

Are We There Yet? Looking for Evidence of a New Economy

Simon van Norden
HEC Montréal, CIRANO & CIREQ
simon.van-norden@hec.ca

September 2007

Abstract

Over the past decade, debates over the impact of new information technologies on trend productivity growth rates have played a key role in the formulation of monetary policy, particularly in the United States. However, the question of whether the *trend* growth rate of aggregate productivity has changed *significantly* is rarely examined formally. This paper examines aggregate data for the US using a new testing approach specifically designed to detect recent changes in trends. In addition to showing the strength of the evidence for the "new" economy, it also considers how large such changes must be before they can be detected, and to what degree detection tends to lag structural change. We also examine the impact data revision has had on inference about structural changes. The results support Gordon's (2003) contention of a further acceleration in US productivity growth after 2000. Perhaps more disturbingly, we find that data revision has at times caused major changes in inferences about the presence or absence of structural breaks in productivity trend growth. These changes were particularly acute when examining the period immediately after the turn of the millenium. We also find that a new test for recent structural breaks can detect breaks rapidly provided that they are large enough; rapid detection typically requires growth rates to more than double. More generally, we estimate that in US data we would detect a structural break at conventional significance levels within 5 years more than 95% of the time with a doubling of the trend growth rate, about 50% of the time with a 50% improvement, and about 20% of the time with a 25% improvement.

KEYWORDS: Productivity growth, detrending, breakpoints, structural change, data revision

The author would like to thank Bob Rasche of the St. Louis Federal Reserve Board for making vintage data available, and Bruce Hansen for sharing code from his 2001 *Journal of Economic Perspectives* article. Thanks also to seminar participants at Rutgers University, the Bank of Canada and CIRANO, and at conferences of the Econometric Society, the SCE and the SNDE. Special thanks to Robert Amano and Jan Jacobs for their encouragement. Portions of this research have been financed by the Bank of Canada and by the New Economic Initiative of the Social Sciences and Humanites Research Council of Canada. The opinions expressed herein are the author's and should not be confused with those of the Bank of Canada. The author retains sole responsibility for all the errors, defects and shortcomings of this work.

1 Introduction

Trends in productivity growth play a key role in the formulation of public policy. They are important factors in determining long-run economic growth and therefore play central roles in the management of public pension systems and government debt. They are an essential component in defining measures of economic slack, and have therefore played a key role in the formulation of monetary policy.¹ International differences such trends in turn have profound influences on the balance of world saving and investment. Not surprisingly, the possibility of a persistent change in aggregate productivity growth casts a long shadow over many of the most important economic policy debates. For all these reasons, great effort is devoted to accurate productivity measurement and to the analysis of sources of productivity growth.

Surprisingly, however, the question of whether the *trend* growth rate of aggregate productivity has changed *significantly* is rarely examined formally. When it is, the evidence of a shift is surprisingly fragile. We argue that at least two potential problems complicate the analysis in this literature. First, very few papers perform statistical tests for changes in productivity growth trends, and most that do use methods that are known to be unreliable close to the end of sample. Second, productivity data are revised over time, with the revisions often causing non-trivial changes in measured growth rates.

This paper examines the aggregate data for the US using a new testing approach specifically designed to detect recent changes in trends. In addition to showing the strength of the evidence for shifts to a new, higher rate of productivity growth in recent years, it also considers how large such changes must be before they can be detected, and to what degree detection tends to lag the structural change. Using real-time data, it then considers the impact of data revision on the detection of trend breaks.

2 Literature Review

Most of the literature on trends in aggregate productivity growth relies on informal methods to characterize trends. On the basis of such analysis, it appears that profession opinion shifted around 2000 in favor of a improvement in US trend productivity growth which occurred about five years before.² The methods used to distinguish "trends" from other movements in productivity in most papers are ad hoc; typical methods include the comparison of growth

¹For example, the May 4th 2004 minutes of the FOMC note the FRB staff's opinion that "*...that the remaining slack in resource utilization and strong productivity growth would keep core inflation at a low level over the forecast period.*" The same minutes also summarize the committee's view that "*... a range of factors was continuing to restrain inflation, including slack in resource utilization, strong productivity gains*"

²For example, see [Stiroh 1999], [Jorgenson et al. 2002, p. 12], [Blinder and Yellen 2001, p. 59], [Gordon 1998, p. 299], [Gordon 2000, p. 49], [Gordon 2003, p. 223], [Baily and Lawrence 2001, p. 308], [Edge et al. 2004], [Borio et al. 2003, p. 17, footnote 33], [Filardo 1995] and [Hansen 2001].

rates for non-overlapping five-year periods, or the use of the Hodrick-Prescott filter or other smoothing technique. Lacking any formal statistical framework, such studies are unable to assess whether the changes in estimated trend rates are statistically significant and therefore whether they are reliable indicators of structural economic changes. Exceptions are fewer and noteworthy; they include [Filardo 1995], [Hansen 2001], [Fair 2004], [Maury and Pluyard 2004], [Erber and Fritsche 2005], and [Kahn and Rich 2004]. However, some of these studies find no reliable evidence to support the professional consensus [Filardo 1995], or place breaks some years later than commonly believed [Maury and Pluyard 2004], or conclude that multiple breaks occurred [Erber and Fritsche 2005]. Results appear to be sensitive to the precise productivity measure used, the statistical tests used, and perhaps others factors as well.³

One problems common to the statistical methods used in most of these studies is that they are not well suited to detecting *recent* changes in growth trends. Unfortunately, it is such recent changes that are precisely of greatest interest to policymakers. Furthermore, the power of such tests to detect recent changes is poorly documented.⁴ To better understand the nature of the testing problem, we briefly review conventional tests for structural breaks, as well as a new test for recent breaks due to [Andrews 2003].

