ESTIMATION OF A CHANGE POINT IN MULTIPLE REGRESSION MODELS

Jushan Bai*

Abstract—This paper studies the least squares estimation of a change point in multiple regressions Consistency, rate of convergence, and asymptotic distributions are obtained The model allows for lagged dependent variables and trending regressors. The error process can be dependent and heteroskedastic For nonstationary regressors or disturbances, the asymptotic distribution is shown to be skewed The analytical density function and the cumulative distribution function for the general skewed distribution are derived The analysis applies to both pure and partial changes. The method is used to analyze the response of market interest rates to discount rate changes

I. Introduction

PARAMETER instability for economic models is a common phenomenon. This is particularly true for time-series data covering an extended period, as it is more likely for the underlying data-generating mechanism to be disturbed over a longer horizon by various factors such as policy-regime shift. For example, for the empirical problem examined in this paper, we find that the response pattern of interest rates to the changes in discount rates varies over time. The timing of variation is consistent with the timing of changes in the Fed's operating procedures. It is well known that failure to take into account parameter changes, given their presence, may lead to incorrect policy implications and predictions. On the other hand, proper treatment of parameter changes can be useful in uncovering the underlying factors that fostered the changes, in identifying misspecification, and in analyzing the effect of a policy change. The purpose of this paper is to study the parameter-change problem in multiple regressions with an unknown change point. The main concern is the estimation of the change point as well as the statistical theory of the change-point estimator. We derive some useful results under fairly general conditions.

For independent and identically distributed observations up to a parameter shift, there is a well-developed theory under maximum-likelihood estimation (see Hinkley (1970), Bhattacharya (1987), and Yao (1987)). For a mean shift in linear processes estimated with the least-squares method, the theory is also worked out (see Bai (1994) and Antoch et al. (1996)). For multiple regression models, the change-point problem is also studied widely. Yet most of the existing work focuses on testing rather than estimation (e.g., Kim and Siegmund (1989) and Gombay and Horváth (1994)). A number of authors investigated the consistency property for the estimated change point, but the rate of convergence is not obtained. These include Krishnaiah and Miao (1988), Horváth (1994), and Nunes et al. and Newbold (1995), among others. Consistency is not enough for deriving the asymptotic distribution. The rate of convergence is needed for this purpose.

In this paper we develop the asymptotic theory for the estimated change point in multiple regressions. These include consistency, rate of convergence, and asymptotic distribution. While we improve the existing consistency result by deriving the rate of convergence, the conditions required are much weaker than those in the existing literature. Our model allows for lagged dependent variables and trending regressors. Furthermore, the disturbances can be heterogeneous and dependent over time. In particular, we assume the disturbances form a sequence of mixingale, which include strong mixing, linear processes, and other dependent structures as special cases, as shown by Hall and Heyde (1980) and Andrews (1988). In addition, the rate of convergence is obtained for both fixed magnitude of shift and shrinking magnitude of shift. As for the asymptotic distribution, we show that it is skewed except for stationary or asymptotically stationary regressors and disturbances. We also derive the analytical density function and cumulative distribution function for the skewed distribution. These are not studied by the existing literature.

Furthermore, the problem is considered in the context of partial change, in which some of the regression parameters hold constant throughout the sample. These parameters are estimated using the entire sample in order to gain efficiency. When all parameters are allowed to change, a pure structural change model is obtained. Therefore a partial structural change amounts to imposing parameter restrictions across regimes. Advantages of a partial change model include more efficient estimation of the regression parameters and better conservation of degrees of freedom. The latter is important for a regression problem with a limited number of observations but many regressors, which is typical for many economic applications.

As an application, we study the relationship between changes in market interest rates and changes in discount rate. This relationship is found to be unstable over time. It appears that changes in the Fed's operating procedures during the period of October 1979–October 1982 alter the response pattern. Through this application, we describe in detail the procedure for implementing the theoretical result. In particular, we discuss how to tackle the problem of multiple breaks for empirical problems in a sequential way, and how to construct confidence intervals.

The rest of the paper is organized as follows. Section II specifies the model and the underlying assumptions. Consistency, rate of convergence, and asymptotic distribution are derived in this section. Construction of confidence intervals is considered. Section III studies the response pattern of market interest rates to changes in the discount rate.

Received for publication July 25, 1995 Revision accepted for publication July 23, 1996
* Massachusetts Institute of Technology
An earlier version of this paper was titled "Estimation of Structural Change Based on Wald Type Statistics." I thank two anonymous referees and an editor for their constructive criticism, which led to a substantial improvement in the presentation This work is supported in part by grants from the Provost Fund and the Dean's Faculty Development Fund of MIT

© 1997 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology

Copyright © 2001. All Rights Reserved.

Procedures for implementing the theoretical results are discussed. Concluding remarks are provided in section IV. Technical materials are collected in the appendices.

II. Theoretical Result

A. Model and Assumptions

Consider the following linear regression with a change point at k_0:

$$\begin{aligned} y_t &= w_t'\alpha + z_t'\delta_1 + \epsilon_t, \qquad t = 1, 2, \ldots, k_0 \\ y_t &= w_t'\alpha + z_t'\delta_2 + \epsilon_t, \qquad t = k_0 + 1, \ldots, T \end{aligned} \tag{1}$$

where y_t is an observation on the dependent variable, w_t and z_t are vectors of regressors, and ϵ_t is an unobservable disturbance. The vectors α, δ_1, and δ_2 are unknown parameters. We assume $\delta_1 \neq \delta_2$, so that a change has taken place. The change point k_0 is unknown. Our purpose is to estimate these parameters. The analysis is based on a reparameterization. Let $x_t = (w_t', z_t')'$, $\beta = (\alpha', \delta_1')'$, and $\delta = \delta_2 - \delta_1$. Then equations (1) can be rewritten as

$$\begin{aligned} y_t &= x_t'\beta + \epsilon_t, \qquad t = 1, 2, \ldots, k_0 \\ y_t &= x_t'\beta + z_t'\delta + \epsilon_t, \qquad t = k_0 + 1, \ldots, T. \end{aligned} \tag{2}$$

Now z_t is a subvector of x_t and $\delta \neq 0$. More generally, let $z_t = R'x_t$, where R is a $p \times q$ known matrix with full column rank. This defines z_t as a linear transformation of x_t. For $R = (0, I)'$, with I an identity matrix, a partial change like equations (1) is obtained. For $R = I$, a pure change model is obtained.

To present the model in a matrix format, further notation is necessary. Let $Y = (y_1, \ldots, y_T)'$, $X = (x_1, x_2, \ldots, x_T)'$, $\epsilon = (\epsilon_1, \epsilon_2, \ldots, \epsilon_T)'$, $X_1 = (x_1, x_2, \ldots, x_k, 0, \ldots, 0)'$, $X_2 = (0, \ldots, 0, x_{k+1}, \ldots, x_T)'$, and $X_0 = (0, \ldots, 0, x_{k_0+1}, \ldots, x_T)'$. The matrices X_1 and X_2 depend on k, but this dependence will not be displayed for notational simplicity. Define Z_1, Z_2, and Z_0 in a similar way (i.e., replace x_t by z_t). Then $Z_1 = X_1R$, $Z_2 = X_2R$, and $Z_0 = X_0R$. Equations (2) can then be rewritten as

$$Y = X\beta + Z_0\delta + \epsilon. \tag{3}$$

The least-squares method is used to estimate the model. Let $S_T(k)$ denote the sum of squared residuals when regressing Y on X and Z_2. The change point $\hat{k}$ is defined as

$$\hat{k} = \arg \min_{1 \le k \le T} S_T(k).$$

In practice, the existence of a break may not be known. A hypothesis testing is often performed. The sup-Wald-type test statistic is $\sup_{k \in [\pi T, (1-\pi)T]} W_T(k)$, where

$$W_T(k) = \frac{\hat{\delta}_k'(Z_2'MZ_2)\hat{\delta}_k}{\hat{\sigma}^2(k)} \tag{4}$$

with $M = I - X(X'X)^{-1}X$, $\hat{\sigma}^2(k) = S_T(k)/(T - p - q)$, $\pi \in (0, \frac{1}{2})$ a small number, and $(\hat{\beta}_k, \hat{\delta}_k)$ being the least-squares estimator of (β, δ) by regressing Y on X and Z_2. Let $\bar{S}$ denote the sum of squared residuals by regressing Y on X alone. From the identity $\bar{S} - S(k) = \hat{\delta}_k(Z_2'MZ_2)\hat{\delta}_k$ (see Amemiya (1985, pp. 31–33)), we see that the Wald test statistic is a monotonic transformation of $S_T(k)$. It follows that

$$\begin{aligned} \hat{k} &= \arg \min_k S_T(k) = \arg \max_k \hat{\delta}_k'(Z_2'MZ_2)\hat{\delta}_k \\ &= \arg \max_k W_T(k). \end{aligned} \tag{5}$$

That is, the estimator obtained by minimizing the sum of squared residuals is the same as maximizing Wald-type statistics. These statistics are used in Andrews (1993) and others for testing the existence of a change point. Christiano (1992) uses these statistics for estimating potential changes in U.S. gross national product (GNP). It is useful to note that, for empirical applications, a break point estimator is obtained automatically when the null hypothesis is rejected, combining the testing and estimation in a single step.[1]

The assumptions needed for the theoretical results are stated below. The notation $\|\cdot\|$ denotes the Euclidean norm, that is, $\|x\| = (\sum_{i=1}^p x_i^2)^{1/2}$ for $x \in \mathscr{R}^p$. For a matrix A, $\|A\|$ represents the vector-induced norm.

A1: $k_0 = [\tau T]$, where $\tau \in (0, 1)$ and $[\cdot]$ is the greatest integer function.

A2: The data $\{(y_{tT}, x_{tT}, z_{tT});\ 1 \le t \le T, T \ge 1\}$ form a triangular array. For notational simplicity, the subscript T will be suppressed. In addition, $z_t = R'x_t$, where R is $p \times q$, rank $(R) = q$, $z_t \in \mathscr{R}^q$, $x_t \in \mathscr{R}^p$, $q \le p$.

A3: The matrices $(1/j)\sum_{t=1}^j x_tx_t'$, $(1/j)\sum_{t=T-j+1}^T x_tx_t'$, $(1/j)\sum_{t=k_0-j+1}^{k_0} x_tx_t'$, and $(1/j)\sum_{t=k_0+1}^{k_0+j} x_tx_t'$ have minimum eigenvalues bounded away from zero in probability for all large j. That is, there exists $\lambda > 0$ such that for every $\epsilon > 0$, there exists j_0 such that $P(\lambda_j > \lambda) > 1 - \epsilon$ for all $j > j_0$, where λ_j denotes the minimum eigenvalue for each of the above matrices. For simplicity, we assume these matrices are invertible when $j \ge p$. In addition, these four matrices have stochastically bounded norms uniformly in j. That is, for example, $\sup_{j\ge1} \|(1/j)\sum_{t=1}^j x_tx_t'\|$ is stochastically bounded.

A4: $(X'X/T)$ converges in probability to a nonrandom and positive definite matrix.

