SOME MONTE CARLO RESULTS ON NONPARAMETRIC CHANGEPOINT TESTS

by Edward Bryden, John B. Carlson, and Ben Craig

Edward Bryden is a statistician at Booze, Allen, and Hamilton, Inc., Cleveland. John B. Carlson is an economist and Ben Craig is an economic advisor at the Federal Reserve Bank of Cleveland. The authors would like to thank Ben Keen for excellent research assistance. This paper was presented at the Midwest Macro Conference, Michigan State University, on September 16, 1995.

December 1995

Abstract

For long periods since 1982, core inflation has behaved as if it were generated by a process with a fixed mean and serially independent error term. Nonparametric changepoint tests proposed by Pettitt (1979) and Lombard (1987) suggest that since 1982, changes in core inflation have been infrequent and rather abrupt. However, little is known about the small-sample properties, the power of the tests, or the robustness of changepoint tests when a series is not i.i.d. This paper uses Monte Carlo analysis to investigate the probabilities of false positive tests under alternative assumptions about the time-series properties of the underlying process.

1. Introduction

In many situations with economic time series, researchers are faced with the question of whether an underlying probability distribution has changed in some distinct way. To illustrate, consider a continuous distribution $F(x, \theta_t)$ on a sequence of independent random variables $x_1, \ldots, x_T$. If $\theta_1 = \ldots = \theta_\tau = \theta$, while $\theta_{\tau+1}, \ldots, \theta_T$ differ in some unknown way, the sequence is said to have a changepoint at τ. Often, the changepoint problem is examined under specific assumptions regarding the form of the underlying distribution. The distributional assumptions, however, may be controversial. Moreover, evidence suggests that some established procedures may be quite sensitive to deviations from assumed distributional forms.

In light of these issues, Lombard (1987) argues that initial analysis requires procedures that are robust against deviations from distributional assumptions. He proposes a nonparametric procedure for identifying changepoints. By replacing data with functions of rank statistics, nonparametric techniques provide a distribution-free test of the null hypothesis of no change. Such methods also offer some protection against potential effects of spurious outliers.

The gains from generality, however, have potential costs, particularly in terms of the power of the test. Without a specific distribution, we have no analytic means for assessing the power lost. For example, how quickly does the Lombard procedure identify a changepoint when it does occur? Moreover, most economic time series exhibit some degree of autocorrelation. It is thus useful to assess how sensitive Lombard's tests are to deviations from the serial independence assumption. To address these issues, we present some results from Monte Carlo experiments designed to estimate significance levels in

small samples, the likelihood that a changepoint will be detected by a given time period, and the probability of a false positive in the presence of serial correlation. Although our analysis is limited to univariate estimators, we offer an application that illustrates their potential usefulness.

The rest of the article is organized as follows. Section 2 describes alternative nonparametric estimators of changepoints. The design of the Monte Carlo analysis is presented in section 3, while the results are discussed in section 4. To illustrate how our results may aid an initial analysis of the changepoint problem, we provide an application of the nonparametric procedures to a measure of core inflation -- the 15 percent trimmed mean. These results are given in section 5. Section 6 presents a discussion of our findings in a more general context. We offer some concluding thoughts in the final section.

2. Nonparametric Changepoint Tests

In his widely cited paper, Pettitt (1979) offers an appealing nonparametric test to detect changepoints based on the Mann-Whitney two-sample test. Using his notation, let

$$U_{t,T} = 2W_t - t(T-1),$$

where

$$W_t = \sum_{i=1}^{t} R_i .$$

$U_{t,T}$ is equivalent to the Mann-Whitney statistic for testing whether two samples are from the same population.

For the hypothesis H_0: no change versus H_A: change, Pettitt proposes the statistic

$$K_T = \max_{1 \leq t < T} |U_{t,T}|.$$

Pettitt shows that the significance value k of K_T is approximated by

$$p_{OA} = \exp\{-6k^2/(T^3 + T^2)\},$$

which is accurate to two decimal places for $p_{OA} \leq 0.5$.

Although some situations may dictate that a changepoint occurred rather abruptly, it is often more realistic to assume that a change occurs smoothly over a period of time. For this purpose, Lombard introduces a smooth-change specification:

$$\theta_i = \begin{cases} \xi_1 & (i \leq \tau_1) \\ \xi_1 + (\iota - \tau_1)(\xi_2 - \xi_1)/(\tau_2 - \tau_1) & (\tau_1 < i \leq \tau_2) \\ \xi_2 & (i > \tau_2), \end{cases} \qquad (1.1)$$

where ξ_1, ξ_2, τ_1, and τ_2 are unknown. Note that the abrupt-change model is a special case where $\tau_2 = \tau_1 + 1$. Moreover, an onset of a trend is a special case characterized by $\tau_2 = T$ and $\tau_1 < \tau_2 - 1$. Using standard rank statistic notation, the rank of x_i is denoted as r_i. The rank score of x_i is given by

$$s(r_i) = [\phi\{r_i/T+1\} - \overline{\phi}]/A \quad (1 \leq i \leq T),$$

where ϕ is an arbitrary score function, $\overline{\phi}$ is the average value, and A is the standard deviation of that function. When changepoints τ_1 and τ_2 are known (e.g., $\tau_1 = t_1$ and $\tau_2 = t_2$),

Lombard suggests

$$v_{t_1, t_2} = \sum_{j=t_1+1}^{t_2} \sum_{i=1}^{j} s(r_i) \tag{1.2}$$

as a rank test statistic to test H_0: $\xi_1 = \xi_2$ of the smooth-change model (1.1). When τ_1 and

τ_2 are unknown, he suggests rejecting H_0 for large values of the statistic

$$q_T = \sum_{t_1=1}^{T-1} \sum_{t_2=t_1+1}^{T} v_{t_1, t_2}^2. \tag{1.3}$$

An interesting special case of the smooth-change model is when $\tau_1 = \tau$ and $\tau_2 = T$.

Lombard calls this the onset-of-trend model because the parameter θ_t is initially stable,

but slowly increases or decreases after time τ. Under this constraint, (1.3) reduces to

$$q_T^* = \sum_{t=1}^{T} v_{t, T}^2$$

as a rank test statistic to test the null hypothesis of no change.

For each of these test statistics, Lombard derives the asymptotic distributions

based on null hypotheses and provides a table of significance points. Asymptotic

significance points are shown to be applicable when sample sizes are at least 30. A

method for estimating both τ_1 and τ_2 is also provided.

A single abrupt-change test emerges as a special case where $\tau_1 = \tau$ and $\tau_2 = \tau + 1$.

Lombard denotes this statistic as $m_{1,T}$, where

$$m_{1,T} = \sum_{t=1}^{T-1} v_{t, t+1}^2.$$

He shows that under the null hypothesis, $T^{-2} m_{1,T}$ converges in distribution to the limiting form of the Cramer-von Mises goodness-of-fit criterion, for which significance points are available in Anderson and Darling (1952, p. 203).

In the case of multiple abrupt changes, Lombard suggests

$$m_{k,T} = \Sigma_{(k)} \sum_{i=1}^{k+1} \left\{ \sum_{j=\tau_{i-1}}^{\tau_i} s(r_j) \right\}^2,$$

where $\Sigma_{(k)}$ denotes summation over indices $1 \le \tau_1 < \ldots < \tau_k < T$, as a test statistic for the null hypothesis that $\xi_1 = \ldots = \xi_k$. This statistic converges in distribution. Lombard (1987) provides asymptotic significance points for cases $k = 2$ and $k = 3$ (table 2, p. 609).

3. Experimental Design

Monte Carlo methods are used here to estimate significance levels in small samples, the power of each test, and the sensitivity of the tests to deviations from the assumption of serial independence. To estimate small-sample significance levels for a given asymptotic critical value of 5 percent, we generate at least 5,000 samples of varying length. The various tests are applied to each of the generated series, and the percentage of trials that reject the null hypothesis is the estimated significance level.

