Journal of Economic Perspectives—Volume 15, Number 4—Fall 2001—Pages 117–128

The New Econometrics of Structural Change: Dating Breaks in U.S. Labor Productivity

Bruce E. Hansen

Applied time series analysis and forecasting is based on the assumption of stationarity—the constancy of parameters like the mean, variance and trend over time. But what happens if the parameters change? A particularly common assertion is that United States labor productivity experienced a "slowdown" around 1973 and a "speedup" in the second half of the 1990s. While one can certainly pick groups of years in the 1970s and 1980s where the average annual productivity growth is lower than it was in the 1960s and the late 1990s, this ad hoc selection of convenient time periods hardly qualifies as serious analysis.

Structural change is a statement about parameters, which only have meaning in the context of a model. To focus our discussion, we will discuss structural change in the simplest dynamic model, the first-order autoregression:

$$y_t = \alpha + \rho y_{t-1} + e_t$$

$$Ee_t^2 = \sigma^2,$$

where e_t is a time series of serially uncorrelated shocks. The parameters are (α, ρ, σ^2). The assumption of *stationarity* implies that these parameters are constant over time. We say that a *structural break* has occurred if at least one of these parameters has changed at some date—the *breakdate*—in the sample period. While it may seem unlikely that a structural break could be immediate and might seem more reasonable to allow a structural change to take a period of time to take effect, we most

■ *Bruce E. Hansen is the Stockwell Professor of Economics, University of Wisconsin, Madison, Wisconsin. His e-mail address is ⟨bhansen@ssc.wisc.edu⟩.*

often focus on the simple case of an immediate structural break for simplicity and parsimony.

A structural break may affect any or all of the model parameters, and these cases have different implications. Changes in the autoregressive parameter ρ reflect changes in the serial correlation in y_t. The intercept α controls the mean of y_t through the relationship $E(y_t) = \mu = \alpha/(1 - \rho)$. Since y_t is the growth rate in labor productivity, changes in μ are identical to changes in the trend and are probably the issue of primary interest. Finally, changes in σ^2 imply changes in the volatility of labor productivity.

The econometrics of structural change looks for systematic methods to identify structural breaks. In the past 15 years, the most important contributions to this literature include the following three innovations: 1) Tests for a structural break of unknown timing; 2) Estimation of the timing of a structural break; and 3) Tests to distinguish between a random walk and broken time trends. These three innovations have dramatically altered the face of applied time series econometrics. We discuss these three topics in turn and use U.S. labor productivity data to illustrate their applicability.

Testing for Structural Change of Unknown Timing

The classical test for structural change is typically attributed to Chow (1960). His famous testing procedure splits the sample into two subperiods, estimates the parameters for each subperiod, and then tests the equality of the two sets of parameters using a classic F statistic. This test was popular for many years and was extended to cover most econometric models of interest. For a recent treatment, see Andrews and Fair (1988).

However, an important limitation of the Chow test is that the breakdate must be known a priori. A researcher has only two choices: to pick an arbitrary candidate breakdate or to pick a breakdate based on some known feature of the data. In the first case, the Chow test may be uninformative, as the true breakdate can be missed. In the second case, the Chow test can be misleading, as the candidate breakdate is endogenous—it is correlated with the data—and the test is likely to indicate a break falsely when none in fact exists. Furthermore, since the results can be highly sensitive to these arbitrary choices, different researchers can easily reach quite distinct conclusions—hardly an example of sound scientific practice.

To illustrate this point, let's take United States labor productivity in the manufacturing/durables sector. We measure this as the growth rate of the ratio of the Industrial Production Index for manufacturing/durables to average weekly labor hours, a monthly time series available from February 1947 to April 2001 (yielding 651 observations).[1] If we compute a Chow statistic using 1973 as the

[1] Average weekly hours is the number of employees multiplied by average weekly hours, as measured by the Bureau of Labor Statistics.

breakdate, we obtain a value of 3.9.[2] The 5 percent critical value from the chi-square distribution is 6.0, so we fail to find evidence of a structural break. If we instead compute a Chow statistic using 1975 as the breakdate—just two years forward—we obtain a value of 7.1. As this exceeds the 5 percent critical value, it appears to provide evidence of a structural break. But how can we be confident, as the two similar breakdates give such different answers?

