Review of Economic Studies (1998) 65, 395–432 0034-6527/98/00180395$02.00


# Testing For and Dating Common Breaks in Multivariate Time Series

JUSHAN BAI
*Massachusetts Institute of Technology,*

ROBIN L. LUMSDAINE
*Brown University*
*and*
*National Bureau of Economic Research,*

*and*

JAMES H. STOCK
*Harvard University*
*and*
*National Bureau of Economic Research*

*First version received August* 1994; *final version accepted June* 1997 (*Eds.*)

This paper develops methods for constructing asymptotically valid confidence intervals for the date of a single break in multivariate time series, including $I(0)$, $I(1)$, and deterministically trending regressors. Although the width of the asymptotic confidence interval does not decrease as the sample size increases, it is inversely related to the number of series which have a common break date, so there are substantial gains to multivariate inference about break dates. These methods are applied to two empirical examples: the mean growth rate of output in three European countries, and the mean growth rate of U.S. consumption, investment, and output.

## 1. INTRODUCTION

The past decade has seen considerable empirical and theoretical research on the detection of breaks in economic time series. Notably, Perron (1989) and Rappoport and Reichlin (1989) provided evidence that aggregate output can be usefully thought of as being subject to two types of shocks: highly persistent shocks, which affect mean growth rates over decades, and transitory shocks, which result in business cycles and other short-run dynamics. Because the permanent shocks occur so rarely, it is useful to model them as one-time events, in the case at hand as one-time changes in the trend growth of real output. Making the assumption that the break dates were known, Perron (1989) and Rappoport and Reichlin (1989) concluded that U.S. output was better modelled as being stationary around a broken trend, or a trend with a change in its slope, than as being integrated of order one. Although subsequent work (Banerjee, Lumsdaine, and Stock (1992), Christiano (1992), and Zivot and Andrews (1992)) which treated the break date as unknown questioned some of these results, there remains evidence of breaks in the mean growth rates of many aggregate economic time series.

The empirical motivation for this study is to build on this evidence of the existence of breaks in mean macroeconomic growth rates by providing point estimates and, more importantly, confidence intervals for the dates of these breaks. Graphical evidence points to growth in the United States and Europe slowing down sometime in the 1970s, although for some countries, especially Germany, this graphical evidence is far from clear. Although Perron (1989), Banerjee, Lumsdaine, and Stock (1992), and others have used time series methods to provide estimates of break dates in mean growth rates, formal measures of the precision of these estimates were unavailable to these researchers. An interval estimate of the break date is considerably more useful to economists attempting to understand this growth slowdown than is a simple point estimate with no measure of sampling uncertainty. Further, many factors which are generally deemed important in this slowdown, such as external supply shocks and the growth of the modern European welfare state, are international and could result in the breaks being contemporaneous. This suggests that gains in precision might be achieved by a multivariate treatment, in which the growth rates are modelled as breaking contemporaneously across series. However, techniques for inference about break dates in multivariate systems are currently unavailable.

This paper therefore develops techniques for inference about breaks, including interval estimation of the break date, in multivariate systems. The econometric literature to date has focused on tests for structural breaks, with recent emphasis on the case that the break date is unknown (see Hansen (1992), Andrews (1993), and Andrews and Ploberger (1994) for recent treatments). However, the problem of inference about the break date itself has received significantly less attention. We therefore develop the econometric theory of interval estimation of the date of a break in a multivariate time series model with otherwise stationary or cointegrated variables. This entails developing asymptotic distribution theory for the maximum likelihood estimator of the break date. As this theory makes precise, the break-point problem is one in which there are substantial payoffs for using multivariate rather than univariate techniques: while the asymptotic confidence interval for the break date does not decrease with the sample size, it is inversely related to the dimension of the time series.

The empirical motivation concerns breaks in the mean growth rate, for which the parameters describing the stationary dependence in the stochastic part of the process (the autoregressive parameters) are treated as nuisance parameters. However, our results are general enough to permit an extension to breaks in any of the coefficients of an $I(0)$ or cointegrated model. Even though this general problem is not the focus of the empirical work in this study, it is arguably of interest in other applications, so we present tests and confidence intervals for the general case.

We next turn to the empirical problem of dating the slowdown in postwar European and U.S. output growth. For France, Germany, and Italy, there is evidence of a break in the univariate growth rates of output (*cf.* Banerjee, Lumsdaine, and Stock (1992)), and the model of a single common break date is found to be consistent with the data. We therefore consider a multivariate system with a single common break date, and find that a 90% confidence interval for the break is the second quarter of 1972 to the second quarter of 1975.

Dating the slowdown in the postwar U.S. is somewhat more difficult; the univariate estimate of the break date for U.S. output is imprecise. However, dynamic economic theories suggest that a discrete productivity slowdown will be reflected in lower growth rates not only of output, but of series that are cointegrated with output, in particular, consumption and investment (*cf.* King, Plosser, and Rebelo (1988)). We therefore examine a trivariate system of real *per capita* output, consumption, and investment in which, following King, Plosser, Stock, and Watson (1991), there are two cointegrating vectors corresponding to the stationarity of the logarithms of the consumption/income and investment/income ratios. When modelled as a system there appears to be a common slowdown in the growth rate that is statistically significant. The 90% confidence interval is centred around the first quarter of 1969 and is very tight when the theoretical cointegrating vectors are imposed, although it is somewhat wider (for example, 1966–1971) when the cointegrating vectors are estimated.

The paper is organized as follows. Section 2 contains the theoretical econometric results concerning multivariate change-point tests and confidence intervals for $I(0)$ dynamic models. Section 3 addresses the change-point problem in a cointegrating system. Section 4 presents a Monte Carlo study of the tests and interval estimates. The empirical results are presented in Sections 5 and 6 for the European and U.S. applications, respectively. Section 7 concludes.

## 2. TESTS AND CONFIDENCE INTERVALS FOR A BREAK IN $I(0)$ DYNAMIC MODELS

### 2.1. *Model and notation*

The system of equations considered is

$$y_t = \mu + \textstyle\sum_{j=1}^{p} A_j y_{t-j} + \Gamma X_{t-1} + d_t(k)(\lambda + \textstyle\sum_{j=1}^{p} B_j y_{t-j} + \Pi X_{t-1}) + \varepsilon_t, \tag{2.1}$$

where $y_t$, $\mu$, $\lambda$, and $\varepsilon_t$ are $n\times 1$ and $\{A_j\}$ and $\{B_j\}$ are $n\times n$; the roots of $\{I - A(L)L\}$ and of $\{I - A(L)L - B(L)L\}$ are outside the unit circle; $d_t(k)=0$ for $t\leqq k$ and $d_t(k)=1$ for $t>k$; and $X_t$ is a matrix of stationary variables. It is convenient to write the system of equations (2.1) in its stacked form

$$y_t = (V_t'\otimes I)\theta + d_t(k)(V_t'\otimes I)\delta + \varepsilon_t, \tag{2.2}$$

where $V_t' = (1, y_{t-1}', \ldots, y_{t-p}', X_{t-1}')$, $\theta = \mathrm{Vec}\,(\mu, A_1, \ldots, A_p, \Gamma)$, $\delta = \mathrm{Vec}\,(\lambda, B_1, \ldots, B_p, \Pi)$, and $I$ is the $n\times n$ identity matrix. Model (2.2) is that of a full structural change in that it allows all coefficients to change. If it is known that only a subset of coefficients such as the intercept has a possible break, a partial structural change model is more appropriate. The unchanged parameters should be estimated using all of the observations to gain efficiency. In addition, tests for partial structural changes will have better power than those for full structural changes. This leads to the consideration of a general partial structural change model

$$y_t = (V_t'\otimes I)\theta + d_t(k)(V_t'\otimes I)S'S\delta + \varepsilon_t, \tag{2.3}$$

where $S$ is a selection matrix, containing 0's and 1's and having full row rank. Note that $S'S$ is idempotent with non zero elements only on the diagonal. The rank of $S$ is equal to the number of coefficients that are allowed to change. For $S=I$, (2.2) is obtained. For $S=(s\otimes I)$ with $s=(1, 0, \ldots, 0)$, we have

$$y_t = (V_t'\otimes I)\theta + \lambda d_t(k) + \varepsilon_t, \tag{2.4}$$

which has a break in the intercept only. The system (2.3) can be rewritten more compactly as

$$y_t = Z_t'(k)\beta + \varepsilon_t, \tag{2.5}$$

where $Z_k'(k)=((V_t'\otimes I), d_t(k)(V_t'\otimes I)S')$ and $\beta=(\theta', (S\delta)')'$. Write $Z_t$ for $Z_t(k)$ for notational simplicity. The errors $\varepsilon_t$ are assumed to satisfy the following assumption:

*Assumption* 2.1. Let $\varepsilon_t$ be a martingale difference sequence with respect to $\mathscr{F}_{t-1}=$ $\sigma$-field $(Z_t, \varepsilon_{t-1}, Z_{t-1}, \varepsilon_{t-2}, \ldots)$ satisfying, for some $\alpha>0$, $\max_i \sup_t E(\varepsilon_{it}^{4+\alpha})<\infty$ and $E(\varepsilon_t\varepsilon_{t-j}'|\mathscr{F}_{t-1})=\Sigma$ for $j=0$ and 0 otherwise. Also suppose that $EX_t=\mu_x$ for all $t$, $\max_i$ $\sup_t E(X_{it}^{4+\alpha})<\infty$, $T^{-1}\sum_{t=1}^T(X_t-\mu_x)(X_t-\mu_x)'\xrightarrow{p}M_{xx}(0)$, $T^{-1}\sum_{t=1}^T X_t y_{t-j}'\xrightarrow{p}EX_t y_{t-j}'=M_{xy}(j)$, $j=-p,\ldots,p$ and $\chi_T(\cdot)\Rightarrow B_x(\cdot)$, where $\chi_T(\tau)=T^{-1/2}\sum_{t=1}^{[T\tau]}(X_t-\mu_x)$, $[x]$ represents the integer part of $x$, and $B_x(\cdot)$ is a Brownian motion with covariance matrix $M_{xx}(0)$.

Throughout, $\|x\|$ represents the Euclidean norm, i.e. $\|x\|=(\sum_{i=1}^p x_i^2)^{1/2}$ for $x\in R^p$. All limits are taken as the sample size, $T$, converges to infinity, unless stated otherwise.

### 2.2. *Tests for a break at an unknown date*

The tests for a break in the coefficients are based on the sequence of $F$-statistics testing $S\delta=0$, for $k=k_*+1,\ldots,T-k_*$, where $k_*$ is some trimming value. The null hypothesis is that no break exists $(S\delta=0)$. For a given $k$, the estimator of $\hat{\beta}(k)$ (by the feasible seemingly unrelated regression method) and the $F$-statistic testing $S\delta=0$, $\hat{F}(k)$, are

$$\hat{\beta}(k)=\{\textstyle\sum_{t=1}^T(Z_t\hat{\Sigma}_k^{-1}Z_t')\}^{-1}\sum_{t=1}^T Z_t\hat{\Sigma}_k^{-1}y_t, \tag{2.6}$$

$$\hat{F}(k)=T\{R\hat{\beta}(k)\}'\{R(T^{-1}\textstyle\sum_{t=1}^T Z_t\hat{\Sigma}_k^{-1}Z_t')^{-1}R'\}^{-1}\{R\hat{\beta}(k)\}, \tag{2.7}$$

where $R=(0, I)$ so that $R\beta=S\delta$, $\hat{\Sigma}_k$ is the estimator of $\Sigma$ based on OLS residuals under the alternative hypothesis, given $k$. The stochastic processes of interest are the estimator process, $T^{1/2}(\hat{\beta}([T\tau])-\beta)$, and the $F$-statistic process, $F_T(\tau)=\hat{F}([T\tau])$.

A variety of tests for a break, based on the Wald $F$-statistic process $F_T$, have been proposed in the literature. For example, the Quandt (1960) likelihood ratio statistic is the maximum of the likelihood ratio statistics, testing for a break at a sequence of possible break dates; the analogous statistic here is to consider the maximum of the $F_T$ process. Hansen (1992) proposed using the mean score test for a break; the Wald test variant used here is the average of $\hat{F}(k)$ over some range. Andrews and Ploberger (1994) examine the question of devising most powerful tests for breaks. In the empirical application, we consider two of these test statistics: the maximum Wald statistic and the logarithm of the Andrews–Ploberger exponential Wald statistic

$$\text{Sup-}W\text{:} \quad \sup_{\tau\in(\tau_*, 1-\tau_*)} F_T(\tau), \tag{2.8}$$

$$\text{Exp-}W\text{:} \quad \ln\{\textstyle\int_{\tau_*}^{1-\tau_*}\exp\{\frac{1}{2}F_T(\tau)\}d\tau\}. \tag{2.9}$$

Here $\tau_*$ refers to an initial fraction of the sample which is trimmed; this is often taken to be either 0·15 or 0·01. The limiting distribution of the $F_T$ process has been well studied for a variety of models; see for example Deshayes and Picard (1986), Andrews (1993), and Andrews and Ploberger (1994). The limiting distributions of these statistics obtain by applying the continuous mapping theorem to the limiting representation of $F_T$. Andrews and Ploberger (1994) provide general conditions under which these limiting distributions will hold, although these conditions are "high level" and must be verified in practice. For the model of interest here, (2.3), the limiting properties of the estimator and $F$-statistic process are summarized by the following theorem. Let " $\Rightarrow$ " denote weak convergence of random elements in a product space of $D[0, 1]$.

**Theorem 1.** *Under the assumption of* 2.1 *and* $S\delta = 0$,

(i) $\sqrt{T}(\hat{\beta}([T\cdot]) - \beta) \doteq H(\cdot)^{-1}G(\cdot)$, *where*

$$H(\tau) = \begin{pmatrix} Q(1)\otimes\Sigma^{-1} & \{(1-\tau)Q(1)\otimes\Sigma^{-1}\}S' \\ S\{(1-\tau)Q(1)\otimes\Sigma^{-1}\} & S\{(1-\tau)Q(1)\otimes\Sigma^{-1}\}S' \end{pmatrix}$$

*and* $G = (G_1', G_2')'$ *with* $G_1(\tau) = B(1)$, *and* $G_2(\tau) = S[B(1) - B(\tau)]$, *where* $B(\tau)$ *is a vector of Brownian motion with variance* $\tau Q(1)\otimes\Sigma^{-1}$, *and* $Q(\tau) = \text{plim}\ T^{-1}\sum_{t=1}^{[T\tau]} V_t V_t' = \tau Q(1)$. *The weak convergence holds in the space* $D[\tau^*, 1-\tau^*]$.

(ii) $F_T \Rightarrow F^*$, *where* $F^*(\tau) = \{\tau(1-\tau)\}^{-1}\|W(\tau) - \tau W(1)\|^2$ *and* $W(\cdot)$ *is a vector of independent standard Brownian motion processes with* $\dim(W) = \text{rank}(S)$.

(iii) $g(F_T) \Rightarrow g(F^*)$, *for* $g(\cdot)$ *denoting the* Sup-$W$ *and* Exp-$W$ *functionals, respectively, given in* (2.8) *and* (2.9).

The proofs of this theorem and Theorem 8 in Section 3 are omitted to conserve space. Proofs of all other theorems are given in the Appendix.[1]

### 2.3. *Inference for breaks in I(0) dynamic models*

If there is in fact a break, then a natural question to raise is how one could construct confidence intervals for the true break date. This problem has been considered by various authors, using a variety of approaches; see, for example, Hinkley (1970), Picard (1985), Yao (1987), Siegmund (1988), and Kim and Siegmund (1989). Most of this work has focused on the change-point problem with i.i.d. Gaussian errors. Picard (1985) provided an asymptotic distribution for the Gaussian MLE of the breakpoint in the case that a univariate process follows a finite order autoregression; also see Yao (1987). Picard's results permit the construction of asymptotic confidence intervals for the break point in the univariate case. These results are extended here in a number of directions: (1) the time series is multivariate rather than univariate; (2) the covariance matrix $\Sigma$ is explicitly treated as unknown and estimated; (3) no normality assumption is made, nor is the underlying density function assumed to be known. We only assume the disturbances form a sequence of martingale differences with some moment conditions, and use pseudo-Gaussian maximum likelihood estimation; (4) we consider partial structural change models, allowing some of regression parameters to be estimated with the full sample to gain efficiency; (5) we further study regression models with $I(1)$ and trending regressors and with serially correlated errors, encompassing a broken-trend stationary model and that of broken cointegrating relationships.

We consider estimating (2.3) or (2.5) by the pseudo-Gaussian MLE. We assume $\|S\delta\| \neq 0$, so that there indeed exists a break. Denote by $L(k, \beta, \Sigma)$ the pseudo-likelihood function admitting a break at $k$ with parameters $\beta$ and $\Sigma$. Let $(k_0, \beta_0, \Sigma_0)$ denote the true parameter with $k_0 = [T\tau_0]$ for $\tau_0 \in (0, 1)$. For each given $k$, denote by $(\hat{\beta}(k), \hat{\Sigma}(k))$ the estimator that maximizes the likelihood function. The break point estimator is defined as

$$\hat{k} = \underset{1\leq k\leq T}{\text{argmax}}\ L(k, \hat{\beta}(k), \hat{\Sigma}(k)).$$

The final estimator is defined as $(\hat{k}, \hat{\beta}(\hat{k}), \hat{\Sigma}(\hat{k}))$, also written as $(\hat{k}, \hat{\beta}_{\hat{k}}, \hat{\Sigma}_{\hat{k}})$.

The asymptotic behaviour of $\hat{k}$ is obtained by considering a sequence which is designed to produce an asymptotic approximation to the finite sample distribution of $\hat{k}$ when the

1. A complete proof of the two omitted theorems is available upon request.

magnitude of the break is small. Specifically, following Picard (1985) we assume that $\beta_0=(\theta_0', \delta_T'S')'$, where $\delta_T$ is a sequence such that $\delta_T=\delta_0 v_T$ with $v_T\to 0$ and $\sqrt{T}v_T/(\log T)\to\infty$, where $v_T>0$ is a scalar. There are two reasons for considering small shifts. First, this framework permits an analytical solution to the density function of the estimated break point, so that confidence intervals can be easily constructed. Second, if we show that a break with a small magnitude of shift can be consistently estimated, it must be the case that we can consistently estimate a break with a larger magnitude of shift, for the larger the magnitude of shifts, the easier to identify a break.

We shall study the joint behaviour of $(\hat{k}, \hat{\beta}_{\hat{k}}, \hat{\Sigma}_{\hat{k}})$, particularly, their rates of convergence and their limiting distributions. The final result is given in Theorem 4 below. Anticipating the rates of convergence for the estimated parameters, we reparameterize the likelihood function, such that $L(k, \beta_0+T^{-1/2}\beta, \Sigma_0+T^{-1/2}\Sigma)$, where $\beta=(\theta', (S\delta)')'$. The break $k$ is reparameterized such that $k=k(v)=k_0+[vv_T^{-2}]$, for $v\in R$. When $v$ varies, $k$ can take on all possible integer values. We define the likelihood function to be zero for $k$ nonpositive and for $k$ greater than $T$. It is clear that maximizing the original likelihood function is equivalent to maximizing the reparameterized likelihood. Define the pseudo-likelihood ratio as

$$\begin{aligned}\Lambda_T(v, \beta, \Sigma) &= \frac{L(k, \beta_0+T^{-1/2}\beta, \Sigma_0+T^{-1/2}\Sigma)}{L(k_0, \beta_0, \Sigma_0)} \\ &= \frac{|\Sigma_0+T^{-1/2}\Sigma|^{-T/2}\exp\{-\frac{1}{2}\sum_{t=1}^T \varepsilon_t(k)'(\Sigma_0+T^{-1/2}\Sigma)^{-1}\varepsilon_t(k)\}}{|\Sigma_0|^{-T/2}\exp\{-\frac{1}{2}\sum_{t=1}^T \varepsilon_t'\Sigma_0^{-1}\varepsilon_t\}}, \end{aligned} \tag{2.10}$$

where $\varepsilon_t(k)=y_t-Z_t(k)'(\beta_0+T^{-1/2}\beta)$.

It is also clear that maximizing the original likelihood function is equivalent to maximizing the likelihood ratio. Suppose that $v^*, \beta^*, \Sigma^*$ maximize the likelihood ratio, then $\beta^*=\sqrt{T}(\hat{\beta}_{\hat{k}}-\beta_0)$, $\Sigma^*=\sqrt{T}(\hat{\Sigma}_{\hat{k}}-\Sigma_0)$, and $v^*=v_T^2(\hat{k}-k_0)$. Thus to show $\sqrt{T}(\hat{\beta}_{\hat{k}}-\beta_0)$, $\sqrt{T}(\hat{\Sigma}_{\hat{k}}-\Sigma_0)$, and $v_T^2(\hat{k}-k_0)$ are all stochastically bounded [i.e., $O_p(1)$], it is sufficient to show $\beta^*$, $\Sigma^*$, and $v^*$ are all stochastically bounded. This, in turn, is equivalent to showing that the likelihood ratio cannot achieve its maximum when any of the parameters, $v$, $\beta$, $\Sigma$, is too large. Because $\Lambda_T(0,0,0)=1$, it suffices to show the likelihood ratio is smaller than 1 for large values of $v$, $\beta$, and $\Sigma$. Formally,

**Theorem 2.** *Under Assumption* 2.1, *if $v_T$ is fixed, or if $v_T\to 0$ and $\sqrt{T}v_T/(\log T)\to\infty$, then for every $\varepsilon>0$, there exists a $v_1>0$, such that*

$$\Pr\left(\sup_{|v|\geqq v_1}\sup_{\beta,\Sigma}\Lambda_T(v, \beta, \Sigma)>\varepsilon\right)<\varepsilon, \tag{2.11}$$

*and for every $\varepsilon>0$, there exists an $M>0$, such that*

$$\Pr\left(\sup_{|v|\leqq v_1}\ \sup_{\|\beta\|>M \text{ or } \|\Sigma\|>M}\Lambda_T(v, \beta, \Sigma)>\varepsilon\right)<\varepsilon. \tag{2.12}$$

This theorem gives rise to the desired rates of convergence. In particular, $v_T^2(\hat{k}-k_0)=O_p(1)$.