[1] While hypothesis testing requires that $k \in [T\pi, T(1 - \pi)]$ for small $\pi > 0$, the estimation theory does not require this restriction, provided that a change point is assumed to exist. In practice, the existence of a break may not be known a priori. If the existence is determined via a test, such a restriction is implicitly imposed for empirical work.

Copyright © 2001. All Rights Reserved.

A5: For random regressors, $\sup_t E\|x_t\|^{4+\delta} \leq K$ for some $\delta > 0$ and $K < \infty$.

A6: The disturbances $\{\epsilon_t\}$ satisfy one of the following two alternatives:

a. $\{\epsilon_t, \mathscr{F}_t\}$ form a sequence of martingale differences for $\mathscr{F}_t = \sigma -$ field $\{\epsilon_s, x_{s+1}, \quad s \leq t\}$. Moreover, for all t, $E|\epsilon_t|^{4+\delta} < K$ for some $K < \infty$ and $\delta > 0$ Or

b. The disturbance ϵ_t is independent of the regressor x_s for all t and s. For an increasing sequence of σ fields $\mathscr{F}_t$, $\{\epsilon_t, \mathscr{F}_t\}$ form a sequence of L^r, $r = 4 + \delta$, mixingales with mixingale numbers $\{\psi_t\}$ and mixingale norms $\{c_t\}$. See Hall and Heyde (1980, p. 21) for the definition of L^2 mixingales and Andrews (1988) for general L^r mixingales. In addition, $\Sigma_j j^{1+\delta}\psi_j < \infty$ for some $\delta > 0$ and $\max_j|c_j| < K < \infty$.

Assumption A1 assumes that the shift point is bounded away from the end points, which is used for asymptotic purpose. Assumption A2 allows for trending regressors written in the form $(t/T)^l$, $l > 0$, or, more generally, written as any function of the time trend $g(t/T)$. Expressing them in this format avoids a scaling matrix that would otherwise be required when deriving limiting distributions. This format is also required by assumptions A3 and A4 because no scaling matrix is used in these two assumptions. A3 requires that there be enough data around the change point and at the beginning and at the end of the sample, so that the change point can be identified. The latter half of A3 is typically implied by the strong law of large numbers for the sequence of $\{x_tx_t'\}$. Assumptions A4 and A5 are standard for linear regressions, and they are used for the central limit theorem. Assumption A6a allows for lagged dependent regressors (e.g., autoregressive models); A6b allows for general serial correlation in the disturbances. A6 together with A4 and A5 imply the strong law of large numbers for $x_t\epsilon_t$, because conditions in Hansen (1991) are satisfied. These conditions are sufficient to obtain the rate of convergence for the estimated break point. Note that no stationarity for regressors or disturbances is required for consistency.

B. *Consistency*

We shall establish the consistency and the rate of convergence for the change-point estimator. The rate of convergence not only describes how fast the estimator converges to the true value, it is also necessary in order to derive the limiting distribution. As will be explained later, it is useful to consider asymptotic distributions under shrinking magnitude of shift such that δ depends on T with $\delta_T \to 0$. Thus we will need the rate of convergence for shrinking shifts. Apparently the magnitude of shift cannot be too small, otherwise it will be impossible to identify the change point. Assume

A7: $\delta_T \to 0$, and $T^{\frac{1}{2}-\alpha}\delta_T \to \infty$ for some $\alpha \in (0, \frac{1}{2})$.

PROPOSITION 1: Assume the conditions of A1–A6. If δ_T is fixed or $\delta_T \to 0$ but satisfying A7, then $\hat{k} = k_0 + O_p(\|\delta_T\|^{-2})$.

Similar results are obtained for identically and independently distributed (i.i.d.) models up to a shift, as in Yao (1987) and Bhattacharya (1987). Bai (1994) derives this result for a mean shift in linear processes and Picard (1985) for a Gaussian autoregressive model. To the author's knowledge, proposition 1 is the first result for multiple regression models.[2] Furthermore, the conditions required in this paper are far less restrictive than those in the existing literature. This rate of convergence holds for models with lagged dependent variables, so that the autoregressive model is a special case. Our result holds for trending regressors. We also allow for dependent and heterogeneous disturbances. The disturbances can be a sequence of mixingales, thus including linear processes and strong mixing processes as special cases. Under much stronger assumptions, Horváth (1994) proves the consistency of $\hat{\tau} = \hat{k}/T$, but without obtaining the rate of convergence. Proposition 1 implies that $T^{\alpha}(\hat{\tau} - \tau)$ converges to zero in probability for any $\alpha < 1$ for δ_T fixed. In a recent work, Nunes et al. (1995) also prove consistency for the estimated break fraction, but without rate of convergence. A fixed magnitude of shift is considered by these authors.

Let $\hat{\beta} = \hat{\beta}(\hat{k})$ and $\hat{\delta} = \hat{\delta}(\hat{k})$ be the estimators of β and δ corresponding to $\hat{k}$. That is, we replace Z_0 by Z_2 with $k = \hat{k}$ and then estimate model (3). We have

COROLLARY 1: Under the assumptions of proposition 1 together with ϵ_t being uncorrelated and $E\epsilon_t^2 = \sigma^2$ for all t, then

$$\begin{bmatrix} \sqrt{T}(\hat{\beta} - \beta) \\ \sqrt{T}(\hat{\delta} - \delta_T) \end{bmatrix} \xrightarrow{d} N(0, \sigma^2 V^{-1})$$

where

$$V = \operatorname{plim} \frac{1}{T} \begin{bmatrix} \sum_{t=1}^{T} x_t x_t' & \sum_{t=k_0}^{T} x_t z_t' \\ \sum_{t=k_0}^{T} z_t x_t' & \sum_{t=k_0}^{T} z_t z_t' \end{bmatrix}. \tag{6}$$

For serially correlated and heteroskedastic disturbances, the variance–covariance matrix of the limiting distribution is

[2] For the least absolute deviation (LAD) method, Bai (1995a) obtains the same result under stronger assumptions

Copyright © 2001. All Rights Reserved.

given by $V^{-1}UV^{-1}$, with

$$U = \lim \frac{1}{T}\begin{bmatrix} \sum_{i,j\geq 1}^{T} E(x_i x_j' \epsilon_i \epsilon_j) & \sum_{i,j\geq k_0}^{T} E(x_i z_j' \epsilon_i \epsilon_j) \\ \sum_{i,j\geq k_0}^{T} E(z_j x_i' \epsilon_i \epsilon_j) & \sum_{i,j\geq k_0}^{T} E(z_i z_j' \epsilon_i \epsilon_j) \end{bmatrix}. \quad (7)$$

Notice that $\hat{\alpha}$ and $\hat{\beta}$ have the same limiting distributions as if k_0 were known. A similar result is obtained by Bai (1994) for a mean shift in linear processes.

C. Asymptotic Distribution

We consider two frameworks of asymptotic distribution. One is based on the fixed magnitude of a shift, the other on a shrinking magnitude of shift. We first consider the case of a fixed shift. To obtain the limiting distribution, additional assumptions are necessary. Assume

A8: The process $\{z_t, \epsilon_t\}_{t=-\infty}^{\infty}$ is strictly stationary.

We define a stochastic process $W^*(m)$ on the set of integers as follows: $W^*(0) = 0$, $W^*(m) = W_1(m)$ for $m < 0$ and $W^*(m) = W_2(m)$ for $m > 0$, with

$$W_1(m) = -\delta' \sum_{i=m+1}^{0} z_i z_i' \delta + 2\delta' \sum_{i=m+1}^{0} z_i \epsilon_i, \quad m = -1, -2, \ldots \quad (8)$$

$$W_2(m) = -\delta' \sum_{i=1}^{m} z_i z_i' \delta - 2\delta' \sum_{i=1}^{m} z_i \epsilon_i, \quad m = 1, 2, \ldots. \quad (9)$$

In case of independence for (z_t, ϵ_t) over t, the process W^* is a two-sided random walk with (stochastic) drift.

PROPOSITION 2: Under assumptions A1–A6 and A8, and assuming that $(\delta' z_t)^2 \pm (\delta' z_t)\epsilon_t$ has a continuous distribution, then

$$\hat{k} - k_0 \xrightarrow{d} \arg\max_m W^*(m).$$

The above limiting distribution can be extended to nonstationary data. For simplicity, assume $\{z_t, \epsilon_t\}$ are i.i.d. $(z^{(1)}, \epsilon^{(1)})$ for $t \leq k_0$ and $\{z_t, \epsilon_t\}$ are i.i.d. $(z^{(2)}, \epsilon^{(2)})$ for $t > k_0$ in model (2). Then the limiting distribution still holds if one regards (z_t, ϵ_t) in equation (8) as i.i.d. $(z^{(1)}, \epsilon^{(1)})$ and (z_t, ϵ_t) in equation (9) as i.i.d. $(z^{(2)}, \epsilon^{(2)})$.

To derive the probability function of the limiting distribution, one must know δ and the distribution of $\{z_t, \epsilon_t\}$. Hinkley (1970) studied the case for $z_t = 1$ and ϵ_t i.i.d. normal. Analytical solutions for general cases are typically difficult to obtain. An alternative asymptotic theory is to consider small shifts, assuming the magnitude of shifts converges to zero as the sample size grows unbounded. The result is that the limiting distribution is invariant to the underlying distribution of z_t and ϵ_t. The resulting distribution can be used as an approximation even for large shifts. This idea can be found in Bhattacharya (1987), Picard (1985), and Yao (1987) for various special models. This framework is extended here to general regression models. Moreover, we allow nonstationary regressors and disturbances. In this case, the limiting distribution turns out to be nonsymmetric. We derive the analytical density function and the cumulative distribution function (cdf) for this general limiting distribution, so that confidence intervals can be constructed. To this end, we amend assumption A7 to

A7′: $\delta_T = \delta_0 v_T$, where v_T is a positive number such that $v_T \rightarrow 0$ and $T^{(1/2)-\alpha} v_T \rightarrow \infty$ for some $\alpha \in (0, \frac{1}{2})$ and $\delta_0 \neq 0$.

This amendment makes the asymptotic argument easier. Further assume

A9a: (z_t, ϵ_t) is second-order stationary within each regime such that $Ez_t z_t' = Q_1$ and $E\epsilon_t^2 = \sigma_1^2$ for $t \leq k_0$ and $Ez_t z_t' = Q_2$ and $E\epsilon_t^2 = \sigma_2^2$ for $t > k_0$.

A9b: A functional central limit theorem holds for $\{z_t \epsilon_t\}$ within each regime. That is,

$$k_0^{-1/2} \sum_{t=1}^{[rk_0]} z_t \epsilon_t \Rightarrow B_1(r),$$

$$(T - k_0)^{-1/2} \sum_{t=k_0+1}^{k_0+[r(T-k_0)]} z_t \epsilon_t \Rightarrow B_2(r)$$

where $B_i(r)$ is a multivariate Gaussian process on [0, 1] with mean zero and covariance $EB_i(u)B_i(v) = \min\{u, v\}\Omega_i$ for $i = 1, 2$, and $\Omega_1 = \lim E(k_0^{-1/2}\Sigma_{t=1}^{k_0} z_t \epsilon_t)^2$ and $\Omega_2 = \lim E[(T - k_0)^{-1/2}\Sigma_{t=k_0+1}^{T} z_t \epsilon_t]^2$.