To assess the power of each test, we generate series of varying initial lengths, each normally distributed with zero mean and unitary standard deviation. To each of these series, we append additional terms generated by the same distribution, but with a different mean. A battery of tests is applied sequentially. The first time a test rejects the null hypothesis, its position in the series is tabulated and no further tests are performed. From these data, we obtain the percentage of trials for which the null is rejected for each

additional term. The cumulative sum of the percentages is our estimate of the percentage of tests that detect a change in mean by a given period, i.e., the power of the test.

Finally, to assess the sensitivity of each test to the assumption of serial independence, we generate a set of autocorrelated series (first order) for each of varying lengths and for alternative values of ρ. The mean of each series is unchanged. The percentage of times the null is rejected is hence a measure of the percentage of false positives when the assumption of serial independence is violated.

4. Results

Table 1 reports the estimated significance levels for simulated series of a standardized normal random variable and for which the asymptotic critical value of the test is 5 percent. Generally, all tests perform better as the sample size increases, i.e., estimated significance levels tend to approach 5 percent. The smooth-change test seems best for the smallest sample size, while the Pettitt test consistently performs less favorably than other tests, but especially in small samples.

Figure 1a compares the powers of the Lombard and the Pettitt tests when there is a one-standard-deviation increase in the mean. It is clear that the Lombard test dominates. The Pettitt test statistic is evaluated at its estimated changepoint. If we were to restrict the test to accept the null hypothesis only if the changepoint estimate were exactly correct, the power difference would be even greater.[1] The Lombard test statistic does not depend on an estimate of the changepoint.

---

[1] For example, by the sixth period the Pettitt test found the correct changepoint only 2.52 percent of the time.

Figure 1b illustrates that the power of both tests improves significantly for two

standard deviations and that the Lombard test still dominates the Pettitt test. It is

interesting that the power of all tests is higher in periods immediately following a change

when the initial sample size is small. It appears as though power is greatest when the

change occurs in the middle of the sample.

Figure 2 illustrates the estimated probabilities of detecting a smooth change with a

smooth-change test. Each panel contrasts results for both one- and two-standard-

deviation changes that occur smoothly over 12 periods. Because the change is spread out

over a year, the smooth-change test takes around 15 months to detect a one-standard-

deviation change with a probability of 0.5. A two-standard-deviation change occurring

within 12 periods, on the other hand, is detected within 10 months. The power of the

smooth-change test also seems highest when a change occurs in the middle of a sample.

Nevertheless, initial sample size does not appear to be especially important.

Figure 3 illustrates the estimated probabilities for detecting an onset of trend with

an onset-of-trend test. Each panel contrasts results for onsets of trend occurring at both

one-half and one standard deviation per 12 periods. Again, initial sample size matters

little.

The powers of both the smooth-change and Lombard one-change tests for

detecting an abrupt one-standard-deviation change are compared in figure 4. We find

little difference between the two tests when a change occurs after an initial sample of 12.

Although the one-change test is slightly better at detecting an abrupt change between 7

and 12 periods, the smooth-change test performs better after 15 periods. We suspect, however, that these differences are due to limited trials.

Similarly, the powers of both the smooth-change and onset-of-trend tests for detecting an onset of trend are compared in figure 5. The simulated trend change occurs at a rate of one standard deviation per year after an initial sample of 24. The onset-of-trend test is slightly better after the tenth period, while the smooth-change test appears to be better immediately after the break.

Finally, table 2 presents estimated significance levels for each of the tests when first-order autocorrelation is present. These values indicate the percentage of false positives associated with each test for five values of ρ and for varying initial sample sizes. Not surprisingly, a high degree of positive serial correlation leads to a high proportion of false positives, especially as sample size increases.

For a ρ of 0.2, the estimated significance levels hover around 10 percent for the Lombard one-change, smooth-change, and onset-of-trend tests for all sample sizes. The same value of ρ leads to considerably higher estimated significance levels for the Lombard three-change tests. In the latter, the significance level increases with sample size.

For negative values of ρ, all estimates of significance levels are below 5 percent. Thus, negative autocorrelation tends to bias tests against the null. Although sample size appears to be irrelevant, significance levels tend to decline with sample size when ρ equals -0.3 and to increase with sample size when ρ equals -0.1.

To summarize, we find that sample size can matter for some nonparametric tests. The smooth-change test suggested by Lombard appears to be the most powerful under most circumstances. Nevertheless, less dominant tests such as the Pettitt can corroborate changepoint dates. Our Monte Carlo results indicate that serial dependence can matter.

5. An Application

Since 1982, core inflation — as measured by the 15 percent trimmed mean — has behaved much differently than it did over the previous 15 years (see figure 6). Indeed, relative to the earlier period, core inflation appears to have become a stationary process with little or no serial correlation. Figure 7 illustrates more clearly that around May 1988, inflation appears to have moved higher. In each of the five months following May, the trimmed mean registered persistently above its previous average. The measure varied around this higher level until February 1991, when its mean appears to have moved permanently lower. The abruptness of these changes suggests that the underlying process experienced at least two permanent changes in its mean over the sample period. Thus, as an initial analysis, it would seem suitable to use parametric changepoint methods to test this hypothesis.

Table 3 presents test statistics for the various changepoint tests in selected samples. These results confirm that there were at least three changepoints: one abrupt changepoint in May 1988, another in January 1991, and a smooth change between June 1991 and July 1993 (see figure 8). When applied to the whole sample, all changepoint tests indicate significant location changes, with the most likely being an abrupt change of about one and a half standard deviations immediately after January 1991.

We split the sample at this point and found the other two changepoints within each subsample. The Lombard estimates for the smooth-change range also indicated an abrupt change in May 1988, the same date that the Pettitt test estimates. Stratification around these changes failed to produce any evidence of additional changes.

As our Monte Carlo results indicate, the presence of autocorrelation can bias the tests toward rejecting the null of no change. We therefore examine the autocorrelation functions for periods both between and across changepoints. Figure 9a, for example, illustrates the estimated autocorrelation functions for selected periods from January 1983 to January 1991. In no case is there any evidence of autocorrelation. The Ljung-Box-Pierce statistic fails to reject the hypothesis that the error process is white noise. Thus, we conclude that the data support the use of our techniques, and we accept the alternative that a changepoint occurred in May 1988.

Similarly, figure 9b illustrates the estimated autocorrelation functions beginning in February 1991. Although there is some indication of negative serial correlation, we suspect that this may reflect the preliminary nature of the seasonal used to adjust the data. Nevertheless, the Ljung-Box-Pierce statistic fails to reject the hypothesis of white noise for either sample. We conclude that the data support the hypothesis of a smooth decline in the mean between June 1991 and July 1993. Since July 1993, the trimmed mean has averaged less than 3 percent.

6. Discussion

We must emphasize that our application of the nonparametric methods is meant to be an initial analysis of the time-series properties of one measure of core inflation.

Nevertheless, we are surprised at how far these techniques can take us. Our results raise some important questions about the conventional wisdom concerning the inflation process. For example, many economists believe that inflation is either highly autocorrelated or nonstationary. The data examined above indicate that for periods as long as 65 quarters, the trimmed mean appears to have been generated by a process with a fixed mean and no serial correlation.

Although we make no claims about what theoretical models may account for the regularities uncovered, we believe the results may provide some guidance in forming modeling strategies. For example, do the periods of stationarity reflect particular monetary regimes? If so, one would clearly want to consider the relevant restrictions on the policy-reaction function implied by the facts. Moreover, what accounts for the abrupt changes? Perhaps S-s type models could account for adjustment in inflation. Our purpose here is not to provide a basis for what we find, but to illustrate how useful some simple empirical techniques are in an initial investigation.

7. Conclusions

On the basis of our Monte Carlo simulations, we conclude that the smooth-change test suggested by Lombard is the preferred test in most situations, particularly when the researcher has no knowledge about the location of the change. The estimated significance levels in small samples were closest to asymptotic values for this test, while its power was at least as good in almost all cases. If one suspects an onset of trend in the series began more than 10 periods earlier, we recommend using the onset-of-trend test.

Although one might expect that nonparametric tests lack power, our experiment reveals that these techniques are not so bad. A one-standard-deviation change in mean is generally detectable half the time within 12 periods. The power improves if the initial sample is even smaller. Not surprisingly, the presence of positive serial correlations biases all tests toward accepting the null.