Having studied the global property of the likelihood ratio (equivalently, the rates of convergence), we now examine the local property of the likelihood ratio in order to obtain the limiting distributions for the estimated parameters. We shall derive the limiting process for the likelihood ratio when the parameter vector $(v, \beta, \Sigma)$ is restricted in an arbitrary compact set. We then use the continuous mapping theorem for the argmax functional to obtain the limiting distributions. Note that the argmax functional is generally only continuous for stochastic processes defined on compact sets (details can be found in Bai (1992)). It is thus necessary to obtain the rates of convergence of estimated parameters in order to invoke the continuous mapping theorem. The rates of convergence guarantee that $(v^*, \beta^*, \Sigma^*)$ will lie in a compact set with a probability arbitrarily close to 1.

**Theorem 3.** *Under the conditions of Theorem* 2, *if* $v_T \to 0$ *and* $\sqrt{T}v_T/(\log T) \to \infty$, *then* $\log \Lambda_T(v, \beta, \Sigma)$ *converges weakly on any compact set of* $(v, \beta, \Sigma)$ *to the process* $\log \Lambda$ *given by*

$$\log \Lambda(v, \beta, \Sigma) = \tfrac{1}{2}\operatorname{tr}(\Sigma_0^{-1}\Sigma\Sigma_0^{-1}(\Psi - \tfrac{1}{2}\Sigma)) + \beta' Q^{1/2}\xi - \tfrac{1}{2}\beta' Q\beta + \sqrt{c}W(v) - \tfrac{1}{2}|v|c, \quad (2.13)$$

*where* $\operatorname{tr}(A)$ *denotes the trace of matrix* $A$ *and* $\Psi$ *is a* $n \times n$ *symmetric matrix of normal random variables. More specifically,* $\Psi$ *is the limiting random matrix of* $T^{-1/2}\sum_{t=1}^{T}(\varepsilon_t\varepsilon_t' - \Sigma_0)$. *Furthermore,* $Q = \operatorname{plim} T^{-1}\Sigma_{t=1}^{T} Z_t(k_0)\Sigma_0^{-1}Z_t(k_0)'$, $\xi$ *is* $N(0, I)$, $c = \delta_0' S' S(Q_1 \otimes \Sigma_0^{-1})S'S\delta_0$ *with* $Q_1 = \operatorname{plim} T^{-1}\Sigma_{t=1}^{T} V_t V_t'$. *The process* $W(\cdot)$ *is a single dimensional two-sided Brownian motion on* $(-\infty, \infty)$. *A two-sided Brownian motion* $W(\cdot)$ *on the real line is defined as* $W(v) = W_1(-v)$ *for* $v < 0$ *and* $W(v) = W_2(v)$ *for* $v \geqq 0$, *where* $W_1$ *and* $W_2$ *are two independent Brownian motion processes on* $[0, \infty)$ *with* $W_1(0) = W_2(0) = 0$.

**Theorem 4.** *Under the assumptions of Theorem* 3, *we have*

$$T^{1/2}(\hat{\beta}_{\hat{k}} - \beta_0) \xrightarrow{d} Q^{-1/2}\xi, \quad (2.14)$$

$$T^{1/2}(\hat{\Sigma}_{\hat{k}} - \Sigma_0) \xrightarrow{d} \Psi, \quad (2.15)$$

$$[\delta_T' S' S(Q_1 \otimes \Sigma_0^{-1})S'S\delta_T](\hat{k} - k_0) \xrightarrow{d} V^*, \quad (2.16)$$

*where* $V^*$ *is distributed as* $\operatorname{argmax}_v(W(v) - \tfrac{1}{2}|v|)$.

**Corollary 4.1.** *Assume the conditions of Theorem* 3.

(i) *For the intercept shift model* (2.4), *we have*

$$\lambda_T'\Sigma_0^{-1}\lambda_T(\hat{k} - k_0) \xrightarrow{d} V^*.$$

(ii) *For the full structural change model* (2.2), *we have*

$$\delta_T'(Q_1 \otimes \Sigma_0^{-1})\delta_T(\hat{k} - k_0) \xrightarrow{d} V^*. \quad (2.17)$$

*In addition,* (i) *and* (ii) *hold when* $\lambda_T$, $\Sigma_0$, $\delta_T$, *and* $Q_1$ *are replaced by their estimates.*

Because $\sqrt{T}v_T/(\log T) \to \infty$, in this formulation $\hat{\tau} = \hat{k}/T$ is consistent for $\tau_0 = k_0/T$, even though $\hat{k}$ itself is not consistent. Picard (1985) provides an explicit expression for the limiting density of $V^*$, $\gamma(x)$, as

$$\gamma(x) = \tfrac{3}{2}e^{|x|}\Phi(-\tfrac{3}{2}\sqrt{|x|}) - \tfrac{1}{2}\Phi(-\tfrac{1}{2}\sqrt{|x|}), \quad (2.18)$$

where $\Phi(\cdot)$ is the cumulative normal distribution function. The density is symmetric and nondifferentiable at $x = 0$. The 90th and 95th percentiles are, respectively, 4·67 and 7·63. It can be shown that all the moments of the density exist and the density has heavy tails, with a variance of 26 and a kurtosis of 14·5.

Because the density of $V^*$ does not depend on any nuisance parameters, the results in Corollary 4.1 can be used to construct asymptotically similar tests of $k=k_0$ or equivalently to construct asymptotic confidence intervals for $k_0$. In general, a confidence interval with asymptotic coverage of at least $100(1-\pi)\%$ is given by

$$I=[\hat{k}-[\Delta k]-1, \hat{k}+[\Delta k]+1]. \tag{2.19}$$

For the case of an intercept shift only,

$$\Delta k = c_{(1/2)\pi}(\hat{\lambda}'\hat{\Sigma}_{\hat{k}}^{-1}\hat{\lambda})^{-1}, \tag{2.20}$$

where $c_{(1/2)\pi}$ is the $(1-\frac{1}{2}\pi)$th quantile of $V^*$.

For the general case

$$\Delta k = c_{(1/2)\pi}[(S\hat{\delta}_T)'S(\hat{Q}_1 \otimes \hat{\Sigma}_{\hat{k}}^{-1})S'(S\hat{\delta}_T)]^{-1}, \tag{2.21}$$

where $\hat{Q}_1 = 1/T\Sigma_{t=1}^T V_tV_t'$.

The expression (2.20) provides an important motivation for using multiple equation systems to construct confidence intervals for $k$. Consider the case that the break date is the same for each of the series, $\Sigma$ is diagonal, and $\lambda_i^2/\Sigma_{ii}$ is the same in each equation. Then, in large samples, the width of the confidence interval declines as $1/n$. This contrasts with the conventional case of constructing confidence intervals for regression coefficients that are the same across equations, in which confidence intervals shrink at rate $1/\sqrt{n}$. In the break date case, even if $\lambda_i=0$ for some equation, so that there is no break in that equation, the asymptotic results indicate that including the additional restriction does not impose any additional cost in terms of the width of the resulting confidence intervals.

## 3. INFERENCE FOR BREAKS IN COINTEGRATING PARAMETERS

Testing for cointegration allowing for a possible break is studied by Gregory and Hansen (1996), and Campos, Ericsson, and Hendry (1996), and we provide no further results on tests for a break. Instead, we study how to estimate the break date when there is indeed a break and investigate the statistical property of the estimated break point. We consider the cointegrated system in triangular form

$$Y_t = AX_t + \gamma t + \mu + Bw_t + \xi_t, \tag{3.1a}$$

$$X_t = X_{t-1} + \Xi_t, \tag{3.1b}$$

where $Y_t$ is $r\times 1$ and $X_t$ is $(n-r)\times 1$, $w_t$ is an observable $I(0)$ process, $\xi_t$ and $\Xi_t$ are $I(0)$ error processes. More specifically, we make the following assumptions:

*Assumption* 3.1. $\xi_t = \sum_{j=0}^{\infty} C_j\varepsilon_{t-j} = C(L)\varepsilon_t$, $\Xi_t = \sum_{j=0}^{\infty} D_je_{t-j} = D(L)e_t$, $C(1)$ and $D(1)$ are full rank; $\sum_{j=0}^{\infty} j\|C_j\| < \infty$ and $\sum_{j=0}^{\infty} j\|D_j\| < \infty$; $(\varepsilon_t, e_t)$ are i.i.d. with finite $4+a$ $(a>0)$ moment. The regressor $w_t$ is a mean-zero second order stationary process with uniformly bounded $4+a$ moment.

For technical reasons, $\xi_t$ is assumed to be independent of the regressors. This is a plausible assumption because $w_t$ can be taken as the leads and lags of changes in $X_t$. (cf. Stock and Watson (1993)).

*Assumption* 3.2. The error process $\xi_t$ is independent of the regressors for all leads and lags.

*Assumption* 3.3. The $I(1)$ regressor $X_t$ satisfies:

$$E\left(\frac{kX_{ik}^2}{X_{i1}^2+X_{i2}^2+\cdots+X_{ik}^2}\right)\leqq M \qquad \text{for all } k\geqq 1 \text{ and } i=1,\ldots,n-r.$$

Note that the expectand is bounded by $k$, so every moment exists. The assumption requires a uniformly bounded first moment over all $k$. Any random walk with i.i.d. normal disturbances satisfies Assumption 3.3.[2]

*Assumption* 3.4. The regressor $w_t$ satisfies:

$$E\left(\frac{kw_{ik}^2}{w_{i1}^2+w_{i2}^2+\cdots+w_{ik}^2}\right)\leqq M, \qquad \text{for all } k\geqq 1,\ i=1,\ldots,\dim(w_t).$$

This assumption is satisfied by any i.i.d. sequence, and in this case, the expected value is identically 1. It is also satisfied by linear Gaussian processes with absolutely summable coefficients.

Incorporating a linear time trend in (3.1a) has two purposes. First, it allows us easily to extend the model to allow for drifts in $X_t$ (to be discussed later). Second, it includes a trend-stationary model as a special case (corresponding to $A=0$). Equation (3.1) may be considered as a general regression model with $I(1)$ regressors.

We consider the extension of (3.1) to the case of a single break in one or more of the coefficients. If all regression coefficients are allowed to change, we can write

$$Y_t=AX_t+\gamma t+\mu+Bw_t+d_t(k_0)(A_1X_t+\gamma_1t+\mu_1+B_1w_t)+\xi_t. \tag{3.2}$$

This can be rewritten as

$$Y_t=(U_t'\otimes I)\theta_0+d_t(k_0)(U_t'\otimes I)\delta_T+\xi_t,$$

where $U_t=(X_t', t, 1, w_t')'$, $\theta_0=\mathrm{Vec}\,(A,\gamma,\mu,B)$, $\delta_T=\mathrm{Vec}\,(A_1,\gamma_1,\mu_1,B_1)$. This is a full structural change model. As in Section 2, if it is known that some of the coefficients do not change, a full structural change model does not give efficient estimation for the regression parameters. Thus we consider a more general setup, allowing the unchanged parameters to be estimated with the entire sample. Such a partial change model has the form

$$Y_t=(U_t'\otimes I)\theta_0+d_t(k_0)(U_t'\otimes I)S'S\delta_T+\xi_t, \tag{3.3}$$

where $S$ is a selection matrix, containing elements 0 or 1. Model (3.2) corresponds to $S=I$. The system of equations (3.3) can be further rewritten as

$$Y_t=Z_t(k_0)\beta_0+\xi_t,$$

where $Z_t(k)=((U_t'\otimes I), d_t(k)(U_t'\otimes I)S')$, $\beta_0=(\theta_0', (S\delta_T)')'$.

We shall assume there is a break in at least one coefficient, so that $\|S\delta_T\|\neq 0$. As in the previous section, we assume $\delta_T$ converges to zero. In addition to the two reasons given in Section 2.2, there is an additional reason for this framework. When $\delta_T$ does not depend on $T$ and if there is a shift in the linear trend or in the cointegrating coefficients (i.e., $\gamma_1\neq 0$ or $A_1\neq 0$), then the estimated break point $\hat{k}$ converges rapidly to $k_0$, so that $P(\hat{k}\neq k_0)\to 0$ as

2. We thank K. Tanaka and C. Z. Wei for providing us with proofs for this claim (private communication). It can be shown that the underlying random sequence has a uniformly bounded moment of an arbitrary order for normal errors.

$T\to\infty$. This implies $\hat{k}$ itself has a degenerate limiting distribution. In practice, because $T$ is finite, $\hat{k}$ varies and has a nondegenerate distribution.

Since the linear trend dominates, in magnitude, the other regressors, if the trend parameter $\gamma_1$ shrinks at the same rate as other parameters, then in the limit, the trend parameter dominates the limiting behaviour of the break point estimator as if there were no shifts in other parameters (equivalent to $A_1=0$, $\mu_1=0$, $B_1=0$). When a linear trend is not included in the model, then the cointegrating coefficients will dominate the $I(0)$ coefficients. In the following, we shall consider the case in which the linear trend and the cointegrating coefficients shrink much faster than the coefficients of $I(0)$ regressors so that in the limit $\hat{k}-k_0$ is influenced by all shifted parameters. This also indicates that we can identify a much smaller shift in the cointegrating coefficients or in trend coefficients than in the case of a shift in the $I(0)$ regressors. Let

$$\delta_T=(a_T', \gamma_T', \mu_T', b_T')'=\left(\frac{1}{\sqrt{T}}a_0', \frac{1}{T}\gamma_0', \mu_0', b_0'\right)'v_T=\delta_0 \operatorname{diag}\left(\frac{1}{\sqrt{T}}I_1, \frac{1}{T}I_2, I_3\right)v_T, \tag{3.4}$$

where $a_T=\operatorname{vec}(A_1)$, $\gamma_T=\gamma_1$, $\mu_T=\mu_1$, $b_T=\operatorname{vec}(B_1)$, $\delta_0=(a_0', \gamma_0', \mu_0', b_0')'$, and $I_1, I_2, I_3$ are identity matrices. We assume $v_T$ is a scalar such that

$$v_T\to 0, \quad \text{and} \quad \sqrt{T}v_T/\log T\to\infty. \tag{3.5}$$

Despite the magnitude of shift in the trend and cointegrating coefficients being much smaller than that of the $I(0)$ regressors, we can consistently estimate the break fraction $k_0/T$, and, in addition, $v_T^2(\hat{k}-k_0)=O_p(1)$.

This rate of convergence for the estimated break point is sufficient to establish that the estimated $A$ and $A_1$ have the $T$ rate of convergence, $\gamma$ and $\gamma_1$ have the $T^{3/2}$ rate of convergence, and that the estimated $\mu$s and $B$s have the $\sqrt{T}$ rate of convergence. In addition, the variance matrix of $\xi_t$, $\Sigma_0$, is estimated with root-$T$ consistency. Anticipating these rates of convergence (but not imposing these rates of convergence), we reparameterize the parameters in the following way. Let

$$\tilde{D}_T=\operatorname{diag}(D_T, SD_TS'),$$

where $D_T=\operatorname{diag}(TI_1, T^{3/2}I_2, \sqrt{T}I_3)$ with $I_1, I_2, I_3$ being identity matrices. The matrices $D_T$ and $SD_TS'$ correspond to the rates of convergence for $\hat{\theta}$ and $\widehat{S\delta_T}$, respectively. The regression parameter is then reparameterized as $\beta_0+\tilde{D}_T^{-1}\beta$, the variance of $\xi_t$, $\Sigma_0$, is reparameterized as $\Sigma_0+T^{-1/2}\Sigma$, and the break is reparameterized as $k=k_0+[vv_T^{-2}]$. The corresponding likelihood ratio is given by

$$\Lambda_T(v, \beta, \Sigma)=L(k, \beta_0+\tilde{D}_T^{-1}\beta, \Sigma_0+T^{-1/2}\Sigma)/L(k_0, \beta_0, \Sigma_0).$$

Unlike the procedure in Section 2, where the break point is obtained via a global maximization of the likelihood for every $k\in[1, T]$, a restriction such that $k\in[T\varepsilon_0, T(1-\varepsilon_0)]$ is imposed in this section, where $\varepsilon_0>0$ is a small number. Without this restriction, the proof of our result would be much more demanding. The restriction is not too stringent from the practical point of view, as $\varepsilon_0>0$ is arbitrary. Under this setup, the estimated break point is defined as

$$\hat{k}=\underset{T\varepsilon_0\leq k\leq T(1-\varepsilon_0)}{\operatorname{argmax}}\left[\sup_{\beta,\Sigma} L(k, \beta_0+\tilde{D}_T^{-1}\beta, \Sigma_0+T^{-1/2}\Sigma)\right].$$

The global behaviour of the likelihood ratio is characterized by:

**Theorem 5.** *Under Assumptions* 3.1–3.4, *the result of Theorem* 2 *still holds for the newly defined* $\Lambda_T(v, \beta, \Sigma)$.

As in the previous section, Theorem 5 implies the following rates of convergence $\tilde{D}_T(\hat{\beta}_{\hat{k}} - \beta_0) = O_p(1)$, $T^{1/2}(\hat{\Sigma}_{\hat{k}} - \Sigma_0) = O_p(1)$, and $v_T^2(\hat{k} - k_0) = O_p(1)$. The theorem implies that the likelihood function achieves its maximum on compact sets of the reparameterized parameters $(v, \beta, \Sigma)$ with large probability. Thus, to study the limiting distribution of the estimated parameters, we only need to focus on the behaviour of the likelihood ratio on compact sets of $(v, \beta, \Sigma)$. This we do in the following theorem.

**Theorem 6.** *Under Assumptions* 3.1–3.4 *and* (3.5), *the log-likelihood ratio converges weakly on any compact set of* $(v, \beta, \Sigma)$ *to*

$$\log \Lambda(v, \beta, \Sigma) = \tfrac{1}{2} \operatorname{tr} (\Sigma_0^{-1} \Sigma \Sigma_0^{-1} (\Psi - \tfrac{1}{2}\Sigma)) + \beta' \kappa - \tfrac{1}{2} \beta' Q \beta + \sqrt{c_1} W(v) - \tfrac{1}{2} |v| c_2,$$

*where*

(1) $\Psi$ *is a* $n \times n$ *random matrix of normal random variables*;
(2) $Q$ *is the random matrix* $\operatorname{plim} (\tilde{D}_T^{-1} \Sigma_{t=1}^T Z_t(k_0) \Sigma_0^{-1} Z_t'(k_0) \tilde{D}_T^{-1})$;
(3) $\kappa$ *is the limiting distribution of* $\tilde{D}_T^{-1} \sum_{t=1}^T Z_t(k_0) \Sigma_0^{-1} \xi_t$;
(4) $c_1$ *is a random variable given by* $c_1 = \delta_0' S' S H_1 S' S \delta_0$,

$$H_1 = \begin{pmatrix} \Phi \otimes \Sigma_0^{-1} C(1) \Omega_\varepsilon C(1)' \Sigma_0^{-1} & 0 \\ 0 & \Gamma \end{pmatrix}$$

*with* $\Gamma = \Sigma_{h=-\infty}^{\infty} [E(w_t w_{t-h}') \otimes E(\xi_t' \xi_{t-h})]$, *and*

$$\Phi = \begin{pmatrix} \tau_0 D(1) \Omega_e^{1/2} Z Z' \Omega_e^{1/2} D(1)' & \tau_0^{3/2} D(1) \Omega_e^{1/2} Z & \sqrt{\tau_0} D(1) \Omega_e^{1/2} Z \\ \tau_0^{3/2} Z' \Omega_e^{1/2} D(1)' & \tau_0^2 & \tau_0 \\ \sqrt{\tau_0} Z' \Omega_e^{1/2} D(1)' & \tau_0 & 1 \end{pmatrix}$$

*where* $Z$ *is a standard normal vector*, $\Omega_\varepsilon = E\varepsilon_t \varepsilon_t'$, $\Omega_e = E e_t e_t'$ *and* $\tau_0 = k_0/T$;
(5) $c_2$ *is random variable given by* $\delta_0' S' S H_2 S' S \delta_0$, *where* $\Omega_w = E w_t w_t'$ *and*

$$H_2 = \begin{pmatrix} \Phi \otimes \Sigma_0^{-1} & 0 \\ 0 & \Omega_w \otimes \Sigma_0^{-1} \end{pmatrix};$$

(6) $W(v)$ *is a two-sided standard Brownian motion on the real line and is independent of the random variables* $c_1$ *and* $c_2$;
(7) $c_1$, $c_2$ *and* $W(v)$ *are independent of* $Q$, $\kappa$, $\Psi$.

The next theorem derives the limiting distributions for the estimated parameters, including the change point:

**Theorem 7.** *Under the assumptions of Theorem* 6,

$$\tilde{D}_T(\hat{\beta}_{\hat{k}} - \beta_0) \xrightarrow{d} Q^{-1} \kappa,$$

$$\sqrt{T}(\hat{\Sigma}_{\hat{k}} - \Sigma_0) \xrightarrow{d} \Psi,$$

$$\frac{(\delta_0' S' S H_2 S' S \delta_0)^2}{\delta_0' S' S H_1 S' S \delta_0} v_T^2(\hat{k} - k_0) \xrightarrow{d} V^*,$$

*where* $V^*$ *is distributed as* $\operatorname{argmax} \{W(v) - |v|/2\}$.

The following corollaries address several leading special cases:

**Corollary 7.1.** *For an intercept shift only,*

$$\frac{(\mu_T'\Sigma_0^{-1}\mu_T)^2}{\mu_T'[\Sigma_0^{-1}C(1)\Omega_\varepsilon C(1)'\Sigma_0^{-1}]\mu_T}(\hat{k}-k_0)\xrightarrow{d}V^*. \tag{3.6}$$

**Corollary 7.2.** *If there is a shift in the trend only* ($A_1=0$, $\mu_1=0$, $B_1=0$, *but* $\gamma_1\neq 0$), *and* $\gamma_T=\gamma_1$, *we have*

$$\frac{(\gamma_T'\Sigma_0^{-1}\gamma_T)^2}{\gamma_T'[\Sigma_0^{-1}C(1)\Omega_\varepsilon C(1)'\Sigma_0^{-1}]\gamma_T}k_0^2(\hat{k}-k_0)\xrightarrow{d}V^*. \tag{3.7}$$

**Corollary 7.3.** *Suppose a shift occurs in the cointegrating coefficients only* ($A_1\neq 0$, *but* $\gamma_1=\mu_1=0$, $\beta_1=0$). *Let* $a_T=\text{vec}\,(A_1)$, *we have,*

$$\frac{[a_T'(X_{k_0}X_{k_0}'\otimes\Sigma_0^{-1})a_T]^2}{a_T'[X_{k_0}X_{k_0}'\otimes\Sigma_0^{-1}C(1)\Omega_\varepsilon C(1)'\Sigma_0^{-1}]a_T}(\hat{k}-k_0)\xrightarrow{d}V^*, \tag{3.8}$$

*where* $X_{k_0}$ *is equal to* $X_t$ *for* $t=k_0$.