Define

$$\xi = \frac{\delta_0' Q_2 \delta_0}{\delta_0' Q_1 \delta_0} \quad \text{and} \quad \phi = \frac{\delta_0' \Omega_2 \delta_0}{\delta_0' \Omega_1 \delta_0}. \quad (10)$$

Let $W_i(s)$, $i = 1, 2$, be two independent standard Wiener processes defined on $[0, \infty)$, starting at the origin when $s = 0$. Let

$$Z(s) = \begin{cases} W_1(-s) - |s|/2, & \text{if } s \leq 0 \\ \sqrt{\phi} W_2(s) - \xi |s|/2, & \text{if } s > 0. \end{cases} \quad (11)$$

PROPOSITION 3: Under assumptions A1–A6, A7′, and A9,

$$\frac{(\delta_T' Q_1 \delta_T)^2}{\delta_T' \Omega_1 \delta_T} (\hat{k} - k_0) \xrightarrow{d} \arg\max_s Z(s). \quad (12)$$

The density function and the cdf of $\arg\max_s Z(s)$ are derived in appendix B. The distribution of $\hat{k}$ is not symmetric when $\xi \neq 1$ or $\phi \neq 1$. When Q_i, Ω_i, and σ_i^2 are the same for

Copyright © 2001. All Rights Reserved.

each regime ($\xi = 1$ and $\phi = 1$), the limiting distribution reduces to (write $Q_1 = Q_2 = Q$ and $\Omega_1 = \Omega_2 = \Omega$)

$$\frac{(\delta_T'Q\delta_T)^2}{\delta_T'\Omega\delta_T}(\hat{k} - k_0) \xrightarrow{d} \arg\max_s \{W(s) - |s|/2\} \tag{13}$$

where $W(s) = W_1(-s)$ for $s \leq 0$ and $W(s) = W_2(s)$ for $s > 0$. Finally, when the errors are uncorrelated, we have $\Omega = \sigma^2 Q$, and the limiting distribution further reduces to

$$\frac{\delta_T'Q\delta_T}{\sigma^2}(\hat{k} - k_0) \xrightarrow{d} \arg\max_s \{W(s) - |s|/2\}. \tag{14}$$

In this case, the distribution is symmetric about the origin. Its density function is studied by Bhattacharya (1987), Picard (1985), and Yao (1987).

Next we consider z_t being a vector of trending regressors such that $z_t = g(t/T) = [g_1(t/T), \ldots, g_q(t/T)]'$. Clearly, assumption A9 will not be satisfied; we need to derive the limiting distribution separately.

PROPOSITION 4: Let $z_t = g(t/T) = [g_1(t/T), \ldots, g_q(t/T)]'$, with $g_i(x)$ having a bounded derivative on [0, 1]. Assume that the functional central limit theorem holds for ϵ_t within each regime. Under assumptions A1–A6,

$$\frac{\delta_T' g(\tau)g(\tau)'\delta_T}{\psi_1^2}(\hat{k} - k_0) \xrightarrow{d} \arg\max_s Z^{\dagger}(s) \tag{15}$$

where

$$Z^{\dagger}(s) = \begin{cases} W_1(-s) - |s|/2, & \text{if } s \leq 0 \\ \sqrt{\psi_2/\psi_1}\,W_2(s) - |s|/2, & \text{if } s > 0. \end{cases} \tag{16}$$

with $\psi_1 = \lim E(k_0^{-1/2}\sum_{t=1}^{k_0}\epsilon_t)^2$ and $\psi_2 = \lim E[(T - k_0)^{-1/2}\sum_{t=k_0+1}^{T}\epsilon_t]^2$.

When ϵ_t is second-order stationary for the whole sample, $\psi_1 = \psi_2 = \psi = \sigma^2 + 2\sum_{j=2}^{\infty}E(\epsilon_1\epsilon_j)$, the limiting distribution reduces to $\delta_T' g(\tau)g(\tau)'\delta_T\psi^{-2}(\hat{k} - k_0) \rightarrow \arg\max\{W(s) - |s|/2\}$. Furthermore, when the ϵ_t are serially uncorrelated, $\psi^2 = \sigma^2 = E\epsilon_t^2$.

D. *Confidence Intervals*

The results derived above allow easy construction of confidence intervals for the change point. All that is needed is to construct consistent estimates for Q_i, Ω_i, σ_i^2, $i = 1, 2$, and δ_T. We discuss various special cases below.

1. *(z_t, ϵ_t) is second-order stationary for the whole sample.* In this case, the limiting distribution is characterized by equation (13) or equation (14). For serially uncorrelated disturbances, we only need estimates for δ_T, Q, and σ^2. An estimate for δ_T is already available; Q and σ^2 are estimated by $T^{-1}\sum_{t=1}^{T} z_t z_t'$ and $T^{-1}\sum_{t=1}^{T}\hat{\epsilon}_t^2$, respectively. Define $\hat{L} = \hat{\delta}_T'\hat{Q}\hat{\delta}_T/\hat{\sigma}^2$. It can be shown that $(\hat{L} - L)(\hat{k} - k_0) \xrightarrow{p} 0$. Thus equation (14) holds when δ_T, Q, and σ^2 are replaced by their estimators. A $100(1 - \alpha)$% confidence interval is given by

$$[\hat{k} - [c/\hat{L}] - 1, \hat{k} + [c/\hat{L}] + 1] \tag{17}$$

where c is the $(1 - \alpha/2)$th quantile of the random variable $\arg\max\{W(s) - |s|/2\}$; $[c/\hat{L}]$ is the integer part of $c/\hat{L}$. The quantile c can be computed from the cdf formula (B.4) in appendix B. For example, for $\alpha = 0.1$, $c = 7.0$ and for $\alpha = 0.5$, $c = 11.0$.

For serially correlated disturbances we need an estimate for Ω, which can be obtained by the method of Newey and West (1987). Define $\hat{L} = (\hat{\delta}_T'\hat{Q}\hat{\delta}_T)^2/(\hat{\delta}_T'\hat{\Omega}\hat{\delta}_T)$. The confidence interval is again given by equation (17) with this new $\hat{L}$.

2. *(z_t, ϵ_t) is stationary within each regime.* In this case, the limiting distribution is described by equation (12). The quantities to be estimated are Q_i, Ω_i, $i = 1, 2$, and δ_T. We estimate Q_1 by $\hat{k}^{-1}\sum_{t=1}^{\hat{k}} z_t z_t'$, and Q_2 by $(T - \hat{k})^{-1}\sum_{t=\hat{k}+1}^{T} z_t z_t'$. The Ω_i are estimated by the Newey and West method, but Ω_1 is computed using the subsample $[1, \hat{k}]$ and Ω_2 using the subsample $[\hat{k} + 1, T]$. For serially uncorrelated ϵ_t, $\Omega_i = \sigma_i^2 Q_i$; σ_1^2 is estimated by $\hat{k}^{-1}\sum_{t=1}^{\hat{k}}\hat{\epsilon}_t^2$ and σ_2^2 by $(T - \hat{k})^{-1}\sum_{t=\hat{k}+1}^{T}\hat{\epsilon}_t^2$. Once these quantities are obtained, we can estimate ξ and ϕ in equation (10) by the plug-in method. Note that δ_0 can be replaced by $\hat{\delta}_T$ in estimating ξ and ϕ. (Multiply both the numerator and the denominator by v_T and use $\delta_T = \delta_0 v_T$.) When ξ and ϕ are known, the quantiles for $\arg\max_z Z(s)$ can be computed from the cdf formulas (B.2) and (B.3). Let c_1 denote the $(\alpha/2)$th quantile, and c_2 the $(1 - \alpha/2)$th quantile. Define $\hat{L} = (\hat{\delta}_T'\hat{Q}_1\hat{\delta}_T)^2/(\hat{\delta}_T'\hat{\Omega}_1\hat{\delta}_T)$. Because $P[c_1 < \hat{L}(\hat{k} - k_0) < c_2] \rightarrow 1 - \alpha$, a $100(1 - \alpha)$% confidence interval is given by

$$[\hat{k} - [c_2/\hat{L}] - 1, \quad \hat{k} - [c_1/\hat{L}] + 1]. \tag{18}$$

3. *Trending regressors.* The limiting distribution is given in proposition 4. First consider the simple case that ϵ_t is stationary. Let $\hat{f}(0)$ denote an estimate of the spectral density of ϵ_t at zero (without 2π). For serially uncorrelated ϵ_t, $\hat{f}(0) = \hat{\sigma}^2 = T^{-1}\sum_{t=1}^{T}\hat{\epsilon}_t^2$. Define $\hat{L} = \hat{\delta}_T' g(\hat{k}/T)g(\hat{k}/T)'\hat{\delta}_T/\hat{f}(0)$. In this case, $\psi_1 = \psi_2$ and $Z^{\dagger}(s) = W(s) - |s|/2$. The quantiles are computed from equation (B.4). The confidence interval is given by equation (17) with the newly defined $\hat{L}$.

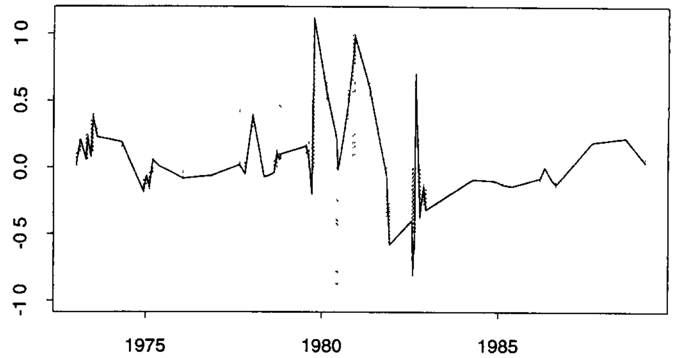
When the ϵ_t are only stationary within each regime, let $\hat{f}_1(0)$ be an estimate of spectral density of ϵ_t for the first regime (using the subsample $[1, \hat{k}]$) and $\hat{f}_2(0)$ for the second regime. We estimate ψ_2/ψ_1 by $\hat{f}_2(0)/\hat{f}_1(0)$. The quantiles of $\arg\max Z^{\dagger}(s)$ are computed from equations (B.2) and (B.3) with $\phi = \psi_2/\psi_1$ and $\xi = 1$. Define $\hat{L} = \hat{\delta}_T g(\hat{k}/T)g(\hat{k}/T)'\hat{\delta}_T/\hat{f}_1(0)$. The confidence interval is then computed using equation (18).

Copyright © 2001. All Rights Reserved.

III. Empirical Result

In this section we analyze the response of market interest rates to the changes in the discount rates. The discount rate is the rate at which the Fed lends and is set by the Fed. It has been noted by many researchers that discount-rate changes of the same size can have different effects on market interest rates (see Dueker (1992) and the references therein). The differentials in the response may reflect different causes that prompted the Fed to change the discount rates or reflect the conditions of the economy. Here we model the response of market interest rates to discount-rate changes by a simple linear model and examine the response pattern over time. As an additional purpose, this application illustrates the procedures for implementing the theoretical results.