2.1 Testing for Structural Breaks

[Chow 1960] proposes F-tests for a one-time structural change in one or more estimated regression coefficients when the date of the break is known. In the case of a simple AR(1) model, the null hypothesis is

$$y_t = \rho \cdot y_{t-1} + \varepsilon_t \text{ for } t = 1, \dots, T \text{ and } \varepsilon_t \sim IN(0, \sigma)$$

and the alternative is

$$y_t = \rho_0 \cdot y_{t-1} + \varepsilon_t \text{ for } t = 1, \dots, \tau \text{ and } y_t = \rho_1 \cdot y_{t-1} + \varepsilon_t \text{ for } t = \tau + 1, \dots, T, \text{ where } \rho_0 \neq \rho_1.$$

Let $\hat{\varepsilon}$, $\hat{\varepsilon}_0$, and $\hat{\varepsilon}_1$ be the OLS residuals for these three equations and S , S_0 and S_1 be their sum of squared residuals. The Wald test statistic for a structural break at τ is then given by

$$W = T \cdot \frac{S - S_0 - S_1}{S_0 + S_1}$$

[Andrews 1993] considers the distribution of this and related statistics when the researcher searches over possible values of τ . He proposes the test statistic

$$\text{Sup } W = \max_{\tau} W \text{ where } \pi \cdot T \leq \tau \leq (1 - \pi) \cdot T \text{ and } \pi \text{ is referred to as a "trim factor".}$$

Andrews shows that this statistic converges to a nonstandard distribution under very general conditions and provides tabulated asymptotic critical values. He also shows that the test will generally have better asymptotic power than

³[Edge et al. 2004] study the influence of data revisions in an *ad hoc* learning model, while [Kahn and Rich 2004] also show some results of changing data vintages.

⁴Benati (2005) finds that state-space methods give very diffuse estimates of recent productivity growth trends, but such results are not directly comparable to the formal tests considered in the above papers.

other stability tests (such as the CUMSUM.) [Andrews and Ploberger 1994] provide stronger optimality results for a closely related statistic,

$$\text{Exp } W = \ln \left[\sum_{\tau} \exp \frac{W}{2} \right] + \ln (T \cdot (1 - 2\pi))$$

[Hansen 1997] provides formulas for approximate p-values for both the Sup W and the Exp W statistics; both statistics are used in [Hansen 2001].

Throughout this literature, the asymptotic theory requires that π is constant and bounded away from 0 as $T \rightarrow \infty$. We will therefore refer below to this test as a *mid-sample* test to distinguish it from the *end-of-sample* test introduced below. [Andrews 1993] suggests using $\pi = 0.15$, and this choice is widespread in the applied literature. As the number of observations after the structural break becomes small, the test statistic tends to produce spurious evidence of structural breaks. In practice, this makes such tests poorly suited when we wish to test for recent structural breaks.⁵ However, [Andrews 2003] proposes a related approach for testing stability close to the end of the sample. Let ε_1^R and X_1 respectively be the entries of $\hat{\varepsilon}$ and the set of regressors X for $t = \tau + 1, \dots, T$, and let $\hat{\sigma} = S'S/(T - 1)$. Andrews' test statistic is then⁶

$$\psi = \hat{\varepsilon}_1^{R'} X_1 (X_1' X_1)^{-1} X_1' \varepsilon_1^R / \hat{\sigma}^2$$

Unlike the Andrews (1993) approach, this end-of-sample statistic does not compare parameter estimates before and after the breakpoint. Instead, it compares full-sample estimates of the residual variance ($\hat{\sigma}^2$) to the size of the (transformed) residuals near the end of sample. Large values of the latter relative to the former are evidence of a structural break. The distribution of the test statistic under the null hypothesis is non-standard; [Andrews 2003] proposes a subsampling-based simulation approach to tabulate appropriate p-values in specific applications. This consists of calculating the test statistic for all possible samples of length $T - \tau$ over the period $t = 1, \dots, \tau$ in order to estimate its distribution under the null hypothesis of stability. Unlike the mid-sample test, the end-of-sample test relies on the *joint* null hypothesis of no structural break and homoscedasticity. As can be seen from the above equation, an increase in the variance of the residuals at the end of the sample will tend to increase the size of the test, while a decrease in their variance will decrease its power. We return to this point below. [Andrews 2003] also finds that this test performs *best* near the end of sample; as π increases, we lose accuracy both in our estimates of σ and in the simulated distribution of ψ under the null.⁷ Finally, while this test statistic assumes that the date of the break is known, it also has power against alternatives with breakpoints in the range $[\tau, T]$.

⁵ An earlier version of this paper provided simulation evidence on the finite sample properties of such tests for various values of π . Results are available from the author on request.

⁶ See Andrews (2003) for a general exposition. Note that when the number of regressors is greater than or equal to the number of post-break observations, the test statistic is instead calculated as $\psi = \hat{\varepsilon}_1^{R'} \varepsilon_1^R / \hat{\sigma}^2$.

⁷ An earlier version of this paper also provided simulation evidence on the finite sample properties of this tests for various values of π . Results are available from the author on request. They suggested that for data sets similar to those considered below, end-of-sample tests were to be preferred over mid-sample tests for values of π below 10-15%.

Its properties suggest applying the end-of-sample break test to productivity data in order to consider the evidence for recent breaks in trend productivity growth. Before doing so, we first discuss the productivity data we will examine and consider simulation evidence on the behaviour of this test.

3 Data

Our data consist of seasonally-adjusted quarterly observations on output per person hour starting in 1947Q1 for the US non-farm business sector as reported by the BLS. As noted above, productivity data are subject to revisions which are sometimes large and may could affect estimated trend growth rates. To examine the effects of data revision, we use all "vintage" data series as published by the BLS from 1968Q3 to 2004Q3.⁸ The most recent (or "final") vintage is shown in Figure 1. Figure 2 shows the importance of some of the revisions in this series by showing both the importance of revisions to the most recent observations as well as the frequent rebenchmarking that the series undergoes. It should be noted that the revisions appear to be relatively more important than for most other macroeconomic aggregates.⁹ Figure 3 compares growth rates calculated from the final data with their revisions (final - first or "real-time" estimate.) Quarterly changes in measured productivity are seen to be very volatile and highly revised. As the period over which changes are measured lengthens, the growth rates become smoother and the recent higher periods of productivity growth stand out more. The revisions also become more muted. These changes are quantified in Table 1, which provides descriptive statistics. It shows that mean productivity growth rates have tended to be revised upwards over this period by 0.3% on average at annual rates. Although revisions are fully as volatile as observed quarterly growth rates, their relative importance falls as the differencing period increases, becoming roughly half as volatile over an 8-year differencing interval and producing a correlation between first and last-observed productivity growth of almost 88%. On the other hand, the correlation and volatility statistics ignore the bias in the revision; incorporating this would give a mean-squared revision of 0.574, or 60% of the observed variation in 8-year growth rates. This suggests that data revision could plausibly complicate the problem of making inferences about changes in rate of productivity growth.

⁸Data were kindly provided by Robert Rasche of the St. Louis Federal Reserve Board. Note that some vintages contain some missing observations for which researchers at the St. Louis Fed. are continuing to search. The *meta*-vintage of this data set is June 2005 and is available from the author upon request. Reporting lags in this data set vary from 1-2 quarters. Missing values are replaced by splicing log values from the last available vintage. The vintages are transformed into log differences for analysis. The 2005Q2 vintage is used as a "final" vintage in this draft.