While the application we propose seems well suited to the techniques applied, it illustrates a need to extend our analysis in several directions. First, we recognize a need to provide some common parametric changepoint test as a benchmark, particularly for assessing the power lost by choosing nonparametric tests. It would be useful to nest our tests into a richer framework that would enable us to discriminate between a changepoint and a series that is autocorrelated. These issues will be addressed in a forthcoming extension of this study.

References

Anderson, T.W., and D.A. Darling. "Asymptotic Theory of Certain 'Goodness of Fit' Criteria Based on Stochastic Processes," *Annals of Mathematical Statistics.* vol. 63 (1952), pp. 193-212.

Box, George E.P., and Gwilym M. Jenkins. *Time Series Analysis: Forecasting and Control.* San Francisco: Holden-Day, 1976.

Bryan, Michael F., and Stephen Cecchetti. "Measuring Core Inflation," Federal Reserve Bank of Cleveland, Working Paper No. 9304, June 1993.

Lombard, F. "Rank Tests for Changepoint Problems," *Biometrica*, vol. 74, no. 3 (1987), pp. 615-24.

Pettitt, A.N. "A Nonparametric Approach to the Changepoint Problem," *Applied Statistician*, vol. 28 (1979), pp. 126-35.

Figure 1a. Estimated Probability that a One-Standard-Deviation Change in Mean Will Be Detected by a Specified Period
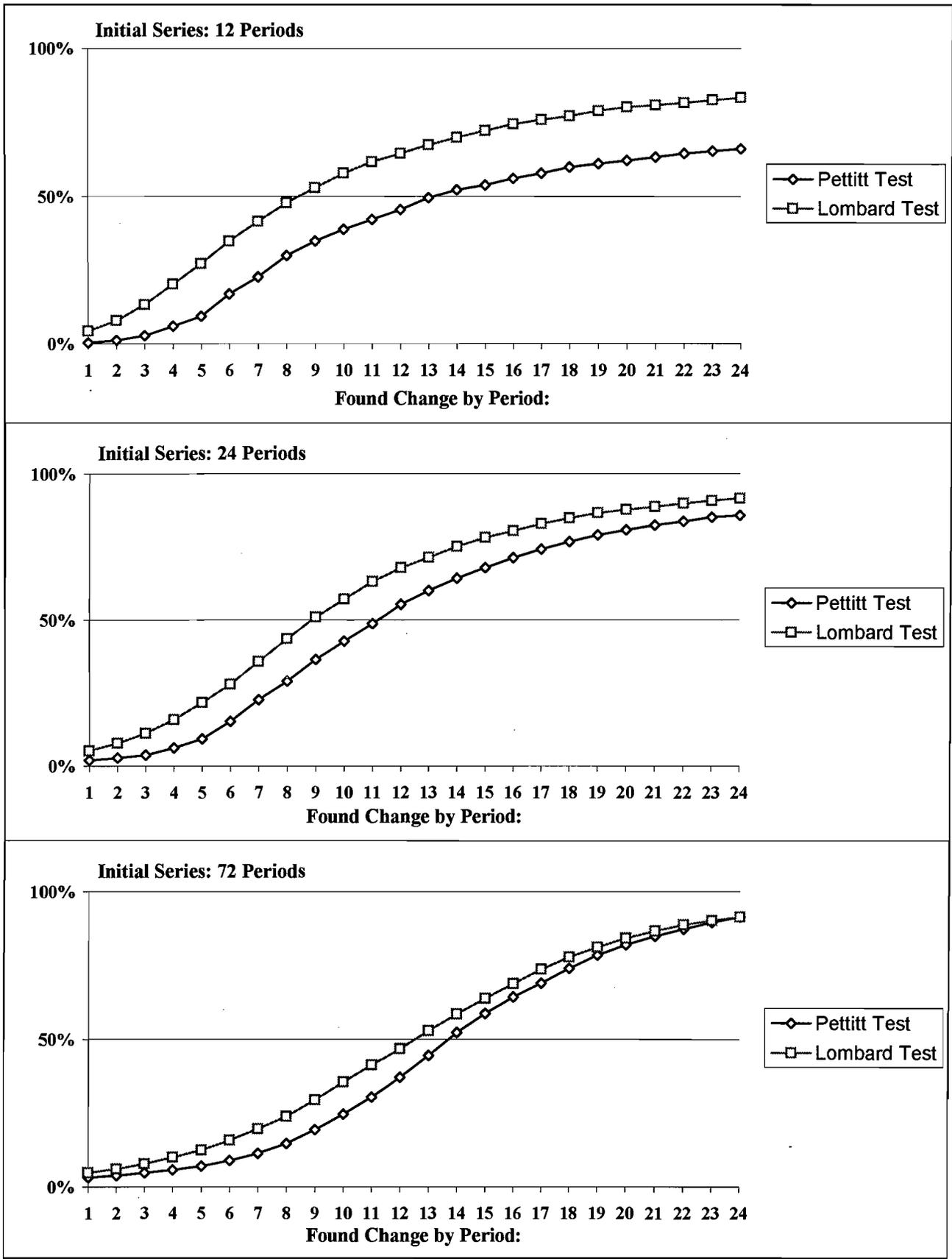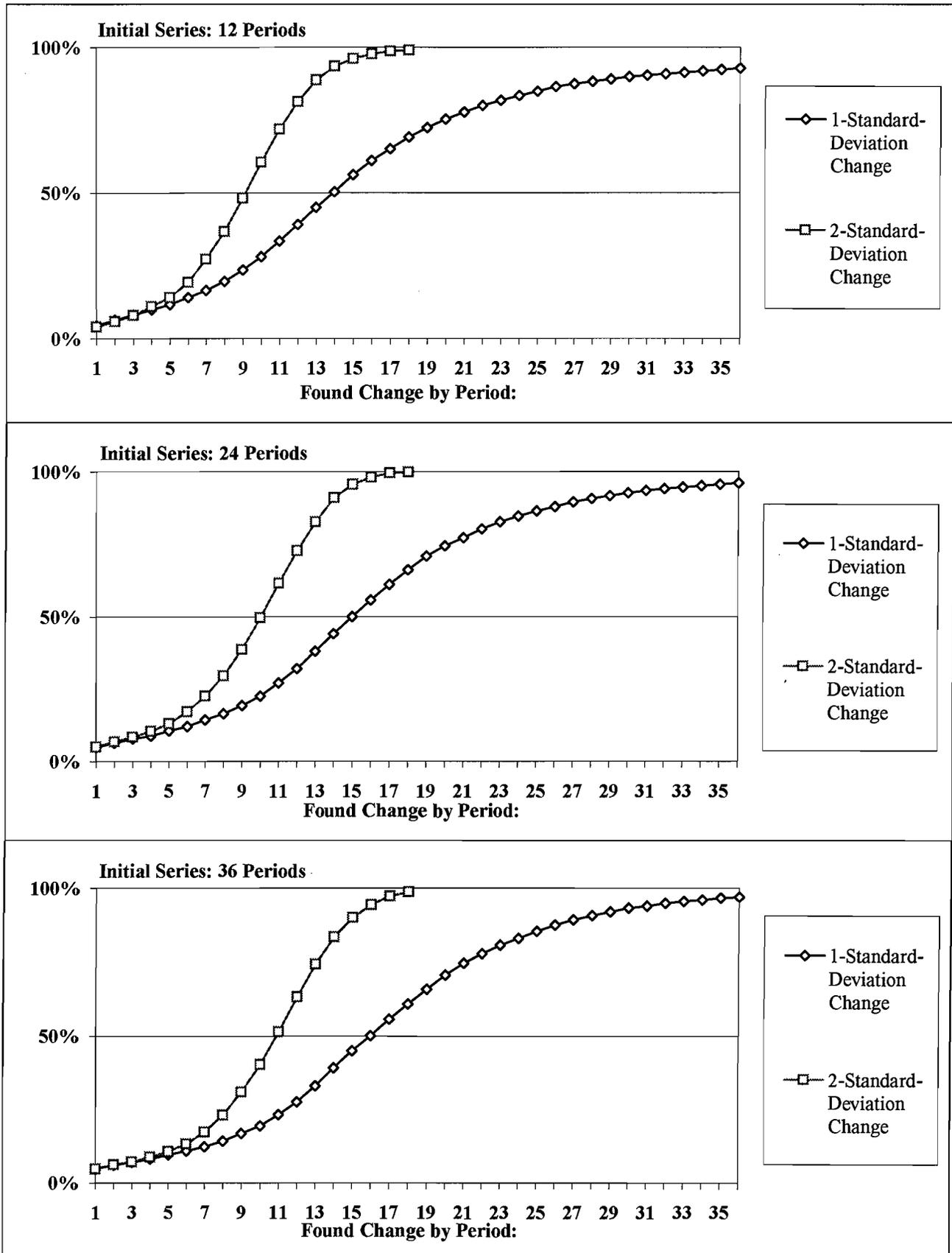

Source: Authors' calculations.