FIGURE 1 —CHANGES IN DISCOUNT AND INTEREST RATES (DOTTED VERTICAL LINES—DISCOUNT-RATE CHANGES, SOLID CURVE—INTEREST-RATE CHANGES)



A. *The Data*

Data used in this analysis can be found in Dueker (1992). The yields of three-month T-bills are used as the market interest rates. The range of data spans 1973 to 1989. Over this period, there were 56 changes in the discount rate made by the Fed. An observation is obtained every time there is a change in the discount rate. A change in interest rate is calculated as the closing T-bill rate on the day of a discount-rate change, minus the closing T-bill rate the day before the change. Thus the data are in daily level (frequency). Because discount rates are changed irregularly, the observations are not equally spaced in calendar time. Changes in the discount rate are plotted in figure 1 along with changes in interest rates. The vertical dotted lines represent the size and direction of the discount-rate changes. The solid line represents the changes in interest rate.

B. *The Model*

The following simple linear regression is used to describe the response relationship:

$$\Delta TB_t = \alpha + \beta \Delta DR_t + \epsilon_t \tag{19}$$

where ΔDR_t denotes the change in the discount rate for the ith observation, ΔTB_t denotes the change in the market interest rate, and ϵ_t is a stochastic disturbance. Because the data have daily frequency and two adjacent observations are generally far apart over time, it is reasonable to assume that the ϵ_t are uncorrelated over time. In particular, under the efficient-market hypothesis, ϵ_t will be a sequence of martingale differences. In this case, equation (19) describes the market's reaction to the "news" of the Fed's discount-rate changes. This model is similar in spirit to the one used by Roley and Wheatley (1990) in their study of interest-rate response to the money announcement surprises. The appropriateness of this model is further discussed in the comment section below.

The above simple linear relationship assumes that a change in the discount rate of the same size in absolute value (regardless of an increase or a decrease) has the same effect on market interest rates. It is possible that an increase in the discount rate has a different impact on interest rates from a decrease, if the direction of discount-rate changes carries economic or financial information. Therefore we suggest a modified response model,

$$\Delta TB_t = \alpha + \beta \Delta DR_t^+ + \gamma \Delta DR_t^- + \epsilon_t \tag{20}$$

where $\Delta DR^+ = \max\{0, \Delta DR\}$ and $\Delta DR^- = \Delta DR^+ - \Delta DR$. When this model is fitted to the data, the adjusted R^2 shows a marked improvement over equation (19). Also, the hypothesis of equal coefficients for ΔDR^+ and ΔDR^- is easily rejected. Both models are to be examined. The first will be referred to as the symmetric response model, and the second as the asymmetric response model. We will compare the estimation results. In particular, we examine how the estimated change points are affected by model specification.

C. *Estimation Procedure*

For empirical problems, more than one break may exist, as evidenced by the empirical application here. We discuss a procedure capable of detecting and locating multiple breaks. The procedure is based on hypothesis testing and works as follows. Starting from the whole sample, perform a parameter constancy test. If the test rejects the null hypothesis, then estimate a break point and divide the sample into two subsamples at the estimated break point. Perform a parameter constancy test for each subsample. Estimate an additional break whenever the subsample fails the constancy test. Divide the subsample at the estimated break point (if any) into nested subsamples and do the same analysis. This step is repeated until all the subsamples do not reject the null hypothesis. If the number of break points is known a priori, then no hypothesis testing is necessary. In this case, one just estimates the break points sequentially in the ordering for which the sum of squared residuals is reduced the most, until the specified number of breaks is obtained. This is discussed in Bai and Perron (1994). Analogously, if it is known that the number of breaks is at least m, then no testing is necessary for the first m break points.

Copyright © 2001. All Rights Reserved.

When all of the break points are obtained, a refinement is needed. An estimated change point should be reestimated if it is obtained from a sample or subsample containing more than one break. This reestimation procedure is called refinement. Suppose there are only two breaks and they are identified as $\hat{k}_1$ and $\hat{k}_2$, $\hat{k}_1 < \hat{k}_2$. If $\hat{k}_1$ is identified in the first place (i.e., from the whole sample), then $\hat{k}_1$ should be reestimated using the subsample $[1, \hat{k}_2]$. Similarly, if $\hat{k}_2$ is identified in advance of $\hat{k}_1$, then the second break should be reestimated using the subsample $[\hat{k}_1, T]$. Each refined estimator has the same asymptotic distribution as if the sample had a single break point, so that our asymptotic theory can be applied, as asserted by Bai (1995b). This procedure was first suggested in an earlier version of this paper. A refinement may be performed prior to the complete estimation of all of the breaks.

It is important to note that the Wald statistic is designed for a single break point; it has less power when multiple breaks exist. This is because $\hat{\sigma}^2(k)$ in equation (4) is not a consistent estimator of σ^2 and is biased upward in the presence of multiple breaks. Although for large samples (or for a large magnitude of shift) one will reject the null hypothesis regardless of the consistency[3] of $\hat{\sigma}^2(k)$, the lack of power may not be negligible for small or moderate samples. Fortunately, the problem can be overcome by a simple modification of the test statistic.[4] In constructing the test statistic (4), the residual variance in the denominator is estimated by allowing for more than one break. For moderate sample sizes, two or three breaks should be good enough in adjusting the denominator of equation (4). In our empirical application, this adjustment is made for the full sample test. Critical values of the hypothesis testing are obtained assuming 15% precent "trimming" and 10% significance level. The size of a test should be large for small samples, as we choose here. The critical values are 10.01 for $q = 2$ and 12.27 for $q = 3$. Here q denotes the number of coefficients that are allowed to change (see Andrews (1993)).

D. *Test and Estimation Results*

Symmetric Response Model: When the whole sample is tested for parameter constancy, the sup-Wald statistic is 8.05, occurring at October 9, 1979 ($k = 28$). The test does not show instability compared with the critical value of 10.01. However, if we proceed to divide the sample into two subsamples at the estimated break date and test for parameter constancy for each subsample, we will find that the second subsample has a sup-Wald statistic of 15.20, occurring at December 4, 1981 ($k = 38$). This indicates the existence of a break. In addition, if we use the subsample [1, 37] to test parameter constancy, the sup-Wald is 58.15, achieved at October 9, 1979 ($k = 28$), strongly rejecting the null. Moreover, when the subsample [38, 56] is used, we uncover another break point at August 27, 1982 ($k = 42$) with sup-Wald 11.44. Further tests for each resulting subsample do not show instability. Our analysis indeed indicates the existence of multiple breaks. The failure to uncover any break with the full sample proves the lack of power for the sup-Wald test in the face of multiple breaks. As a verification, if we use the adjustment mentioned earlier by estimating the variance using residuals based on two breaks (at observations 28 and 38), the sup-Wald statistic for the whole sample becomes 11.96, thereby rejecting the null.

The final model has three breaks, occurring at the dates October 9, 1979, December 4, 1981, and August 27, 1982. (In terms of observation numbers, they are 28, 38, and 42, respectively.) Because the middle break is estimated with the subsample [28, 56], a refinement is obtained by reestimating the middle break with the subsample [28, 41]. When this is performed, the same break point is obtained, but the sup-Wald becomes 33.00. The other two breaks are refined already. These estimates are summarized in table 1. The first break point marks the beginning of the Fed's October 1979–October 1982 operating procedure changes. The last break is near the end of the Fed's procedure changes. It remains an open question as to what economic activity induces the second break. The estimated regression coefficients are reported in table 2.

Asymmetric Response Model: The adjusted sup-Wald statistic for the full sample is 13.56, rejecting the null hypothesis[5] (the critical value is 12.27 for $q = 3$). The maximum is again achieved on October 9, 1979 ($k = 28$). For the subsample [1, 27], the sup-Wald statistic is 9.23, which is not significant compared with the critical value of 12.26. For the subsample [28, 56], only two of the three coefficients (the intercept and ΔDR^+) are allowed to change, because of the lack of variation in the data. The sup-Wald statistic is 10.001, which is just shy of significance compared with the critical value of 10.01 (note that $q = 2$). The maximum is attained on August 27, 1982 ($k = 42$). This date will be treated as a break point for an additional reason besides the closeness of the test statistic to the critical value. When the subsample [1, 41] is used to refine the first break, the sup-Wald statistic is 37.07, more than triple the value computed for the full sample (see footnote 5). This is a sign of multiple breaks. The refined break point is on September 19, 1979 ($k = 27$).

Further hypothesis testing for each resulting subsample does not reveal additional breaks. In particular, the middle break point ($k = 38$) associated with the symmetric response model does not show up here. The sup-Wald statistic for the subsample [28, 41] is only 6.55, as opposed to 33.00 under the symmetric response model. Thus the middle break point identified in the first model is accounted for by the asymmetric response. In summary, two break points are identified.

[3] Because $\bar{S} - S_T(k)$ converges to infinity as the sample size increases and $\hat{s}^2(k)$ is stochastically bounded, it follows that the Wald test statistic becomes arbitrarily large as the sample size grows

[4] When the null hypothesis can be rejected without this modification, it will imply a stronger evidence of instability

[5] The adjustment assumes two breaks in estimating the variance Without this adjustment, the sup-Wald statistic for the full sample is 10 25

Copyright © 2001. All Rights Reserved.

TABLE 1 —TEST STATISTICS, BREAK POINTS, AND CONFIDENCE INTERVALS SYMMETRIC RESPONSE MODEL

Sample	Data[a]	Sup-Wald	Critical[b] Value	$\hat{k}$	95% Confidence Interval Symmetric[c]	Skewed[d]
[1, 56]	1/15/73–2/24/89	11 96[e]	10 01	28		
[1, 27]	1/15/73–9/19/79	8 59	10 01	—		
[28, 56]	10/9/79–2/24/89	15 20	10 01	38		
[38, 56]	12/4/81–2/24/89	11 44	10 01	42		
[42, 56]	8/27/82–2/24/89	2 21	10.01	—		
			Refinement[f]			
[1, 37]	1/15/73–11/2/81	58 15	10 01	28	[26, 30]	[25, 29]
[28, 41]	10/9/79–8/16/82	33 00	10 01	38	[37, 39]	[37, 39]
[38, 56]	12/4/81–2/24/89	11 44	10 01	42	[39, 45]	[39, 44]

Notes [a] Beginning and ending dates covered by sample
[b] Critical values corresponding to 15% trimming and 10% significance level For $q = 2$, the value is 10 01, and for $q = 3$, the value is 12 27 (See Andrews (1993))
[c] Based on symmetric distribution
[d] Based on skewed distribution
[e] Adjusted value Without adjustment, the value is 8 05 See section IIIC for explanation
[f] See section IIIC for explanation

These two breaks correspond to the October 1979–October 1982 change in the Fed's operating procedures. These results are reported in table 3. The estimated regression coefficients for the final model are provided in table 4. The middle regime exhibits a higher response to discount-rate changes, particularly to discount-rate increases.