⁹See Aruoba (2006).

3.1 The Great Moderation

The volatility of quarterly productivity growth appears to be much lower after the 1980s (see Figure 1) than before, a feature common to many macroeconomic series that is often referred to as "the Great Moderation." In our final data vintage, for example, the annualized standard deviation of quarterly log productivity growth fell from 5.0% to 3.1%.¹⁰ This poses a potential problem for our end-of-sample structural break test, which assumes homoscedasticity. As discussed in the previous section, such a decrease in the volatility of productivity should cause the end-of-sample test to lose power, as relatively larger breaks are now required to generate significant outliers. For that reason, we modified the application of the [Andrews 2003] test to adjust for the maintained assumption of a drop in volatility in 1983Q1.¹¹

4 Test Properties

To assess the reliability of the [Andrews 2003] test when applied to our data set (and when using our adjustment for the break in volatility), we performed a set of bootstrap experiments. In addition to verifying whether the test is properly sized (i.e. does it give spurious evidence of trend breaks when none exist), we also document the test's power (i.e. the size of structural changes that the test can reasonably detect.) Throughout, we follow [Hansen 2001] and others by modeling the final vintage of quarterly productivity growth as $q_t = \alpha + \rho \cdot q_{t-1} + \varepsilon_t$, and then test for the joint significance of a structural break in α and ρ .

4.1 Test Size

To investigate the size of the test, we apply the same test to data simulated under the null hypothesis of no structural break. Specifically,

1. We estimate $(\hat{\alpha}, \hat{\rho})$ in the above equation using OLS on the full sample period
2. We draw a random value of t and set $q_0^i = q_t$.
3. From the OLS residuals $\hat{\varepsilon}$, we randomly draw (with replacement) T observations ε^i .
4. Using $(\hat{\alpha}, \hat{\rho})$, q_0^i and ε^i , we simulate a new series q^i .

¹⁰The Andrews (1991) test found evidence of a break in the variance of shocks to quarterly productivity growth in both countries that was significant at the 0.1% level. The estimated break date was 1983Q1.

¹¹Specifically, OLS regression residuals before 1983Q1 were rescaled so that their mean squared errors equalled that of the OLS residuals after that period. In simulation experiments, bootstrapped residuals were then drawn with replacement from these rescaled residuals. Ignoring the break in volatility tended to produce somewhat similar results but with considerably lower test power.

5. We use q^i to calculate the Andrews end-of-sample test statistics $\text{Exp } W^i$ (for a break at $\tau = 0.85 \cdot T$.)
6. We use the simulated series q^i to estimate the p-value of the test statistic using Andrew's parameter sub-sampling procedure and store the results.
7. We repeat steps 3 through 7 5,000 times.
8. We increment τ , repeating steps 3 through 7, until $\tau = T$.

If the tests are correctly sized, then the resulting p-values should have a uniform distribution. Table 2 presents the frequency with which low p-values are observed, and the overall distribution of p-values is summarized in Figure 4.

The results show that the test is most accurately sized at the end of the sample, and tends to become increasingly liberal as the breakpoint moves away from the sample end. Size for breaks in the last few observations is very close to its nominal values, and it drifts up by less than 5% as we move further from the end of sample. This suggests that our testing procedure should not be expected to give spurious evidence of structural breaks. Before proceeding to apply the data to our data set, however, we also consider what kinds of structural breaks it can reasonably detect.

4.2 Test Power

It would therefore be useful to know whether breaks are likely to be recognized within a few quarters, or whether several years pass before enough evidence is available to reliably conclude that a change in trend has occurred. However, as discussed above, little has been documented about the power of structural break tests, particularly when applied to recent structural breaks. To answer this question, we estimated the power of the end-of-sample test using a series of simulation experiments similar to those used above to establish its size.

To generate data under the alternative hypothesis of a break in trend, we modified the above procedure only by introducing a multiplicative constant k to the intercept term in the AR representation for the data. In all cases, $k = 1$ until $t = \tau$, after which it takes on a new constant value. End-of-sample break tests were then run on each of resulting 5000 artificial data sets and estimated p-values were tabulated. This was repeated for all possible break dates over the last 15% of the data sample. Tests were always correctly specified in the sense that the break date tested always corresponded to the true break date.¹²

Results are summarized in Figure 5, where the annualized rate of productivity growth under the null of no breaks is 2.2% annually. By way of comparison, a 4-year moving average of the growth rate of productivity rose from a low of less than 1% around 1990 to a high of just under 4% by 2005, an overall improvement of roughly 3%. The four panels show the frequency with which the test produced p-values lower than or equal to the value shown on the vertical

¹²Misspecification of the break date should lower the power.

axis while the horizontal axis indicates the number of periods after the break which are being tested. Blue values indicate that low p-values were infrequent (i.e. low power), red values indicate frequent low p-values (high power) while greens and yellows indicate intermediate results. For reference, the panel in the bottom right simulates the data under the null hypothesis of no breaks and provides evidence on the size of the test similar to that reported previously in Figure 4.

The four panels show that as the magnitude of the breaks increase, contour lines move down and the left. This means that the probability of detecting a break after a given number of periods increases with the size of the break. It also implies that the time required to detect a break with a given probability level falls with break size. For example, the top left panel shows that after 5 years (20 quarters) the probability of detecting a break at the 5% significance level is over 95% for a doubling of the trend growth rate (i.e. an increase of 2.2% annually). The detection probability is roughly 50% for a 50% rise in the growth rate, and is about 20% for a 25% improvement. Put another way, the time required to detect an improvement in productivity growth at the 5% significance level with a probability of at least 50% is about 6 quarters in the case of a doubling of trend growth, about 5 years for a 50% rise and more than 9 years (if ever) for a 25% improvement. These results imply that the end-of-sample test can be quite powerful for large enough changes; improvement of over 2% per year will be detected with high probability in about 3 years at conventional significance levels. However, smaller but still economically significant improvements (on the order of 0.5 to 1%) require substantially longer periods before convincing statistical evidence of a change is likely to be found.

5 Recent Evidence of Structural Breaks

Now that we understand the properties of our tests, we examine our data on the log change in real Output per Hour for the US non-farm business sector for recent evidence of changes in productivity growth trends.

5.1 Final data

As a preliminary step, we used the conventional [Andrews 1993] mid-sample tests on the final data series. Perhaps surprisingly, for conventional values (15%) of π , we found no evidence of structural breaks *anywhere* in the sample that was significant at a 5% level.¹³ Since our test results could be sensitive to assumptions regarding trend breaks earlier in the sample, we produced test results under three different assumptions.