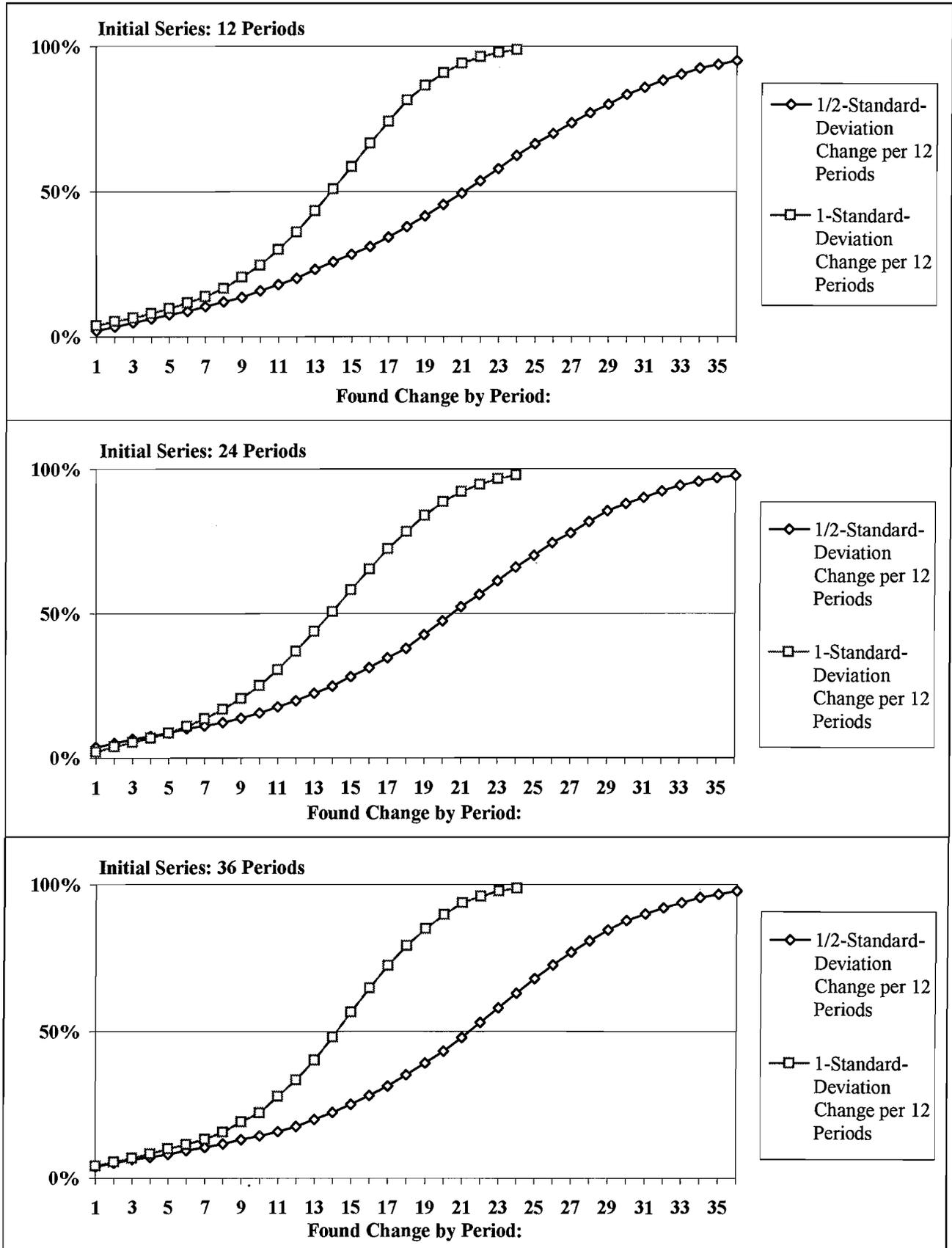**Estimated Probability that a Two-Standard-Deviation Change in Mean Will Be Detected by a Specified Period**
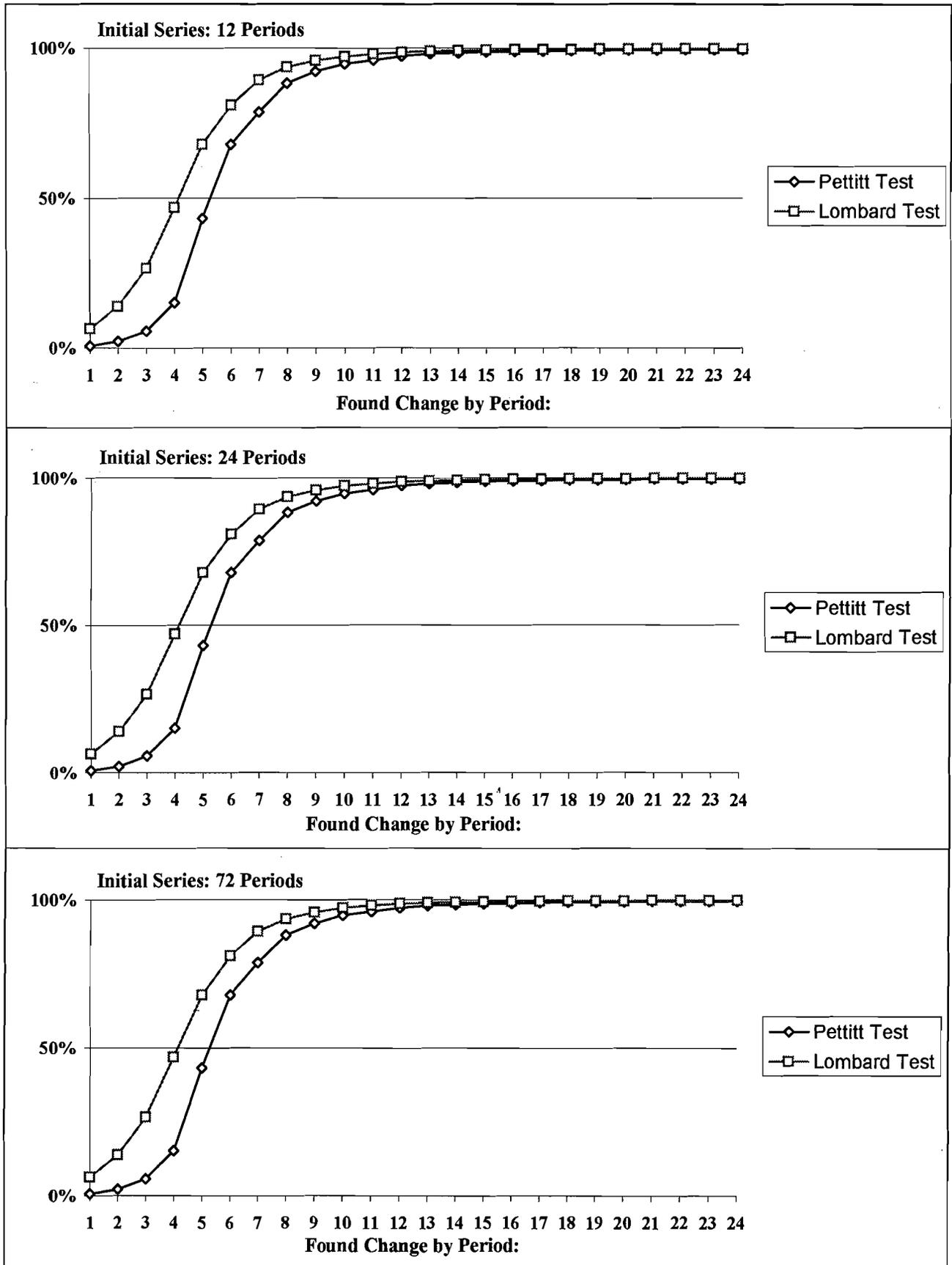
Figure 2: Estimated Probability that a Smooth Change over 12 Periods Will Be Detected by a Specified Period with a Smooth-Change Test
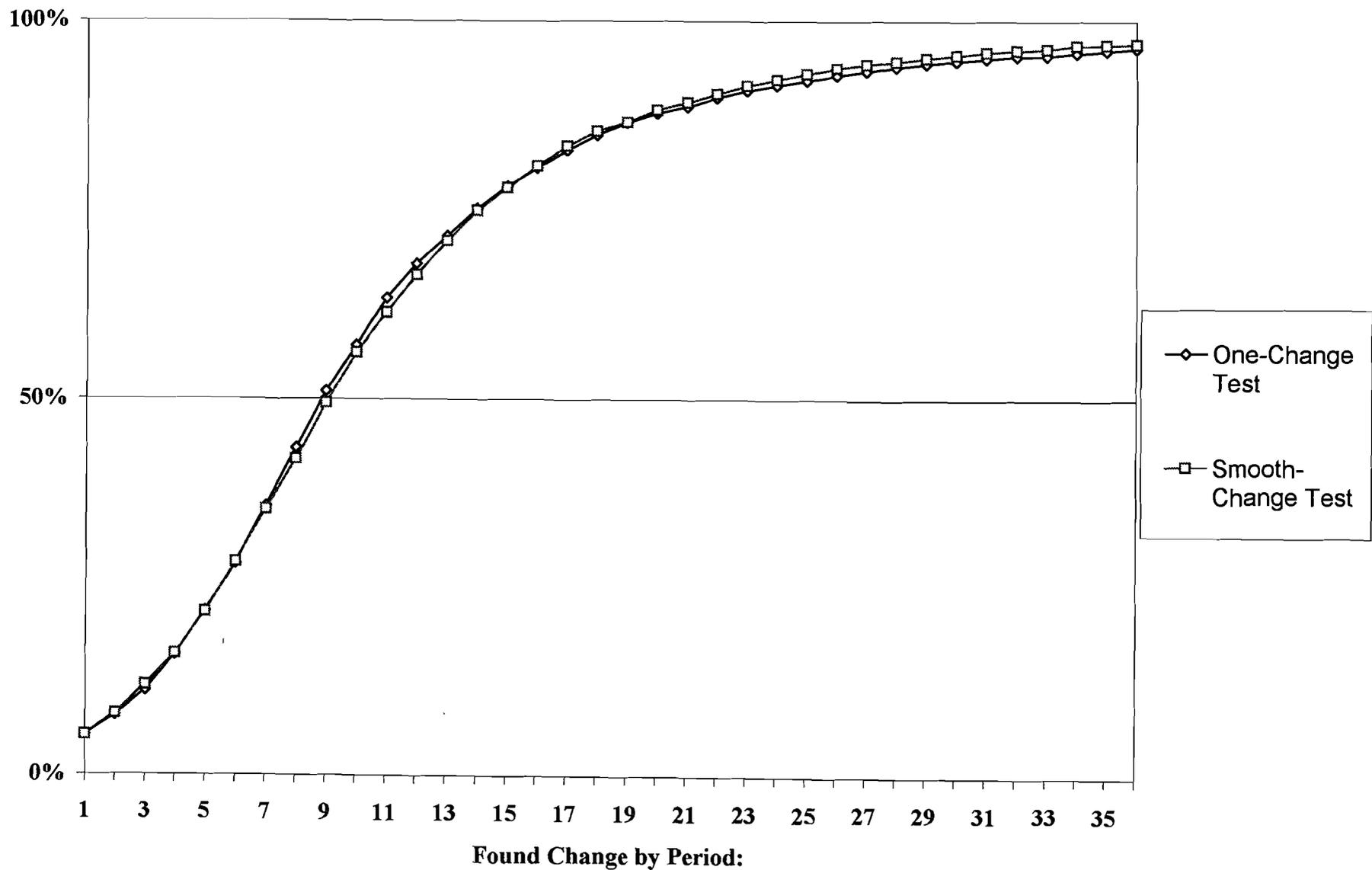