*Remarks.*

(1) When $\xi_t$ are serially uncorrelated, so that $C(1)=I$ and $\Sigma_0=\Omega_\varepsilon$, then (3.6), (3.7) and (3.8) are simplified to

$$(\mu_T'\Sigma_0^{-1}\mu_T)(\hat{k}-k_0)\xrightarrow{d}V^*, \tag{3.9}$$

$$(\gamma_T'\Sigma_0^{-1}\gamma_T)k_0^2(\hat{k}-k_0)\xrightarrow{d}V^*, \tag{3.10}$$

$$a_T'(X_{k_0}X_{k_0}'\otimes\Sigma_0^{-1})a_T(\hat{k}-k_0)\xrightarrow{d}V^*, \tag{3.11}$$

respectively.

(2) We note that $a_T'(X_{k_0}X_{k_0}'\otimes\Sigma_0^{-1})a_T\equiv X_{k_0}'A_1'\Sigma_0^{-1}A_1X_{k_0}$. Thus the scaling factors in (3.8) and (3.11) may be written in this form.

(3) Corollary 7.3 can be modified to apply to various other special cases.

   (i) For a shift in the intercept and cointegrating coefficients, replace $a_T$ by $(a_T', \mu_T')'$ and replace $X_{k_0}$ by $(X_{k_0}', 1)'$.

   (ii) For a shift in cointegrating coefficients and the linear trend, replace $a_T$ by $(a_T', \gamma_T')'$ and replace $X_{k_0}$ by $(X_{k_0}', k_0)'$.

   (iii) For a shift in the intercept, cointegrating coefficients, and linear trend, replace $a_T$ by $(a_T', \gamma_T', \mu_T')'$ and replace $X_{k_0}$ by $(X_{k_0}', k_0, 1)'$.

   (iv) Results corresponding to a shift in the $I(0)$ regressor $w_t$ can be derived easily from Theorem 7. Because $H_1$ and $H_2$ are block diagonal, modifications take a simple form. For example, in Corollary 7.1, if $B_1\neq 0$ in addition to $\mu_1\neq 0$, then (3.6) still holds with the following modification: adding $b_T'[E(w_tw_t')\otimes\Sigma_0^{-1}]b_T$ to the numerator before taking the squared value, and adding $b_T'\Gamma b_T$ to the denominator, where $b_T=\text{vec}\,(B_1)$.

*Confidence Intervals*

The preceding results can be used to construct confidence intervals for the true break point. Let $c_{(1/2)\pi}$ denote the $1-\frac{1}{2}\pi$-th quantile of $V^*$. In general, a confidence interval for $k_0$ with asymptotic coverage of at least $100(1-\pi)\%$ is given by

$$I=[\hat{k}-[\Delta k]-1, \hat{k}+[\Delta k]+1]. \tag{3.12}$$

For a shift in the intercept only

$$\Delta k = c_{(1/2)\pi}\frac{\hat{\mu}'_T[\hat{\Sigma}_0^{-1}\hat{f}_\xi(0)\hat{\Sigma}_0^{-1}]\hat{\mu}_T}{(\hat{\mu}'_T\hat{\Sigma}_0^{-1}\hat{\mu}_T)^2}, \tag{3.13}$$

with $\hat{f}_\xi(0)$ being the estimate of $C(1)\Omega_\varepsilon C(1)'$, the spectral density of $\xi_t$ at zero.

For a shift in the trend only, $\Delta k$ becomes

$$\Delta k = c_{(1/2)\pi}\frac{1}{(\hat{k})^2}\frac{\hat{\gamma}'_T[\hat{\Sigma}_0^{-1}\hat{f}_\xi(0)\hat{\Sigma}_0^{-1}]\hat{\gamma}_T}{(\hat{\gamma}'_T\hat{\Sigma}_0^{-1}\hat{\gamma}_T)^2}. \tag{3.14}$$

For a shift in the cointegrating coefficients only [see Remarks (2)]

$$\Delta k = c_{(1/2)\pi}\frac{X'_{\hat{k}}\hat{A}'_1[\hat{\Sigma}_0^{-1}\hat{f}_\xi(0)\hat{\Sigma}_0^{-1}]\hat{A}_1 X_{\hat{k}}}{(X'_{\hat{k}}\hat{A}'_1\hat{\Sigma}_0^{-1}\hat{A}_1 X_{\hat{k}})^2}. \tag{3.15}$$

In the general case of Theorem 7,

$$\Delta k = c_{(1/2)\pi}\frac{(S\hat{\delta}_T)'S\hat{H}_1S'(S\hat{\delta}_T)}{[(S\hat{\delta}_T)'S\hat{H}_2S'(S\hat{\delta}_T)]^2}, \tag{3.16}$$

where $S\hat{\delta}_T$ is an estimate of $(S\delta_T)$ and $\hat{H}_1$ and $\hat{H}_2$ are estimates of $H_1$ and $H_2$, respectively. In particular, the upper-left block of $\hat{H}_1$ contains $\hat{\Phi}\otimes\hat{\Sigma}^{-1}\hat{f}_\xi(0)\hat{\Sigma}^{-1}$, with $\hat{\Phi}=(X_{\hat{k}}, \hat{k}, 1)(X_{\hat{k}}, \hat{k}, 1)'$; the lower-right block of $\hat{H}_1$ contains $\hat{\Gamma}$, an estimate of $\Gamma$. The matrix $\hat{H}_2$ is defined similarly.

Simpler expressions are available for serially uncorrelated $\xi_t$. In this case, confidence intervals should be based on (3.9)–(3.11).

*Extension to drifted $X_t$.*

Thus far, we assume the process $X_t$ is driftless. However, the same results hold even if some or all of the components of $X_t$ have drifts. Consider

$$Y_t = AX_t + \mu + Bw_t + (A_1X_t + \mu_1 + B_1w_t)d_t(k_0) + \xi_t, \tag{3.17a}$$

$$X_t = \Upsilon + X_{t-1} + \Xi_t. \tag{3.17b}$$

We assume $\Upsilon \neq 0$, so at least one component has a nonzero drift. In model (3.17), the linear trend is not included.[3] If more than one component of $X_t$ has a drift, then asymptotic multicollinearity exists. Write $\Upsilon=(\Upsilon_1, \ldots, \Upsilon_q)'$, where $q=n-r$. Without loss of generality, assume $\Upsilon_q \neq 0$. The final result does not depend on which component has a nonzero drift, and there is no need to estimate the drift coefficients. Assume that $w_t=(\Delta X_{t+l}, \ldots, \Delta X_{t-l})$. Let $A_1=(\Gamma_1, \ldots, \Gamma_q)$ so $\Gamma_i$ is the $i$-th column of $A_1$. We can rewrite (3.17a) as follows (see, e.g., Hamilton (1994), p. 627)

$$Y_t = A^*X_t^* + \gamma X_{qt} + \mu^* + Bw_t^* + d_t(k_0)(A_1^*X_t^* + \gamma_1 X_{qt} + \mu_1^* + B_1w_t) + \xi_t, \tag{3.18}$$

3. A linear trend can be added without affecting the analysis.

where $A_1^*$ is equal to $A_1$ with the last column deleted, $X_{it}^* = X_{it} - \Upsilon_i/\Upsilon_q X_{qt}$ $(i = 1, \ldots, q-1)$,

$$\gamma_1 = \Gamma_q + \Gamma_1 \frac{\Upsilon_1}{\Upsilon_q} + \cdots + \Gamma_{q-1} \frac{\Upsilon_{q-1}}{\Upsilon_q},$$

$$\mu_1^* = \mu_1 + B_1(1 \otimes \Upsilon),$$

$$w_t^* = (\Delta X'_{t+l} - \Upsilon', \ldots, \Delta X'_{t-l} - \Upsilon')'.$$

Note that $X_t^*$ is driftless, and $w_t^*$ has mean zero.[4] Thus the transformed model satisfies all the conditions imposed to (3.2), except the linear trend is replaced by $X_{qt}$. But $X_{qt}$ has the same behaviour as a linear trend, as far as asymptotic behaviour is concerned. Since the transformation is rank preserving, the pseudo likelihood function based on (3.17) is identical to that based on (3.18). This implies that the estimated break point is identical irrespective of which equation is used. Thus we can apply the earlier result for the driftless model to the transformed model. The difference is that we must assume $A_1$ shrinks faster than the driftless model, because $X_t$ behaves like a time trend $t$. Assume

$$A_1 = T^{-1} A_0 v_T, \quad \text{and} \quad A_0 \Upsilon \neq 0. \tag{3.19}$$

Under this assumption, Corollary 7.3 still holds. It is not necessary to assume all coefficients decrease at the rate $T^{-1} v_T$. Let $X_t = (X_t'^{(1)}, X_t'^{(2)})'$, where $X_t^{(1)}(h \times 1)$ is driftless, and $X_t^{(2)}$ $((q-h) \times 1)$ has a drift for each of its components. Let $\Upsilon = (0, \Upsilon'^{(2)})'$ and $A_1 = (A_T^{(1)}, A_T^{(2)})$, partitioned conformably. Assume for some fixed matrices $A_0^{(1)}$ and $A_0^{(2)}$,

$$A_T^{(1)} = T^{-1/2} A_0^{(1)} v_T, \; A_T^{(2)} = T^{-1} A_0^{(2)} v_T, \text{ either } A_0^{(1)} \neq 0, \text{ or } A_0^{(2)} \Upsilon^{(2)} \neq 0, \text{ or both.} \tag{3.20}$$

Then Corollary 7.3 still holds. These results are stated in the following theorem.

**Theorem 8.** *For a shift in the intercept only* (*equation* (3.17a) *with* $A_1 = 0$ *and* $B_1 = 0$), *Corollary* 7.1 *still holds. For a shift in the cointegrating coefficients only* (*equation* (3.17a) *with* $\mu_1 = 0$ *and* $B_1 = 0$) *together with* (3.19) *or* (3.20), *Corollary* 7.3 *still holds.*

It should be emphasized that Theorem 8 is stated in terms of the original model (3.17), rather than the transformed model. This is useful because the transformed model is of theoretical interest only, and is not directly estimable.

In summary, if one is only interested in a shift in the intercept, or in the cointegrating coefficients or both, the break point estimator has the same distribution (more specifically, the scaling factor of $\hat{k} - k_0$ has the same expression) regardless of the existence or absence of drifts. The confidence intervals for the true break point have the same form. The estimated regression parameters, however, will have limiting distributions different from those in the absence of drifts.

## 4. MONTE CARLO ANALYSIS

This section presents the results of a Monte Carlo study of break date statistics in two models. The first, which is motivated by our empirical applications and based on the theory developed in Section 2, is an $I(0)$ model with a break in the intercept only. The second is a cointegrated specification with a break in the intercept and the cointegrating coefficient. This latter model is less relevant for the subsequent empirical work, but is of

4. If $w_t$ does not contain lags or leads of $\Delta X_t$, then $w_t^* = w_t$ and $\mu_1^* = \mu_1$.

methodological interest because it studies the finite sample performance of statistics proposed in Section 3. Because Monte Carlo evidence on the test statistics (2.8)–(2.9) has been documented elsewhere, attention is restricted here to coverage rates of the asymptotic confidence intervals for $k_0$.[5]

### 4.1. *Breaks in intercepts, I*(0) *models*

For considering a change in the intercept, the data were generated according to the following Gaussian autoregression

$$y_t = (\lambda \iota_n) d_t([T\delta_0]) + (\beta I_n) y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \text{ i.i.d } N(0, \Sigma_\varepsilon) \tag{4.1}$$

where $\iota_n$ is an $n$-vector of 1's and $\varepsilon_t$ is $n \times 1$. Both univariate and trivariate models were studied. In the univariate case, $\Sigma_\varepsilon = 1$. In the trivariate case, $\Sigma_\varepsilon$ was set to have 1 on the diagonal and $\rho$ off the diagonal.

Monte Carlo coverage probabilities for central 80% and 90% asymptotic confidence intervals for $k_0$, constructed using (2.20), are reported in Table 1 for univariate models and Table 2 for trivariate models,[6] with $T = 100$ and $T = 400$. These tables also report summary statistics (the median and the range between the 5% and 95% points) for the Gaussian MLE of $k_0$.

The results for the Monte Carlo distributions of $\hat{k}$ suggest five conclusions. First, the precision of the MLE of $k_0$ depends strongly on the true value of $\lambda$. This is consistent with the theoretical rate discussed in Section 2, that is, $\hat{k} - k_0 = O_p((\lambda' \Sigma_\varepsilon^{-1} \lambda)^{-1})$. For example, for $n = 1$ and $\lambda = 0{\cdot}75$, the 90% range is 33 for $T = 100$; the median of the Monte Carlo draws is $\hat{k} = 50$ (the true value) and 90% of the estimates approximately fall in $50 \pm 16$. Second, there does not appear to be substantial median bias in the Monte Carlo results, either for $\delta_0 = 0{\cdot}5$ or in the one case in which $\delta_0 = 0{\cdot}25$. Third, the result in Corollary 4.1 suggests that, for fixed parameters, increasing the sample will not affect the precision of the estimator, measured in terms of $k$; this prediction is confirmed by noting that the 90% ranges change little, or not at all, when the sample size is quadrupled. Fourth, the precision of $\hat{k}$ does not appear to be strongly affected by $\beta$. Fifth, as discussed in Section 2, moving from a univariate to a trivariate system should reduce the range of $\hat{k}$ by one-third, at least if $\Sigma_\varepsilon$ is diagonal and $\lambda_i / (\Sigma_\varepsilon)_{ii}$ are the same for each equation. This prediction is borne out; for example, for $\lambda = 0{\cdot}75$, $\beta = 0$, and $T = 100$, the 90% range is 33 for $n = 1$ and is 11 for $n = 3$ ($p = 0$ case). These observations, along with the theoretical results in Section 2, support a conclusion with important practical implications. For a fixed $\lambda$, more precise estimates of $k_0$ cannot be obtained by acquiring longer data sets; one must instead use additional series that break at the same date. We explore these conclusions in the empirical examples of Sections 5 and 6.

The results suggest that, in general, the asymptotic confidence intervals tend to have coverage rates less than their confidence coefficient; that is, the confidence intervals are too tight. This effect is more pronounced for small than large values of $\lambda$, and for large than small values of $\beta$ (for large values of $\beta$, the least squares estimate of $\beta$ tends to be downward biased, resulting in the break point being inaccurately estimated). For example, the coverage rate of the 90% interval in the univariate system with $\lambda = 1{\cdot}5$, $\beta = 0$, and $p =$

5. An earlier version of this paper contained results assessing the size and power of the various test statistics for a change in the mean in univariate and trivariate models. Also see Andrews (1993) and Andrews, Lee, and Ploberger (1996).

6. Our reporting of three decimal places does not reflect this level of accuracy for the exact distributions but rather estimates from the limited number of Monte Carlo simulations.

TABLE 1

*Monte Carlo results: finite sample coverage rates of asymptotic confidence intervals for the break date $k_0$, univariate case ($n=1$).*

| | | | | $T=100$ | | | | $T=400$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Conf. | Intervals | Estimator | | Conf. | Intervals | Estimator | |
| $\lambda$ | $\delta_0$ | $\beta$ | $p$ | 80% | 90% | Med | 90% range | 80% | 90% | Med | 90% range |
| 0·75 | 0·50 | 0·0 | 0 | 0·744 | 0·845 | 50 | 33 | 0·811 | 0·895 | 200 | 30 |
| 0·75 | 0·50 | 0·0 | 1 | 0·728 | 0·829 | 50 | 33 | 0·806 | 0·891 | 200 | 30 |
| 0·75 | 0·50 | 0·0 | 4 | 0·682 | 0·788 | 50 | 33 | 0·798 | 0·882 | 200 | 30 |
| 0·75 | 0·50 | 0·0 | BIC | 0·739 | 0·839 | 50 | 33 | 0·811 | 0·895 | 200 | 30 |
| 0·75 | 0·50 | 0·4 | 0 | 0·585 | 0·680 | 50 | 33 | 0·614 | 0·726 | 201 | 28 |
| 0·75 | 0·50 | 0·4 | 1 | 0·706 | 0·791 | 50 | 33 | 0·790 | 0·876 | 201 | 28 |
| 0·75 | 0·50 | 0·4 | 4 | 0·669 | 0·763 | 50 | 33 | 0·777 | 0·866 | 201 | 28 |
| 0·75 | 0·50 | 0·4 | BIC | 0·698 | 0·780 | 50 | 33 | 0·790 | 0·876 | 201 | 28 |
| 0·75 | 0·50 | 0·8 | 0 | 0·254 | 0·314 | 53 | 30 | 0·267 | 0·357 | 203 | 27 |
| 0·75 | 0·50 | 0·8 | 1 | 0·626 | 0·719 | 53 | 30 | 0·784 | 0·869 | 203 | 27 |
| 0·75 | 0·50 | 0·8 | 4 | 0·588 | 0·688 | 53 | 30 | 0·770 | 0·855 | 203 | 27 |
| 0·75 | 0·50 | 0·8 | BIC | 0·623 | 0·717 | 53 | 30 | 0·783 | 0·869 | 203 | 27 |
| 0·25 | 0·50 | 0·0 | 0 | 0·668 | 0·810 | 50 | 67 | 0·668 | 0·791 | 201 | 225 |
| 0·50 | 0·50 | 0·0 | 0 | 0·698 | 0·806 | 50 | 57 | 0·764 | 0·858 | 200 | 75 |
| 1·00 | 0·50 | 0·0 | 0 | 0·796 | 0·876 | 50 | 19 | 0·821 | 0·896 | 200 | 18 |
| 1·50 | 0·50 | 0·0 | 0 | 0·872 | 0·927 | 50 | 7 | 0·880 | 0·939 | 200 | 6 |
| 2·00 | 0·50 | 0·0 | 0 | 0·913 | 0·947 | 50 | 4 | 0·927 | 0·950 | 200 | 4 |
| 0·75 | 0·25 | 0·0 | 0 | 0·775 | 0·873 | 25 | 41 | 0·801 | 0·890 | 100 | 30 |

*Notes*: Coverage rates for asymptotic confidence intervals are respectively reported under the "80%" and "90%" columns, for the applicable sample size of either $T=100$ or $T=400$, where $T$ denotes the number of observations in each regression. The median and 90% range (the range between the 5% and 95% points) of the distribution of the estimator of $k$, $\hat{k}$, are reported under the "Med" and "90% range" columns. Point estimates $\hat{k}$ for the break date were computed by maximum likelihood estimation for $k=k_*+1, \ldots, T-k_*$, where $k_*=[T\delta_*]$, $\delta_*=0\cdot15$. The confidence intervals are computed as described in the text. "$p$" denotes the autoregressive order; "BIC" means that the lag length was selected by minimizing the BIC over $p=0, 1, \ldots, 6$ (univariate models) or $p=0, 1, 2, 3$ (trivariate models). 4000 replications were performed for each experiment.

0 is 93%; for $\lambda=0\cdot75$, $\beta=0\cdot8$, and $p=1$, the coverage rate is only 72%. The choice of lag length $p$ has at most a moderate effect on the coverage rates, except in the case in which the regression is misspecified (that is, $p=0$ when $\beta\neq0$), in which case the coverage rates are very low. In particular, the use of the BIC choice of $p$ rather than the true lag length has little effect on the coverage rates. The performance of all the univariate intervals (except for the intervals constructed using a misspecified regression) improves substantially when $T$ is quadrupled. The confidence intervals perform less well in the trivariate model than in the univariate model, particularly when $\beta$ is large. Still, for all values of $\beta$ considered, the coverage rates of the 90% intervals for the trivariate models exceed 78% (except in the misspecified models) when the sample size is increased to 400.

### 4.2. *Breaks in $I(1)$ models*

The data are generated according to (3.17), with $w_t$ omitted, $\mu=0$, $A=$ a vector of ones, $\sigma_\Xi=0\cdot1$ and $\Sigma_\xi=I_r$, where $I_r$ is an $r$-dimensional identity matrix, and $\xi_t$ and $\Xi_t$ are independent Gaussian errors, with 4000 replications. We consider values of $\mu_1\in(0, 1\cdot0, 1\cdot5, 2\cdot0)$, $A_1\in(0, 0\cdot5, 1\cdot0, 1\cdot5)$, $\tau_0\in(0\cdot25, 0\cdot5, 0\cdot75)$ and $\Upsilon\in(0, 0\cdot1, 0\cdot2, 0\cdot4)$. Monte Carlo coverage probabilities for central 80% and 90% asymptotic confidence intervals for $k_0$, constructed using (3.16), are reported in Table 3 for univariate models ($n=1$) and trivariate ($n=3$) models, with $T=100$ and $T=400$. We focus on the confidence intervals in (3.16)

TABLE 2

*Monte Carlo results: finite sample coverage rates of asymptotic confidence intervals for the break date $k_0$, trivariate case ($n=3$)*

| | | | | | $T=100$ | | | | $T=400$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Conf. Intervals | | Estimator | | Conf. Intervals | | Estimator | |
| $\lambda$ | $\delta_0$ | $\beta$ | $\rho$ | $p$ | 80% | 90% | Med | 90% range | 80% | 90% | Med | 90% range |
| 0·50 | 0·50 | 0·0 | 0·0 | 0 | 0·692 | 0·793 | 50 | 31 | 0·797 | 0·886 | 200 | 23 |
| 0·50 | 0·50 | 0·0 | 0·0 | 1 | 0·658 | 0·757 | 50 | 31 | 0·788 | 0·881 | 200 | 23 |
| 0·50 | 0·50 | 0·0 | 0·0 | BIC | 0·685 | 0·785 | 50 | 31 | 0·797 | 0·887 | 200 | 23 |
| 0·50 | 0·50 | 0·4 | 0·0 | 0 | 0·555 | 0·637 | 50 | 30 | 0·625 | 0·722 | 200 | 25 |
| 0·50 | 0·50 | 0·4 | 0·0 | 1 | 0·625 | 0·719 | 50 | 30 | 0·776 | 0·865 | 200 | 25 |
| 0·50 | 0·50 | 0·4 | 0·0 | BIC | 0·613 | 0·707 | 50 | 30 | 0·776 | 0·865 | 200 | 25 |
| 0·50 | 0·50 | 0·8 | 0·0 | 0 | 0·237 | 0·262 | 53 | 28 | 0·247 | 0·314 | 203 | 21 |
| 0·50 | 0·50 | 0·8 | 0·0 | 1 | 0·446 | 0·525 | 53 | 28 | 0·722 | 0·821 | 203 | 21 |
| 0·50 | 0·50 | 0·8 | 0·0 | BIC | 0·436 | 0·515 | 53 | 28 | 0·722 | 0·820 | 203 | 21 |
| 0·50 | 0·50 | 0·4 | 0·5 | 0 | 0·421 | 0·527 | 51 | 52 | 0·563 | 0·673 | 200 | 51 |
| 0·50 | 0·50 | 0·4 | 0·5 | 1 | 0·512 | 0·619 | 51 | 52 | 0·726 | 0·827 | 200 | 51 |
| 0·50 | 0·50 | 0·4 | 0·5 | BIC | 0·501 | 0·609 | 51 | 52 | 0·725 | 0·825 | 200 | 51 |
| 0·25 | 0·50 | 0·0 | 0·0 | 0 | 0·483 | 0·611 | 50 | 64 | 0·671 | 0·779 | 200 | 127 |
| 0·50 | 0·50 | 0·0 | 0·0 | 0 | 0·705 | 0·791 | 50 | 30 | 0·802 | 0·891 | 200 | 23 |
| 0·75 | 0·50 | 0·0 | 0·0 | 0 | 0·815 | 0·890 | 50 | 11 | 0·856 | 0·917 | 200 | 10 |
| 1·00 | 0·50 | 0·0 | 0·0 | 0 | 0·883 | 0·930 | 50 | 6 | 0·898 | 0·944 | 200 | 6 |
| 1·50 | 0·50 | 0·0 | 0·0 | 0 | 0·948 | 0·976 | 50 | 2 | 0·949 | 0·983 | 200 | 2 |

*Notes*: 4000 replications were performed for each experiment. See the notes to Table 1.

because they are the most general in that they allow for a change in any or all of the coefficients in the regression.