- 1) No prior structural breaks in the data.

¹³These results were reported in detail in an earlier version of the paper and are available from the author on request. Smaller values of π produced estimates of a break in 1997Q2, suggesting that the evidence for a recent increase in productivity may be as or more compelling than the evidence of the productivity slowdown in the 1970s.

- 2) One prior structural break in (α, ρ) in 1973Q2.
- 3) Two prior structural breaks in 1973Q2 and 1997Q2.

We then apply the [Andrews 2003] end-of-sample breakpoint test for all possible break dates in the last 15% of the sample. The resulting estimated p-values are shown in Figure 6. All three assumptions provide similar conclusions. They provide no hard evidence of trend breaks in US Output per Hour in the late 1990s, with estimated p-values never falling below 15%. However, they find evidence of a break after 2000, with p-values dropping below 1% regardless of the number of previous breaks on which we condition. The change in size is economically significant, with average growth rising from 2.4% in the 1996Q4-2001Q3 period to 4.0% thereafter.

The conclusion that aggregate productivity growth has accelerated by 50-100% accords well with the professional consensus mentioned at the outset, and may at first sight seem banal. However, several features are at odds with the consensus. For example, we found no statistically significant evidence of the 1970s productivity slowdown. The timing of the acceleration appears to be about 5 years later than has previously been suggested.¹⁴ Before suggesting that the consensus should be revised in the face of such evidence, we turn to consider how robust our conclusions are to data revision.

5.2 Real-Time Data

To put the above results into perspective, it would be useful to know the extent to which such inferences may change over time as new observations become available and old observations are revised. For that reason, we repeated the analysis in the above section using various vintages of the productivity series.¹⁵ The resulting p-values are presented in the form of contour plots for the last 15% of the observations in each vintage in Figure 7. Dark blue areas include low p-values and therefore are evidence of breakpoints. Consistent detection of the same breakdate in subsequent vintages should appear as a blue area with a horizontal edge corresponding to the break date. This edge will extend horizontally to the upper edge of the tested region if the break is detected the first time it is tested; however if more observations are required to detect a break then the blue zones would be to the left and below the upper edge of

¹⁴It is also odd that evidence of a structural break appears to be very sensitive to the breakpoint used; although the test should have power to detect breaks after the specified breakdate, p-values drop abruptly after 2000.

¹⁵The earliest of the data vintage pre-dates the break in the variance of productivity growth noted above (1983Q1.) Proper real-time analysis around this period is therefore complicated by the need to model the detection of the variance break in real time. However, our primary interest in this paper is the behaviour of productivity growth in the 1990s and thereafter, so we restrict our attention to this later period and assume that the break in variance has been detected and corrected for prior to the start of our analysis. Specifically, we work here only with data vintages produced at least 3 years after the break in the variance. Results closest to the variance break date therefore rely on the implicit and untested assumption that the variance break would have been detected by that date and should be interpreted with caution.

the tested region. Vertical features are associated with data revisions; vertical contours indicate that the use of different vintages causes important changes in the perceived probability of a trend break.

Figure 7 shows several distinct episodes during which end-of-sample tests would have provided evidence of a structural break in the trend productivity growth rate. One of these occurs from 1991-92, when the tests gave substantial evidence of a trend break in the late 1980s. As time passes and the data are revised, however, the evidence becomes weaker and effectively disappears with data revisions in 1995. Another such episode occurs from 1996-1997 when there is evidence a trend shift in 1993. This shifts abruptly in late 1999 with new evidence of a break in 1999 instead. Evidence of this new break is itself short-lived and disappears after little more than a year. Finally, in late-2001 evidence of break at the end of the sample emerges, which increases sharply in 2003 to imply a break anywhere after 1994.

To better understand the reasons for this changing evidence, Figure 8 performs the same experiment, but now replaces the data used in the previous figure with the same observations taken from the 2005Q2 ("Final") data vintage. By removing the effects of data revision from the analysis, this isolates the effects on the end-of-sample tests of increasing the length of the data series. For comparison, Figure 9 shows the difference in p-values between the two figures. Figure 9 tends to show somewhat more consistent results across time; in particular, the evidence for breaks around 1986, 1993 and 1999 is now more persistent. However, the most striking difference lies with samples ending after 2000, where the Final data provide clear and consistent evidence of a recent break from late 1999 onwards. The time of the break is imprecisely specified, however, with the range stretching from the mid-90s onwards to 2000. As Figure 9 makes clear, there are several instances where the magnitude of the change in p-value exceeds 0.7; reflecting instances where data revision both caused previously significant breaks to become insignificant, and where it caused breaks to become apparent only in retrospect.¹⁶

6 Conclusions

This paper has examined the problem of detecting changes in productivity trends from a policy-maker's perspective, emphasizing the problem of detecting recent changes with data that may later be revised. In addition to carefully applying a novel test that is designed for the detection of recent structural breaks, it also finds some statistical evidence to support [Gordon 2003]'s contention of a

¹⁶One puzzling feature of these results, however, is that they are inconsistent with the changing consensus in the literature reviewed above. Although published analysis suggests that a break in the late 1990s became apparent by early 2001, these tests are able to detect such breaks only after data revisions published in 2003. Surprisingly, when using the most recent data vintage, breaks are first detected around the time that the professional consensus changed.

further acceleration in US productivity growth after 2000, with estimated trend growth rates rising from 1.5% to 3.0% annually.

More generally, however, many of the results presented here suggest that such claims need to be interpreted with considerable caution. Documenting some of the properties of revisions in the productivity data, we find that revisions in quarterly growth rates are fully as important as the final growth rate estimates, although growth rates measured over longer periods of time tended to reduce the relative importance of the revisions. Our statistical tests appear to have good size properties and sufficient power to find economically important changes in growth with a few years of data. Despite this, we find that data revision can cause major changes in inferences about the presence or absence of structural breaks in productivity trend growth. These changes were particularly acute in the US data in examining the period immediately after the turn of the millennium, but statistical evidence about structural breaks in trend growth generally appears to be quite fragile. Reconciling these results with the apparent shift in professional opinion around early 2001 remains an issue for future research. While more reliable methods of estimating (and forecasting) productivity growth with data subject to revision are developed, policymakers should treat recent findings with a degree of skepticism.

