) Testing for and Dating Common Breaks in Multiple Time Series. Manuscript, Kennedy School of Government, Harvard University (forthcoming in *Review of Economic Studies*).

Durlauf, S.N. & P.C.B. Phillips (1988) Trends versus random walks in time series analysis. *Econometrica* 56, 1333–1354.

Granger, C.W.J. & P. Newbold (1974) Spurious regression in econometrics. *Journal of Econometrics* 2, 111–120.

Nunes, L.C., C.M. Kuan, & P. Newbold (1995) Spurious break. *Econometric Theory* 11, 736–749.

Perron, P. (1989) The great crash, the oil price shock and the unit root hypothesis. *Econometrica* 57, 1361–1401.

Phillips, P.C.B. (1986) Understanding spurious regression in econometrics, *Journal of Econometrics* 33, 311–340.

Vogelsang, T.J. (1994a) Wald-type tests for detecting shifts in the trend function of a dynamic time series. CAE Working paper 94-12, Cornell University.

Vogelsang, T.J. (1994b) Testing for a shift in mean without having to estimate serial correlation parameters. Manuscript, Department of Economics, Cornell University.