The main conclusions from Tables 1 and 2 hold in Table 3. In addition, in this DGP, when the data generating process contains only a shift in the mean, the precision of the confidence interval deteriorates when the break occurs near the ends of the sample; this sensitivity is less apparent for the trivariate case (as seen by comparing lines 1, 4 and 5 of Table 3). Also, when the break occurs in the regression coefficients, $A_1$, the theoretical predictions of (3.11) suggest that the break date should be estimated more precisely the later the break occurs in sample. This is because the magnitude of the regressors is larger when $k_0$ is larger. This prediction is confirmed in the simulations; this is seen by comparing lines 12 and 13 or lines 18 and 19 of Table 3.

The results suggest that, for breaks of small magnitude, the asymptotic confidence intervals tend to have coverage rates less than their confidence coefficient; that is, the confidence intervals are too tight. For example, the coverage rate of the 90% interval in the univariate system with $\mu_1=1{\cdot}5$, $A_1=0$, and $\Upsilon=0$ is 89%; for $\mu_1=1{\cdot}0$, the coverage rate is only 78%. However the performance of all the univariate intervals improves substantially when $T$ is quadrupled.

In summary, these results indicate that the asymptotic theory provides a good basis for the construction of break date confidence intervals in these designs, when the break is of moderate size.

## 5. EUROPEAN OUTPUT GROWTH

Although a slowdown in the growth of output in European economies is widely acknowledged, formal dating of this slowdown has been hampered by the lack of appropriate

TABLE 3

*Monte Carlo results: finite sample coverage rates of asymptotic confidence intervals for the break date $k_0$, univariate case ($n=1$) and trivariate case ($n=3$) with $I(1)$ Regressors.*

| | | | | $n=1$ | | | | $n=3$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Conf. | Intervals | Estimator | | Conf. | Intervals | Estimator | |
| $\mu_1$ | $A_1$ | $\tau_0$ | $\Upsilon$ | 80% | 90% | Med | 90% range | 80% | 90% | Med | 90% range |
| $T=100$ | | | | | | | | | | | |
| 1·5 | 0·0 | 0·50 | 0·00 | 0·831 | 0·887 | 50 | 9 | 0·929 | 0·939 | 50 | 2 |
| 1·0 | 0·0 | 0·50 | 0·00 | 0·710 | 0·780 | 50 | 37 | 0·770 | 0·804 | 50 | 8 |
| 2·0 | 0·0 | 0·50 | 0·00 | 0·901 | 0·938 | 50 | 4 | 0·983 | 0·985 | 50 | 1 |
| 1·5 | 0·0 | 0·25 | 0·00 | 0·803 | 0·854 | 25 | 17 | 0·909 | 0·919 | 25 | 3 |
| 1·5 | 0·0 | 0·75 | 0·00 | 0·810 | 0·860 | 75 | 17 | 0·912 | 0·921 | 75 | 3 |
| 1·5 | 0·0 | 0·50 | 0·01 | 0·833 | 0·883 | 50 | 10 | 0·932 | 0·944 | 50 | 2 |
| 1·5 | 0·0 | 0·50 | 0·02 | 0·819 | 0·874 | 50 | 11 | 0·928 | 0·936 | 50 | 2 |
| 1·5 | 0·0 | 0·50 | 0·04 | 0·830 | 0·879 | 50 | 11 | 0·930 | 0·936 | 50 | 2 |
| 0·0 | 1·0 | 0·50 | 0·00 | 0·824 | 0·854 | 50 | 20 | 0·847 | 0·867 | 50 | 9 |
| 0·0 | 0·5 | 0·50 | 0·00 | 0·730 | 0·787 | 50 | 50 | 0·723 | 0·754 | 50 | 33 |
| 0·0 | 1·5 | 0·50 | 0·00 | 0·877 | 0·898 | 50 | 10 | 0·890 | 0·901 | 50 | 4 |
| 0·0 | 1·0 | 0·25 | 0·00 | 0·813 | 0·850 | 25 | 32 | 0·813 | 0·842 | 25 | 13 |
| 0·0 | 1·0 | 0·75 | 0·00 | 0·833 | 0·861 | 75 | 21 | 0·857 | 0·873 | 75 | 8 |
| 0·0 | 1·0 | 0·50 | 0·01 | 0·841 | 0·868 | 50 | 19 | 0·862 | 0·877 | 50 | 8 |
| 0·0 | 1·0 | 0·50 | 0·02 | 0·862 | 0·885 | 50 | 17 | 0·872 | 0·885 | 50 | 7 |
| 0·0 | 1·0 | 0·50 | 0·04 | 0·915 | 0·929 | 50 | 8 | 0·925 | 0·937 | 50 | 3 |
| 1·0 | 1·0 | 0·50 | 0·00 | 0·838 | 0·863 | 50 | 18 | 0·865 | 0·883 | 50 | 8 |
| 1·0 | 1·0 | 0·25 | 0·00 | 0·820 | 0·852 | 25 | 30 | 0·833 | 0·855 | 25 | 11 |
| 1·0 | 1·0 | 0·75 | 0·00 | 0·842 | 0·868 | 75 | 18 | 0·866 | 0·878 | 75 | 8 |
| 1·0 | 1·0 | 0·50 | 0·01 | 0·859 | 0·885 | 50 | 16 | 0·886 | 0·898 | 50 | 7 |
| 1·0 | 1·0 | 0·50 | 0·02 | 0·883 | 0·903 | 50 | 12 | 0·903 | 0·914 | 50 | 6 |
| 1·0 | 1·0 | 0·50 | 0·04 | 0·939 | 0·951 | 50 | 4 | 0·957 | 0·962 | 50 | 0 |
| $T=400$ | | | | | | | | | | | |
| 1·5 | 0·0 | 0·5 | 0·00 | 0·867 | 0·932 | 200 | 7 | 0·949 | 0·956 | 200 | 2 |
| 1·5 | 0·0 | 0·5 | 0·04 | 0·863 | 0·922 | 200 | 8 | 0·952 | 0·955 | 200 | 2 |
| 0·0 | 1·0 | 0·5 | 0·00 | 0·906 | 0·919 | 200 | 8 | 0·916 | 0·929 | 200 | 3 |
| 0·0 | 1·0 | 0·5 | 0·04 | 0·984 | 0·987 | 200 | 0 | 0·990 | 0·990 | 200 | 0 |
| 1·0 | 1·0 | 0·5 | 0·00 | 0·906 | 0·923 | 200 | 8 | 0·926 | 0·935 | 200 | 4 |
| 1·0 | 1·0 | 0·5 | 0·04 | 0·987 | 0·988 | 200 | 0 | 0·990 | 0·991 | 200 | 0 |

*Notes*: 4000 replications were performed for each experiment. See the notes to Table 1.

statistical techniques. We therefore turn to the task of dating this slowdown. The starting point for this investigation is the observation by Banerjee, Lumsdaine and Stock (1992) that output in France, Germany, and Italy each appeared to be difference stationary, but that there appeared to be a break in the mean growth rate for each country during the sample. Their analysis was strictly univariate, and the results of the previous sections show that there can be substantial gains from using multivariate inference about the break dates. Specifically, we are now in a position to address three empirical questions. First, what are confidence intervals for the break date when the series are treated individually? Second, is there evidence that these breaks occurred at the same time? Third, if so, what are the interval estimates of the break date, when the date is modelled as common across these three countries?

We use Banerjee, Lumsdaine and Stock's (1992) data for comparability to their study. The three European series are the logarithms of quarterly GDP for France and Italy and GNP for Germany. We also examine the logarithm of quarterly GDP for the U.S. Because the data are available over different periods, the system results examine the joint behaviour of output over only a short common period, 1962:1 to 1982:4. The data and data sources are described in more detail in Banerjee, Lumsdaine, and Stock (1992).

Banerjee, Lumsdaine, and Stock (1992) tested the null hypothesis that each of these series had a unit root, against the alternative that the series was stationary around a linear time trend, possibly with a break in the time trend at an unknown date. The univariate analysis of these European output data provided no evidence against the unit root null hypothesis; based on these earlier findings, we proceed under the assumption that each series is $I(1)$, possibly with a change in drift, so that each differenced series is modelled as having the univariate stationary autoregressive representation (2.1), where $y_t$ is the growth rate of output, $X_t$ is omitted (there are no exogenous variables), and $\beta_j=0$, $j=1,\ldots,p$; the break term corresponds to a shift in the mean growth rate of output. The series are modelled as jointly having the stationary autoregressive representation (2.1), where $y_t$ is interpreted as the vector of growth rates of output of the various countries and $X_t$ is omitted.

TABLE 4

*Empirical results: European output*

| Country | Sample | Sup-$W$-15% | Exp-$W$-15% | $\hat{k}$ | 90% Conf. Int. |
|---|---|---|---|---|---|
| *A. Univariate* | | | | | |
| France | 64:4–89:2 | 23·68 | 9·15 | 74:2 | (72:4, 75:4) |
| | | (0·00) | (0·00) | | |
| Germany | 51:4–89:2 | 21·68 | 8·28 | 61:2 | (59:1, 63:3) |
| | | (0·00) | (0·00) | | |
| Italy | 53:4–82:4 | 10·30 | 2·83 | 74:3 | (70:2, 78:4) |
| | | (0·03) | (0·03) | | |
| U.S. | 48:4–89:2 | 1·42 | 0·25 | 68:4 | (<47:1, >89:2) |
| | | (0·91) | (0·71) | | |
| *B. Bivariate and Trivariate VAR Systems* | | | | | |
| F, G | 64:2–89:2 | 26·00 | 10·14 | 75:1 | (73:3, 76:3) |
| | | (0·00) | (0·00) | | |
| F, I | 64:2–82:4 | 17·97 | 6·24 | 73:4 | (72:1, 75:3) |
| | | (0·00) | (0·00) | | |
| G, I | 53:2–82:4 | 14·98 | 5·32 | 74:1 | (71:1, 77:1) |
| | | (0·02) | (0·01) | | |
| F, G, I | 64:1–82:4 | 19·43 | 6·98 | 73:4 | (72:2, 75:2) |
| | | (0·01) | (0·00) | | |

*Notes*: $p$-values, computed using the asymptotic distributions of the relevant test statistic, are given in parentheses; 0·00 denotes a $p$-value less than 0·005. The sample period denotes the period over which the regressions were run; earlier observations were used for initial conditions. All lag lengths were selected using the BIC, with a minimum of one lag and a maximum of 6, 4, and 3 lags, respectively for the 1-, 2-, and 3-variable systems. In each model the BIC criterion picked $p=1$.

Change-point statistics for European and U.S. output are presented in Table 4. For France and Germany, treated as univariate series, both of the test statistics rejects at the 1% level; for Italy, both reject at the 5% level. The point estimates of the break date are in 1974 for France and Italy, although for Italy the estimate is imprecise. For Germany, the 90% confidence interval for the break is (59:1, 63:3). In contrast, for U.S. output the hypothesis of a constant mean growth rate cannot be rejected at the 10% level using any of the tests. The 90% confidence interval for the U.S. is so wide that it contains the entire sample period. Because of the insignificance of the break statistics for the U.S., we postpone further analysis of the U.S. data until the next section.

The univariate evidence is consistent with France and Italy having a contemporaneous break and with there being no identifiable break in the U.S. It is less clear whether Germany has a break at the same time as France and Italy; although the confidence intervals for the break date do not overlap, the $F$-stastic for Germany is large throughout the mid-1970s. Indeed, this $F$-statistic takes on a local maximum of 17·52 in 1970:2, which exceeds the 5% critical value for the sup $W$ statistic. Although the results are therefore somewhat ambiguous, we interpret this as broadly supporting an exploration of multivariate models including these three countries with a common break date.

Results of the multivariate analysis are reported in panel B of Table 4. Consider first the France–Italy system, for which the univariate evidence is most consistent with a single common break date. The test statistics reject the hypothesis of no break in the mean growth rate against the alternative of a break in the mean at a common break date; the 90% confidence interval of 1972:1 to 1975:3 is similar to that for France alone and tighter than the interval for Italy. The other bivariate systems also reject the null of no break against the common-date alternative, providing support for proceeding to construct interval estimates for a common break date including Germany in the system. The tests reject the null of no break against the common-date alternative in the trivariate system and, consistent with econometric theory and Monte Carlo results, inference is most precise in this case: the break is estimated at 1973:4 with a 90% confidence interval of 1972:2 to 1975:2.

This multivariate analysis points to a slowdown in European output which occurred approximately simultaneously in France and Italy and, arguably, in Germany as well. When data on output growth in all three countries are used at once, the break date is sharply estimated to be in the early 1970s. Of course, this dating coincides with conventional wisdom; the contribution here is that this date can now be associated with the formal measure of uncertainty provided by a tight 90% confidence interval spanning slightly more than three years.

## 6. U.S. OUTPUT, CONSUMPTION, AND INVESTMENT

Because the univariate results for U.S. output do not provide sharp evidence either in favour of or opposing the one-break model, in this section we extend the investigation to include multiple time series, in particular, consumption, income and investment. One rationale for using this trivariate system is that a range of models of long-run stochastic growth suggest that a permanent shift in the average growth rate of productivity will result in permanent shifts in the mean growth rates in each of the series. This arises from the three series sharing a common stochastic trend (productivity) and thus being cointegrated. Even though the shift in the mean growth rate might be statistically insignificant using just output, the break might be more readily detected and estimated if consumption and investment are used as well.

### 6.1. *Theoretical preliminaries*

Theoretical arguments for the cointegration of these series are made in King, Plosser, and Rebelo (1988). In brief, in their model, a representative firm produces according to a Cobb–Douglas production function with constant returns to scale and total factor productivity follows a random walk with drift. Then the logarithm of output, consumption, and investment each inherits the stochastic trend in productivity; each series is $I(1)$, and the logarithm of the consumption/output and investment/output ratios are stationary. As discussed in King, Plosser, Stock, and Watson (1991), this system has the common trends representation

$$Y_t = \phi + D\tau_t + u_t, \tag{6.1}$$

where $Y_t$ is the vector of the logarithms of output, investment, and consumption, and $\phi$ and $D$ are $3\times 1$ vectors, $\tau_t$ is the common stochastic trend, and $u_t$ is an additional $I(0)$ disturbance. The common trend $\tau_t$ is a scalar, which in the King–Plosser–Rebelo (1988) model is the logarithm of total factor productivity. The vector $D$ is defined (up to scale) by $\alpha' D=0$, where $\alpha$ is the matrix of cointegrating vectors ($\alpha=(\alpha_1, \alpha_2)$); in the King–Plosser–Rebelo (1988) model, $\alpha_1=(1, -1, 0)$, $\alpha_2=(1, 0, -1)$, and $D=(1, 1, 1)'$. In general, for $Y_t$ an $n$-vector with $r$ cointegrating vectors, $D$ is $n\times(n-r)$ and is defined by $\alpha' D=0$ up to postmultiplication by a nonsingular $(n-r)\times(n-r)$ matrix. The $I(0)$ disturbance $u_t$ represents measurement error and/or additional unmodelled short-run dynamics of the system.[7]

We extend the King–Plosser–Rebelo (1988) model to incorporate the possibility of a one-time shift in the average growth rate of productivity from $\tilde{\mu}$ for $t<k_0$ to $\tilde{\mu}+\tilde{\lambda}$ for $t\geq k_0$. Specifically, $\tau_t$ is assumed to evolve according to

$$\tau_t=\tilde{\mu}+\tilde{\lambda} d_t(k_0)+\tau_{t-1}+\eta_t, \tag{6.2}$$

where $\eta_t$ is a martingale difference sequence. This results in a change in the growth rates in each of the series at date $k_0$.

To apply the tools developed in Section 2, (6.1) and (6.2) must be re-expressed in the form (2.1). The univariate representation for the $i$-th series is obtained by combining (6.1) and (6.2) in first differences, which yields $\Delta Y_{it}=D_i\tilde{\mu}+D_i\tilde{\lambda} d_t(k_0)+v_{it}$, where $v_{it}=D_i\eta_t+\Delta u_{it}$. We further suppose that the $I(0)$ disturbance $v_{it}$ follows an autoregressive representation, $v_{it}=(1-a_i(L)L)^{-1}\zeta_{it}$, where $\zeta_{it}$ is serially uncorrelated. Then $\Delta Y_{it}$ has the univariate representation

$$\Delta Y_{it}=\mu_i+\lambda_i d_t(k_0)+a_i(L)\Delta Y_{it-1}+d_{it}^*(k_0)D_i\tilde{\lambda}+\zeta_{it}, \tag{6.3}$$

where $\mu_i=(1-a_i(1))D_i\tilde{\mu}$, $\lambda_i=(1-a_i(1))D_i\tilde{\lambda}$, and $d_{it}^*(k)=a_i^*(L)\Delta d_t(k)$, where $a_{ij}^*=-\sum_{m=j}^{\infty} a_{im}$. It is assumed that $\sum_{j=1}^{\infty}\sum_{m=j}^{\infty}|a_{im}|<\infty$ for each $i$. Under this assumption, the term $d_{it}^*(k_0)D_i\tilde{\lambda}$ has asymptotically negligible effects on the regression estimates; we omit it and thus have

$$\Delta Y_{it}=\mu_i+\lambda_i d_t(k_0)+a_i(L)\Delta Y_{it-1}+\zeta_{it}. \tag{6.3$'$}$$

In the empirical application, $a_i(L)$ is assumed to have finite order ($p$).

Just as there are multiple representations of cointegrated systems (*cf.* Engle and Granger (1987)), there are various ways to rewrite (6.1) and (6.2) in the form (2.1). One such representation is the vector error correction model (VECM). Following a standard derivation of the VECM model (*cf.* Watson (1994)), modified for the break model (6.2), we obtain

$$\Delta Y_t=\mu+\lambda d_t(k_0)+A(L)\Delta Y_{t-1}+\gamma\alpha' Y_{t-1}+\varepsilon_t, \tag{6.4}$$

which is of the form (2.1) with $X_t=\alpha' Y_{t-1}$. (To obtain (6.4) we have dropped the transient term which is the multivariate counterpart of the univariate term dropped to obtain (6.3$'$) from (6.3).) Thus (6.4) is a VECM modified to admit the possibility of a break in the

7. The King–Plosser–Rebelo model is not the only model that produces a representation such as (2.3) for the major aggregates. For example, Sargent (1989) derives (2.3) as the reduced form of a linear-quadratic optimization model with an investment accelerator. In Sargent's (1989) model, $Y_t$ is income, consumption, and investment, $u_t$ is modelled explicitly as measurement error, and $\tau_t$ is a smooth average of past productivity innovations. In Sargent's model $\tau_t$ is stationary with an autoregressive root close to but less than one. In general $u_t$ and $\Delta\tau_t$ can be correlated. In Sargent's model, this correlation arises from the data construction operations of the economic statistics agencies.

growth rate of the common stochastic trends. Note that, in this representation, the break enters only through a shift in the intercept term.