7 Bibliography

References

- Andrews, Donald W. K. (1993) "Tests for Parameter Instability and Structural Change with Unknown Change Point." *Econometrica*, 61(4), 821-56.
- Andrews, Donald W. K. (2003) "End-of-Sample Instability Tests." *Econometrica*, 71(6), 1661-94.
- "Optimal Test When a Nuisance Parameter is Present Only Under the Alternative" *Econometrica*, 62:6, 1383-414.
- Bai, Jushan and Pierre Perron (1998) "Estimating and Testing Linear Models with Multiple Structural Changes." *Econometrica*, 66(1), 47-78.
- Bailey, Martin Neil, "Comments and Discussion on 'Exploding Productivity Growth: Context, Causes and Implications' by Robert Gordon" *Brookings Papers on Economic Activity*, 2:2003, 280-287.
- Baily, Martin Neil and Robert Z. Lawrence (2001) "Do We Have a New Economy?" *The American Economic Review*, 19(2), 308-312.
- Blinder, Alan S. and Janet L. Yellen, *The Fabulous Decade: Macroeconomic Lessons from the 1990s*, The Century Foundation Press, New York, 2001.
- Borio, Claudio, William English and Andrew Filardo (2003) "A tale of two perspectives: old or new challenges for monetary policy?" *BIS Papers*, 19.
- Chow, Gregory C. (1960) "Tests of Equality Between Sets of Coefficients in Two Linear Regressions" *Econometrica*, 28:3, 591-605.
- Crawford, Allan (2002) "Trends in Productivity Growth in Canada" *Bank of Canada Review*, Spring 2002, 19-32.
- Edge, Rochelle M, Thomas Laubach and John C. Williams (2004) "Learning and Shifts in Long-Run Productivity Growth" *Federal Reserve Board Finance and Economics Discussion Series*, 2004-21.
- Erber, Georg and Ulrich Fritsche (2005) "Estimating and Forecasting Aggregate Productivity Growth Trends in the US and Germany." *DIW Berlin, Discussion paper* 471.
- Fair, Ray C. (2004) "Testing for a New Economy in the 1990s" *Business Economics*, January 2004, 43-53.
- Filardo, Andrew J. (1995) "Has the Productivity Trend Steepened in the 1990s?" *Federal Reserve Bank of Kansas City Economic Review*, 1995Q4, 41-59.
- Gordon, Robert J., "Does the 'New Economy' Measure up to the Great Inventions of the Past?" *The Journal of Economic Perspectives*, 14(4), 49-74.
- Gordon, Robert J., "Exploding Productivity Growth: Context, Causes and Implications", *Brookings Papers on Economic Activity*, 2:2003, 207-279.

- Gordon, Robert J., "Foundations of the Goldilocks Economy: Supply shocks and the Time-Varying NAIRU" *Brookings Papers on Economic Activity*, 2:1998, 297-346.
- Hansen, Bruce E. (1997) "Approximate Asymptotic P Values for Structural-Change Tests" *Journal of Business and Economic Statistics*, 15:1, 60-67.
- Hansen, Bruce E., "The New Econometrics of Structural Change: Dating Breaks in U.S. Labor Productivity", *Journal of Economic Perspectives*, 15(4), 117-128.
- Jorgenson, Dale W., Mun S. Ho and Kevin J. Stiroh (2002) "Projecting Productivity Growth: Lessons from the U.S. Growth Resurgence" *Federal Reserve Bank of Atlanta Economic Review*, III2002, 1-13.
- Kahn, James A. and Robert W. Rich (2004) *Tracking the New Economy: Using Growth Theory to Detect Changes in Trend Productivity*, Federal Reserve Bank of New York mimeo, March 2004.
- Macklem, Tiff (2004) "Future Productivity Growth in Canada: comparing to the United States" *International Productivity Monitor*, 7(Fall), 50-57.
- Maury, Tristan-Pierre and Bertrand Pluyaud (2004) "The Breaks in Per Capita Productivity Trends in a Number of Industrial Countries." *Notes d'études et de recherche de la Banque de France*, 111.
- Robidoux, Benôit and Bing-Sun Wong (2003) "Has Trend Productivity Growth Increased in Canada?" *International Productivity Monitor*, 6(Spring), 47-55.
- Sharpe, Andrew (2004) "Recent Productivity Developments in Canada and the United States: Productivity Growth Deceleration versus Acceleration." *International Productivity Monitor*, 8(Spring), 16-26.
- Stiroh, Kevin (1999), "Is There a New Economy?" *Challenge*, 42(4), 82-101.

8 Tables

Table 1: Revisions in Quarterly Change of Log US Output per Hour in Non-farm Business Sector

Final data is 2005Q3 Vintage; Realtime Vintages are 1968Q3-2004Q3
All series start 1947Q1; all changes calculated at annual rates x100.

Statistic	1Q Change	4Q Change	16Q Change	32 Q Change
Realtime Mean	1.66	1.65	1.60	1.55
Realtime Std. Dev.	2.81	1.84	1.41	0.96
Revision Mean	0.32	0.30	0.31	0.33
Revision Std. Dev	2.89	1.19	0.74	0.47
ρ (Realtime,Final)	0.552	0.780	0.854	0.877
ρ (Revision _t ,Revision _{t-1})	-0.101	0.693	0.806	0.877

Table 2: Joint S Tests for Parameter Stability

Frequency of Parametrically Resampled p-Values Under the Null

Periods from EOS	10% p-value	5% p-value	1% p-value
2	10.1%	5.1%	1.9%
4	9.9%	5.7%	1.7%
8	11.2%	6.2%	1.7%
16	11.3%	6.9%	2.3%
32	14.5%	9.0%	5.1%