E. *Confidence Intervals*

Symmetric Response Model: For this model, three breaks are identified, at the dates of October 9, 1979, December 4, 1981, and August 27, 1982. (The observation numbers are 28, 38, and 42, respectively.) We first consider constructing confidence intervals using the symmetric limiting distribution (derived assuming homogeneous data). Those using the skewed limiting distribution are given later. The first break is estimated with the subsample [1, 37]. From table 2 we see that $\hat{\delta} = (-0.403, -0.171)'$. For $z_t = (1, \Delta DR_t)'$, the estimated scale factor $\hat{L}$ is $\hat{\delta}(\frac{1}{37}\Sigma_1^{37} z_t z_t')\hat{\delta}/s^2 = 7.95$, where s^2 is the estimated error variance. The 97.5% quantile of the limiting distribution is 11.0, obtained from equation (B.4). By equation (17), the upper boundary of the 95% confidence interval will be $28 + [11.0/7.95] + 1 = 30$. By symmetry, the confidence interval is [26, 30]. The second break is estimated with the sample [28, 41] with $\hat{\delta} = (1.028, 0.373)'$. The estimated scale factor is $\hat{L} = \hat{\delta}(\frac{1}{14}\Sigma_{28}^{41} z_t z_t')\hat{\delta}/s^2 = 24.35$.

TABLE 2 —ESTIMATED REGRESSION COEFFICIENTS, DEPENDENT VARIABLE ΔTR SYMMETRIC RESPONSE MODEL

Sample	Date	Intercept	ΔDR_t	$d_t(k)$	$d_t(k)\Delta DR_t$	$\bar{R}^2$
[1, 56]	1/15/73–2/24/89	0 046 (1 23)	0 338 (5 57)			0 35
[1, 37] $k = 28$	1/15/73–11/2/81 10/9/79	0 415 (7 98)	0.340 (6 54)	−0 403 (6 36)	−0 171 (1 84)	0 73
[38, 56] $k = 42$	12/4/81–2/24/89 8/27/82	0 008 (0 11)	0 164 (1 13)	−0 622 (1 62)	−0 197 (0 33)	0 45

Notes t-statistics are provided in parentheses The variable $d_t(k)$ is a dummy variable with $d_t(k) = 1$ for $t < k$ and 0 for $t \geq k$, $d_t(k)\Delta DR_t$ is the interactive term

Thus the upper boundary of the confidence interval is $38 + [11.0/24.35] + 1 = 39$. Again by symmetry, the confidence interval is [37, 39]. Using the same method, the confidence interval for the third break point is found to be [39, 45]. These results are reported in table 1. Confidence intervals in calendar time are provided in table 5. The first two breaks are estimated with more precision, the last one has a larger interval.

In the above construction it is assumed that the two neighboring regimes surrounding each break have homogeneous data. By visual inspection of figure 1 we see that this is not the case. For example, the magnitude of the regressors for the second regime, October 79–August 89, is typically twice the magnitude for the first regime. This suggests that Q_1 is not equal to Q_2. Thus Q_1 and Q_2 should be estimated separately. In addition, the disturbances may not have common variance cross regimes, thus confidence intervals based on a skewed limiting distribution might be more appropriate. For the first break point, we estimate Q_1 by $\frac{1}{27}\Sigma_{t=1}^{27} z_t z_t'$ and Q_2 by $\frac{1}{10}\Sigma_{t=28}^{37} z_t z_t'$. Then ξ is estimated by the ratio of $\hat{\delta}'(\frac{1}{10}\Sigma_{t=28}^{37} z_t z_t')\hat{\delta}'$ and $\hat{\delta}'(\frac{1}{27}\Sigma_1^{27} z_t z_t')\hat{\delta}'$. This gives $\xi = 1.08$. The ratio σ_2^2/σ_1^2 is estimated by s_2^2/s_1^2, yielding 2.556. Thus the estimated ϕ is $(s_2/s_1)^2\xi = 2.771$. Using equations (B.2) and (B.3) for $\xi = 1.085$ and $\phi = 2.771$, the 97.5% quantile is 28.0, the 2.5% quantile is −9.2. The scale factor $L = \delta' Q_1 \delta/\sigma_1^2$ is estimated to be 10.72. From equation (18), the lower boundary of the interval is $28 - [28/10.72] - 1 = 25$. The upper boundary is $28 - [9.2/10.72] + 1 = 29$. Thus an asymmetric 95% confidence interval is [25, 29]. Using the same method, the confidence interval for the second break is [37, 39], and for the third break it is [39, 44]. The skewed confidence intervals are similar to the symmetric ones, except that the first is wider.

Asymmetric Response Model: For this model, two break points are identified, the first is September 19, 1979, and the second is August 27, 1982. (The observation numbers are 27 and 42, respectively.) Assuming homogeneous data, the confidence interval for the first break is found to be [25, 29],

Copyright © 2001. All Rights Reserved.

TABLE 3 —TEST STATISTICS, BREAK POINTS, AND CONFIDENCE INTERVALS ASYMMETRIC RESPONSE MODEL

Sample	Date	Sup-Wald	Critical Value	$\hat{k}$	95% Confidence Interval: Symmetric	95% Confidence Interval: Skewed
[1, 56]	1/15/73–2/24/89	13.56[a]	12 27	28		
[1, 27]	1/15/73–9/19/79	9 23	12 27	—		
[28, 56]	10/9/79–2/24/89	10 00	10 01	42		
[28, 41]	10/9/79–8/16/89	6 55	10 01			
[42, 56]	18/27/82–2/24/89	2 01	10 01	—		
			Refinement			
[1, 41]	1/15/73–8/16/89	37 07	12 27	27	[25, 29]	[22, 28]
[27, 56]	9/19/79–2/24/89	12.30	10 01	42	[38, 46]	[38, 47]

Notes See footnotes to table 1

[a] Adjusted value Without adjustment, the value is 10 25

TABLE 4 —ESTIMATED REGRESSION COEFFICIENTS, DEPENDENT VARIABLE ΔTR ASYMMETRIC RESPONSE MODEL

Sample	Date	Intercept	ΔDR_t^+	ΔDR_t^-	$d_t(k)$	$d_t(k)\Delta DR_t^+$	$\bar{R}^2$
[1, 56]	1/15/73–2/24/89	−0 253	0 847	0 205			0 48
		(2 90)	(5 76)	(1 32)			
[1, 26]	1/15/73–8/17/79	0 024	0 178	−0 201			0 23
		(0 32)	(1 13)	(0 95)			
[27, 56]	9/19/79–2/24/89	−0 617	1 414	1 087	−0 525	0 483	0 69
$k = 42$	8/27/82	(3.36)	(3 13)	(3 27)	(3 40)	(1 41)	

and for the second it is [38, 46]. For nonhomogeneous data, the confidence intervals are found to be [22, 28] and [38, 47], respectively. These intervals are given in table 3. The corresponding intervals in calendar time are given in table 5.

F. Some Comments

1. Changes in the interest rates may be affected by many other variables, such as money growth, unemployment, and industrial production. These macroeconomic variables are relatively smooth (lower frequency data) compared with the daily interest rates. We believe that these variables are reflected more in the level of the interest rates than in the changes. Changes in the daily interest rates may be treated as white noise in the absence of "news" under the efficient market hypothesis. On the day that the discount rate is altered, the change in the interest rate consists of two components: a response to a change in the discount rate, and an idiosyncratic variation, at least in the absence of other "news." This is the rationale for considering models (19) and (20). Nevertheless, the condition of the economy may affect the strength of a response. One might argue, for example, that high unemployment spurs a stronger reaction. If this is the case, then an interactive term of ΔDR_t and the unemployment rate may be added as a separate explanatory variable. In this case, the simple model (19) amounts to an omitted-variable specification. However, one aspect of the structural change method is (or can be viewed as) identifying misspecification. Our result shows that the misspecification is due, perhaps, more to the omission of the Fed's policy as a variable than to the lack of inclusion of other macroeconomic variables.

2. We point out that observation 42 (August 27, 1982) is unusual. The discount rate was cut down by a half-point, whereas the market interest rate rose by 0.7 point. Three cuts were made in the month of August, each by a half-point. This is the only month in which the Fed changed the discount rate so frequently. The market reacted strongly to the first two cuts. Even though the timing of discount-rate changes is generally not easily predicted (see Dueker (1992)), the third cut seems to have been anticipated by the market. The day before the cut, the market interest rate was the lowest of that week.

When a dummy variable is made to account for this observation, the result is not altered much for either the symmetric or the asymmetric response models. The number

TABLE 5 —CONFIDENCE INTERVALS OF BREAK DATES (IN CALENDAR TIME)

Symmetric Response Model: Break Date	Symmetric Response Model: 95% Confidence Interval: Symmetric	Symmetric Response Model: 95% Confidence Interval: Skewed	Asymmetric Response Model: Break Date	Asymmetric Response Model: 95% Confidence Interval: Symmetric	Asymmetric Response Model: 95% Confidence Interval: Skewed
10/9/79	9/19/79–2/15/80	7/20/79–2/15/80	9/19/79	7/20/79–2/15/80	10/16/78–10/9/79
12/4/81	11/2/81–7/20/82	11/2/81–7/20/82			
8/27/82	7/20/82–12/14/82	7/20/82–11/22/82	8/27/82	12/4/81–4/9/84	12/4/81–11/23/84

Copyright © 2001. All Rights Reserved.

of breaks and their locations are the same, except that the last break is identified at observation 44 (November 22, 1982) rather than at 42. This change is not dramatic and is not unexpected, because observation 42 is the boundary of a regime in the absence of this dummy. Tests and estimation results for this case are not reported here.

3. The 1979 and 1982 breaks are identified under both the symmetric and the asymmetric response models, suggesting a stronger evidence in supporting these two breaks. The 1981 break appears only under the symmetric response, illustrating the sensitivity of break point estimation to model specification. The asymmetric response model appears to be a reasonable one. In the absence of breaks, the asymmetric model fits the data better in terms of the adjusted R^2. The regression coefficients of ΔDR^+ and ΔDR^- are significantly different. With breaks, direct comparison of R^2 is not appropriate, as regression equations cover different time periods. An asymmetric response model will be meaningful if the sign of the discount-rate changes provides information about the Fed's monetary policy objectives. At any rate, the hypothesis of asymmetric response deserves further examination in future research.

IV. Conclusion

We have studied the change-point problem in multiple regressions where some or all of the coefficients have a shift occurring at an unknown time. Consistency and the limiting distribution are established for the change-point estimator computed by the least-squares method. The results are obtained under very general conditions. In particular our results hold for a wide range of regressors such as deterministic and stochastic regressors, lagged dependent variables, and time trend. Dependent and heteroskedastic disturbances are allowed. We derived a skewed and parameter-embedded analytical density function and a cumulative distribution function for the asymptotic distribution of the change-point estimator. The result includes the symmetric distribution in the prior literature as a special case and is useful for constructing confidence intervals for the true break. The empirical analysis examined the relationship between the interest-rate responses to the changes in the discount rate. This response pattern is found to be unstable. The instability seems to be caused by the Fed's change in operating procedures during the period of October 1979–October 1982. In addition, we documented the asymmetry in the response pattern. The regression result suggests a larger response to an increase in the discount rate than to a decrease.