An alternative representation is obtained by noting that because $\alpha' D = 0$, (6.1) and (6.2) can be rewritten in terms of $\alpha' Y_t$ and $\Delta D' Y_t$:

$$\alpha' Y_t = \alpha'\phi + v_t^1, \tag{6.5a}$$

$$\Delta D' Y_t = D'D\tilde{\mu} + D'D\tilde{\lambda} d_t(k_0) + v_t^2, \tag{6.5b}$$

where $v_t^1 = \alpha' u_t$ and $v_t^2 = D'D\eta_t + \Delta D' u_t$. The form used here adopts a triangular decomposition of $v_t$ as in Phillips (1991) and Stock and Watson (1993). Let $v_t^2 = P(v_t^2 | v_t^1, v_{t\pm1}^1, \ldots) + \tilde{v}_t^2$, where $P(\cdot|\cdot)$ denotes the linear projection, so that $\tilde{v}_t^2$ is uncorrelated with all leads and lags of $v_t^1$. In general, $\tilde{v}_t^2$ has a Wold representation, which is written here in the autoregressive form $\tilde{v}_t^2 = \{I - F(L)L\}^{-1}\omega_t$. In addition, the linear projection can be written, $P(v_t^2 | v_t^1, v_{t\pm1}^1, \ldots) = \tilde{B}(L)v_t^1 = \tilde{B}(L)(\alpha' Y_t - \alpha'\phi)$, where $\tilde{B}(L)$ is in general two-sided. Upon substituting these expressions into (6.5b) and letting $Y_t^\dagger = D' Y_t$, one obtains the modified triangular representation

$$\Delta Y_t^\dagger = \mu + \lambda d_t(k_0) + D_t^\dagger(k_0)\tilde{\lambda} + F(L)\Delta Y_{t-1}^\dagger + B(L)(\alpha' Y_{t-1}) + \omega_t, \tag{6.6}$$

where $\mu = \{I - F(1)\}(D'D\tilde{\mu} - \tilde{B}(1)\alpha'\phi)$, $B(L) = \{I - F(L)L\}\tilde{B}(L)$, $\lambda = \{I - F(1)\}D'D\tilde{\lambda}$, and $D_t^\dagger(k) = F^*(L)D'D\Delta d_t(k)$, where $F_j^* = -\sum_{i=j}^{\infty} F_i$. As in the univariate and VECM representations, the transient $D_t^\dagger(k_0)$ is omitted. Thus, the modified triangular form considered is the system composed of (6.5a) and

$$\Delta Y_t^\dagger = \mu + \lambda d_t(k_0) + F(L)\Delta Y_{t-1}^\dagger + B(L)(\alpha' Y_{t-1}) + \omega_t. \tag{6.7}$$

In the King–Plosser–Rebelo (1988) model, $D = (1, 1, 1)'$ so $Y_t^\dagger$ is the sum of the logarithms of output, consumption, and investment. The system (6.5a) and (6.7) will be referred to as the modified triangular representation.

In the modified triangular form, the break appears in only the second block of equations, (6.7). By construction, $\omega_t$ is serially uncorrelated, is uncorrelated with $v_t^1$ at all leads and lags, and is uncorrelated with the regressors. The Gaussian MLE for $\lambda$ and $k_0$ is obtained, asymptotically, simply by estimating (6.7) by ordinary least squares. Because (6.7) is of the form (2.1) with $X_{t-1}$ being $\alpha' Y_{t-1}$ and its lags, the tools of Section 2 can be applied to (6.7). A practical advantage of this representation over the VECM is that the number of equations in (6.7) is the dimension of $\tau_t$ (in our case, one), less than the number of equations in the VECM representation (6.4).[8]

The discussion so far has assumed that the matrix of cointegrating vectors $\alpha$ is known. If the cointegrating vectors are unknown, $\alpha$ can be replaced by any $T$-consistent estimator $\hat{\alpha}$ and the asymptotic results of Theorem 1 will be unchanged. The argument here is standard and relies on the fast rate of convergence of estimators of the cointegrating vector (e.g. Stock (1987)). Similarly, because $\hat{\alpha}$ is consistent for $\alpha$, the estimator $\hat{D}$ formed by $\hat{D}'\hat{\alpha} = 0$ will be consistent and $Y_t^\dagger = \Delta D' Y_t$ in (6.7) can be replaced by $\hat{Y}_t^\dagger = \Delta\hat{D}' Y_t$ and the same asymptotic results obtain.

### 6.2. *Empirical results*

The data used here are total real *per capita* GDP, total real *per capita* personal consumption expenditures, and total real *per capita* gross private domestic fixed investment for the

8. The representation (6.7) could alternatively have been derived starting with any linear combinations of $Y_t$ that are integrated and not cointegrated, rather than $Y_t^\dagger$. Because leads and lags of $\alpha' Y_t$ are included as regressors in (6.7), however, any such representation can be rearranged to yield the representation (6.7).

United States, quarterly, 1959:I–1995:IV. All real series are chain-weighted quantity indexes. The quantity indexes are put on a *per capita* basis by dividing by the total civilian noninstitutional population aged sixteen and over. Historical revisions of the quantity indexes before 1959 are not available at the date of this writing, so the full sample is 1959:I–1995:IV. Logarithms of all series are used throughout. Regressions are all run over shorter periods to allow for initial conditions.

The empirical analysis proceeds in two steps. First, we investigate the unit root and cointegration properties of the series, taking into account the possibility that there might be shifts in mean growth rates and further investigating whether the cointegrating coefficients (if any) were stable over this period. We conclude that the evidence is consistent with these series being individually integrated and jointly cointegrated, and that the cointegrating coefficients appear to have been constant. Second, we turn to an investigation of whether there were breaks in the mean growth rates and, if so, when they occurred. This is done both for the individual series and as a system, both using estimated cointegrating coefficients and using the unit coefficients suggested by the King–Plosser–Rebelo (1988) model. The analysis of this section is based on the $I(0)$ framework of Section 2 rather than Section 3 because no shift in the cointegrating relationship is allowed.

*Unit root and cointegration analysis.*

The preliminary unit root and cointegration analysis here mainly uses standard techniques, so we provide only a brief summary of the results. Because we are investigating the possibility of a change in the mean growth rate in the series, as Perron (1989) pointed out, standard unit root tests such as the Dickey–Fuller (1979) test are inappropriate and can lead to spurious acceptance. Therefore univariate unit root tests were performed using the Banerjee–Lumsdaine–Stock (1992)/Zivot–Andrews (1992) minimal sequential ADF test, maintaining the hypothesis of a possible break in the mean growth rate. Using either a lag length chosen by the BIC or a fixed lag length of 4, for GDP and consumption, the tests failed to reject the unit root null at the 10% level; for investment they rejected at the 10% but not 5% level.[9]

Conventional multivariate and system-based tests for cointegration have well documented and pervasive problems with low power and large size distortions (Haug (1996)), making them unreliable. On the other hand, the standard univariate augmented Dickey–Fuller (1979) $t$-test with lags chosen by BIC has good size properties; *cf.* Stock (1994). Because suitable candidate cointegrating vectors are available in our application, this provides a desirable alternative to conventional system-based tests for cointegration (as do the tests developed by Horvath and Watson (1995), although those are not pursued here). We therefore tested for cointegration by using the Dickey–Fuller (1979) $t$-statistic (including a constant and time trend) to test the null of a unit root for $c-y$, $i-y$, and $c-i$, which are the cointegrating relations implied by the King–Plosser–Rebelo theory. Using either lags selected by BIC or a fixed lag length of four, the unit root null is rejected at the 10% level for each of the three error correction terms, with the exception of $i-y$ with BIC lags. Using the demeaned Dickey–Fuller (1979) $t$-test, the unit root null is rejected at the 10% level using either lag choice for $i-y$ and $c-i$, but not for $c-y$. Although these tests do not incorporate modifications for possibly broken time trends in

9. This test statistic is the $t$-ratio testing the hypothesis of a unit root when the deterministic terms include a constant, a time trend $t$, and $(t-k)1(t-k)$, minimized over all $k$ with 15% trimming at the beginning and end of the sample. For lags selected by the BIC, the test statistics were $-3{\cdot}62$ (GDP), $-3{\cdot}91$ (consumption), and $-4{\cdot}38$ (investment); the 10% (5%) critical value is $-4{\cdot}20$ ($-4{\cdot}48$).

the series, Perron's (1989) problem of spurious acceptance would work against finding cointegration using these tests. Because, for these data, the tests typically reject in favour of cointegration, this concern is not relevant.

The cointegrating coefficients were estimated by the Stock–Watson (1993) Dynamic OLS (DOLS) estimator, both in single equation and system forms.[10] The results used here are for the system normalized so that the coefficients are of the form $c-\theta_1 y$ and $i-\theta_2 y$. For the DOLS estimator with 2 leads and lags with an autoregressive estimator of the spectral density (with 4 lags), we obtain the estimated coefficients and standard errors, $\hat{\theta}_1 = 1{\cdot}187$ (0·032) and $\hat{\theta}_2 = 1{\cdot}242$ (0·164). Although the hypothesis that $\theta_2 = 1$ cannot be rejected at the 10% level, it is easily rejected that $\theta_1 = 1$. This gives rise to some ambiguity in these results; the DOLS procedure rejects $\theta_1 = 1$, but the series $y-c$ appears integrated of order zero using the Dickey–Fuller $t$-test. This ambiguity might be explained by the results of Horvath and Watson (1995), which suggest that tests for non-cointegration with a prespecified cointegrating vector can have good power even when the cointegrating vector is slightly misspecified.

Finally, the Gregory–Hansen (1996) statistic was used to test the null of a cointegrating relation with constant coefficients against the alternative that either the intercept or the slope coefficient changed; for both the $(y, c)$ and $(y, i)$ systems, these tests fail to reject the null at the 10% level, using either four lags or lags selected by BIC. We conclude that these three variables can be modelled as cointegrated with two stable cointegrating vectors.

*Evidence of breaks in mean growth rates.*

Break test statistics are summarized in Table 5. All lag lengths were determined using the BIC, searching between 1 and 6 lags for the univariate models, 1 and 4 lags for the bivariate models, and 1 and 3 lags for the trivariate models.

When considered individually, the hypothesis of a constant mean growth rate cannot be rejected at the 5% level for any of the three series. Accordingly, the associated estimates of the break date are very imprecise, with confidence intervals spanning at least twelve years.

Results for multivariate systems, with cointegrating coefficients estimated using the DOLS estimator, appear in panel B of Table 5. Both the mean shift tests reject at the 10% level for the $(c, i)$ system and the Sup-$W$ test rejects at the 10% level for the $(y, i)$ system. Both tests reject at the 10% level in the trivariate VECM, and at the 5% level using the $(y, c, i)$ triangular form. Although the estimated intervals are imprecise in the bivariate systems, the trivariate estimators yield shorter confidence intervals, consistent with the theoretical predictions. Moreover, the point estimates of the break dates are the same, and the confidence intervals are comparable, using either the triangular form or the VECM. The confidence intervals suggest that the break date occurred in the late 1960's or early 1970's. Consistent with the results of the Gregory–Hansen tests, no mean shift is apparent in the estimated error correction terms, a finding that further supports the interpretation of the breaks being in the mean growth rates of the series rather than in the intercepts or slopes of the cointegrating equations.

As a check on these results, break statistics were also computed when unit cointegrating coefficients were imposed; the results for the trivariate models are summarized in panel C of Table 5. The break dates estimated in the trivariate systems are the same as when

10. Although this estimator is efficient only if there is no intercept shift, it is consistent whether or not there is an intercept shift.

TABLE 5

*Empirical results: U.S. output, consumption and investment*

| Series | $p$ | Sup-$W$-15% | Exp-$W$-15% | $\hat{k}$ | 90% Conf. Int. |
|---|---|---|---|---|---|
| *A. Univariate* | | | | | |
| $y$ | 1 | 5·51 (0·19) | 0·97 (0·20) | 66:3 | (60:2, 72:4) |
| $c$ | 1 | 6·79 (0·11) | 1·61 (0·09) | 69:1 | (62:3, 75:3) |
| $i$ | 1 | 2·43 (0·67) | 0·27 (0·68) | 66:3 | (<59:1, 80:4) |
| *B. Multivariate, with estimated cointegrating coefficients* | | | | | |
| $c-1{\cdot}187y$ | 1 | 3·25 (0·49) | 0·38 (0·55) | 69:3 | (<59:1, 84:2) |
| $i-1.242y$ | 3 | 3·14 (0·52) | 0·47 (0·46) | 88:1 | (75:4, >95:4) |
| $c-0{\cdot}881i$ | 3 | 6·07 (0·16) | 1·46 (0·11) | 88:1 | (82:3, 93:3) |
| $y$, $c$, VECM | 1 | 9·48 (0·12) | 2·49 (0·12) | 69:3 | (64:4, 74:2) |
| $y$, $i$, VECM | 1 | 10·69 (0·08) | 2·56 (0·11) | 67:3 | (64:1, 71:1) |
| $c$, $i$, VECM | 2 | 11·10 (0·07) | 3·24 (0·06) | 89:2 | (86:4, 91:4) |
| $y$, $c$, $i$, VECM | 1 | 14·12 (0·05) | 3·97 (0·07) | 69:1 | (66:2, 71:4) |
| $y$, $c$, $i$, Triangular form | 1 | 10·22 (0·03) | 2·47 (0·04) | 69:1 | (65:1, 73:1) |
| *C. Multivariate, with unit cointegrating coefficients imposed* | | | | | |
| $y$, $c$, $i$, VECM | 3 | 22·61 (0·01) | 7·79 (0·01) | 69:1 | (68:3, 69:3) |
| $y$, $c$, $i$, Triangular form | 1 | 14·48 (0·01) | 3·61 (0·01) | 69:1 | (68:1, 70:4) |

*Notes*: $y$, $c$, and $i$ refer to the logarithms of real *per capita* GDP, personal consumption expenditures and gross domestic private fixed investment. Lag lengths $p$ for the regression, chosen by BIC, are given in the "$p$" column. In the triangular form, the BIC chose $p=1$ with one lead, one lag, and a contemporaneous term of the co-integrated series $\alpha' Y_{t-1}$, for both the estimated and the imposed unit cointegrating vectors. See the notes to Table 4.

the cointegrating coefficients were themselves estimated, but the confidence intervals are notably tighter. A caveat on interpreting the results in panel C as strong evidence of a precisely estimated break in the late 1960s is that the consumption/income ratio drifted up during 1959–1995 from 63% to 67%. This increase is reflected in a cointegrating coefficient in the consumption/income relation that is statistically significantly greater than one, a value inconsistent with economic theory in the long run. This upward drift in the consumption/income ratio might result in the break becoming spuriously more significant in the system VECM. This upward drift is not evident, however, in the cointegrating residual $c-1{\cdot}187y$. We therefore interpret the results in panel C as consistent with the results in panel B, although in our judgment the wider confidence intervals in panel B are more reliable.

Taken together these results provide evidence of a break in the mean growth rates of U.S. income, consumption and investment. From 1959:I to 1969:I, U.S. income *per capita* grew at 3.0% annually; from 1969:II–1995:IV, it grew at 1.1% annually. When the series are considered individually, this break is not statistically significant once one allows for an estimated break date, but it is when the series are treated as a cointegrated system. The system estimators date this break with some precision as occurring in the late 1960s or the early 1970s. In contrast to the results for European output, all the 90% confidence intervals for the trivariate systems fall before the oil shocks of the mid-1970s.

## 7. CONCLUSIONS

In both the European and U.S. cases, the use of multiple series sharpened the inference about the existence and dates of shifts in the mean levels. In the European data, there is strong evidence of a slowdown in the average growth rates of real output in the early 1970s, with a confidence interval that includes the first OPEC oil shock. This slowdown appears to have occurred approximately simultaneously in France, Germany, and Italy. The interpretation of the results for the U.S. is, however, less clear. While most of the test statistics reject the no-break hypothesis in the four trivariate specifications considered, the estimated confidence interval is centred around 1969.[11] This evidence argues against conventional associations of the slowdown in growth in the U.S. with the oil shock.

## APPENDIX

*Proof of Theorem* 2.

To prove Theorem 2, we first establish a series of properties for sequential pseudo-likelihood ratios and sequential estimators to be defined below in the absence of structural change. We then show that Theorem 2 can be derived as a consequence of these properties. To begin with, let

$$y_t=(V_t'\otimes I)\theta_0+\varepsilon_t,$$

where $V_t'=(1',y_{t-1}',\ldots,y_{t-p}',X_{t-1}')$, $\theta_0=\operatorname{Vec}(\mu,A_1,\ldots,A_p,\Gamma)$, and the $\varepsilon_t$ are martingale differences with variance $\Sigma_0$. This model is the same as (2.2) but without a break.

Let $(\theta_0,\Sigma_0)$ denote the true parameter. Consider the pseudo-Gaussian likelihood ratio based on the first $k$ observations

$$\mathscr{L}(1,k;\theta,\Sigma)=\frac{\prod_{t=1}^k f(y_t|y_{t-1},\ldots;\theta_0+T^{-1/2}\theta,\Sigma_0+T^{-1/2}\Sigma)}{\prod_{t=1}^k f(y_t|y_{t-1},\ldots;\theta_0,\Sigma_0)}$$
$$=\frac{|\Sigma_0+T^{-1/2}\Sigma|^{-k/2}\exp\{-\frac{1}{2}\sum_{t=1}^k\varepsilon_t(\theta)'(\Sigma_0+T^{-1/2}\Sigma)^{-1}\varepsilon_t(\theta)\}}{|\Sigma_0|^{-k/2}\exp\{-\frac{1}{2}\sum_{t=1}^k\varepsilon_t'\Sigma_0^{-1}\varepsilon_t\}},\tag{A.1}$$

where $\varepsilon_t(\theta)=y_t-(V_t'\otimes I)(\theta_0+T^{-1/2}\theta)=\varepsilon_t-T^{-1/2}(V_t'\otimes I)\theta$. We shall call the above the pseudo-sequential likelihood ratio. Denote by $\hat\theta_{(k)}$ and $\hat\Sigma_{(k)}$ the values of $\theta$ and $\Sigma$ that $\mathscr{L}(1,k;\theta,\Sigma)$ achieves its maximum. Then we have:

**Property 1.** *For each* $\delta\in(0,1]$,

$$\sup_{T\delta\leq k\leq T}(\|\hat\theta_{(k)}\|+\|\hat\Sigma_{(k)}\|)=O_p(1),$$

$$\sup_{T\delta\leq k\leq T}\mathscr{L}(1,k;\hat\theta_{(k)},\hat\Sigma_{(k)})=O_p(1).$$

This property says that the sequential likelihood ratios and the sequential estimators are bounded in probability if a positive fraction of observations are used. This result is a direct consequence of the functional central limit theorem for martingale differences. We thus omit the proof. The next property is concerned with the supremum of the likelihood ratios over all $k$ and over the whole parameter space.

11. Also, there is some evidence that the source of the detected break is not a decline in the mean growth rate but instead highly persistent shifts in the share of output allocated to consumption and, possibly, investment.

**Property 2.** *For each $\varepsilon>0$, there exists a $B>0$ such that for large $T$*

$$\Pr\left(\sup_{1\leq k\leq T} T^{-B}\mathscr{L}(1,k;\hat{\theta}_{(k)},\hat{\Sigma}_{(k)})>1\right)<\varepsilon.$$

This property says that the log-valued pseudo-sequential likelihood ratio has its maximum value bounded by $O_p(\log T)$.

*Proof.* The likelihood ratio evaluated at $\hat{\theta}_{(k)}$ and $\hat{\Sigma}_{(k)}$ can be rewritten as

$$\log\mathscr{L}(1,k;\hat{\theta}_{(k)},\hat{\Sigma}_{(k)})=-k/2(\log|\Sigma^*_{(k)}|-\log|\Sigma_0|)+\tfrac{1}{2}\left(\textstyle\sum_{t=1}^k \varepsilon_t'\Sigma_0^{-1}\varepsilon_t-kn\right), \tag{A.2}$$

where

$$\Sigma^*_{(k)}=\frac{1}{k}\sum_{t=1}^k \varepsilon_t\varepsilon_t'-\left(\frac{1}{k}\sum_{t=1}^k \varepsilon_tV_t'\right)\left(\frac{1}{k}\sum_{t=1}^k V_tV_t'\right)^{-1}\left(\frac{1}{k}\sum_{t=1}^k V_t\varepsilon_t'\right).$$

Thus by adding and subtracting an identity matrix, we obtain

$$-\frac{k}{2}(\log|\Sigma^*_{(k)}|-\log|\Sigma_0|)=-\frac{k}{2}\log\left|I+\frac{1}{k}\sum_{t=1}^k(\eta_t\eta_t'-I)-\left(\frac{1}{k}\sum_{t=1}^k\eta_tV_t'\right)\left(\frac{1}{k}\sum_{t=1}^k V_tV_t'\right)^{-1}\left(\frac{1}{k}\sum_{t=1}^k V_t\eta_t'\right)\right|,$$

where $\eta_t=\Sigma_0^{-1/2}\varepsilon_t$ with $E\eta_t=0$ and $\operatorname{Var}(\eta_t)=I$. The above is equal to, upon Taylor expansion

$$-\tfrac{1}{2}\operatorname{tr}\left(\textstyle\sum_{t=1}^k(\eta_t\eta_t'-I)\right)+\tfrac{1}{2}\operatorname{tr}(\Phi_k), \tag{A.3}$$

$$+\frac{k}{4}\operatorname{tr}\left\{\left(\frac{1}{k}\sum_{t=1}^k(\eta_t\eta_t'-I)-\frac{1}{k}\Phi_k\right)^2\right\}+O_p(1), \tag{A.4}$$

where

$$\Phi_k=\left(k^{-1/2}\textstyle\sum_{t=1}^k\eta_tV_t'\right)\left((1/k)\textstyle\sum_{t=1}^k V_tV_t'\right)^{-1}\left(k^{-1/2}\textstyle\sum_{t=1}^k V_t\eta_t'\right),$$

and $O_p(1)$ is uniform in $k$. The first term of (A.3) is cancelled out with the last term of (A.2). Thus

$$\log\mathscr{L}(1,k;\hat{\theta}_{(k)},\hat{\Sigma}_{(k)})=\tfrac{1}{2}\operatorname{tr}(\Phi_k)+\tfrac{1}{4}\operatorname{tr}\left\{\left(k^{-1/2}\textstyle\sum_{t=1}^k(\eta_t\eta_t'-I)\right)^2\right\}$$
$$-\tfrac{1}{2}\operatorname{tr}\left\{\left(\frac{1}{k}\sum_{t=1}^k(\eta_t\eta_t'-I)\right)\Phi_k\right\}+\tfrac{1}{4}\operatorname{tr}\left(\frac{1}{k}\Phi_k^2\right)+O_p(1).$$

To prove Property 2, it suffices to show the above is $O_p(\log T)$ uniformly in $k$. By the strong law of large numbers, $1/k\sum_{t=1}^k V_tV_t'$ converges to a positive definite matrix as $k\to\infty$; this implies $\sup_{k\geq k_1}\|(1/k\sum_{t=1}^k V_tV_t')^{-1}\|=O_p(1)$, for some fixed $k_1>0$. Because $\max_{1\leq k\leq k_1}\mathscr{L}(1,k;\hat{\theta}_{(k)},\hat{\Sigma}_{(k)})=O_p(1)$, without loss of generality, we may assume $k\geq k_1$. By the law of iterated logarithms for martingale differences, $\|k^{-1/2}\sum_{t=1}^k\eta_tV_t'\|=O_p((\log T)^{1/2})$ and $\|k^{-1/2}\sum_{t=1}^k(\eta_t\eta_t'-I)\|=O_p((\log T)^{1/2})$ uniformly in $k\in[1,T]$. Thus $\|\Phi_k\|=O_p(\log T)$ uniformly in $k\in[k_1,T]$. In addition, from $k^{-1}\sum_{t=1}^k\eta_tV_t'=O_p(1)$ uniformly in $k$, we have $k^{-1}\Phi_k^2=O_p(\log(T))$ uniformly in $[k_1,T]$. This proves Property 2. ‖

The next property states that the value of the pseudo-likelihood ratio, when the parameters are evaluated away from zero, is arbitrarily small for large $T$, assuming a positive fraction of observations are used. We assume that $\Sigma_0+T^{-1/2}\Sigma$ is positive definite so that the likelihood ratio is well defined.