The necessary solution is to treat the breakdate as unknown. This idea—and solution—goes back to Quandt (1960), who proposed taking the largest Chow statistic over all possible breakdates.[3] For illustration, imagine a devilish trickster who is willing to go to any extreme to prove the existence of a structural break in labor productivity. This devilish trickster performs every possible Chow test imaginable, searching across possible breakdates.[4] Through this search, this trickster finds the worst-case Chow statistic, the breakdate where the test is largest. This is Quandt's statistic.

One way to see the construction of this statistic is to plot the sequence of Chow statistics as a function of candidate breakdates. I have done so in Figure 1 for our labor productivity example. The candidate breakdates are along the x-axis; the values of the Chow statistics on the y-axis. To compute these Chow statistics for a particular breakdate, you split the sample at that breakdate and estimate the model parameters separately on each subsample, as well as their covariance matrices. If the true parameters are constant, the subsample estimates should be (roughly) constant across candidate breakdates—subject to estimation error. On the other hand, if there is a structural break, then the subsample estimates will vary systematically across candidate breakdates, and this will be reflected in the Chow test sequence.

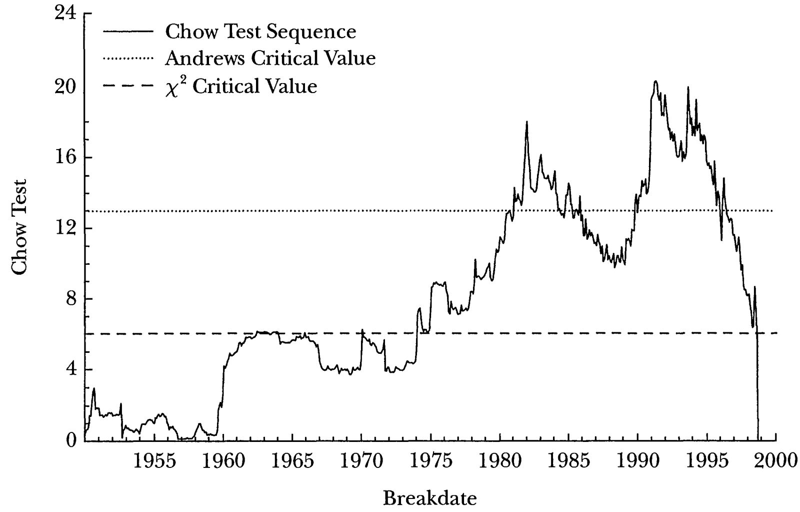
In Figure 1, we can see considerable variation of the Chow test sequence across candidate breakdates, reaching a high of 20.2 in May 1991. This value—20.2—is the Quandt statistic.

If the breakdate is known a priori, then the chi-square distribution can be used to assess statistical significance. We sketch this 5 percent critical value with dashes in Figure 1. For breakdates where the Chow test sequence lies below this critical value, the test appears to be "insignificant," and conversely for breakdates where the Chow test lies above the critical value.

However, if the breakdate is unknown a priori, then the chi-square critical values are inappropriate. What critical values should be used instead? For many years, this question remained unanswered, and the Quandt statistic had no practical application. In the early 1990s, the problem was solved simultaneously by several sets of authors, with the most elegant and general statements given by

[2] I use the Wald form of the statistic computed with a heteroskedasticity-consistent covariance matrix.
[3] This is the likelihood ratio test under normality.
[4] Actually, we do not consider literally "all" possible breakdates. We cannot consider breakdates too close to the beginning or end of the sample, as there are not enough observations to identify the subsample parameters. The conventional solution is to consider all breakdates in the interior τ percent to $(1 - \tau)$ percent of the sample, where the trimming parameter τ is typically between 5 percent and 15 percent. In the examples presented here, I use 5 percent trimming.