REFERENCES

Amemiya, T, *Advanced Econometrics* (Cambridge, MA Harvard University Press, 1985)

Andrews, D W K, "Laws of Large Numbers for Dependent Nonidentically Distributed Random Variables," *Econometric Theory* 4 (1988), 458–467

——— "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica* 61 (1993), 821–856

Antoch, J, M Huskova, and Z Praskova, "Effect of Dependence on Statistics for Determination of Change," *Journal of Statistical Planning and Inference* (1996)

Bai, J., "Least Squares Estimation of a Shift in Linear Processes," *Journal of Time Series Analysis* 15 (1994), 453–472

——— "Least Absolute Deviation Estimation of a Shift," *Econometric Theory* 11 (1995a), 403–436

——— "Estimation of Multiple Breaks One at a Time," Department of Economics, MIT, Working Paper 95-18 (1995b)

Bai, J, and P. Perron, "Testing for and Estimation of Multiple Structural Changes," manuscript, Department of Economics, MIT (1994)

Bhattacharya, P K, "Maximum Likelihood Estimation of a Change-Point in the Distribution of Independent Random Variables, General Multiparameter Case," *Journal of Multivariate Analysis* 23 (1987), 183–208

Bhattacharya, P K., and P J Brockwell, "The Minimum of an Additive Process with Applications to Signal Estimation and Storage Theory," *Z Wahrschein Verw Gebiete* 37 (1976), 51–75

Christiano, L J, "Searching for a Break in GNP," *Journal of Business and Economic Statistics* 10 (1992), 237–250

Dueker, M J, "The Response of Market Interest Rates to Discount Rate Changes," *Federal Reserve Bank of St Louis Review* 74·4 (1992), 78–91

Gombay, E, and L Horváth, "Limit Theorem for Change in Linear Regression," *Journal of Multivariate Analysis* 48 (1994), 43–69

Hájek, J, and A Rényi, "Generalization of an Inequality of Kolmogorov," *Acta Math Acad Sci Hungar* 6 (1955), 281–283

Hall, P, and C. C Heyde, *Martingale Limit Theory and Its Applications* (New York: Academic Press, 1980)

Hansen, B, "Strong Laws for Dependent Heterogeneous Processes," *Econometric Theory* 7 (1991), 213–221

Hinkley, D, "Inference about the Change Point in a Sequence of Random Variables," *Biometrika* 57 (1970), 1–17

Horváth, L, "Detecting Changes in Linear Regressions," Department of Mathematics, University of Utah (1994)

Kim, H J, and D Siegmund, "The Likelihood Ratio Test for a Change Point in Simple Linear Regression," *Biometrika* 76 (1989), 409–423

Krishnaiah, P R, and B Q Miao, "Review about Estimation of Change Points," in P R Krishnaiah and C R Rao, (eds), *Handbook of Statistics*, vol 7 (New York Elsevier, 1988), 375–401

Newey, W, and K West, "A Simple, Positive Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55 (1987), 703–708

Nunes, L, C M Kuan, and P Newbold, "Spurious Break," *Econometrics Theory* 11 (1995), 736–749

Picard, D, "Testing and Estimating Change-Point in Time Series," *Advances in Applied Probability* 17 (1985), 841–867

Roley, V, and S. M Wheatley, "Temporal Variation in the Interest Rate Response to Money Announcements," NBER, no. 3471 (1990)

Yao, Yi-Ching, "Approximating the Distribution of the ML Estimate of the Change-Point in a Sequence of Independent R V's," *Annals of Statistics* 3 (1987), 1321–1328

APPENDIX A

Proof of Proposition 1

Define $V_T(k) = \hat{\delta}_k'(Z_2'MZ_2)\hat{\delta}_k$ By equation (5), $\hat{k} = \arg\max_k V_T(k)$ Notice that $\hat{\delta}_k = (Z_2'MZ_2)^{-1}(Z_2'MZ_0)\delta_T + (Z_2'MZ_2)^{-1} Z_2'M\epsilon$ and $\hat{\delta}_{k_0} = \delta_T + (Z_0'MZ_0)^{-1} Z_0'M\epsilon$ It follows that

$$V_T(k) - V_T(k_0) = \delta_T'\{(Z_0'MZ_2)(Z_2'MZ_2)^{-1}(Z_2'MZ_0) - (Z_0'MZ_0)\}\delta_T + H_T(k)$$

Copyright © 2001. All Rights Reserved.

where

$$H_T(k) = 2\delta_T'(Z_0'MZ_2)(Z_2'MZ_2)^{-1}Z_2'M\epsilon - 2\delta_T'Z_0'M\epsilon \tag{A.1}$$

$$+ \epsilon'MZ_2(Z_2'MZ_2)^{-1}Z_2'M\epsilon - \epsilon'MZ_0(Z_0'MZ_0)^{-1}Z_0'M\epsilon. \tag{A.2}$$

Define for $k \neq k_0$,

$$G_T(k) = \frac{\delta_T'\{(Z_0'MZ_0) - (Z_0'MZ_2)(Z_2'MZ_2)^{-1}(Z_2'MZ_0)\}\delta_T}{|k_0 - k|} \tag{A.3}$$

When $k = k_0$, define $G_T(k_0) = \delta_T'\delta_T$ We have the following identity

$$V_T(k) - V_T(k_0) = -|k_0 - k|G_T(k) + H_T(k), \qquad \text{for all } k \tag{A.4}$$

Let $X_\Delta = X_2 - X_0 = (0, \ldots, 0, x_{k+1}, \ldots, x_{k_0}, 0, \ldots, 0)'$ for $k < k_0$, $X_\Delta = -(X_2 - X_0) = (0, \ldots, 0, x_{k_0+1}, \ldots, x_k, 0, \ldots, 0)'$ for $k > k_0$, and $X_\Delta = 0$ for $k = k_0$. It follows that $X_2 = X_0 + X_\Delta \operatorname{sgn}(k_0 - k)$. Denote $Z_\Delta = X_\Delta R$

LEMMA A 1 The following two inequalities hold·

$$\begin{aligned}&(Z_0'MZ_0) - (Z_0'MZ_2)(Z_2'MZ_2)^{-1}(Z_2'MZ_0)\\ &\quad \geq R'(X_\Delta'X_\Delta)(X_2'Z_2)^{-1}(X_0'X_0)R, \qquad k < k_0\end{aligned} \tag{A.5}$$

$$\begin{aligned}&(Z_0'MZ_0) - (Z_0'MZ_2)(Z_2'MZ_2)^{-1}(Z_2'MZ_0)\\ &\quad \geq R'X_\Delta'X_\Delta(X'X - X_2'X_2)^{-1}(X'X - X_0'X_0)R, \qquad k \geq k_0.\end{aligned} \tag{A.6}$$

For a proof see the previous version of this paper, which is available upon request.

LEMMA A.2 Assume the conditions of A2–A4. There exists $\lambda > 0$ such that for every $\epsilon > 0$, there exists $C < \infty$ such that $\inf_{|k-k_0|>C\|\delta_T\|^{-2}} G_T(k) \geq \lambda\|\delta_T\|^2$ with probability at least $1 - \epsilon$

Proof· Suppose $k \leq k_0$. Let $A(k) = (k_0 - k)^{-1}X_\Delta'X_\Delta(X_2'X_2)^{-1}(X_0'X_0)$. Note that $A(k)$ is symmetric and is positive definite when $X_\Delta'X_\Delta$ is invertible. This is because $A(k)$ can be written as $[1/(k_0 - k)[(X_0'X_0)^{-1} + (X_\Delta'X_\Delta)^{-1}]^{-1}$ Then by lemma A 1, $G_T(k) \geq \delta_T'R'A(k)R\delta_T \geq \lambda_T(k)\|\delta_T\|^2$, where $\lambda_T(k)$ is the minimum eigenvalue of $R'A(k)R$ It is sufficient to argue that, with probability tending to 1, $\lambda_T(k)$ is bounded away from zero as $k_0 - k$ increases For large $k_0 - k$, $X_\Delta'X_\Delta = \sum_{t=k+1}^{k_0} x_t x_t'$ will be positive definite with large probability by A3 Now $\|A(k)^{-1}\| \leq \|[1/(k_0 - k)X_\Delta'X_\Delta]^{-1}\| \, \|(X_2'X_2)(X_0'X_0)^{-1}\|$. But $\|(X_2'X_2)(X_0'X_0)^{-1}\| \leq \|X'X\|\|(X_0'X_0)^{-1}\|$ is bounded by A3 and A4 In addition, the minimum eigenvalue of $[1/(k_0 - k)]X_\Delta'X_\Delta$ is bounded away from zero by A3 with large probability, so $\|[1/(k_0 - k)X_\Delta'X_\Delta]^{-1}\|$ is bounded with large probability for all large $k_0 - k$ Thus $\|A(k)^{-1}\|$ is bounded with large probability for all large $k_0 - k$ This implies that the minimum eigenvalue of $A(k)$ is bounded away from zero for all large $k_0 - k$. This is also true for $R'A(k)R$ because R has full common rank

LEMMA A 3· Under assumptions A4–A6 there exists a $B < \infty$ such that for every $c > 0$ and $m > 0$,

$$P\left(\sup_{T \geq k \geq m} \frac{1}{k}\left\|\sum_{t=1}^{k} z_t\epsilon_t\right\| > c\right) \leq \frac{B}{c^2 m} \tag{A.7}$$

This lemma generalizes the inequality of Hájek and Rényi (1955) for martingale differences to mixingales. A similar result for ϵ_t being a linear process is proved in Bai (1994) A complete proof is available upon request

LEMMA A 4: Under assumptions A1–A6, for every $\epsilon > 0$ and $\eta > 0$, there exists $T_0 > 0$, so that when $T > T_0$, $P(|\hat{k} - k_0| > T\eta) < \epsilon$

For a proof see Bai and Perron (1994) as well as an earlier version of this paper

We are now in the position to prove proposition 1 Because $V_T(\hat{k}) \geq V_T(k_0)$ by definition, it suffices to show that for each $\epsilon > 0$, there exists $C > 0$ such that

$$P\Big(\sup_{|k-k_0|>C\|\delta_T\|^{-2}} V_T(k) \geq V_T(k_0)\Big) < \epsilon \tag{A.8}$$

By lemma A 4 it is sufficient to show that

$$P\Big(\sup_{k\in K(C)} V_T(k) \geq V_T(k_0)\Big) < \epsilon$$

where $K(C) = \{k. \; |k - k_0| > C\|\delta_T\|^{-2}$ and $T\eta \leq k \leq (1 - \eta)T\}$ for a small number $\eta > 0$ By equation (A 4), $V_T(k) \geq V_T(k_0)$ is equivalent to $H_T(k)/|k_0 - k| \geq G_T(k)$ By lemma 2 it is sufficient to prove that

$$P\left(\sup_{k\in K(C)} \left|\frac{H_T(k)}{k_0 - k}\right| > \lambda\|\delta_T\|^2\right) < \epsilon \tag{A.9}$$