**Property 3.** *Let $S_T=\{(\theta,\Sigma);\ \|\theta\|\geq\log T$ or $\|\Sigma\|\geq\log T\}$. For any $\delta\in(0,1)$, $D>0$, $\varepsilon>0$, the following holds when $T$ is large*

$$\Pr\left(\sup_{k\geq T\delta}\ \sup_{(\theta,\Sigma)\in S_T} T^D\mathscr{L}(1,k;\theta,\Sigma)>1\right)<\varepsilon.$$

*Proof.* The sequential log-likelihood ratio can be written as

$$\log\mathscr{L}(1,k;\theta,\Sigma)=\mathscr{L}_{1,T}+\mathscr{L}_{2,T},$$

where

$$\mathcal{L}_{1,T}=-\frac{k}{2}\log|I+\Psi_T|-\frac{k}{2}\left[\frac{1}{k}\sum_{t=1}^{k}\eta_t'(I+\Psi_T)^{-1}\eta_t-\frac{1}{k}\sum_{t=1}^{k}\eta_t'\eta_t\right], \tag{A.5}$$

and

$$\mathcal{L}_{2,T}=T^{-1/2}\theta'(I\otimes(I+\Psi_T)^{-1})\sum_{t=1}^{k}(V_t\otimes\eta_t)-\frac{1}{2}\frac{k}{T}\theta'\left(\frac{1}{k}\sum_{t=1}^{k}V_tV_t'\otimes(I+\Psi_T)^{-1}\right)\theta, \tag{A.6}$$

with $\eta_t=\Sigma_0^{-1/2}\varepsilon_t$ and $\Psi_T=T^{-1/2}(\Sigma_0^{-1/2}\Sigma\Sigma_0^{-1/2})$. Let $S_T=S_1\cup S_2$ with

$$S_1=\{(\theta,\Sigma);\ \|\Sigma\|\geqq\log T,\ \theta\text{ arbitrary}\},$$

and

$$S_2=\{(\theta,\Sigma);\ \|\theta\|\geqq\log T\text{ and }\|\Sigma\|\leqq\log T\}.$$

Then on $S_1$, $\mathcal{L}_{2,T}$ is maximized with respect to $\theta$ at ($\mathcal{L}_{2,T}$ is quadratic in $\theta$)

$$\hat{\theta}_{(k)}=1/k\sqrt{T}(H_k^{-1}\otimes I)\sum_{t=1}^{k}(V_t\otimes\eta_t),$$

where $H_k=1/k\sum_{t=1}^{k}V_tV_t'$. Thus

$$\sup_{\theta}\mathcal{L}_{2,T}=(k/2)\zeta_k'(I\otimes(I+\Psi_T)^{-1})\zeta_k,$$

where

$$\zeta_k=T^{-1/2}(H_k^{1/2}\otimes I)\hat{\theta}_{(k)}=(H_k^{-1/2}\otimes I)(1/k)\sum_{t=1}^{k}(V_t\otimes\eta_t). \tag{A.7}$$

Consequently, concentrating $\theta$ out, we obtain

$$\log\mathcal{L}(1,k;\theta,\Sigma)\leqq-\frac{k}{2}\left[\log|I+\Psi_T|+\frac{1}{k}\left(\sum_{t=1}^{k}\eta_t'\{(I+\Psi_T)^{-1}-I\}\eta_t\right)-\zeta_k'(I\otimes(I+\Psi_T)^{-1})\zeta_k\right].$$

We next show that the term inside the bracket is positive on $S_1$ when $T$ is large and the term is of the order of magnitude $(\log T)^2/T$. Because $(I+\Psi_T)^{-1}$ is a symmetric matrix, there exists an orthogonal matrix $U$ such that $U(I+\Psi_T)^{-1}U'=\operatorname{diag}(1/(1+\lambda_i),\ i=1,\ldots,n)$ where $\lambda_i$ are the eigenvalues of $\Psi_T$. Since $\operatorname{tr}(A)=\operatorname{tr}(UAU')$ for any orthogonal matrix $U$, we have

$$\frac{1}{k}\sum_{t=1}^{k}\eta_t'\{(I+\Psi_T)^{-1}-I\}\eta_t=\operatorname{tr}\left(\operatorname{diag}\left\{\frac{1}{1+\lambda_i}-1\right\}\left(\frac{1}{k}U\sum_{t=1}^{k}\eta_t\eta_t'U'\right)\right), \tag{A.8}$$

for some $U$ which diagonalizes $(I+\Psi_T)^{-1}$. We shall derive a lower bound for (A.8). Notice

$$\left\|\frac{1}{k}U\sum_{t=1}^{k}\eta_t\eta_t'U'-I\right\|=\left\|\frac{1}{k}U\sum_{t=1}^{k}(\eta_t\eta_t'-I)U'\right\|\leqq\frac{1}{k}\left\|\sum_{t=1}^{k}(\eta_t\eta_t'-I)\right\|,$$

because $\|U\|=1$. Furthermore, let $b_T=T^{-1/2}\log T$, then for any $a>0$,

$$\Pr\left(\sup_{k\geqq T\delta}\frac{1}{k}\left\|\sum_{t=1}^{k}(\eta_t\eta_t'-I)\right\|>ab_T\right)\leqq\Pr\left(\sup_{k\geqq T\delta}\delta^{-1}T^{-1/2}\left\|\sum_{t=1}^{k}(\eta_t\eta_t'-I)\right\|>a\log T\right)<\varepsilon, \tag{A.9}$$

for large $T$. This is because $T^{-1/2}\|\sum_{t=1}^{k}(\eta_t\eta_t'-I)\|$ is uniformly (in $k$) bounded in probability by the functional central limit theorem. Thus the diagonal elements of $U(1/k)\sum_{t=1}^{k}\eta_t\eta_t'U'$ are bounded above by $1+ab_T$ and below by $1-ab_T$ with probability at least $1-\varepsilon$. Also note the sign of $1/(1+\lambda_i)-1$ is opposite to the sign of $\lambda_i$. Thus we have, from (A.8)

$$\frac{1}{k}\sum_{t=1}^{k}\eta_t'\{(I+\Psi_T)^{-1}-I\}\eta_t\geqq\sum_{i=1}^{n}\left(\frac{1}{1+\lambda_i}-1\right)(1+\operatorname{sign}(\lambda_i)ab_T), \tag{A.10}$$

with probability at least $1-\varepsilon$. Similar to (A.9), we have for any $\gamma>0$,

$$\Pr\left(\sup_{k\geqq T\delta}\|\zeta_k\|>\gamma b_T\right)\leqq\varepsilon, \tag{A.11}$$

when $T$ is large. This is because $\zeta_k$ involves sums of martingale differences and $H_k^{-1/2}=O_p(1)$ for all large $k$, see (A.7). Now $((I+\Psi_T)^{-1}\otimes I)$ has $q$ repeated eigenvalues of $1/(1+\lambda_i)$ $(i=1,2,\ldots,n)$ with $q=\dim(H_k)$ (same as the dimension of the second identity matrix). We have from (A.11)

$$\zeta_k'(I\otimes(I+\Psi_T)^{-1})\zeta_k \leqq \textstyle\sum_{i=1}^n \frac{q}{1+\lambda_i}\gamma^2 b_T^2,$$

with probability not less than $1-\varepsilon$ when $T$ is large. Therefore

$$\log \mathscr{L}(1,k;\theta,\Sigma)\leqq -\frac{k}{2}\left[\textstyle\sum_{i=1}^n\left(\log(1+\lambda_i)+\left(\frac{1}{1+\lambda_i}-1\right)(1+\operatorname{sign}(\lambda_i)ab_T)-\frac{1}{1+\lambda_i}q\gamma^2 b_T^2\right)\right], \tag{A.12}$$

with probability not less than $1-2\varepsilon$. On $S_1$, $\|\Sigma\|\geqq\log T$. This implies $\|\Psi_T\|\geqq cT^{-1/2}\log T=cb_T$ for some $c>0$ (taking $c=\|\Sigma_0^{1/2}\|^{-2}$ is enough). This further implies that there exists an $i$ such that $|\lambda_i|\geqq Cb_T$ for some $C>0$, because there exist $c_1>0$ and $c_2>0$ such that for any symmetric matrix $B$

$$c_1\|B\|\leqq\max_i|\lambda_i(B)|\leqq c_2\|B\|, \tag{A.13}$$

where $\lambda_i(B)$ are eigenvalues of matrix $B$. Without loss of generality, assume $|\lambda_1|\geqq Cb_T$. Define

$$f(x)=\log(1+x)+\left(\frac{1}{1+x}-1\right)(1+\operatorname{sign}(x)ab_T)-\frac{1}{1+x}q\gamma^2b_T^2.$$

The function $f(x)$ has two local minima and $f(x)\to+\infty$ when either $x\to+\infty$ or $x\to-1$ (we need $1+\operatorname{sign}(x)ab_T-q\gamma^2b_T^2>0$. This is true for large $T$ since $b_T\to0$. This is also true for $b_T=O(1)$, as in Property 5 below, by choosing small $a$ and $\gamma$). Furthermore

$$\inf_{-1<x<\infty} f(x)=-\tfrac{1}{2}a^2b_T^2-q\gamma^2b_T^2+o(b_T^2)=(a^2+\gamma^2)O(b_T^2), \tag{A.14}$$

and $f(x)$ satisfies

$$f(x)\geqq\tfrac{1}{2}C^2b_T^2-Cab_T^2-q\gamma^2b_T^2+o(b_T^2),\qquad \text{when } |x|\geqq Cb_T.$$

Thus the right-hand side of (A.12) is not larger than

$$-\frac{k}{2}\left[\inf_{|\lambda_1|\geqq Cb_T} f(\lambda_1)+\inf_{\lambda_i}\textstyle\sum_{i=2}^n f(\lambda_i)\right]\leqq-\frac{k}{2}\left[\tfrac{1}{2}C^2b_T^2-Cab_T^2-q\gamma^2b_T^2+(n-1)(a^2+\gamma^2)O(b_T^2)\right]. \tag{A.15}$$

If we take $a$ and $\gamma$ sufficiently small, then (A.15) is less than

$$-kb_T^2C^2/8\leqq-(\log T)^2\delta C^2/8,\qquad \text{for all } k\geqq T\delta,$$

which is further less than $-D\log T$ for any given $D>0$ for large $T$. Thus we have shown that for large $T$

$$\Pr\left(\sup_{k\geqq T\delta}\sup_{(\theta,\Sigma)\in S_1}\log\mathscr{L}(1,k;\theta,\Sigma)>-D\log T\right)<2\varepsilon. \tag{A.16}$$

That is, Property 3 holds on $S_1$. We next obtain the corresponding result on $S_2$.

We will first derive upper bounds for $\mathscr{L}_{1,T}$ and $\mathscr{L}_{2,T}$ separately on $S_2$ and then combine the bounds to derive another upper bound with respect to $k$. Using previous arguments, we can show that for any $a>0$, with probability at least $1-\varepsilon$, the following holds

$$\begin{aligned}\mathscr{L}_{1,T}&\leqq-\frac{k}{2}\left[\textstyle\sum_{i=1}^n\left(\log(1+\lambda_i)+\left(\frac{1}{1+\lambda_i}-1\right)(1+\operatorname{sign}(\lambda_i)ab_T)\right)\right]\\&\leqq n\frac{k}{4}a^2b_T^2(1+o(1))\\&\leqq na^2\frac{k}{8}b_T^2\qquad\text{for all }(\theta,\Sigma)\in S_2.\end{aligned} \tag{A.17}$$

The first inequality makes use of (A.10) and the second makes use of (A.14). Next consider $\mathscr{L}_{2,T}$. By Property 1, for any $\varepsilon>0$,

$$\Pr\left(\sup_{k\geqq T\delta}\|\hat\theta_{(k)}\|>\log T\right)<\varepsilon,$$

when $T$ is large. Thus we can assume $\|\theta\| = \log T$ on $S_2$. Also on $S_2$, $\|\Psi_T\| \leqq C \log T/\sqrt{T} \to 0$ for some $C>0$; this implies $(I+\Psi_T) > \frac{1}{2}I$ or $(I+\Psi_T)^{-1} < 2I$ for large $T$. This further implies that $\|(I+\Psi_T)^{-1}\|$ is bounded on $S_2$. Thus similar to the inequality (A.9), we have for any $\varepsilon>0$ and any $\alpha>0$

$$\Pr\left(\sup_{k\geqq T\delta}\sup_{S_2}\frac{1}{k}\|(I\otimes(I+\Psi_T)^{-1})\textstyle\sum_{t=1}^{k}(V_t\otimes\eta_t)\| > \alpha b_T\right) < \varepsilon.$$

This is because $V_t\otimes\eta_t$ is a sequence of martingale differences. Next, by the law of large numbers, there exists a positive definite matrix $\Omega$, such that $1/k\sum_{t=1}^{k} V_tV_t' \to \Omega$. Therefore with probability not less than $1-3\varepsilon$,

$$\mathscr{L}_{2,T} \leqq T^{-1/2}k\|\theta\|\alpha b_T - \frac{1}{4}\frac{k}{T}\theta'(\Omega\otimes I)\theta \leqq k\alpha b_T^2 - \frac{k}{4}b_T^2 C, \tag{A.18}$$

for some $C>0$ (note on $S_2$, we assume $\|\theta\| = \log T$ as noted earlier).

Combining (A.17) and (A.18) and choosing $\alpha$ sufficiently small, we obtain with probability at least $1-4\varepsilon$,

$$\begin{aligned}\log \mathscr{L}(1,k;\theta,\Sigma) &= \mathscr{L}_{1,T} + \mathscr{L}_{2,T}\\ &\leqq -kb_T^2C/16\\ &\leqq -(\log T)^2\delta C/16, \qquad \text{for all } k\geqq T\delta \text{ and all } (\theta,\Sigma)\in S_2.\end{aligned} \tag{A.19}$$

The above is less than $-D\log T$ for any $D>0$ when $T$ is large. This together with (A.16) implies Property 3. ||

For a given $M>0$, define $S_M = \{(\theta,\Sigma);\ \|\theta\|\geqq M \text{ or } \|\Sigma\|\geqq M\}$. Then similar arguments lead to

**Property 4.** *For any $\varepsilon>0$, there exists a $M>0$ such that*

$$\Pr\left(\sup_{k\geqq T\delta}\sup_{S_M}\mathscr{L}(1,k;\theta,\Sigma) > \varepsilon\right) < \varepsilon.$$

This property says that the value of pseudo-likelihood ratios evaluated outside a bounded set is small, assuming a positive fraction of observations are involved. The next property is similar to Property 4, but with no positive fraction of observations being required. This is compensated by moving $\theta$ and $\Sigma$ further away from zero ($M\to\infty$ in the notation of Property 4).

**Property 5.** *Let $h_T$ and $d_T$ be positive sequences such that $h_T$ is nondecreasing, $d_T\to+\infty$, and $(h_Td_T^2)/T\to h>0$, where $h<\infty$. Let $S_T = \{(\theta,\Sigma);\ \|\theta\|\geqq d_T \text{ or } \|\Sigma\|\geqq d_T\}$. Then for any $\varepsilon>0$, there exists an $A>0$, such that when $T$ is large*

$$\Pr\left(\sup_{k\geqq Ah_T}\ \sup_{(\theta,\Sigma)\in S_T}\mathscr{L}(1,k;\theta,\Sigma) > \varepsilon\right) < \varepsilon.$$

*Proof.* Define $b_T = T^{-1/2}d_T$. Then by assumption, $b_T = O(1)$ if $h_T$ stays bounded and $b_T\to 0$ if $h_T\to\infty$. Furthermore, $h_Tb_T^2\to h$. As in proving Property 3, we decompose $S_T$ into two subsets $S_1$ and $S_2$, where $S_1$ and $S_2$ are defined as in the earlier proof with $\log T$ replaced by $d_T$. On $S_1$, all arguments in Property 3 go through if inequalities (A.9) and (A.11) still hold true when $k\geqq T\delta$ is replaced with $k\geqq Ah_T$ and for the newly defined $b_T$. However, these are the immediate consequences of the Hajek and Renyi (1955) type of inequalities because by their inequality

$$\Pr\left(\sup_{k\geqq Ah_T}\frac{1}{k}\|\textstyle\sum_{t=1}^{k}(\eta_t\eta_t' - I)\| > \alpha b_T\right) \leqq \frac{C}{Ah_T\alpha^2b_T^2} < \frac{2C}{A\alpha^2h}, \tag{A.20}$$

for some $C>0$; the above is small if $A$ is large. Similarly, applying the Hajek and Renyi inequality to $1/k\sum_{t=1}^{k}(V_t\otimes\eta_t)$ together with $H_k^{-1} = O_p(1)$ uniformly in large $k$, we obtain, for any $\varepsilon>0$ and $\gamma>0$, there exists an $A>0$ such that

$$\Pr\left(\sup_{k\geqq Ah_T}\|\zeta_k\| > \gamma b_T\right) < \varepsilon. \tag{A.21}$$

Where $\zeta_k$ is given by (A.7). Using the inequalities (A.20) and (A.21) and the same arguments as in Property 3, we obtain, with probability at least $1-2\varepsilon$, $\mathscr{L}(1,k;\theta,\Sigma)\leqq -kb_T^2C^2/8$ for all $k\geqq Ah_T$ and all $(\theta,\Sigma)\in S_1$, which is further bounded by $-Ah_T b_T^2 C^2/8 < -AC^2h/16 < \log \varepsilon$ if $A$ is large. The proof on $S_2$ is almost the same as in Property 3 with only minor changes. Thus we omit it. ||

**Property 6.** *Under the same hypotheses as in Property 5, we have for any $A>0$,*

$$\sup_{k \leq Ah_T} \sup_{(\theta,\Sigma)\in S_T^c} \mathscr{L}(1,k;\theta,\Sigma)=O_p(1),$$

*where $S_T^c$ is the complement of $S_T$ with $S_T$ given in Property 5.*

This property asserts that when evaluated not too far away from zero and with the number of observations increasing not too fast, the likelihood ratio is simply bounded.

*Proof.* It suffices to prove the *log-valued* likelihood ratio is bounded in probability. The log-likelihood ratio consists of two expressions $\mathscr{L}_{1,T}$ and $\mathscr{L}_{2,T}$ given in (A.5) and (A.6), respectively. First consider $\mathscr{L}_{2,T}$. It is enough to prove the first term of $\mathscr{L}_{2,T}$ is bounded because the second term of $\mathscr{L}_{2,T}$ is negative (the exponential of a negative value is less than 1). The norm of the first term is bounded by

$$T^{-1/2}(d_T\sqrt{Ah_T}) \sup_{S_T^c} \|(I+\Psi_T)^{-1}\| \sup_{k\leq Ah_T} \frac{1}{\sqrt{Ah_T}} \|\textstyle\sum_{t=1}^k (V_t\otimes\eta_t)\|. \tag{A.22}$$

Note that $\|(I+\Psi_T)^{-1}\|$ is uniformly bounded on $S_T^c$ because $\|\Psi_T\|=O(T^{-1/2}d_T)<1$ (always possible because we can redefine $d_T$ by multiplying by a small constant). The second supremum is bounded by the functional central limit theorem for martingale differences. Combined with the boundedness of $T^{-1/2}(d_T\sqrt{Ah_T})$ (because its squared value is bounded by assumption), we see that (A.22) is $O_p(1)$. Next consider $\mathscr{L}_{1,T}$. Because

$$(I+\Psi_T)^{-1}=I-\Psi_T+\Psi_T^2(I+\Psi_T)^{-1},$$

$\mathscr{L}_{1,T}$ can be written as

$$\mathscr{L}_{1,T}=-k/2(\log|I+\Psi_T|-\operatorname{tr}(\Psi_T))+\tfrac{1}{2}\operatorname{tr}[\Psi_T \textstyle\sum_{t=1}^k(\eta_t\eta_t'-I)]-\sum_{t=1}^k \eta_t'\Psi_T^2(I+\Psi_T)^{-1}\eta_t.$$

The last term is nonpositive, so it is enough to consider the first two terms on the right. The first term is equal to

$$-\frac{k}{2}\textstyle\sum_{i=1}^n(\log(1+\lambda_i)-\lambda_i),$$

where again the $\lambda_i$'s are the eigenvalues of $\Psi_T$. By Taylor expansion, it becomes

$$\frac{k}{2}\textstyle\sum_{i=1}^n(\frac{1}{2}\lambda_i^2+o(\lambda_i^2))\leq kn\max_i \lambda_i^2 \leq knC\|\Psi_T\|^2\leq CAh_Td_T^2/T \qquad \text{for all } k\leq Ah_T,$$

which is bounded by assumption. We have utilized the relationship between a symmetric matrix and its eigenvalues, see (A.13). Next, consider the second term

$$\|\Psi_T \textstyle\sum_{t=1}^k(\eta_t\eta_t'-I)\|\leq\|\Psi_T\|\sqrt{Ah_T}\sup_{k\leq Ah_T}\left\|\frac{1}{\sqrt{Ah_T}}\sum_{t=1}^k(\eta_t\eta_t'-I)\right\|=C(T^{-1/2}d_T\sqrt{Ah_T})O_p(1)$$

which is bounded in probability. ||

We are now in the position to prove Theorem 2. We only consider the case in which $v\leq 0$, i.e., $k\leq k_0$. The case for $v>0$ is similar. The likelihood ratio $\Lambda_T(v,\beta,\Sigma)$ in (2.10) is based on the whole sample $[1,T]$. We can write it as the product of likelihood ratios for the three subsamples, $[1,k]$, $[k+1,k_0]$, and $[k_0+1,T]$. In this way, the likelihood ratio will have $(V_t\otimes I)$ rather than $Z_t(k)$ as regressors. Recall $\beta=(\theta',(S\delta)')'$. Let $\psi=\theta+S'S\delta$, which is the combined coefficients of $(V_t'\otimes I)$ for the second regime.