Figure 1
Testing for Structural Change of Unknown Timing: Chow Test Sequence as a Function of Breakdate



Andrews (1993) and Andrews and Ploberger (1994).[5] These authors provide tables of critical values, and Hansen (1997) provides a method to calculate *p*-values.

These asymptotic critical values are considerably larger than the comparable chi-square critical values, depending on the number of parameters in the model and other factors. In our example (which has two parameters), the Andrews 5 percent critical value is 12.9, just over twice the chi-square critical value. We have sketched this critical value as well (dotted line) in Figure 1. A visual way to assess significance is to see if the Chow test sequence breaks above the critical value (as this is equivalent to the maximum of the sequence exceeding the critical value). We can see that the maximum (20.2) easily exceeds the Andrews critical value (the *p*-value is 0.0016), so we are easily able to reject the hypothesis of no structural break.[6] We are therefore quite confident that this time series has a structural break.

If we find evidence of one structural break, could there be more than one? Bai and Perron (1998) develop tests for multiple structural changes. Their method is sequential, starting by testing for a single structural break. If the test rejects the null hypothesis that there is no structural break, the sample is split in two (based on the breakdate estimate presented in the next section) and the test is reapplied to each

[5] Andrews and Ploberger (1994) show that improved power can be obtained by taking exponential averages of the Chow test sequence.
[6] Hansen (2000) shows that Andrews' critical values are not robust to structural change in the *marginal* distribution of the regressors, which is undesirable in tests focusing on *conditional* relationships. He shows how to simulate robust critical values on a case-by-case basis. In this example, the robust *p*-value (calculated with 10,000 bootstrap replications) is 0.0044, yielding the same conclusion.

subsample. This sequence continues until each subsample test fails to find evidence of a break.

The Quandt-Andrews and Andrews-Ploberger family of statistics have essentially replaced the Chow statistic in recent econometric practice. One comprehensive application is Stock and Watson (1996), who apply the tests systematically to 76 monthly time series using both univariate and bivariate regressions. They reject stability at the 10 percent level for over half of their models.

Another interesting application is Ben-David and Papell (1998), who look for evidence of "slowdowns" (a decrease in the trend function) in the Summers-Heston GDP data from 74 countries. They find statistically significant evidence of a slowdown in 46 countries. In 21 of these cases, the postbreak trend function is actually negative.

A final example getting considerable recent attention is McConnell and Perez-Quiros (2000). They test for the stability of the volatility of United States GDP growth rates and find overwhelming evidence of a substantial decrease in volatility around 1984.

Estimating the Timing of Structural Change

In many applications, it is useful to know when the structural change occurred. Treating the date of structural change—the breakdate—as an unknown parameter, the issues are how to estimate the breakdate and how to obtain confidence intervals for the breakdate.

An obvious candidate for a breakdate estimate is the date that yields the largest value of the Chow test sequence (in our labor productivity example, May 1991). It turns out that this is known to be a good estimate only in one special case—in linear regressions when the Chow test is constructed with the "homoskedastic" form of the covariance matrix.[7]

In regression models, an appropriate method to estimate the parameters—including the breakdate—is least squares. Operationally, the sample is split at each possible breakdate, the other parameters estimated by ordinary least squares and the sum of squared errors calculated and stored. The least squares breakdate estimate is the date that minimizes the full-sample sum of squared errors (equivalently, minimizes the residual variance).

A theory of least squares estimation has been developed in a sequence of papers by Jushan Bai, both alone and with coauthors. Bai (1994, 1997a) derives the asymptotic distribution of the breakdate estimator and shows how to construct

[7] When the Wald test is calculated using the homoskedastic covariance matrix, it is algebraically a transformation of the sum of squared residuals. Thus, the date that maximizes this Wald test sequence is algebraically identical to the date that minimizes the sum of squared residuals—the least squares estimator.

confidence intervals for the breakdate.[8] These confidence intervals are easy to calculate and hence are very useful in applications, as they indicate the degree of estimation accuracy. Bai, Lumsdaine and Stock (1998) extend this analysis to multiple time series with simultaneous structural breaks. They show that using multiple time series improves estimation precision. Bai and Perron (1998) discuss simultaneous estimation of multiple breakdates.