Note that $H_T(k)$ is defined in expressions (A 1) and (A 2) Consider expression (A 1) Use $Z_0 = Z_2 - Z_\Delta \operatorname{sgn}(k_0 - k)$ twice, and we can rewrite expression (A 1) as

$$\begin{aligned}&2\delta_T'(Z_0'MZ_2)(Z_2'MZ_2)^{-1}Z_2'M\epsilon - 2\delta_T'Z_0'M\epsilon\\ &\quad = 2\delta_T'Z_2'M\epsilon - 2\delta_T'(Z_\Delta'MZ_2)(Z_2'MZ_2)^{-1}\\ &\qquad \times Z_2'M\epsilon \operatorname{sgn}(k_0 - k) - 2\delta_T'Z_0'M\epsilon\\ &\quad = [2\delta_T'Z_\Delta'M\epsilon - 2\delta_T'(Z_\Delta'MZ_2)(Z_2'MZ_2)^{-1}Z_2'M\epsilon] \operatorname{sgn}(k_0 - k)\end{aligned} \tag{A.10}$$

Now $Z_\Delta'M\epsilon = Z_\Delta\epsilon - Z_\Delta'X(X'X)^{-1}X'\epsilon$ Note that $Z_\Delta'X = |k_0 - k|O_p(1)$ and $(X'X)^{-1}X'\epsilon = T^{-1/2}O_p(1)$. For example, for $k_0 - k > 0$, $Z_\Delta'X = \sum_{t=k+1}^{k_0} z_t x_t' = |k_0 - k|O_p(1)$ Thus $\delta_T'Z_\Delta'M\epsilon = \delta_T'Z_\Delta\epsilon - |k_0 - k|T^{-1/2}\|\delta_T\|O_p(1)$ Similarly, $\|(Z_\Delta'MZ_2)\| = |k_0 - k|O_p(1)$, and $(Z_2'MZ_2)^{-1} Z_2'M\epsilon = T^{-1/2} O_p(1)$ uniformly on $K(C)$ This implies that equation (A.10) is

$$2\delta_T'Z_\Delta'\epsilon \operatorname{sgn}(k_0 - k) + |k_0 - k|T^{-1/2}\|\delta_T\|O_p(1)$$

Furthermore, it is easy to argue that expression (A.2) is $O_p(1)$ uniformly on $K(C)$ In summary,

$$\begin{aligned}H_T(k) &= 2\delta_T'Z_\Delta'\epsilon \operatorname{sgn}(k_0 - k)\\ &\quad + |k_0 - k|T^{-1/2}\|\delta_T\|O_p(1) + O_p(1).\end{aligned} \tag{A.11}$$

Thus,

$$\begin{aligned}\frac{H_T(k)}{|k_0 - k|} &= \delta_T'\frac{2}{|k_0 - k|}Z_\Delta'\epsilon \operatorname{sgn}(k_0 - k)\\ &\quad + T^{-1/2}\|\delta_T\|O_p(1) + \frac{O_p(1)}{|k_0 - k|}\end{aligned} \tag{A.12}$$

We can now prove expression (A.9) using equation (A 12). Consider $k < k_0$. First,

$$\begin{aligned}&P\left(\sup_{k\in K(C)} \left\|\delta_T'\frac{2}{k_0 - k}\sum_{t=k+1}^{k_0} z_t\epsilon_t\right\| > \frac{\lambda\|\delta_T\|^2}{3}\right)\\ &\quad \leq P\left(\sup_{k\leq k_0 - C\|\delta_T\|^{-2}} \left|\frac{1}{k_0 - k}\sum_{t=k+1}^{k_0} z_t\epsilon_t\right| > \frac{\lambda\|\delta_T\|}{6}\right)\end{aligned}$$

By lemma A.3 (applied with reversed data order), the right-hand side above is bounded by

$$B\frac{36}{\lambda^2\|\delta_T\|^2}\frac{1}{C\|\delta_T\|^{-2}} = B\frac{36}{\lambda^2 C} < \frac{\epsilon}{3}$$

for large C Next, $P(T^{-1/2}\|\delta_T\|O_p(1) > \lambda\|\delta_T\|^2/3) = P(O_p(1)/(T^{1/2}\|\delta_T\|) > \lambda/3) < \epsilon/3$ because $(T^{1/2}\|\delta_T\|)^{-1} \to 0$ Finally, for $k \leq k_0 - C\|\delta_T\|^{-2}$, we have $O_p(1)/|k_0 - k| \leq \|\delta_T\|^2 O_p(1)/C$ Thus, $P(\sup_{k\in K(C)} O_p(1)/|k_0 - k| > \lambda\|\delta_T\|^2/3) \leq P(O_p(1)/C > \lambda/3) < \epsilon/3$ for large C

Copyright © 2001. All Rights Reserved.

Summarizing these results yields expression (A.9). The proof of proposition 1 is complete

Proof of Corollary 1. Let $\hat{Z}_0$ denote Z_0 when k_0 is replaced by $\hat{k}$ Then the estimators $\hat{\beta}$ and $\hat{\delta}$ are obtained by regressing Y on X and $\hat{Z}_0$. The true model can be rewritten as $Y = X\beta + \hat{Z}_0\delta + \epsilon^*$ with $\epsilon^* = \epsilon + (Z_0 - \hat{Z}_0)\delta$ Proceeding in the usual way, we have

$$\sqrt{T}\begin{pmatrix}\hat{\beta} - \beta \\ \hat{\delta} - \delta_T\end{pmatrix} = \left[\frac{1}{T}\begin{pmatrix}X'X & X'\hat{Z}_0 \\ \hat{Z}_0'X & \hat{Z}_0'\hat{Z}_0\end{pmatrix}\right]^{-1} \times \frac{1}{\sqrt{T}}\begin{pmatrix}X'\epsilon + X'(Z_0 - \hat{Z}_0)\delta_T \\ \hat{Z}_0'\epsilon + \hat{Z}_0'(Z_0 - \hat{Z}_0)\delta_T\end{pmatrix}.$$

All we have to show is that the limit of the right-hand side is the same as the limit when $\hat{Z}_0$ is replaced by Z_0. Let us show plim $(1/\sqrt{T})X'(Z_0 - \hat{Z}_0)$ $\delta = 0$. Consider $\hat{k} \le k_0$. Then

$$\left\|\frac{1}{\sqrt{T}}X'(Z_0 - \hat{Z}_0)\delta_T\right\| \le \frac{1}{\sqrt{T}\|\delta_T\|}\left\|\sum_{t=\hat{k}+1}^{k_0} x_t z_t'\right\|\|\delta_T\|^2 = \frac{1}{\sqrt{T}\|\delta_T\|}O_p(1) = o_p(1)$$

Note that the sum involves $O_p(\|\delta_T\|^{-2})$ terms, so $\|\sum_{t=\hat{k}+1}^{k_0} x_t z_t'\|\|\delta_T\|^2 = O_p(1)$. Next,

$$\left\|\frac{1}{T}Z_0'Z_0 - \frac{1}{T}\hat{Z}_0'\hat{Z}_0\right\| = \frac{1}{T\|\delta_T\|^2}\left\|\sum_{t=\hat{k}+1}^{k_0} z_t z_t'\right\|\|\delta_T\|^2 = \frac{1}{T\|\delta_T\|^2}O_p(1) = o_p(1)$$

All other entries are handled similarly The normality follows from the central limit theorem

Proposition 1 implies that $\hat{k}$ will not lie in $K(C)$ for large C with large probability Let $D(C)$ denote the complement set of $K(C)$ such that $D(C) = \{k.\ |k - k_0| \le C\|\delta_T\|^{-2}\}$ To study the limiting distribution, we study the behavior of $V_T(k)$ on $D(C)$.

LEMMA A.5: Under assumptions A1–A6, for $\delta_T = \delta$ fixed or for δ_T satisfying A7,

$$V_T(k) - V_T(k_0) = -\delta_T'Z_\Delta'Z_\Delta\delta_T + 2\delta_T'Z_\Delta'\epsilon \operatorname{sgn}(k_0 - k) + o_p(1)$$

where $o_p(1)$ is uniform on $D(C)$

Proof: We prove first that $|k_0 - k|G_T(k) = \delta_T'Z_\Delta'Z_\Delta\delta_T + o_p(1)$ We use $Z_0 = Z_2 - Z_\Delta \operatorname{sgn}(k_0 - k)$ to obtain

$$\begin{aligned}|k_0 - k|G_T(k) &= \delta_T'[(Z_0'MZ_0) - (Z_0'MZ_2) \\ &\quad \times(Z_2'MZ_2)^{-1}(Z_2'MZ_0)]\delta_T \\ &= \delta_T'(Z_\Delta'MZ_\Delta)\delta_T - \delta_T'(Z_\Delta'MZ_2) \\ &\quad \times(Z_2'MZ_2)^{-1}(Z_2'MZ_\Delta)\delta_T.\end{aligned} \tag{A.13}$$

Now $\|(Z_\Delta'MZ_2)\| = O_p(1)\|\delta_T\|^{-2}$ and $(Z_2'MZ_2)^{-1} = O_p(T^{-1})$. Thus the second term of equation (A.13) is bounded by $O_p(1)/(T\|\delta_T\|^2) = o_p(1)$. Next,

$$\delta_T'(Z_\Delta'MZ_\Delta)\delta_T = \delta_T'Z_\Delta'Z_\Delta\delta_T + \delta_T'Z_\Delta'X(X'X)^{-1}X'Z_\Delta\delta_T$$

The second term above is also $O_p(1)/(T\|\delta_T\|^2) = o_p(1)$. This proves that $|k_0 - k|G_T(k) = \delta_T Z_\Delta'Z_\Delta\delta_T + o_p(1)$ Next consider $H_T(k)$ Because $|k_0 - k| \le C\|\delta_T\|^{-2}$, the second term on the right-hand side of equation (A.11) is bounded by $CT^{-1/2}\|\delta_T\|^{-1}O_p(1) = o_p(1)$ The last term of equation (A 11), $O_p(1)$, which stands for expression (A.2), can be easily shown to be $o_p(1)$ uniformly on $D(C)$. Combining these results together with equations (A 4) and (A.11), we obtain the lemma

Proof of Proposition 2

For $k < k_0$,

$$-\delta'Z_\Delta'Z_\Delta\delta + 2\delta'Z_\Delta'\epsilon \operatorname{sgn}(k_0 - k) = -\delta'\left(\sum_{t=k+1}^{k_0} z_t z_t'\right)\delta + 2\delta'\sum_{t=k+1}^{k_0} z_t\epsilon_t. \tag{A 14}$$