Source: Authors' calculations.

Figure 3: Estimated Probability that an Onset-of-Trend Change Will Be Detected by a Specified Period with an Onset-of-Trend Test
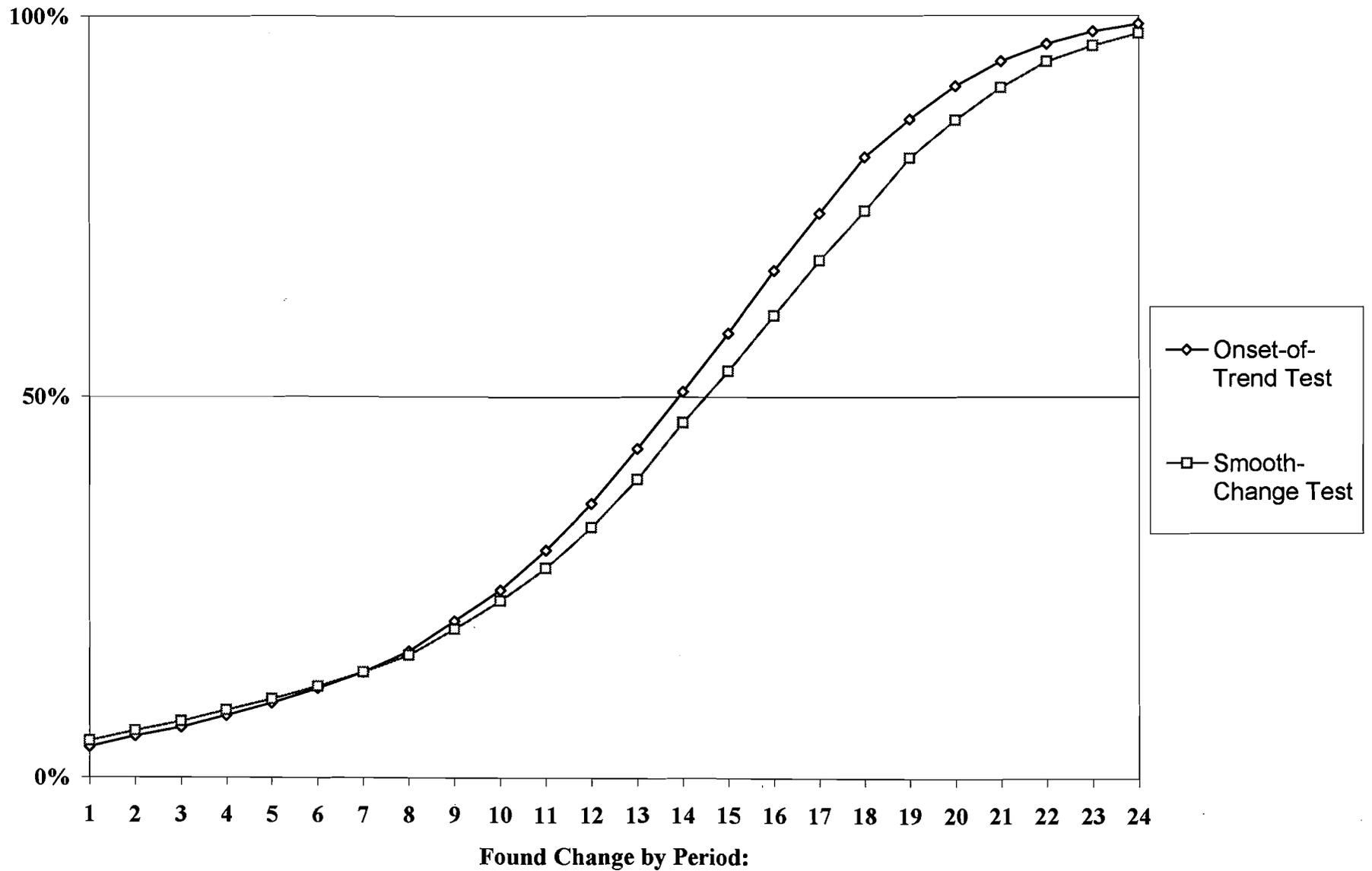
Source: Authors' calculations.