Chong (1995) and Bai (1997b) show how to estimate multiple breakdates sequentially. The key insight is that when there are multiple structural breaks, the sum of squared errors (as a function of the breakdate) can have a local minimum near each breakdate. Thus, the global minimum can be used as a breakdate estimator, and the other local minima can be viewed (cautiously) as candidate breakdate estimators. The sample is then split at the breakdate estimate, and analysis continues on the subsamples. Bai (1997b) shows that important improvements are obtained by iterative refinements: reestimation of breakdates based on refined samples.

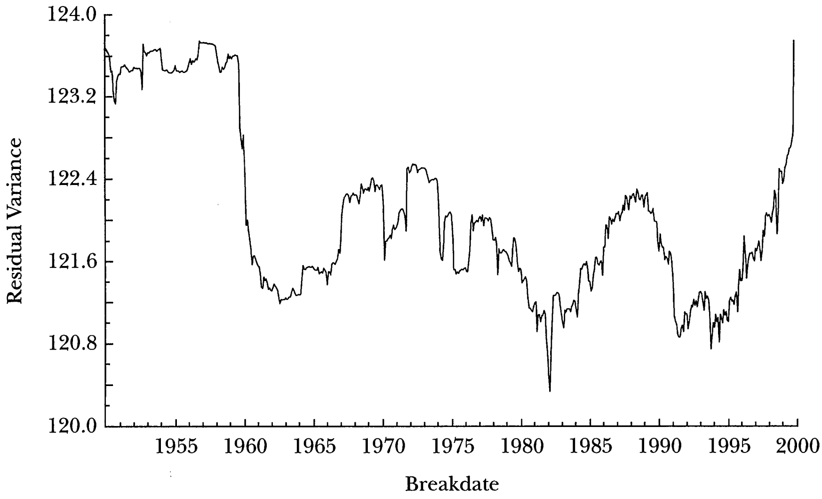
These methods can be best illustrated through our application to labor productivity. In Figure 2, we plot the residual variance (sum of squared errors divided by sample size) as a function of a single breakdate. The breakdates are on the x-axis and the residual variance on the y-axis. The sample is split at each breakdate and the regression parameters estimated separately on each subsample. The sum of squared errors is calculated (for the entire sample) and the residual variance plotted. If the true parameters are constant, the subsample estimates (and hence, sum of squared errors) will vary randomly and erratically across candidate breakdates. On the other hand, if there is a structural break, then the subsample estimates will vary systematically across candidate breakdates, and the sum of squared errors will have a well-defined minimum near the true breakdate.

The plot in Figure 2 is partially erratic, but we can discern three well-defined minima: a global minimum in January 1982 and two local minima in July 1962 and September 1993. The visual evidence suggests that two or three structural breaks are possible in this sample period.

We break the sample at the estimated breakdate (January 1982) and test for structural breaks on the two subsamples. We find no evidence for a break in the period [1947, 1982], but find evidence for a break in the period [1982, 2001]. For this latter period, the residual variance has a strong *V* shape as a function of the breakdate, indicating good identification, and the minimum is obtained in December 1994. Now we split the sample in December 1994 and reestimate on the sample period [1947, 1994]. The Quandt test rejects the hypothesis of parameter constancy at the 5 percent level, indicating a structural break, and the least squares estimate of the breakdate is December 1963, with a second local minimum in January 1982. Now taking the sample [1964, 1994], the Quandt test fails to find evidence for a structural break. The point estimate of the breakdate again is January 1982. Finally,

[8] This builds on earlier work by Picard (1985).

Figure 2
Least Squares Breakdate Estimation: Residual Variance as a Function of Breakdate



taking the sample [1964, 2001], the Quandt statistic finds evidence of a break with a breakdate estimate of April 1994.