Under strict stationarity, the above has the same distribution as $W_1(k - k_0)$ defined in equation (8) Similarly, for $k > k_0$, the left-hand side above is $-\delta'(\sum_{t=k_0+1}^{k} z_t z_t')\delta - 2\delta'\sum_{t=k_0+1}^{k} z_t\epsilon_t$, which has the same distribution as $W_2(k - k_0)$ Thus lemma A 5 implies that $V_T(k) - V_T(k_0)$ converges in distribution to $W^*(k - k_0)$ and the convergence is uniform on any given bounded set of integers around k_0 Now the process $W^*(m)$ has a unique maximum with probability 1 when $(\delta'z_t)^2 + 2\delta'z_t\epsilon_t$ has a continuous distribution, because $P(W^*(m) = W^*(m')) = 0$ for $m \ne m'$ Let $m^* = \arg\max_m W^*(m)$, then m^* is $O_p(1)$ because $W^*(m) \to -\infty$ with probability tending to 1 as $|m| \to \infty$ We thus have that for every $\epsilon > 0$, there exists $K < \infty$ such that $P(|m^*| > K) < \epsilon$ Similarly by proposition 1, $P(|\hat{k} - k_0| > K) < \epsilon$. Let $\hat{k}_K = \arg\max_{|k-k_0| \le K} V_T(k)$ and $m_K^* = \arg\max_{|m| \le K} W^*(m)$ The uniform convergence of $V_T(k) - V_T(k_0)$ to $W^*(k - k_0)$ on any bounded set of integers (the difference $|k - k_0|$ is bounded) implies that $\hat{k}_K - k_0 \xrightarrow{d} m_K^*$ That is, for large T, $|P(\hat{k}_K - k_0 = j) - P(m_K^* = j)| < \epsilon$ for all $|j| \le K$. However, if $|\hat{k} - k_0| \le K$, then $\hat{k} = \hat{k}_K$ and if $|m^*| \le K$, then $m^* = m_K^*$ Thus the difference $|P(\hat{k} - k_0 = j) - P(m^* = j)|$ is bounded by $|P(\hat{k}_K - k_0 = j) - P(m_K^* = j)| + P(|\hat{k} - k_0| > K) + P(|m^*| > K) < 3\epsilon$ Since ϵ can be arbitrarily small and K can be arbitrary large, the proposition is proved

Proof of Proposition 3

Because $\delta_T = \delta_0 v_T$, proposition 1 implies that $\hat{k} = k_0 + O_p(v_T^{-2})$ For any given $C > 0$, we shall derive the limiting process of $V_T(k) - V_T(k_0)$ for $k = k_0 + [sv_T^{-2}]$ and $s \in [-C, C]$ Consider $s \le 0$ (i e , $k \le k_0$) By lemma A.5,

$$V_T(k) - V_T(k_0) = -\delta_0'\left(v_T^2\sum_{t=k+1}^{k_0} z_t z_t'\right)\delta_0 + 2\delta_0'\left(v_T\sum_{t=k+1}^{k_0} z_t\epsilon_t\right) + o_p(1)$$

By assumptions A4 and A9, for $k = k_0 + [sv_T^{-2}]$, $v_T^2\sum_{t=k+1}^{k_0} z_t z_t' \to |s|Q_1$. In addition, $v_T\sum_{t=k+1}^{k_0} z_t\epsilon_t \Rightarrow B_1(-s)$, where $B_1(x)$ is a Brownian motion process on $[0, \infty)$ with variance $x\Omega_1$ Because $\delta_0'B_1(-s)$ has the same distribution as $(\delta_0'\Omega_1\delta_0)^{1/2} W_1(-s)$, where $W_1(\ )$ is a standard Wiener process on $[0, \infty)$, it follows that, for $s \le 0$, $V_T(k_0 + [sv_T^{-2}]) - V_T(k_0) \Rightarrow -|s|\delta_0'Q_1\delta_0 + 2(\delta_0'\Omega_1\delta_0)^{1/2} W_1(-s)$. Similarly, for $s > 0$, $V_T(k_0 + [sv_T^{-2}]) - V_T(k_0) \Rightarrow -s\delta_0'Q_2\delta_0 + 2(\delta_0'\Omega_2\delta_0)^{1/2} W_2(s)$, where W_2 is another Wiener process on $[0, \infty)$ independent of W_1. In summary,

$$V_T(k_0 + [sv_T^{-2}]) - V_T(k_0) \Rightarrow G(s) = \begin{cases}-|s|\delta_0'Q_1\delta_0 + 2(\delta_0'\Omega_1\delta_0)^{1/2}W_1(-s), & \text{if } s < 0 \\ -s\delta_0'Q_2\delta_0 + 2(\delta_0'\Omega_2\delta_0)^{1/2}W_2(s), & \text{if } s \ge 0\end{cases} \tag{A 15}$$

By the continuous mapping theorem, $v_T^2(\hat{k} - k_0) \xrightarrow{d} \arg\max_s G(s)$ However, by a change in variable $s = bv$ with $b = (\delta_0'\Omega_1\delta_0)/(\delta_0'Q_1\delta_0)^2$, it can be shown that $\arg\max_s G(s) = b \arg\max_v Z(v)$, where $Z(v)$ is defined in equation (11) This implies that $b^{-1}v_T^2(\hat{k} - k_0) \xrightarrow{d} \arg\max_v Z(v)$. But $b^{-1}v_T^2 = (\delta_T'Q_1\delta_T)^2/(\delta_T'\Omega_1\delta_T)$ in view of $\delta_T = \delta_0 v_T$

Proof of Proposition 4

The proof is similar to that of Bai (1995a, p 434) The details are omitted

Copyright © 2001. All Rights Reserved.

APPENDIX B

Density Function of arg $\max_v Z(v)$ (see equation (11))

Recall that the process $Z(v)$ involves two independent Brownian motions with different scales ($\phi \neq 1$) and two different drifts $\xi \neq 1$. To obtain the density function of $l^* = \arg\max_v Z(v)$, we use the argument of Bhattacharya and Brockwell (1976), who considered the case in which the two Brownian motions have the same scale In addition to the density function, we also derive the cdf of l^*, which is useful for obtaining quantiles when constructing confidence intervals Let m be the maximum and l the location of the maximum of $X(t) = \lambda B(t) - \theta t/2$, $t \geq 0$, where $B(t)$ is a standard Brownian motion process on $[0, \infty)$ That is, $l = \arg\max_t X(t)$ and $m = \max_t X(t)$ Using the result of Bhattacharya and Brockwell (1976) with a scale transformation, we obtain the joint density of (l, m) as

$$f_{l,m}(x, y) = \frac{\theta}{\lambda}(2\pi)^{-1}\exp\left(-\theta^2\frac{x}{8}\right) \times \exp\left[-\frac{1}{2}\left(\frac{y^2}{\lambda^2}x^{-1} + \frac{\theta}{\lambda}y\right)\right], \quad x > 0, \quad y > 0 \tag{B.1}$$

The marginal distribution of m is exponential with mean λ^2/θ Now let (l_1, m_1) be the location and the maximum of $\{W_1(-v) - |v|/2\}$, $v < 0$ The joint density of (l_1, m_1) is $f_1(|x|, y)$, where f_1 is the same as equation (B 1) with $\theta = \lambda = 1$ Let (l_2, m_2) be the location and the maximum of $\{(\sqrt{\phi})W_2(v) - \xi v/2\}$, $v > 0$, with joint density $f_2(x, y)$ (given by equation (B 1) with $\lambda = \sqrt{\phi}$ and $\theta = \xi$). Because (l_1, m_1) is independent of (l_2, m_2), the joint density of (l_1, m_1, l_2, m_2) is $f_1(|x|, y)f_2(x', y')$. Let $l^* = \arg\max_v Z(v)$. Then $l^* = l_1$ if $m_1 > m_2$ and $l^* = l_2$ if $m_1 < m_2$ Thus the density function $g(x)$ of l^* for $x < 0$ is

$$\begin{aligned} g(x)\,dx &= P(l_1 \in (x, x + dx), m_1 > m_2) \\ &= dx\int_0^\infty \left[f_1(|x|, y)\int_0^\infty\int_0^y f_2(x', y')\,dx'\,dy'\right]dy \\ &= dx\int_0^\infty f_1(|x|, y)P(m_2 < y)\,dy \\ &= dx\int_0^\infty f_1(|x|, y)\left[1 - \exp\left(-\frac{\theta}{\lambda^2}y\right)\right]dy \end{aligned}$$

Carrying out the integration we obtain

$$g(x) = 2^{-1}\Phi(-2^{-1}\sqrt{|x|}) + 2^{-1}(1 + 2\alpha)\exp[2^{-1}\alpha(1 + \alpha)|x|] \times \Phi[-2^{-1}(1 + 2\alpha)\sqrt{|x|}], \quad x < 0$$

where $\alpha = \xi\phi^{-1}$ and $\Phi(\ )$ is the cdf of a standard normal random variable. Similarly, for $x > 0$, $g(x) = \int_0^\infty f_2(x, y)P(m_1 < y)\,dy = \int_0^\infty f_2(x, y)\,(1 - e^{-y})\,dy$ Carrying out the integration, we obtain

$$g(x) = -2^{-1}\beta^2\Phi(-2^{-1}\beta\sqrt{x}) + (\xi + 2^{-1}\beta^2)\exp[2^{-1}(\phi + \xi)x] \times \Phi[-(\phi^{1/2} + 2^{-1}\beta)\sqrt{x}], \quad x > 0$$

where $\beta = \xi/\phi^{1/2}$ The cdf $G(x)$ for $x < 0$ is then obtained as

$$\begin{aligned} G(x) &= -(2\pi)^{-1/2}|x|^{1/2}\exp(-8^{-1}|x|) \\ &\quad - c\exp(a|x|)\Phi(-b|x|^{1/2}) \\ &\quad + (d - 2 + 2^{-1}|x|)\Phi(-2^{-1}|x|^{1/2}), \quad x < 0 \end{aligned} \tag{B.2}$$

where

$$a = 2^{-1}\frac{\xi}{\phi}\left(1 + \frac{\xi}{\phi}\right), \quad b = \frac{1}{2} + \frac{\xi}{\phi}, \quad c = \frac{\phi(\phi + 2\xi)}{\xi(\phi + \xi)}, \quad d = \frac{(\phi + 2\xi)^2}{(\phi + \xi)\xi}$$

Similarly, for $x > 0$,

$$\begin{aligned} G(x) &= 1 + \xi\phi^{-1/2}(2\pi)^{-1/2}x^{1/2}\exp(-8^{-1}\xi^2\phi^{-1}x) \\ &\quad + c\exp(ax)\Phi(-bx^{1/2}) + (-d + 2 - 2^{-1}\xi^2\phi^{-1}x) \\ &\quad \times \Phi(-2^{-1}\xi\phi^{-1/2}x^{1/2}), \quad x > 0 \end{aligned} \tag{B 3}$$

where

$$a = \frac{\phi + \xi}{2}, \quad b = \frac{2\phi + \xi}{2\sqrt{\phi}}, \quad c = \frac{\xi(2\phi + \xi)}{(\phi + \xi)\phi}, \quad d = \frac{(2\phi + \xi)^2}{(\phi + \xi)\phi}.$$

The density function and the cdf specialize for the symmetric case when $\phi = 1$ and $\xi = 1$, yielding (see Yao (1987))

$$G(x) = 1 + (2\pi)^{-1/2}\sqrt{x}e^{-x/8} - \frac{1}{2}(x + 5) \times \Phi\left(-\frac{\sqrt{x}}{2}\right) + \frac{3}{2}e^x\Phi\left(-\frac{3\sqrt{x}}{2}\right) \tag{B 4}$$

for $x > 0$ and $G(x) = 1 - G(-x)$

Copyright © 2001. All Rights Reserved.