The likelihood ratio (2.10) can be rewritten as

$$\frac{L(k,\beta_0+T^{-1/2}\beta,\Sigma_0+T^{-1/2}\Sigma)}{L(k_0,\beta_0,\Sigma_0)}$$
$$=\mathscr{L}(1,k;\theta,\Sigma)\cdot\mathscr{L}(k+1,T\tau_0;\sqrt{T}S'S\delta_T+\psi,\Sigma)\cdot\mathscr{L}(T\tau_0+1,T;\psi,\Sigma), \tag{A.23}$$

where $T\tau_0=k_0$ with $\tau_0\in(0,1)$, and $\mathscr{L}(l,j;\cdot,\cdot)$ is defined as in (A.1) but using observations from $l$ to $j$. For simplicity of notation, we assume $T\tau_0$ is an integer, that is $T\tau_0=[T\tau_0]$. Only the middle term of (A.23) needs some explanation. For $t\in[k+1, T\tau_0]$, $\varepsilon_t(k)=y_t-Z_t(k)'(\beta_0+T^{-1/2}\beta)=\varepsilon_t-(V_t'\otimes I)S'S\delta_T-T^{-1/2}(V_t'\otimes I)(\theta+S'S\delta)=\varepsilon_t-T^{-1/2}(V_t'\otimes I)(\sqrt{T}S'S\delta_T+\psi)$. By the definition of $\mathscr{L}$ in (A.1), the segment $[k+1, T\tau_0]$ involves the parameter value $\sqrt{T}S'S\delta_T+\psi$.

Proof of (2.11). Let $k_T(v_1)=[k_0+v_1v_T^{-2}]$. For some $\varepsilon_0>0$ and $\varepsilon_0<\tau_0$, define

$$B_{1,T}=\{(k,\beta,\Sigma);\ \|\psi\|\leqq\tfrac{1}{2}\sqrt{T}\|S'S\delta_T\|,\ T\varepsilon_0\leqq k\leqq k_T(v_1)\},$$

$$B_{2,T}=\{(k,\beta,\Sigma);\ \|\psi\|\leqq\tfrac{1}{2}\sqrt{T}\|S'S\delta_T\|,\ 0\leqq k\leqq T\varepsilon_0\},$$

$$B_{3,T}=\{(k,\beta,\Sigma);\ \|\psi\|\geqq\tfrac{1}{2}\sqrt{T}\|S'S\delta_T\|,\ 0\leqq k\leqq k_T(v_1)\}.$$

On $B_{1,T}$, both $\mathscr{L}(1,k;\theta,\Sigma)$ and $\mathscr{L}(T\tau_0+1,T;\psi,\Sigma)$ are $O_p(1)$ from Property 1, since both use a positive fraction of observations. Next consider $\mathscr{L}(k+1,T\tau_0;\sqrt{T}S'S\delta_T+\psi,\Sigma)$ which involves $T\tau_0-k=-vv_T^{-2}$ observations. Since $\|\sqrt{T}S'S\delta_T+\psi\|\geqq\|\sqrt{T}S'S\delta_T\|-\|\psi\|\geqq\tfrac{1}{2}\sqrt{T}\|S'S\delta_T\|$, we apply Property 5 with $\theta=\sqrt{T}S'S\delta_T+\psi$, $d_T=\tfrac{1}{2}\sqrt{T}\|S'S\delta_T\|$, $h_T=v_T^{-2}$, and $A=-v_1$ (applied with the reversed data order, i.e. treating $T\tau_0$ as the first observation) to conclude that $\mathscr{L}(k+1,T\tau_0;\sqrt{T}S'S\delta_T+\psi,\Sigma)$ can be arbitrarily small in probability if $-v_1$ is large.

We now assume that $\sqrt{T}\|S'S\delta_T\|\geqq\log T$. Then on $B_{2,T}$, $\mathscr{L}(1,k;\theta,\Sigma)$ is less than $T^B$ for some $B>0$ with probability at least $1-\varepsilon$ from Property 2 and $\mathscr{L}(T\tau_0+1,T;\psi,\Sigma)$ is $O_p(1)$ from Property 1. However, by Property 3 with $\theta=\sqrt{T}S'S\delta_T+\psi$, $\mathscr{L}(k+1,T\tau_0;\sqrt{T}S'S\delta_T+\psi,\Sigma)$ (which involves a positive fraction of the data set, and $\|\theta\|\geqq\tfrac{1}{2}\sqrt{T}\|S'S\delta_T\|$) is less than $T^{-D}$ for any $D>0$ with probability at least $1-\varepsilon$ when $T$ is large. Thus the product of these three terms can be no larger than $\varepsilon$ with probability at least $1-2\varepsilon$ when $T$ is large.

Next on $B_{3,T}$, Property 2 is applicable to both $\mathscr{L}(1,k;\theta,\Sigma)$ and $\mathscr{L}(k+1,T\tau_0;\sqrt{T}S'S\delta_T+\psi,\Sigma)$ and Property 3 is applicable to $\mathscr{L}(T\tau_0+1,T;\psi,\Sigma)$. Thus their product can be arbitrarily small.

Proof of (2.12). Let

$$D_{1,T}=\{(k,\beta,\Sigma);\ \|\beta\vee\Sigma\|\geqq M,\ \|\sqrt{T}S'S\delta_T+\psi\|\leqq2\sqrt{T}\|S'S\delta_T\|,\ \|\Sigma\|\leqq2\sqrt{T}\|S'S\delta_T\|,\ k_T(v_1)\leqq k\leqq T\tau_0\},$$

$$D_{2,T}=\{(k,\beta,\Sigma);\ \|(\sqrt{T}S'S\delta_T+\psi)\vee\Sigma\|\geqq2\sqrt{T}\|S'S\delta_T\|,\ k_T(v_1)\leqq k\leqq T\tau_0\}.$$

where $\|x\vee y\|\geqq M$ means either $\|x\|\geqq M$ or $\|y\|\geqq M$.

On $D_{1,T}$, apply Property 6 to $\mathscr{L}(k+1,T\tau_0;\ \sqrt{T}S'S\delta_T+\psi,\Sigma)$ with reversed data order and with $d_T=2\sqrt{T}\|S'S\delta_T\|$, $h_T=v_T^{-2}$, and $A=-v_1$ to conclude that it is bounded. From Property 1, both $\mathscr{L}(1,k;\theta,\Sigma)$ and $\mathscr{L}(T\tau_0+1,T;\psi,\Sigma)$ are bounded in probability (again, both of them use a positive fraction of the observations). However one of them must be small if $M$ is large. This is because if $\|\beta\|>M$ we then have either $\|\theta\|>M/4$ or $\|\psi\|>M/4$, so that we can apply Property 4 to one of them; if $\|\Sigma\|>M$, then we can apply Property 4 to both of them.

The situation on $D_{2,T}$ is similar to that on $B_{3,T}$. The behaviour of $\mathscr{L}(1,k;\theta,\Sigma)$ and $\mathscr{L}(k+1,T\tau_0;\sqrt{T}S'S\delta_T+\psi,\Sigma)$ is controlled by Property 2. If $\|\sqrt{T}S'S\delta_T+\psi\|\geqq2\sqrt{T}\|S'S\delta_T\|$ then $\|\psi\|\geqq\|\psi+\sqrt{T}S'S\delta_T\|-\|\sqrt{T}S'S\delta_T\|\geqq\sqrt{T}\|S'S\delta_T\|$. Thus on $D_{2,T}$, $\|\psi\vee\Sigma\|\geqq\sqrt{T}\|S'S\delta_T\|$. Consequently, the behaviour of $\mathscr{L}(T\tau_0+1,T;\psi,\Sigma)$ is controlled by Property 3. The product of the three components again is arbitrarily small if $T$ is large. The proof of Theorem 2 is now complete. ||

*Proof of Theorem* 3.

We consider the case of $v\leqq0$ (i.e. $k\leqq k_0$). The case of $v\geqq0$ can be analyzed similarly. Note that for $t\in[1,k]\cup[k_0+1,T]$, $Z_t(k)\equiv Z_t(k_0)$. Thus $\varepsilon_t(k)=y_t-Z_t(k)'(\beta_0+T^{-1/2}\beta)=\varepsilon_t-Z_t(k_0)'T^{-1/2}\beta$. But for $t\in[k+1,k_0]$, $Z_t(k)=Z_t(k_0)+(0',[V_t'\otimes I]S')'$. This implies that $\varepsilon_t(k)=\varepsilon_t-Z_t(k_0)'T^{-1/2}\beta-(V_t'\otimes I)S'S(\delta_T+T^{-1/2}\delta)$. Thus the log-valued pseudo-likelihood ratio can be written as:

$$\log\frac{L(k,\beta_0+T^{-1/2}\beta,\Sigma_0+T^{-1/2}\Sigma)}{L(k_0,\beta_0,\Sigma_0)}$$

$$=-T/2\{\log|\Sigma_0+T^{-1/2}\Sigma|-\log|\Sigma_0|\}-\tfrac{1}{2}\textstyle\sum_{t=1}^{T}(\varepsilon_t'[(\Sigma_0+T^{-1/2}\Sigma)^{-1}-\Sigma_0^{-1}]\varepsilon_t)\tag{A.24}$$

$$+\beta'T^{-1/2}\textstyle\sum_{t=1}^{T}Z_t(k_0)(\Sigma_0+T^{-1/2}\Sigma)^{-1}\varepsilon_t\tag{A.25}$$

$$-\tfrac{1}{2}\beta'T^{-1/2}\textstyle\sum_{t=1}^{T}Z_t(k_0)(\Sigma_0+T^{-1/2}\Sigma)^{-1}Z_t(k_0)'T^{-1/2}\beta\tag{A.26}$$

$$+\textstyle\sum_{t=k+1}^{k_0}(\delta_T+T^{-1/2}\delta)'S'S(V_t\otimes I)(\Sigma_0+T^{-1/2}\Sigma)^{-1}\varepsilon_t\tag{A.27}$$

$$-\tfrac{1}{2}\textstyle\sum_{t=k+1}^{k_0}(\delta_T+T^{-1/2}\delta)'S'S(V_t\otimes I)(\Sigma_0+T^{-1/2}\Sigma)^{-1}(V_t'\otimes I)S'S(\delta_T+T^{-1/2}\delta)\tag{A.28}$$

$$-\textstyle\sum_{t=k+1}^{k_0}(\delta_T+T^{-1/2}\delta)'S'S(V_t\otimes I)(\Sigma_0+T^{-1/2}\Sigma)^{-1}Z_t(k_0)'\beta T^{-1/2}.\tag{A.29}$$

Expression (A.24) can be rewritten as, upon Taylor expansion,

$$\tfrac{1}{2}\operatorname{tr}(\Sigma_0^{-1}\Sigma\Sigma_0^{-1}(\Psi_T-\tfrac{1}{2}\Sigma))+o_p(1).$$

where $\Psi_T=T^{-1/2}\sum_{t=1}^{T}(\varepsilon_t\varepsilon_t'-\Sigma_0)$ and $o_p(1)$ is uniform over $\Sigma$ such that $\|\Sigma\|\leq M$, with $M$ an arbitrary fixed positive number. The above converges in distribution to the first term of (2.13). Expression (A.26) converging in probability to $-\tfrac{1}{2}\beta'Q\beta$ follows from the law of large numbers, and (A.25) converging in distribution to $\beta'Q^{1/2}\xi$ follows from the central limit theorem for martingale differences. For bounded $\delta$ and $\Sigma$, the limit of (A.28) is determined by

$$-\tfrac{1}{2}\delta_TS'S(\textstyle\sum_{t=k+1}^{k_0}V_tV_t'\otimes\Sigma_0^{-1})S'S\delta_T=-\tfrac{1}{2}\delta_0S'S(v_T^2\sum_{t=k+1}^{k_0}V_tV_t'\otimes\Sigma_0^{-1})S'S\delta_0. \qquad (A.30)$$

Note that $\delta_T=\delta_0v_T$. Since $k=k_0+[vv_T^{-2}]$, $v_T^2\sum_{t=k+1}^{k_0}V_tV_t'\to|v|Q_1$ where $Q_1=\operatorname{plim}(1/k_0)\sum_{t=1}^{k_0}V_tV_t'$. This implies that (A.30) converges to $-\tfrac{1}{2}|v|\delta_0'S'S(Q_1\otimes\Sigma_0^{-1})S'S\delta_0=-\tfrac{1}{2}|v|c$. For bounded $\delta$ and $\Sigma$, the limit of (A.27) is determined by

$$\textstyle\sum_{t=k+1}^{k_0}\delta_T'S'S(V_t\otimes I)\Sigma_0^{-1}\varepsilon_t=\delta_0S'S(I\otimes\Sigma_0^{-1})v_T\sum_{t=k+1}^{k_0}(V_t\otimes\varepsilon_t). \qquad (A.31)$$

By the functional central limit theorem for martingale difference, $v_T\sum_{t=k_0+[vv_T^{-2}]}^{k_0}(V_t\otimes\varepsilon_t)\Rightarrow(Q_1^{1/2}\otimes\Sigma_0^{1/2})\eta(v)$, where $\eta(v)$ is a vector of independent Brownian motion processes on $(-\infty,0]$, originated at the origin with reverse time. Thus (A.31) converges weakly to $\delta_0S'S(I\otimes\Sigma_0^{-1})(Q_1^{1/2}\otimes\Sigma_0^{1/2})\eta(v)$. This limit has the same distribution as $[\delta_0S'S(Q_1\otimes\Sigma_0^{-1})S'S\delta_0]^{1/2}W_1(v)=\sqrt{c}W_1(v)$, where $W_1(v)$ is a standard Brownian motion process on $(-\infty,0]$. This is because $b'Y$ has the same distribution as $\sqrt{b'Ab}\cdot N(0,1)$ for an arbitrary $Y\sim N(0,A)$ and an arbitrary constant vector $b$. Thus

$$\textstyle\sum_{t=k+1}^{k_0}\delta_T'S'S(V_t\otimes I)\Sigma_0^{-1}\varepsilon_t\;\Rightarrow\;[\delta_0S'S(Q_1\otimes\Sigma_0^{-1})S'S\delta_0]^{1/2}W_1(v)=\sqrt{c}W_1(v). \qquad (A.32)$$

If $v>0$, we will obtain another Brownian motion process, say $W_2(v)$. The two Brownian motion processes will be independent because they are the limits of non-overlapping martingale differences. Thus we may define a two-sided Brownian motion process on $(-\infty,\infty)$ and for $v\in[-M,M]$, the underlying partial sums converge to $\sqrt{c}W(v)$.

Finally, we show that (A.29) converges to zero in probability uniformly over bounded $\delta$, $\beta$, $\Sigma$ and $v$. For bounded parameters, the limit of (A.29) is determined by

$$\|\textstyle\sum_{t=k+1}^{k_0}\delta_T'S'S(V_t\otimes I)\Sigma_0^{-1}Z_t(k_0)'\beta T^{-1/2}\|\leq T^{-1/2}(k_0-k)\|\delta_T\|\cdot\|\Sigma_0^{-1}\|\cdot\|\beta\|\dfrac{1}{k_0-k}\sum_{t=k+1}^{k_0}\|V_t\|\,\|Z_t(k_0)\|.$$

The right-hand side above is bounded by $T^{-1/2}(k_0-k)\|\delta_T\|O_p(1)\leq MT^{-1/2}v_T^{-2}\|\delta_T\|O_p(1)=(\sqrt{T}v_T)^{-1}O_p(1)$, which is $o_p(1)$. The proof of Theorem 3 is complete. ||

*Proof of Theorem* 4.

The limiting process is quadratic in $\beta$ and in $\Sigma$, and thus is maximized at $\beta^*=Q^{-1/2}\xi$ and $\Sigma^*=\Psi$. With respect to $v$, it is maximized at $v^*=\operatorname{argmin}_\mu\{\sqrt{c}W(\mu)-c|\mu|/2\}$. By the continuous mapping theorem, $\sqrt{T}(\hat\beta_{\hat k}-\beta_0)\to Q^{-1/2}\xi$, $\sqrt{T}(\hat\Sigma_{\hat k}-\Sigma_0)\to\Psi$, and $v_T^2(\hat k-k_0)\to\operatorname{argmin}_\mu\{\sqrt{c}W(\mu)-c|\mu|/2\}$. By a change in variable, it can be shown that $\operatorname{argmin}_\mu\{\sqrt{c}W(\mu)-c|\mu|/2\}=c^{-1}\operatorname{argmin}_s\{W(s)-|s|/2\}$. Thus $cv_T^2(\hat k-k_0)\to\operatorname{argmin}_s\{W(s)-|s|/2\}$. But $cv_T^2=\delta_0S'S(Q_1\otimes\Sigma_0^{-1})S'S\delta_0v_T^2=\delta_TS'S(Q_1\otimes\Sigma_0^{-1})S'S\delta_T$, because $\delta_0v_T=\delta_T$ by definition. We thus proved (2.14)–(2.16). ||

*Proof of Corollary* 4.1.

(i) An intercept shift corresponds to $S=(s\otimes I)$ with $s=(1,0,\ldots,0)$ and $\delta_0S'S(Q_1\otimes\Sigma_0^{-1})S'S\delta_0=\lambda_0\Sigma_0^{-1}\lambda_0$. Thus, by Theorem 4,

$$(\lambda_0\Sigma_0^{-1}\lambda_0)v_T^2(\hat k-k_0)\xrightarrow{d}V^*$$

Now from $\lambda_0v_T=\lambda_T$, we obtain (i). Part (ii) follows from $S=I$ and $\delta_0v_T=\delta_T$. Finally, $T^{1/2}(\hat\lambda_T-\lambda_T)=O_p(1)$ by equation (2.14). Together with (2.15), it is easy to verify that $[(\hat\lambda_T'\hat\Sigma_{\hat k}^{-1}\hat\lambda_T)-(\lambda_T'\Sigma_0^{-1}\lambda_T)](\hat k-k_0)\xrightarrow{p}0$. This implies $(\hat\lambda_T'\hat\Sigma^{-1}\hat\lambda_T)(\hat k-k_0)\to V^*$. Similarly, part (ii) also holds when estimated $\delta_T$, $Q_1$, and $\Sigma_0^{-1}$ are used. ||

*Proof of Theorem* 5.

To prove Theorem 5, it is sufficient to establish the six properties (as in the proof of Theorem 2) for the sequential pseudo-likelihood ratios with $I(1)$ regressors. In the absence of a structural change, the data generating process is

$$Y_t = AX_t + \gamma t + \mu + Bw_t + \xi_t.$$

We may write it as $Y_t = (U_t'\otimes I)\theta_0 + \xi_t$. With $I(1)$ and trending regressors, the new parameterization for $\theta_0$ takes the form $\theta_0 + D_T^{-1}\theta$ (i.e. replacing $T^{-1/2}$ by $D_T^{-1}$, *cf.* (A.1)). The pseudo-likelihood ratio is given by

$$\mathscr{L}(1,k;\theta,\Sigma) = \frac{|\Sigma_0 + T^{-1/2}\Sigma|^{-k/2}\exp\{-\frac{1}{2}\sum_{t=1}^{k}\xi_t(\theta)'(\Sigma_0+T^{-1/2}\Sigma)^{-1}\xi_t(\theta)\}}{|\Sigma_0|^{-k/2}\exp\{-\frac{1}{2}\sum_{t=1}^{k}\xi_t'\Sigma_0^{-1}\xi_t\}}, \tag{A.33}$$

where $\xi_t(\theta) = Y_t - (U_t'\otimes I)(\theta_0 + D_T^{-1}\theta) = \xi_t - D_T^{-1}(U_t'\otimes I)\theta$. If we define $V_t = ((1/\sqrt{T})X_t', (1/T)(t), 1, w_t')'$, then $\xi_t(\theta) = \xi_t - T^{-1/2}(V_t'\otimes I)\theta$. In this way, the new likelihood ratio (A.33) has the same form as (A.1), except that $\varepsilon_t$ is replaced by $\xi_t$. All notations in the proof of Theorem 2 can be maintained, as long as the $I(1)$ regressors are considered as divided by $\sqrt{T}$ and the linear trend is divided by $T$. We shall adopt this convention here. Note that $V_t$ is now a triangular array, and the usual strong law does not apply to $(1/k)\sum_{t=1}^{k}V_tV_t'$, as $V_t$ contains $I(1)$ components. We only outline the major differences. Note that the search of a break is limited in the region $k\in[T\varepsilon_0, T(1-\varepsilon_0)]$ for some $\varepsilon_0>0$.

Property 1 is a standard result, which still holds regardless of the presence of $I(1)$ regressors or not. With the restriction $k\in[T\varepsilon_0, T(1-\varepsilon_0)]$, Property 2 becomes

$$\sup_{T\geqq k\geqq T\varepsilon_0}\log\mathscr{L}(1,k;\hat\theta_{(k)},\hat\Sigma_{(k)}) = O_p(\log T).$$

However, this is implied by Property 1 because $k\geqq T\varepsilon_0$.

Next consider property 3. The previous proof needs (A.11), whose proof requires that $H_k = (1/k)\sum_{t=1}^{k}V_tV_t'$ and its inverse be $O_p(1)$ for all large $k$. It is easy to prove that $H_k$ and $H_k^{-1}$ are $O_p(1)$ uniformly in $k$ such that $k\geqq T\varepsilon_0$. The rest of the proof is the same as before. Similarly, Property 4 still holds.