Figure 4: Estimated Probability that a One-Standard-Deviation Abrupt Change Will Be Detected by a Smooth-Change Test Versus a Lombard One-Change Test

Found Change by Period:

Source: Authors' calculations.

Figure 5: Estimated Probability that an Onset of Trend of One Standard Deviation per 12 Periods Will Be Detected by a Smooth-Change Test Versus an Onset-of-Trend Test

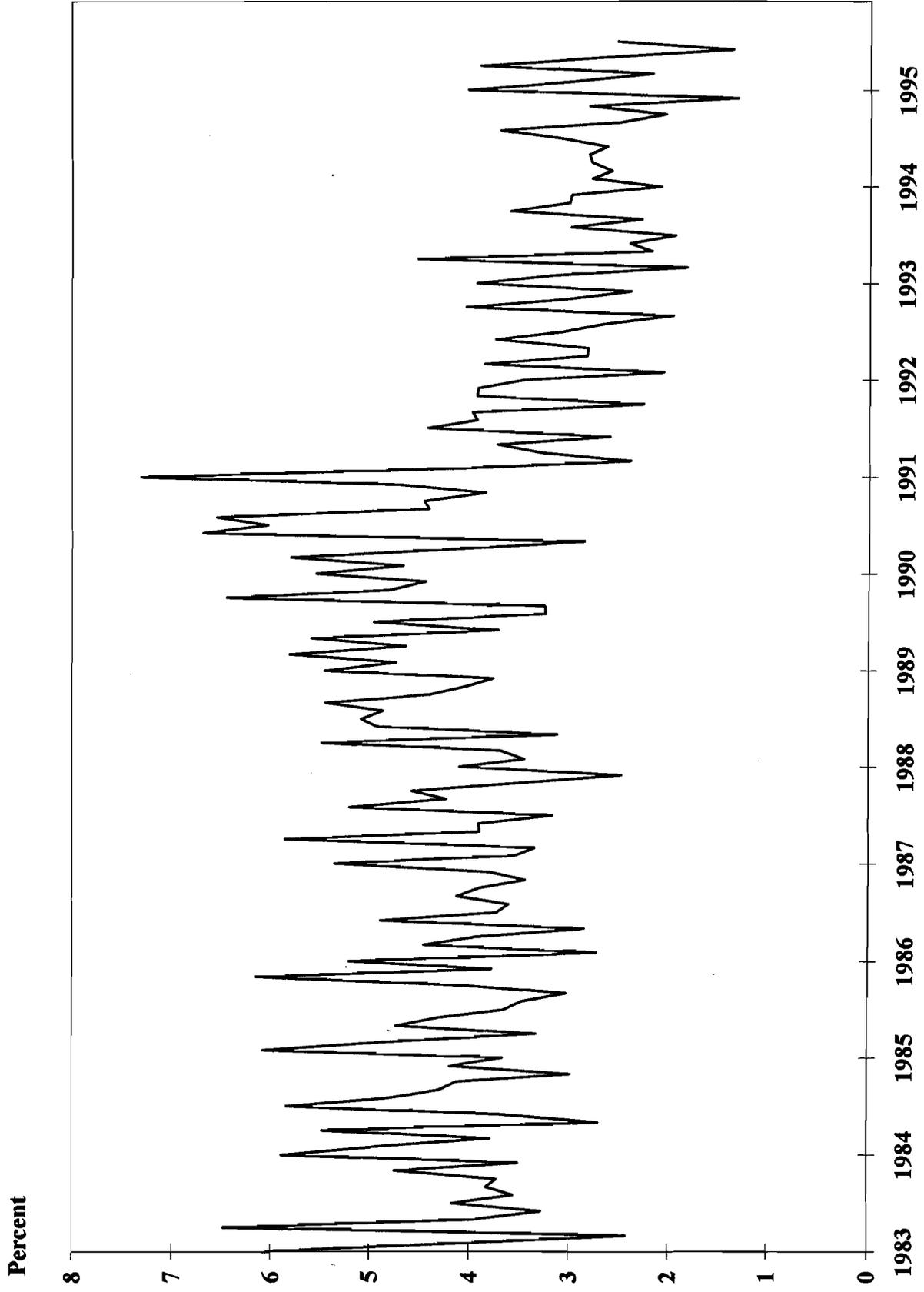
Found Change by Period:

Source: Authors' calculations.

Figure 6: Core Inflation as Measured by the 15 Percent Trimmed Mean
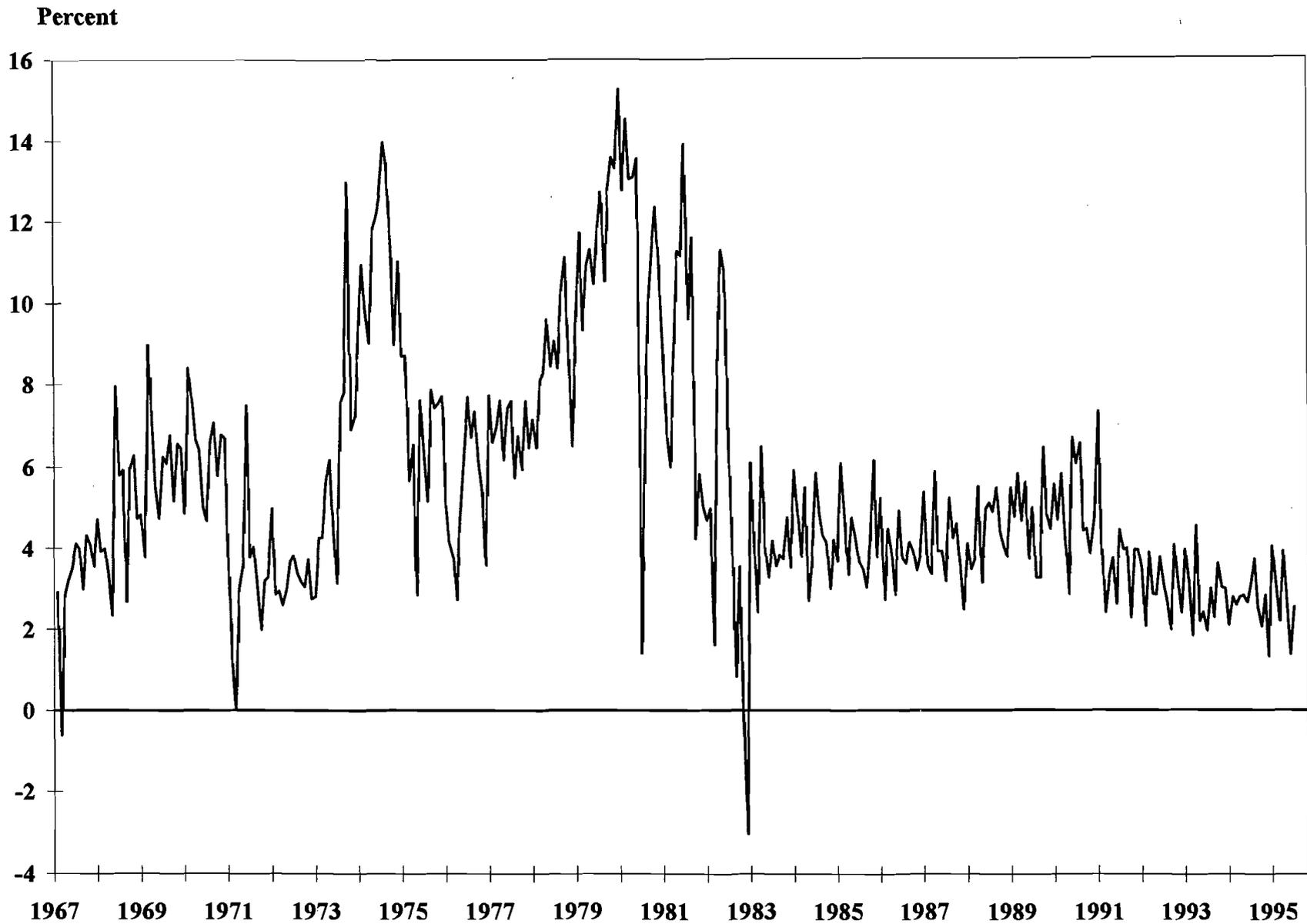
Percent

Source: Federal Reserve Bank of Cleveland.