Figure 1: Output per Person-Hour
US Non-Farm Business Sector

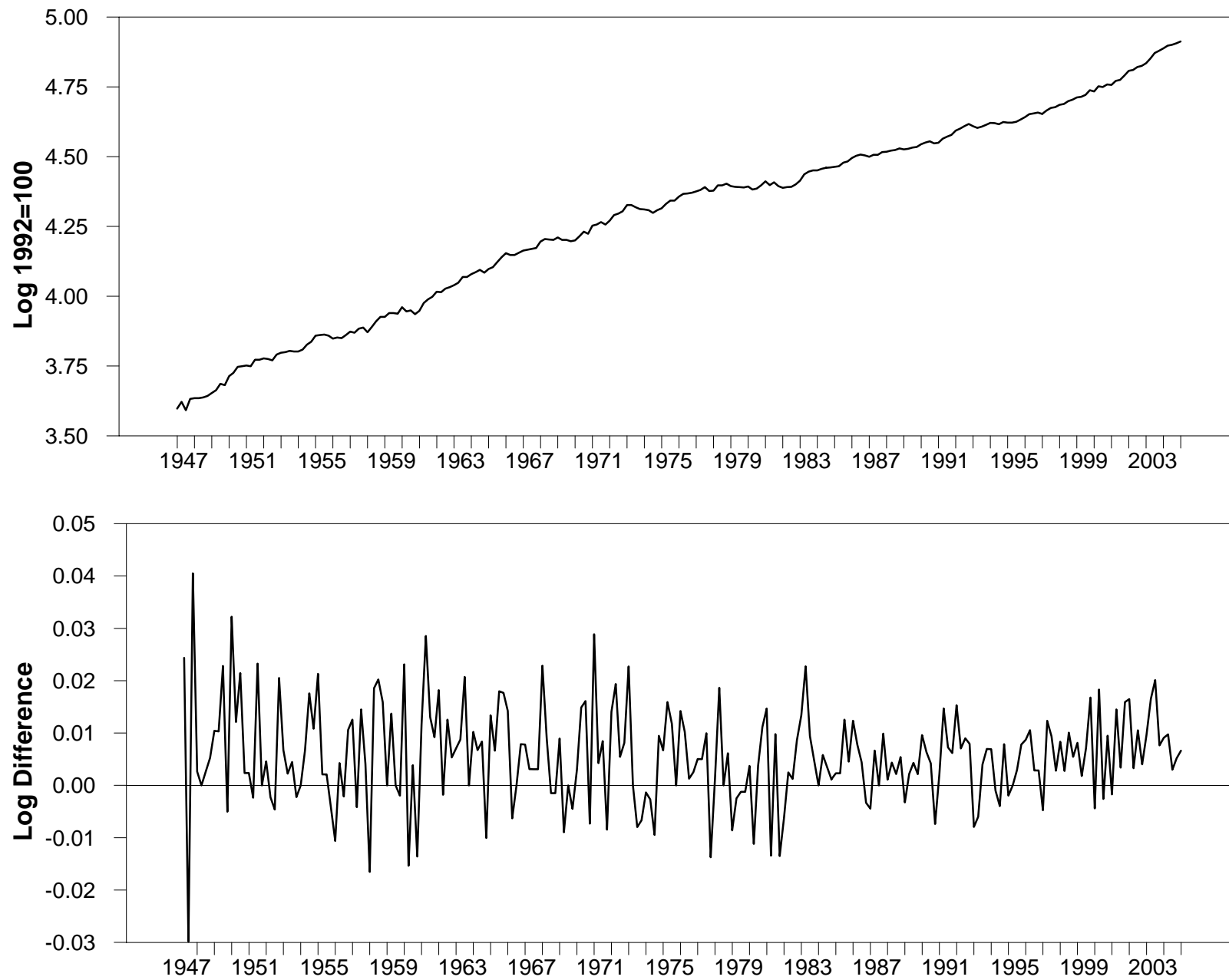


Figure 2: Output per Person-Hour
US Non-Farm Business Sector

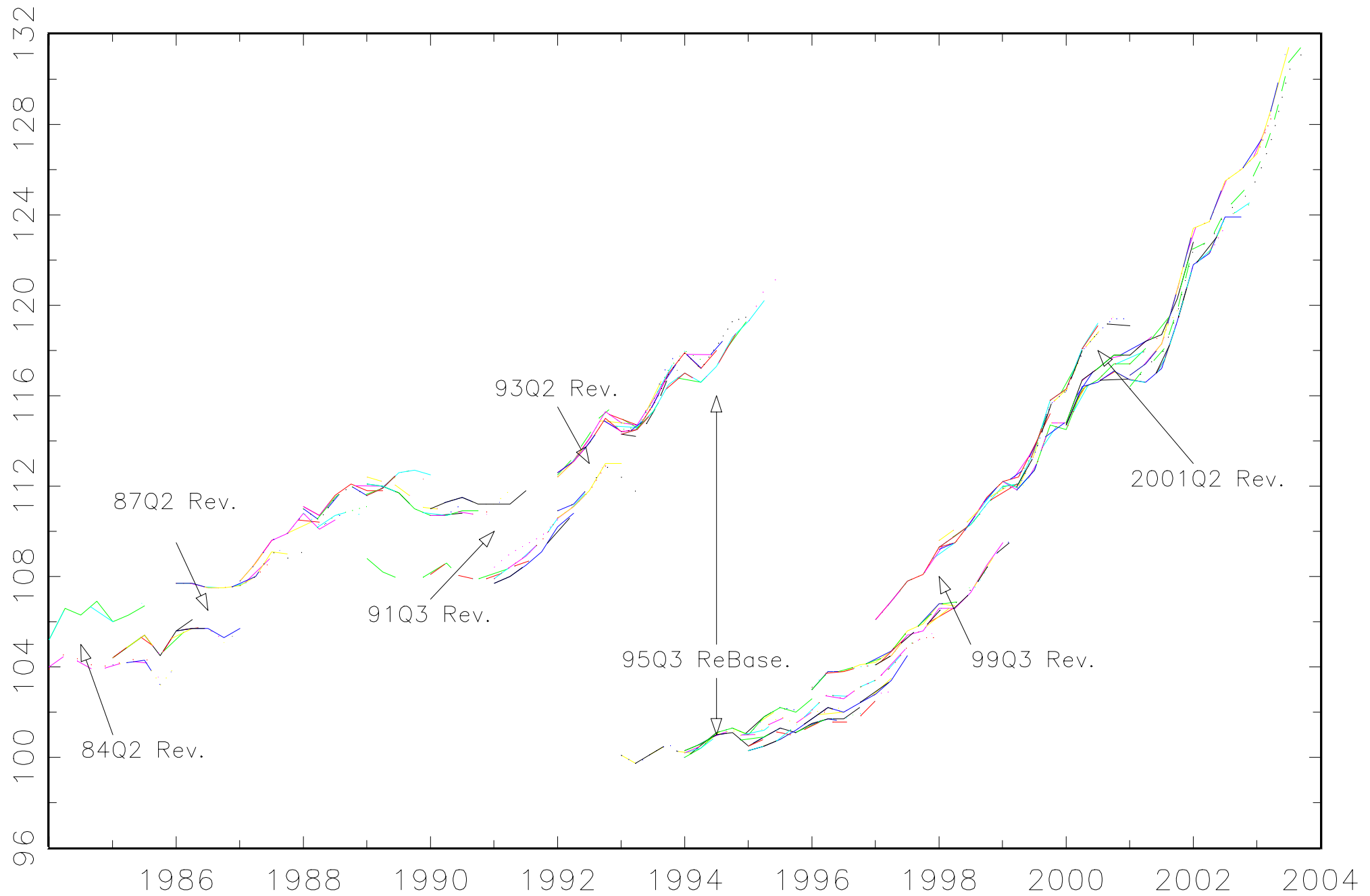


Figure 3: US Productivity Growth