Based on this evidence, there appears to have been a structural break in this series in 1994, and possibly breaks in December 1963 and January 1982. The Bai 90 percent confidence interval for the first breakdate is [1992, 1996], which is fairly tight. The 90 percent confidence intervals for the other two breakdates are [1959, 1971] and [1977, 1988], respectively, neither of which is very precise.

As yet, there have been few rigorous applications of the methodology described here. An important exception is Bai, Lumsdaine and Stock (1998), who attempt to date the alleged slowdown of the early 1970s. Using U.S. quarterly data for 1959 through 1995 on real output, consumption and investment, they find no evidence of structural change when examining the individual series with univariate models, but find strong evidence in a joint vector autoregression, in which the output, consumption and investment variables are regressed on lagged values of output, consumption and investment. Their estimate of the breakdate is the first quarter of 1969, and their 90 percent confidence interval puts the breakdate between the second quarter of 1966 and the fourth quarter of 1971.

An interesting common feature of our breakdate estimates for labor productivity and the Bai, Lumsdaine and Stock (1998) estimate for real output is that they are both quite different than the widely accepted breakdate of 1973. (Neither set of confidence intervals include 1973.) Another common feature is that both sets of estimates are consistent with a structural break in the late 1960s.

Choosing Between Random Walk and Structural Change

Time series are often usefully described as being composed of a trend and a cycle. Before the work of Nelson and Plosser (1982), it was commonplace to assume that the trend was linear. Nelson and Plosser challenged that assumption by providing evidence that for many widely used aggregate macroeconomic time series, the trend could be characterized as a random walk. That is, instead of being a fixed trend to which the time series would revert over the business cycle, the trend would be moved by random shocks—and then would stay at the new level until disturbed by another random shock.

While this result prompted many counterchallenges, the most constructive was mounted by Perron (1989). Perron argued that the movement of the trend could be explained by a parsimonious single structural break in an otherwise constant linear trend. This explanation is plausible, since a trend break produces serial correlation properties that are similar to those of a random walk.

Perron (1989) showed how to test the random walk hypothesis against the trend-break model. This is achieved by estimating a linear autoregression augmented with dummy interactions to capture the desired broken trend specification. The hypothesis of a random walk trend implies that the sum of the autoregressive coefficients equals one (that is, a "unit root" in the autoregressive polynomial), so this can be easily assessed with a *t*-ratio statistic. The distribution of the *t*-ratio is non-normal,[9] but Perron provided a distribution theory and critical values.

Perron (1989) applied this test to the Nelson-Plosser macroeconomic time series, specifying the breakdate as 1929 for the annual series and 1973 for the postwar quarterly series. He was able to reject the random walk model for most of the series at the 5 percent significance level, suggesting that the series were stationary after accounting for structural change in the trend.

However, Perron's (1989) analysis was disputed by a collection of papers, most notably Christiano (1992), Zivot and Andrews (1992), Banerjee, Lumsdaine and Stock (1992) and even Perron and Vogelsang (1992). These authors argue that it is inappropriate to specify the breakdate as known, as it is not reasonable to believe that the choice has been made independently of the data. These authors collectively suggest that an appropriate procedure is to select the breakdate that provides the most evidence against the random walk hypothesis—the breakdate that produces the largest *t*-ratio.

This procedure for choosing a breakdate can produce the same numerical value for the test statistic, as Perron's choices of 1929 and 1973 were quite judicious. However, the test is constructed using a different *procedure*, and hence it has a different sampling distribution. The critical values for the modified test are much larger,[10] making it harder to reject the null hypothesis of a random walk. Armed

[9] It is also different from the Dickey-Fuller (1979) distribution.

[10] The authors also found that asymptotic critical values should only be used as a crude guide and that finite-sample bootstrap critical values are much preferred.

with this new theory, the evidence against the hypothesis of a random walk had evaporated.