Next consider Property 5. The previous proofs require inequalities (A.20) and (A.21). Inequality (A.20) still holds for the linear process $\eta_t = \Sigma_0^{-1}\xi_t$, see Bai (1994) for a proof. Because $k$ is only required to be larger than $Ah_T$ (not a positive fraction of observations), the inverse matrix $H_k^{-1}$ is no longer uniformly bounded over $k$ (in fact, the first $r$ rows and the first $r$ columns of $H_k$ converge to zero for $k=Ah_T$ as $T$ increases), so the proof of (A.21) must be modified. First note that $V_t'(\sum_{i=1}^{k}V_iV_i')^{-1}V_t \equiv U_t'(\sum_{i=1}^{k}U_iU_i')^{-1}U_t$. Now

$$\begin{aligned}
\|\zeta_k\|^2 &= \frac{1}{k^2}\sum_{t=1}^{k}(\eta_t'\otimes V_t')(I\otimes H_k^{-1})\sum_{t=1}^{k}(\eta_t\otimes V_t)\\
&= \frac{1}{k}\sum_{t=1}^{k}(\eta_t'\otimes V_t')(I\otimes[\textstyle\sum_{t=1}^{k}V_tV_t']^{-1})\sum_{t=1}^{k}(\eta_t\otimes V_t)\\
&= \sum_{j=1}^{r}\frac{1}{k}(\textstyle\sum_{t=1}^{k}\eta_{tj}U_t)'(\sum_{t=1}^{k}U_tU_t')^{-1}(\sum_{t=1}^{k}\eta_{tj}U_t)\\
&\leqq m\sum_{j=1}^{r}\sum_{l=1}^{m}\frac{1}{k}(\textstyle\sum_{t=1}^{k}\eta_{tj}U_{tl})^2/(\sum_{t=1}^{k}U_{tl}^2),
\end{aligned}$$

where $\eta_{tj}$ is the $j$-th component of $\eta_t$, $U_{tl}$ is the $l$-th component of $U_t$, $m=\dim(U_t)$. The last inequality follows from the fact that the projection length (squared) of a vector projected onto a $m$-dimensional space is no larger than the sum of the projection length of that vector projected onto each of the $m$ one-dimensional spaces multiplied by $m$. Because $r$ and $m$ are fixed, to prove (A.21), it is sufficient to prove, that for each $j$ and $l$,

$$P\left(\sup_{k\geqq Ah_T}\frac{1}{k}(\textstyle\sum_{t=1}^{k}\eta_{tj}U_{tl})^2/(\sum_{t=1}^{k}U_{tl}^2) > \gamma b_T^2\right) < \varepsilon.$$

We shall use the following result: $P(g(Y,X)>c) = EP(g(Y,X)>c|X)$. For each $j$, $\eta_{tj}$ is a linear process. To conserve space, we only give proofs here for $\eta_t$ being an i.i.d sequence. The general case can be proved using the argument in Bai (1994). By the inequality of Hajek and Renyi (1955), the conditional probability, conditional on the $U$'s, is bounded by

$$\gamma^{-1}b_T^{-2}\sum_{k=Ah_T}^{\infty}\frac{1}{k}\frac{U_{kj}^2}{U_{1j}^2+U_{2j}^2+\cdots+U_{kj}^2}\sigma_{\eta,j}^2,$$

where $\sigma_{\eta,j}^2 = E\eta_{tj}^2$. Thus the unconditional probability satisfies

$$P\left(\sup_{k\geqq Ah_T}\frac{1}{k}(\textstyle\sum_{t=1}^{k}\eta_{kj}U_{kl})^2/(\sum_{t=1}^{k}U_{kl}^2) > \gamma b_T^2\right) \leqq \gamma^{-1}b_T^{-2}\sum_{k=Ah_T}^{\infty}\frac{1}{k^2}E\left(\frac{kU_{kj}^2}{U_{1j}^2+U_{2j}^2+\cdots+U_{kj}^2}\right)\sigma_{\eta,j}^2. \tag{A.34}$$

The expected value involved is uniformly bounded in $k$ by Assumptions 3.3 and 3.4. It is also bounded for $U_{tl}=t$, the trending regressor. Thus the left-hand side of (A.34) is bounded by, for some $M, M'<\infty$,

$$M\gamma^{-1}b_T^{-2}\sum_{k=Ah_T}^{\infty}\frac{1}{k^2}\leq M'\gamma^{-1}\frac{1}{b_T^2Ah_T}<\varepsilon, \qquad \text{for large } A,$$

because $b_T^2h_T$ has a positive limit by assumption. This proves (A.21). The rest of proof is the same.

Next, consider Property 6. The only place that involves $V_t$ is equation (A.22). Using the fact that $V_t=(T^{-1/2}X_t', T^{-1}t, 1, w_t')'$, we see that (A.22) is $O_p(1)$. ‖

*Proof of Theorem* 6.

We consider $v\leq 0$ (i.e. $k\leq k_0$). Analogous to (A.24)–(A.29), the log likelihood ratio can be written as

$$-T/2\,\{\log|\Sigma_0+T^{-1/2}\Sigma|-\log|\Sigma_0|\}-\tfrac{1}{2}\sum_{t=1}^{T}(\xi_t'[(\Sigma_0+T^{-1/2}\Sigma)^{-1}-\Sigma_0^{-1}]\xi_t) \tag{A.35}$$

$$+\beta'\tilde{D}_T^{-1}\sum_{t=1}^{T}Z_t(k_0)(\Sigma_0+T^{-1/2}\Sigma)^{-1}\xi_t \tag{A.36}$$

$$-\tfrac{1}{2}\beta'\tilde{D}_T^{-1}\sum_{t=1}^{T}Z_t(k_0)(\Sigma_0+T^{-1/2}\Sigma)^{-1}Z_t(k_0)'\tilde{D}_T^{-1}\beta \tag{A.37}$$

$$+\sum_{t=k+1}^{k_0}(\delta_T+D_T^{-1}\delta)'S'S(U_t\otimes I)(\Sigma_0+T^{-1/2}\Sigma)^{-1}\xi_t \tag{A.38}$$

$$-\tfrac{1}{2}\sum_{t=k+1}^{k_0}(\delta_T+D_T^{-1}\delta)'S'S(U_t\otimes I)(\Sigma_0+T^{-1/2}\Sigma)^{-1}(U_t'\otimes I)S'S(\delta_T+D_T^{-1}\delta) \tag{A.39}$$

$$-\sum_{t=k+1}^{k_0}(\delta_T+D_T^{-1}\delta)'S'S(U_t\otimes I)(\Sigma_0+T^{-1/2}\Sigma)^{-1}Z_t'(k_0)\tilde{D}_T^{-1}\beta. \tag{A.40}$$

The convergence of (A.35) is already studied. The weak convergence of (A.36) and (A.37) is standard because they do not depend on $k$ but $k_0$ (cointegration regression with fixed interactive dummy variables). Their limit is a quadratic form in $\beta$ as specified in the Theorem. The matrix $Q$ is random (it is easy to derive its concrete expression, but we omit the details). For bounded $\delta$ and $\Sigma$, the limit of (A.38) is determined by

$$\begin{aligned}\sum_{t=k+1}^{k_0}\delta_T'S'S(U_t\otimes I)\Sigma_0^{-1}\xi_t&=\delta_T'S'S(I\otimes\Sigma_0^{-1})\sum_{t=k+1}^{k_0}(U_t\otimes\xi_t)\\&=\delta_0'S'S(I\otimes\Sigma_0^{-1})v_T\sum_{t=k+1}^{k_0}((T^{-1/2}X_t', T^{-1}t, 1, w_t')'\otimes\xi_t).\end{aligned} \tag{A.41}$$

The last equality follows from the assumption on $\delta_T$ and $U_t=(X_t', t, 1, w_t')'$. Now consider the limiting process of $v_T\sum_{t=k+1}^{k_0}(T^{-1/2}X_t\otimes\xi_t)$

$$v_T\sum_{t=k+1}^{k_0}(T^{-1/2}X_t\otimes\xi_t)=T^{-1/2}X_{k_0}\otimes v_T\sum_{t=k+1}^{k_0}\xi_t-T^{-1/2}v_T\sum_{t=k+1}^{k_0}(X_{k_0}-X_t)\otimes\xi_t. \tag{A.42}$$

Now, $T^{-1/2}X_{k_0}=T^{-1/2}\sum_{t=1}^{k_0}\Xi_t\xrightarrow{d}\sqrt{\tau_0}D(1)\Omega_e^{1/2}Z$, where $Z$ is $N(0,I)$, $\tau_0=k_0/T$, $\Omega_e=Ee_te_t'$; $v_T\sum_{t=k_0+[vv_T^{-2}]}^{k_0}\xi_t\Rightarrow C(1)\Omega_\varepsilon^{1/2}\eta(v)$, where $\eta(\cdot)$ is a vector of independent standard Brownian motion processes on $(-\infty,0]$, starting at the origin with reverse time. The process $\eta(v)$ is independent of $Z$. The second term on the r.h.s. of (A.42) is $o_p(1)$. Thus (A.42) converges weakly to $\sqrt{\tau_0}D(1)\Omega_e^{1/2}Z\otimes C(1)\Omega_\varepsilon^{1/2}\eta(v)$. Next

$$v_T\sum_{t=k+1}^{k_0}\left(\frac{t}{T}\right)\xi_t=v_T\frac{k_0}{T}\sum_{t=k+1}^{k_0}\xi_t-v_T\sum_{t=k+1}^{k_0}\left(\frac{k_0-t}{T}\right)\xi_t\Rightarrow\tau_0C(1)\Omega_\varepsilon^{1/2}\eta(v).$$

We have used the fact that, for bounded $v$, $v_T\sum_{t=k+1}^{k_0}T^{-1}(k_0-t)\xi_t=o_p(1)$. Finally,

$$v_T\sum_{t=k+1}^{k_0}(w_t\otimes\xi_t)\Rightarrow[\textstyle\sum_{h=-\infty}^{\infty}E(w_tw_{t-h}')\otimes(E\xi_t\xi_{t-h}')]^{1/2}W^*(v),$$

where $W^*(v)$ is a vector of independent standard Brownian motion processes defined on $(-\infty,0]$, which is also independent of $\eta(v)$. Combining these results together with (A.41), we have

$$\sum_{t=k+1}^{k_0}\delta_TS'S(U_t\otimes I)\Sigma_0^{-1}\xi_t$$

$$\Rightarrow\delta_0'S'S(I\otimes\Sigma_0^{-1})\begin{pmatrix}\begin{pmatrix}\sqrt{\tau_0}D(1)\Omega_e^{1/2}Z\\ \tau_0\\ 1\end{pmatrix}\otimes C(1)\Omega_\varepsilon^{1/2}\eta(v)\\ [\sum_{h=-\infty}^{\infty}E(w_tw_{t-h}')\otimes(E\xi_t\xi_{t-h}')]^{1/2}W^*(v)\end{pmatrix}. \tag{A.43}$$

Using the fact that $b'Y$ has the same distribution as $\sqrt{b'Ab}\cdot N(0,1)$ for $Y\sim N(0,A)$ and $b$ an arbitrary random vector independent of $Y$, we see that the right-hand side of (A.43) has the same distribution as

$\{\delta_0'S'SH_1S'S\delta_0\}^{1/2}W(v)$, where $H_1$ is given in Theorem 7 and $W(v)$ is a scalar standard Brownian motion on $(-\infty, 0]$. Thus

$$\sum_{t=k+1}^{k_0} \delta_T'S'S(U_t\otimes I)\Sigma_0^{-1}\xi_t \Rightarrow \{\delta_0'S'SH_1S'S\delta_0\}^{1/2}W(v)=\sqrt{c_1}W(v).$$

Next, consider (A.39). For bounded $\delta$ and $\Sigma$, the limiting process of (A.39) is determined by

$$-\tfrac{1}{2}\sum_{t=k+1}^{k_0} \delta_T'S'S(U_t\otimes I)\Sigma_0^{-1}(U_t'\otimes I)S'S\delta_T=-\tfrac{1}{2}\delta_T'S'S(\textstyle\sum_{t=k+1}^{k_0} U_tU_t'\otimes\Sigma_0^{-1})S'S\delta_T.$$

Using similar arguments as above, the limiting process is shown to be $-\frac{1}{2}|v|\delta_0'S'SH_2S'S\delta_0=-\frac{1}{2}|v|c_2$, where $H_2$ is defined in Theorem 6.

Expression (A.40) can be shown to be $o_p(1)$. Combining these results, we obtain the limiting process of the log-likelihood ratio. Finally note that $c_1, c_2$ and $W(v)$ are determined by $O(v_T^{-2})$ number of observations whereas $\kappa$, $Q$ and $\Psi$ are determined by the entire set of observations. The latter will not be changed if we delete $O(v_T^{-2})$ observations that determine $c_1, c_2$, and $W(v)$. This gives rise to the asymptotic independence of $(c_1, c_2, W(v))$ and $(\kappa, Q, \Psi)$. The independence of $(c_1, c_2)$ and $W(v)$ follows from the independence of $\xi_t$ and the regressors. ||

*Proof of Theorem* 7.

The limiting process is maximized for $\beta$ at $Q^{-1}\kappa$, for $\Sigma$ at $\Psi$, and for $v$ at $\text{argmax}_s\{\sqrt{c_1}W(s)-|s|c_2/2\}$. From the reparameterization and the continuous mapping theorem we have $\tilde{D}_T(\hat{\beta}-\beta_0)\xrightarrow{d}Q^{-1}\kappa$, $T^{1/2}(\hat{\Sigma}-\Sigma_0)\xrightarrow{d}\Psi$, and $v_T^2(\hat{k}-k_0)\xrightarrow{d}\text{argmax}_s\{\sqrt{c_1}W(s)-|s|c_2/2\}$. By a change in variable, we have $\text{argmax}_s\{\sqrt{c_1}W(s)-|s|c_2/2\}=\text{argmax}_s\{W(s)-(c_2/\sqrt{c_1})|s|/2\}=(c_1/c_2^2)\,\text{argmax}_r\{W(r)-|r|/2\}$. This implies that

$$(c_2^2/c_1)v_T^2(\hat{k}-k_0)\Rightarrow\text{argmax}_s\{W(s)-|s|/2\}.$$

The last part of Theorem 7 follows from the definition of $c_1$ and $c_2$. ||

*Proof of Corollary* 7.1.

This corresponds to a special $S$ such that $S=(S_1\otimes I)$, with $S_1U_t=1$, which is the constant regressor. For this $S$, we have $S\delta_0=\mu_0$, and $SH_2S'=1\otimes\Sigma_0^{-1}=\Sigma_0^{-1}$. Thus $\delta_0'S'SH_2S'S\delta_0=\mu_0'\Sigma_0^{-1}\mu_0$. Similarly, $\delta_0'S'SH_1S'S\delta_0=\mu_0'\Sigma_0^{-1}C(1)\Omega_\varepsilon C(1)'\Sigma_0^{-1}\mu_0$. By Theorem 7

$$\frac{(\mu_0'\Sigma_0^{-1}\mu_0)^2}{\mu_0'[\Sigma_0^{-1}C(1)\Omega_\varepsilon C(1)'\Sigma_0^{-1}]\mu_0}v_T^2(\hat{k}-k_0)\xrightarrow{d}V^*.$$

By definition, $\mu_T=\mu_0v_T$. Replacing $\mu_0v_T$ by $\mu_T$ yields (3.6). ||

*Proof of Corollary* 7.2.

The proof is similar to that of Corollary 7.1. The $S$ has the form $S=(S_1\otimes I)$ with $S_1U_t=t$. In this case $SH_2S'=\tau_0^2\otimes\Sigma_0^{-1}=\tau_0^2\Sigma_0^{-1}$. Similar to Corollary 7.1,

$$\tau_0^2\frac{(\gamma_0'\Sigma_0^{-1}\gamma_0)^2}{\gamma_0'[\Sigma_0^{-1}C(1)\Omega_\varepsilon C(1)'\Sigma_0^{-1}]\gamma_0}v_T^2(\hat{k}-k_0)\xrightarrow{d}V^*.$$

By definition, $\gamma_0v_T=T\gamma_T$. Replacing $\gamma_0v_T$ by $T\gamma_T$, and replacing $\tau_0$ by $k_0/T$ yields (3.7). ||

*Proof of Corollary* 7.3.

This corresponds to another special $S$ such that $S=(S_1\otimes I)$ with $S_1U_t=X_t$. This implies that $S\delta_0=a_0$. In addition, $SH_2S'=G\otimes\Sigma_0^{-1}$, where $G=\tau_0D(1)\Omega_e^{1/2}ZZ'\Omega_e^{1/2}D(1)'$, which is the upper left block of the matrix $\Phi$ defined in Theorem 7. Similarly, $SH_1S'=G\otimes\Sigma_0^{-1}C(1)\Omega_\varepsilon C(1)'\Sigma_0^{-1}$. Thus Theorem 7 implies that

$$\frac{[a_0'(G\otimes\Sigma_0^{-1})a_0]^2}{a_0'[G\otimes\Sigma_0^{-1}C(1)\Omega_\varepsilon C(1)'\Sigma_0^{-1}]a_0}v_T^2(\hat{k}-k_0)\xrightarrow{d}V^*.$$

By definition, $a_0v_T=\sqrt{T}a_T$. In addition, $G$ is the limit of $T^{-1}X_{k_0}X_{k_0}'$. Thus, replacing $a_0v_T$ *by* $\sqrt{T}a_T$ and replacing $G$ by $T^{-1}X_{k_0}X_{k_0}'$ gives (3.8). ||

*Acknowledgements.* The authors are grateful to Anil Bera, John Campbell, Neil Ericsson, Andrew Harvey, David Hendry, Ken Kasa, Deborah Nolan, Pierre Perron, Thomas Rothenberg, Neil Shephard, Katsuto Tanaka, C. Z. Wei, the editor, and three anonymous referees for helpful comments.

REFERENCES

ANDREWS, D. W. K. (1993), "Tests for Parameter Instability and Structural Change with Unknown Change Point", *Econometrica*, **61**, 821–856.

ANDREWS, D. W. K., LEE, I. and PLOBERGER, W. (1996), "Optimal Changepoint Tests for Normal Linear Regression", *Journal of Econometrics*, **70**, 9–38.

ANDREWS, D. W. K. and PLOBERGER, W. (1994), "Optimal Tests when a Nuisance Parameter is Present only under the Alternative", *Econometrica*, **62**, 1383–1414.

BAI, J. (1992), "Estimation of Structural Change in Econometric Models" (Unpublished Ph.D. dissertation, University of California, Berkeley).

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

BANERJEE, A., LUMSDAINE, R. L. and STOCK, J. H. (1992), "Recursive and Sequential Tests of the Unit Root and Trend Break Hypotheses: Theory and International Evidence", *Journal of Business and Economic Statistics*, **10**, 271–288.

CAMPOS, J., ERICCSON, N. R. and HENDRY, D. F. (1996), "Cointegration Tests in the Presence of Structural Breaks", *Journal of Econometrics*, **70**, 187–220.

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

DESHAYES, J. and PICARD, D. (1986), "Off-line Statistical Analysis of Change-Point Models Using Non-Parametric and Likelihood Methods", in Basseville, M. and Benveniste, A. (eds.), *Detection of Abrupt Changes in Signals and Dynamical Systems* (Berlin: Springer Verlag) 103–168.

DICKEY, D. A. and FULLER, W. A. (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", *Journal of the American Statistical Association*, **74**, 427–431.

ENGLE, R. F. and GRANGER, C. W. J. (1987), "Cointegration and Error Correction: Representation, Estimation, and Testing", *Econometrica*, **55**, 251–276.

GREGORY, A. W. and HANSEN, B. E. (1996), "Residual-Based Tests for Cointegration in Models with Regime Shifts", *Journal of Econometrics*, **70**, 99–126.

HAJEK, J. and RENYI, A. (1955), "Generalization of an Inequality of Kolmogorov", *Acta. Math. Acad. Sci.*, **6**, 281–283.

HALL, P. and HEYDE, C. C. (1980) *Martingale Limit Theory and its Applications* (New York: Academic Press).

HAMILTON, J. D. (1994) *Time Series Analysis* (Princeton: Princeton University Press).

HANSEN, B. E. (1992), "Testing for Parameter Instability in Regressions with $I(1)$ Processes", *Journal of Business and Economic Statistics*, **10**, 321–336.

HAUG, A. A. (1996), "Tests for Cointegration: A Monte Carlo Comparison", *Journal of Econometrics*, **71**, 89–116.

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

HORVATH, M. T., and WATSON, M. W. (1995), "Testing for Cointegration when Some of the Cointegrating Vectors are Prespecified", *Econometric Theory*, **11**, 984–1014.

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

KING, R. G., PLOSSER, C. I. and REBELO, S. P. (1988), "Production, Growth, and Business Cycles: II, New Directions", *Journal of Monetary Economics*, **21**, 309–342.

KING, R. G., PLOSSER, C. I., STOCK, J. H. and WATSON, M. W. (1991), "Stochastic Trends and Economic Fluctuations", *American Economic Review*, **81**, 819–840.

PERRON, P. (1989), "The Great Crash, the Oil Price Shock and the Unit Root Hypothesis", *Econometrica*, **57**, 1361–1401.

PHILLIPS, P. C. B. (1991), "Optimal Inference in Cointegrated Systems", *Econometrica*, **59**, 283–306.

PICARD, D. (1985), "Testing and Estimating Change-Points in Time Series", *Advances in Applied Probability*, **176**, 841–867.

QUANDT, R. E. (1960), "Tests of the Hypothesis that a Linear Regression System Obeys Two Separate Regimes", *Journal of the American Statistical Association*, **55**, 324–330.

RAPPOPORT, P. and REICHLIN, L. (1989), "Segmented Trends and Non-Stationary Time Series", *Economic Journal*, **99**, 168–177.

SARGENT, T. (1989), "Two Models of Measurements and the Investment Accelerator", *Journal of Political Economy*, **97**, 251–287.

SIEGMUND, D. (1988), "Confidence Sets in Change-point Problems", *International Statistical Review*, **56**, 31–48.

STOCK, J. H. (1987), "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors", *Econometrica*, **55**, 1035–1056.

STOCK, J. H. (1994), "Unit Roots, Structural Breaks, and Trends", in Engle, R. and McFadden, D. (eds.), *Handbook of Econometrics*, Volume IV (Amsterdam: Elsevier), 2740–2843.

STOCK, J. H. and WATSON, M. W. (1993), "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems", *Econometrica*, **61**, 783–820.

WATSON, M. W. (1994), "Vector Autoregression and Cointegration", in Engle, R. and McFadden, D. (eds.), *Handbook of Econometrics*, Volume IV (Amsterdam: Elsevier) 2844–2918.

YAO, Y-C. (1987), "Approximating the Distribution of the Maximum Likelihood Estimate of the Change-Point in a Sequence of Independent Random Variables", *Annals of Statistics*, **15**, 1321–1328.

ZIVOT, E. and ANDREWS, D. W. K. (1992), "Further Evidence on the Great Crash, the Oil Price Shock, and the Unit Root Hypothesis", *Journal of Business and Economic Statistics*, **10**, 251–270.