Output per Hour for All Persons, Non-Farm Business Sector

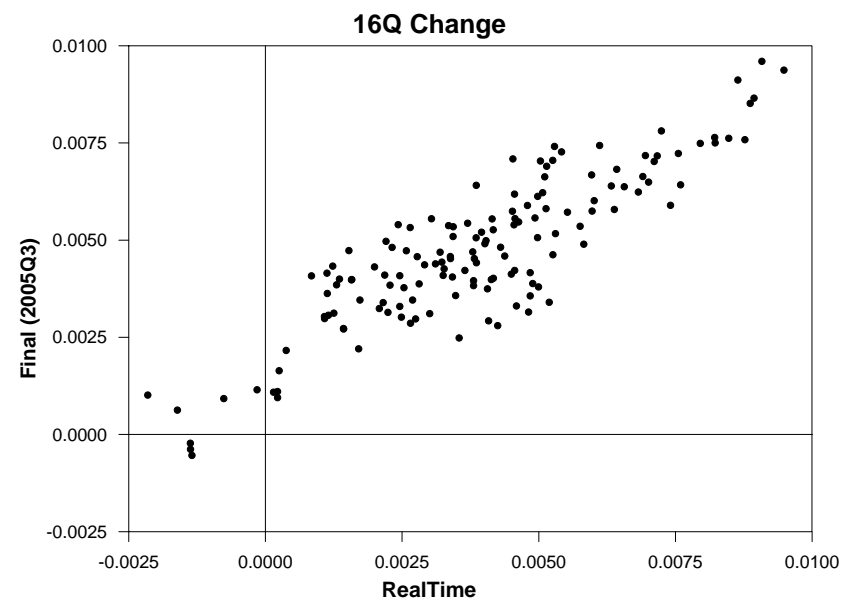
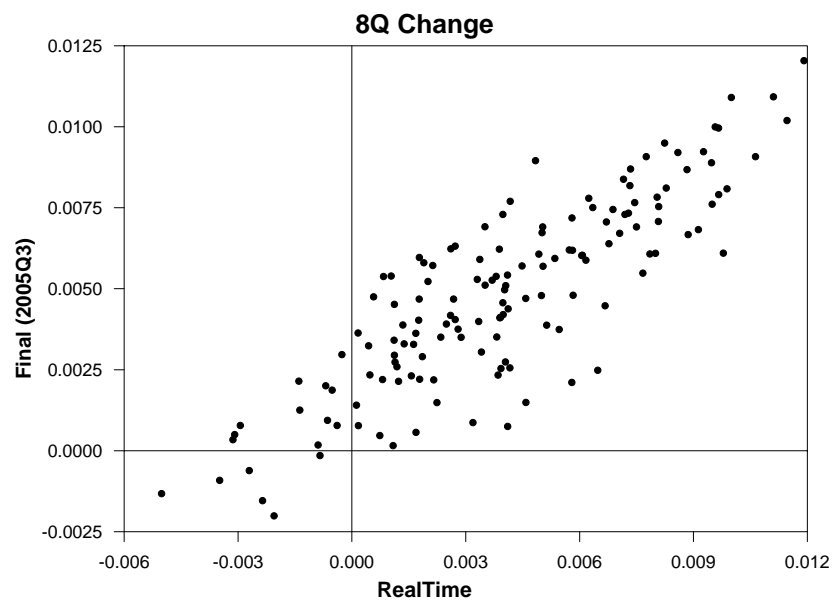
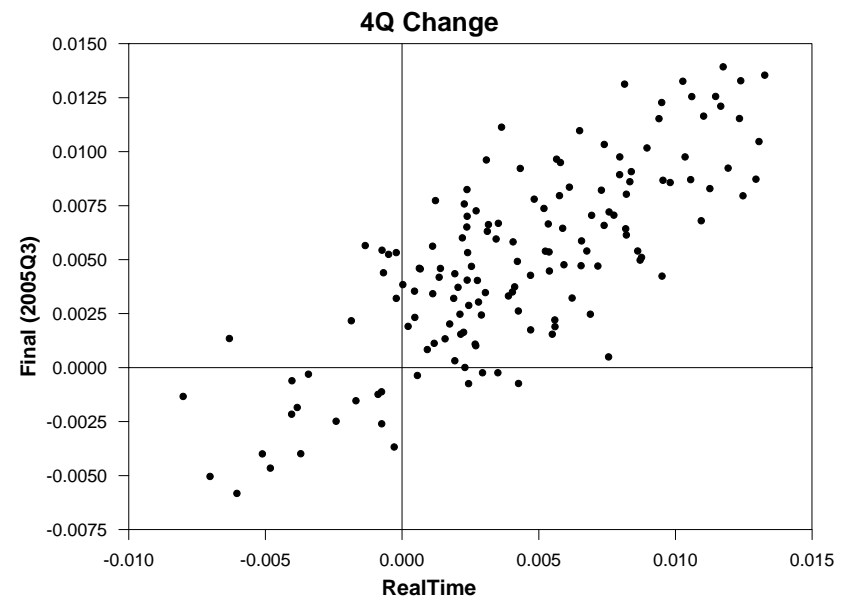
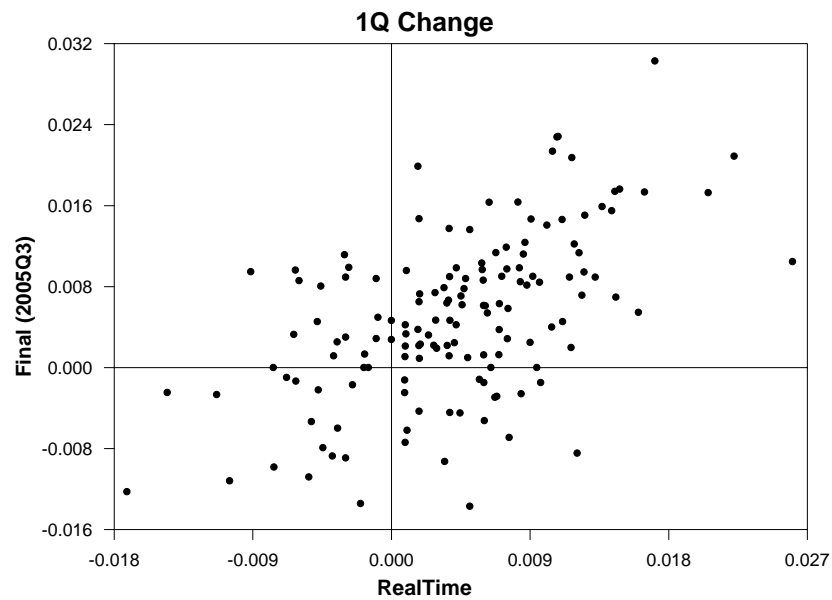
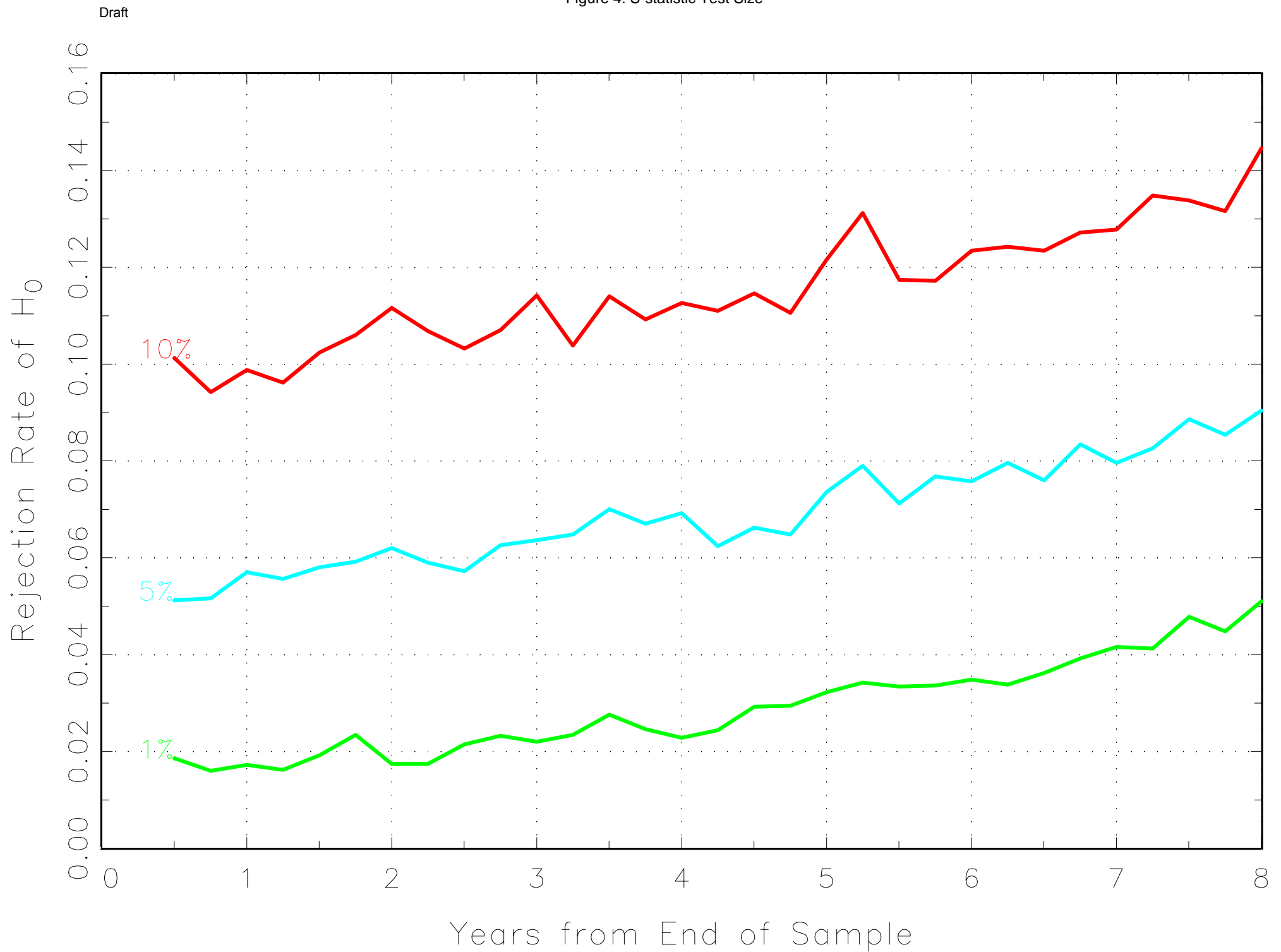