Will longer samples of data settle the debate? Perron (1997) revisited the issue, extending the sample to 1991:III and using different methods to select the autoregressive lag order. He found slightly stronger evidence against the random walk model, but the evidence was not conclusive. The key problem is that the trend functions specified earlier in Perron (1989) do not predict well out of sample. For the annual series, the trend function severely overpredicts the 1970s and 1980s. For the quarterly series, it underpredicts for the period 1987–2000. This is consistent with a random walk trend.

The contributions to this family of tests continue to grow. Of particular note, Lumsdaine and Papell (1997) allow two breakdates rather than one and find that the case against the random walk is strengthened. However, the need for two structural breaks also reduces the distinction between the trend-break and random walk models.[11]

The Perron (1989) idea has had a large and well-deserved impact on empirical analysis and has focused attention on the time series properties of the trend. As we now understand, the distinction between a random walk and a trend break largely concerns the frequency of permanent shocks to the trend. In a random walk process, such shocks occur frequently, while in a trend-break process, they occur infrequently (once or twice in a sample). Future work may attempt to find alternative ways to narrow the difference between these models.

Perron's (1989) idea and its variants have seen many applications. One creative example of this work is Fernandez (1997). His focus is on the question of whether changes in money help to forecast output, even after conditioning on lagged output. An earlier literature had shown that the results depend on whether or not interest rates are included in the regression and whether a time trend is included to detrend the series linearly. Fernandez uses tests from Perron (1997) to argue that output is well represented by a stationary process about a time trend with a single trend break, and he uses the estimated broken trend function to detrend output. Fernandez finds that when the sample period is confined to pre-1985, this produces very robust results; however, he is unable to produce robust results when data after 1985 is included.

Another application is Papell, Murray and Ghiblawi (2000). These authors are concerned with hysteresis in unemployment rates in 16 OECD countries. Hysteresis is the theory that a one-time change in unemployment can have permanent effects; thus, it is closely related to the idea that trend unemployment can be described as a random walk. Using the Perron-Vogelsang (1992) tests, the authors are able to reject the random walk hypothesis for ten of the 16 countries in favor of a one-time

[11] In a recent contribution, Saikkonen and Lutkepohl (2000) propose tests with optimal power through efficient estimation of the model parameters. Their testing methods are related to Elliott, Rothenberg and Stock (1996) in that they use efficient generalized least squares.

break in time trend. This finding suggests a very different economic interpretation about hysteresis.

U.S. Productivity

We now return to our empirical investigation of labor productivity in the U.S. manufacturing/durables sector. To review, we found strong evidence of a structural break sometime between 1992 and 1996, and weaker evidence of a structural break in the 1960s and the early 1980s. What is the nature of these changes in labor productivity? Our subsample estimates of mean growth rates (in annualized units) are 3.4 percent for 1947–1964, 2.5 percent for 1964–1982, 4.2 percent for 1982–1995 and 7.7 percent for 1995–2001. Clearly, the growth rate in the final period is quite large relative to previous history.

As there are many possible measures of labor productivity, it is natural to ask if our results are robust to alternative choices. We consider the quarterly labor productivity index issued by the Bureau of Labor Statistics, which measures output per hour of all persons for the manufacturing/durables sector. Applying the same model and methods to this series (1947 through 2000), we again find strong evidence for a break in the 1990s, with a point estimate of 1997. However, with this series, there is no evidence of a second break.

To investigate further, we disaggregate the manufacturing/durables sector by two-digit SIC industry group, as monthly series are available at this level of disaggregation for both the industrial production index and weekly labor hours. The ten industries are the following: Lumber (SIC 24); Furniture (SIC 25); Stone, clay and glass products (SIC 32); Primary metal industries (SIC 33); Fabricated metal products (SIC 34); Industrial machinery and equipment (SIC 35); Electronic equipment (SIC 36); Transportation equipment (SIC 37); Instruments (SIC 38); and Miscellaneous (SIC 39).