Figure 7: 15 Percent Trimmed Mean Since 1982

Percent

Figure 8: 15 Percent Trimmed Mean Since 1982 and Selected Trend Rates

Source: Federal Reserve Bank of Cleveland.
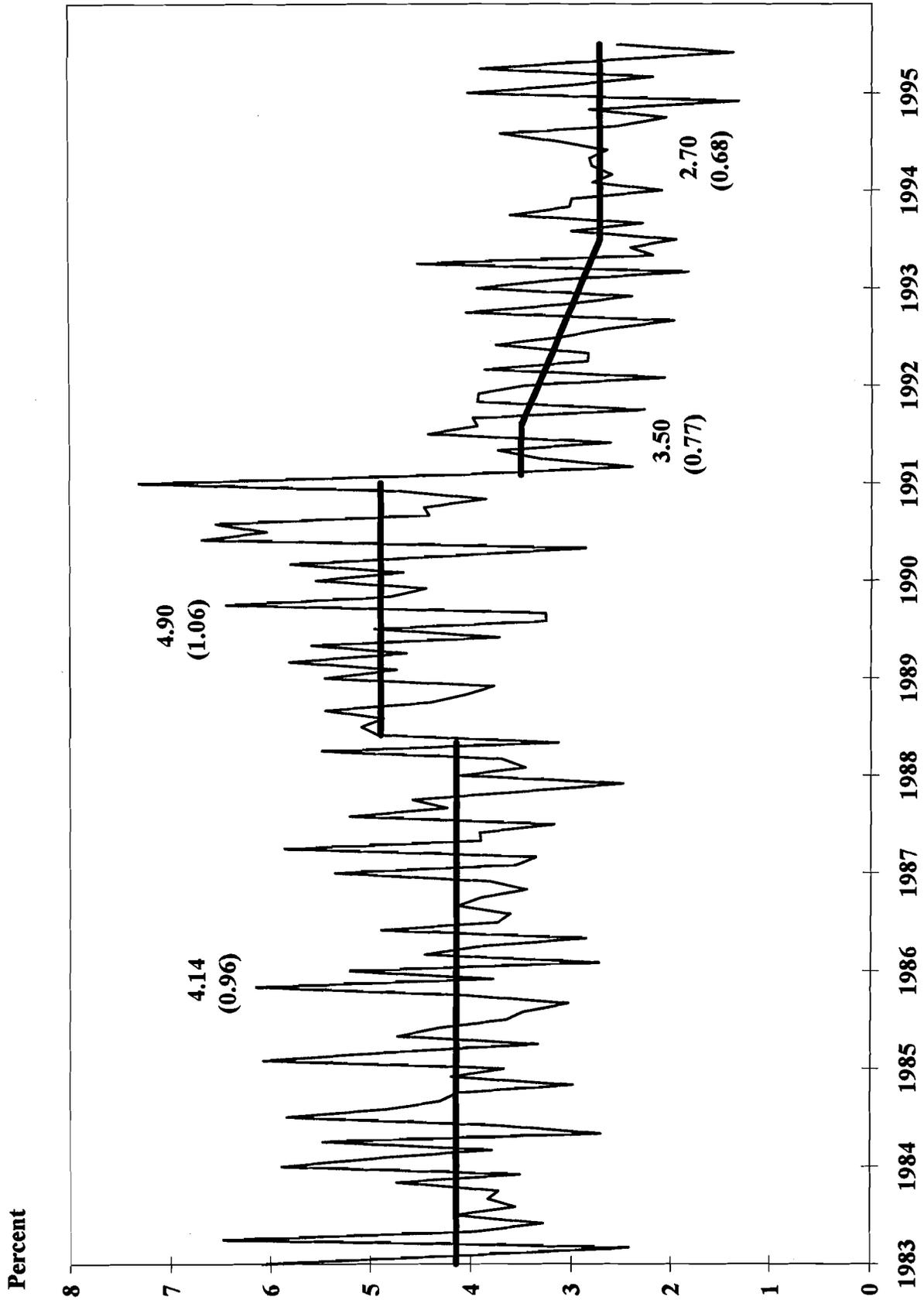
Figure 9a: Autocorrelation Functions of Selected Samples

January 1983 to January 1991

Coefficient

Lags

January 1983 to May 1988

Coefficient

Lags

June 1988 to January 1991

Coefficient

Lags
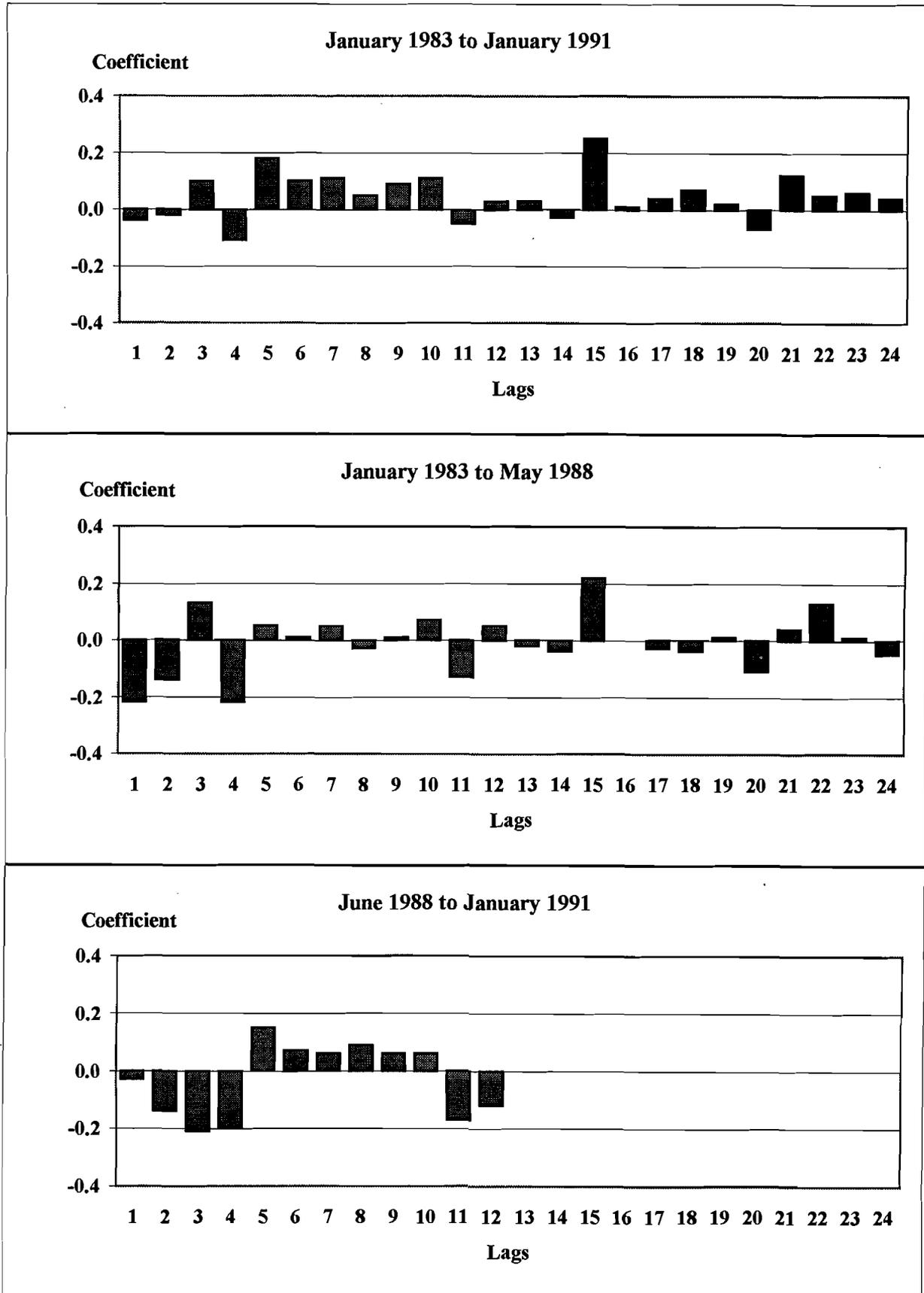
Source: Authors' calculations.