We apply the same empirical methods as described earlier in this paper for the manufacturing/durables sector. There is evidence of structural change in the regression function for seven of the ten industry groups, but only two—Industrial machinery and Electronic equipment—show evidence of a structural break in the mean growth rate. For the other five industry groups—Furniture, Primary Metals, Fabricated Metals, Instruments and Miscellaneous—there is a statistically significant change in the autoregressive parameter, but not in the mean growth rate. This means that the response of labor productivity to shocks has changed, but the long-run impact is unaltered.

As we stated above, the industrial machinery sector has a statistically significant break in the mean growth rate. The breakdate estimate is February 1992, and the Bai 90 percent confidence interval is October 1990 to June 1994. The model implies a mean growth rate (in annualized units) of 3.3 percent before 1992 and 7.8 percent after the break. There is no evidence of a second structural break in this series.

The other series with a break in the mean growth rate is electronic equipment. The breakdate estimate is December 1993, and the Bai 90 percent confidence interval is June 1993 to January 1995. The estimated mean growth rate increased from 5.3 percent before 1993 to 17.8 percent after 1993.

There is a remarkable coincidence of the breakdates for the manufacturing/durables sector as a whole and the industrial machinery and electronic equipment industry groups. Apparently, the structural break in the sector is due to roughly simultaneous breaks in these two industry groups.

This investigation raises many questions: Is this break permanent or transitory? Can further disaggregation identify the sources of the productivity break? Is the timing of the structural break simultaneous across sectors? Why are there no spillover effects into other industries? Can the labor productivity gain be explained by increased utilization of other factors? Are these increases unique to the United States, or have they occurred in other countries as well? I expect that these and related questions will be exciting avenues for future research.

Conclusion

Structural change is pervasive in economic time series relationships, and it can be quite perilous to ignore. Inferences about economic relationships can go astray, forecasts can be inaccurate, and policy recommendations can be misleading or worse. The new tools developed in the past few years are useful aids in econometric model specification, analysis and evaluation.

■ *I thank the editors—Alan Krueger, Brad De Long, Michael Waldman and Timothy Taylor—for detailed and helpful comments on a previous draft, which greatly improved the paper. I also thank John Kennan and Ken West for insightful suggestions about labor productivity. This research was supported by a grant from the National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author, and they do not necessarily reflect the views of the NSF. The data and Gauss programs used in this paper are available on the author's webpage, ⟨http://www.ssc.wisc.edu/~bhansen⟩.*

References

Andrews, Donald W.K. 1993. "Tests for Parameter Instability and Structural Change with Unknown Change Point." *Econometrica.* July, 61:4, pp. 821–56.

Andrews, Donald W.K. and Ray C. Fair. 1988. "Inference in Nonlinear Econometric Models with Structural Change." *Review of Economic Studies.* October, 55:4, pp. 615–39.

Andrews, Donald W.K. and Werner Ploberger. 1994. "Optimal Tests When a Nuisance Parameter is Present Only Under the Alternative." *Econometrica.* November, 62:6, pp. 1383–414.

Bai, Jushan. 1994. "Least Squares Estimation of a Shift in Linear Processes." *Journal of Time Series Analysis.* 15:5, pp. 453–72.

Bai, Jushan. 1997a. "Estimation of a Change Point in Multiple Regression Models." *Review of Economics and Statistics.* November, 79:4, pp. 551–63.

Bai, Jushan. 1997b. "Estimating Multiple Breaks One at a Time." *Econometric Theory.* June, 13:3, pp. 315–52.

Bai, Jushan and Pierre Perron. 1998. "Estimating and Testing Linear Models with Multiple Structural Changes." *Econometrica.* January, 66:1, pp. 47–78.

Bai, Jushan, Robin L. Lumsdaine and James H. Stock. 1998. "Testing for and Dating Common Breaks in Multivariate Time Series." *Review of Economic Studies.* July, 64:3, pp. 395–432.

Banerjee, Anindya, Robin L. Lumsdaine and James H. Stock. 1992. "Recursive and Sequential Tests of the Unit-Root and Trend-Break Hypotheses: Theory and International Evidence." *Journal of Business and Economic Statistics.* July, 10:3, pp. 271–87.