Figure 9b: Autocorrelation Functions of Selected Samples

February 1991 to July 1995

Coefficient

Lags

June 1993 to May 1995

Coefficient

Lags
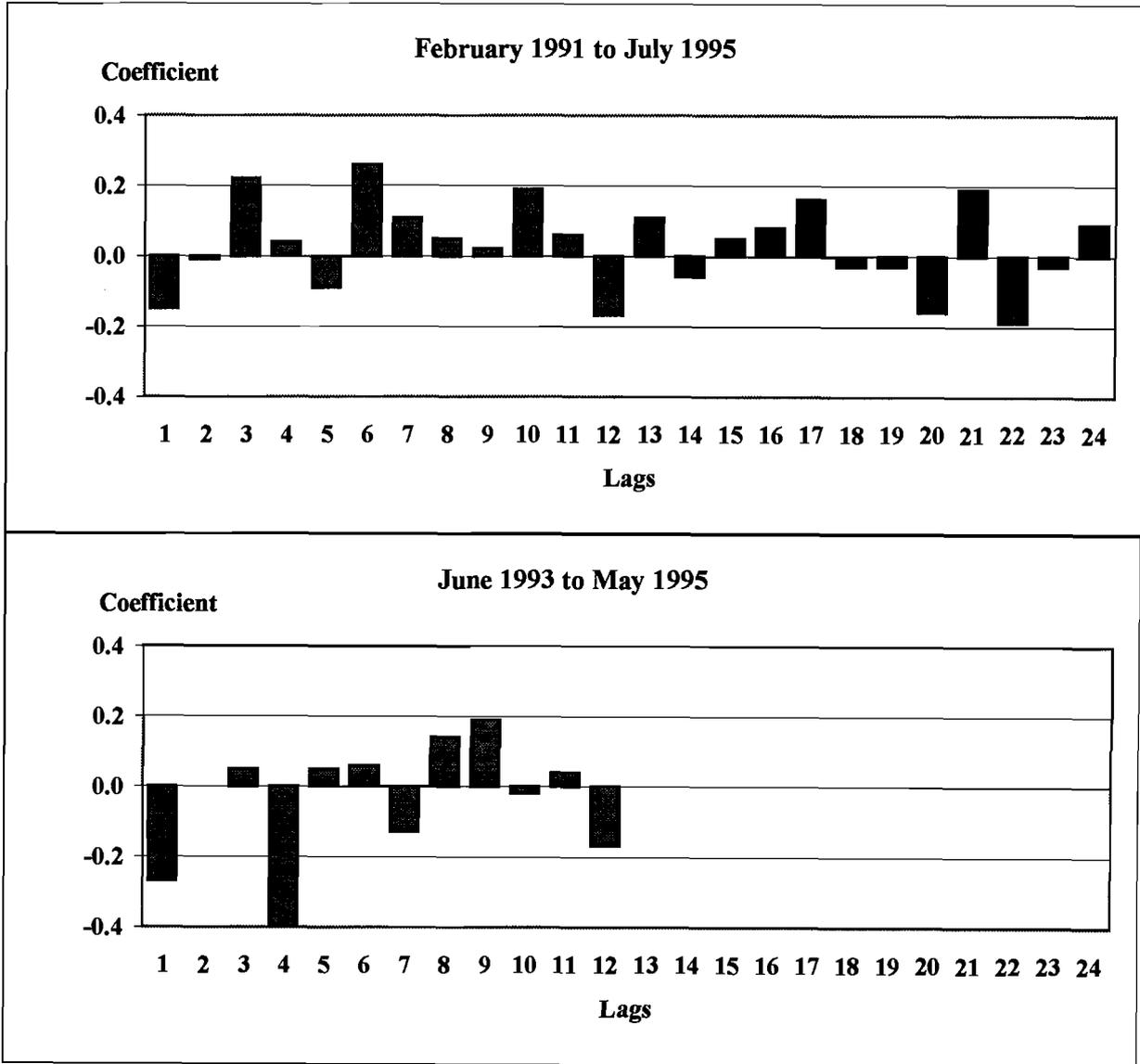
Source: Authors' calculations.

Table 1: Estimated Significance Levels for N(0,1)
(Percent)

Sample Size	Pettitt Statistic	Lombard Test Statistics				
		Number of Changepoints				
		One	Two	Three	Smooth	Trend
12	0.72	3.73	3.36	1.40	4.42	1.95
24	1.92	4.88	4.29	3.36	4.74	3.59
36	2.47	4.40	4.30	3.30	4.51	4.03
48	2.92	4.81	4.55	4.42	4.80	4.26
60	3.02	5.00	4.44	4.54	5.00	4.50
72	3.30	4.52	4.49	4.96	5.08	4.21
120	3.85	5.07	4.90	4.42	4.70	4.92

Source: Authors' calculations.

Table 2. Estimated Significance Levels in Presence of First-Order Autocorrelation

		ρ				
Test	Sample size	-0.3	-0.1	0.2	0.5	0.9
Pettitt	24	0.28	1.44	6.48	22.10	64.54
	36	0.24	1.14	8.28	26.40	76.32
	60	0.32	1.68	9.08	33.10	88.54
	120	0.22	1.60	10.72	39.74	95.94
Lombard 1	24	0.74	2.62	10.92	27.62	69.42
	36	0.84	2.72	10.88	28.90	79.90
	60	0.74	2.74	11.68	33.14	87.68
	120	0.78	2.90	11.66	35.82	94.38
Lombard 3	24	0.04	0.30	2.80	15.36	63.04
	36	0.02	0.44	11.04	21.82	79.46
	60	0.24	1.94	12.54	44.49	79.46
	120	0.44	2.68	48.80	49.28	99.06
Smooth-change	24	0.98	2.60	9.72	23.32	56.86
	36	1.14	3.22	10.32	25.42	62.76
	60	0.84	3.36	10.70	26.02	66.78
	120	0.78	2.88	10.40	25.14	70.62
Onset-of-trend	24	0.78	2.48	8.26	21.56	53.82
	36	0.92	3.02	8.68	20.30	59.68
	60	0.84	3.10	10.18	23.62	64.86
	120	0.64	2.82	9.72	24.50	67.52

Source: Authors' calculations.

Table 3
Changepoint Test Results

| Sample | Pettitt Statistics | | Lombard Test Statistics | | | | | | |
| | Max U | Date | Number of Changepoints | | | | | | |
			One	Two	Three	Trend	Smooth	t₁	t₂
Jan-83 to Jul-95	0.0000*	Feb-91	3.5511*	7.1021*	1.5915*	1.8101*	.3149*	Oct-90	Aug-92
Jan-83 to Jan-91	0.0150*	May-88	0.6005*	1.2010*	.2906*	0.1964*	0.0504*	May-88	Jun-88
Jan-83 to May-88	0.8933	Jun-85	0.102	0.204	0.0483	0.0231	0.0088	Feb-83	Mar-83
Jun-88 to Jul-95	0.0000*	Feb-91	4.4261*	8.8522*	1.7425*	1.2207*	0.4434*	Dec-90	Jan-91
Jun-88 to Jan-91	0.8994	Sep-89	0.0763	0.1526	0.047	0.0159	0.0047	Feb-95	Mar-95
Feb-91 to Jul-95	0.0562	Apr-92	0.0562*	0.6427*	0.2479*	0.1637*	0.0617*	Aug-91	Jul-93
Aug-93 to Jul-95	1.0268	Aug-94	0.0761	0.1522	0.0396	0.0165	0.0045	May-95	Jun-95

Here t_1 and t_2 denote the columns labeled t_1 and t_2.

* Significant at the 5 percent confidence level.

Source: Authors' calculations.