Ben-David, Dan and David H. Papell. 1998. "Slowdowns and Meltdowns: Postwar Growth Evidence from 74 Countries." *Review of Economics and Statistics.* November, 80:4, pp. 561–71.

Chong, Terence Tai-leung. 1995. "Partial Parameter Consistency in a Misspecified Structural Change Model." *Economics Letters.* October, 49:4, pp. 351–57.

Chow, Gregory C. 1960. "Tests of Equality Between Sets of Coefficients in Two Linear Regressions." *Econometrica.* 28:3, pp. 591–605.

Christiano, Lawrence J. 1992. "Searching for a Break in GNP." *Journal of Business and Economic Statistics.* July, 10:3, pp. 237–50.

Dickey, David A. and Wayne A. Fuller. 1979. "Distribution of the Estimators for Autoregressive Time Series with a Unit Root." *Journal of the American Statistical Association.* June, 74:366, pp. 427–31.

Elliott, Graham, Thomas J. Rothenberg and James H. Stock. 1996. "Efficient Tests for an Autoregressive Unit Root." *Econometrica.* July, 64:4, pp. 813–36.

Fernandez, David G. 1997. "Breaking Trends and the Money-Output Correlation." *Review of Economics and Statistics.* November, 79:4, pp. 674–79.

Hansen, Bruce E. 1997. "Approximate Asymptotic *P* Values for Structural-Change Tests." *Journal of Business and Economic Statistics.* January, 15:1, pp. 60–67.

Hansen, Bruce E. 2000. "Testing for Structural Change in Conditional Models." *Journal of Econometrics.* July, 97:1, pp. 93–115.

Lumsdaine, Robin L. and David H. Papell. 1997. "Multiple Trend Breaks and the Unit-Root Hypothesis." *Review of Economics and Statistics.* May, 79:2, pp. 212–18.

McConnell, Margaret M. and Gabriel Perez-Quiros. 2000. "Output Fluctuations in the United States: What Has Changed Since the Early 1980s?" *American Economic Review.* December, 90:5, pp. 1464–76.

Nelson, Charles R. and Charles I. Plosser. 1982. "Trends and Random Walks in Macroeconomic Time Series." *Journal of Monetary Economics.* September, 10:2, pp. 139–62.

Papell, David H., Christian J. Murray and Hala Ghiblawi. 2000. "The Structure of Unemployment." *Review of Economics and Statistics.* May, 82:2, pp. 309–15.

Perron, Pierre. 1989. "The Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis." *Econometrica.* November, 57:6, pp. 1361–401.

Perron, Pierre. 1997. "Further Evidence on Breaking Trend Functions in Macroeconomic Variables." *Journal of Econometrics.* October, 80:2, pp. 355–85.

Perron, Pierre and Timothy J. Vogelsang. 1992. "Nonstationarity and Level Shifts with an Application to Purchasing Power Parity." *Journal of Business and Economic Statistics.* July, 10:3, pp. 301–20.

Picard, D. 1985. "Testing and Estimating Change-Points in Time Series." *Advances in Applied Probability.* 17:4, pp. 841–67.

Quandt, Richard. 1960. "Tests of the Hypothesis that a Linear Regression Obeys Two Separate Regimes." *Journal of the American Statistical Association.* 55, pp. 324–30.

Saikkonen, Pentti and Helmut Lutkepohl. 2000. "Testing for a Unit Root in a Time Series with a Level Shift at Unknown Time." Contributed paper, Economic Society World Congress. *Econometric Theory.* Forthcoming.

Stock, James H. and Mark W. Watson. 1996. "Evidence on Structural Instability in Macroeconomic Time Series Relations." *Journal of Business and Economic Statistics.* July, 14:3, pp. 11–30.

Zivot, E. and D.W.K. Andrews. 1992. "Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis." *Journal of Business and Economic Statistics.* 10, pp. 251–70.