Figure 4: S-statistic Test Size



**Figure 5: EOS Test Power
US OPH - NFB**

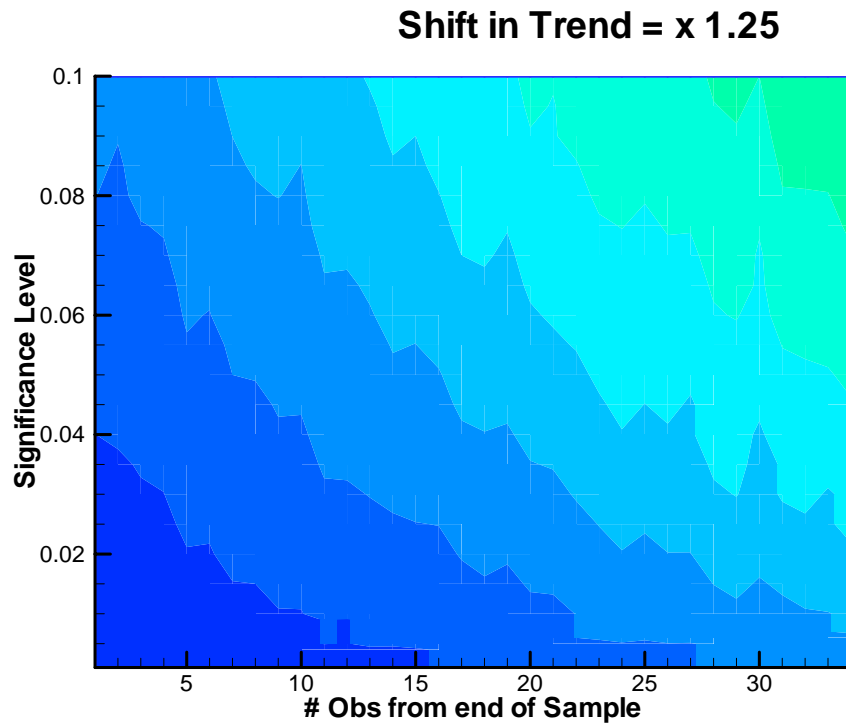
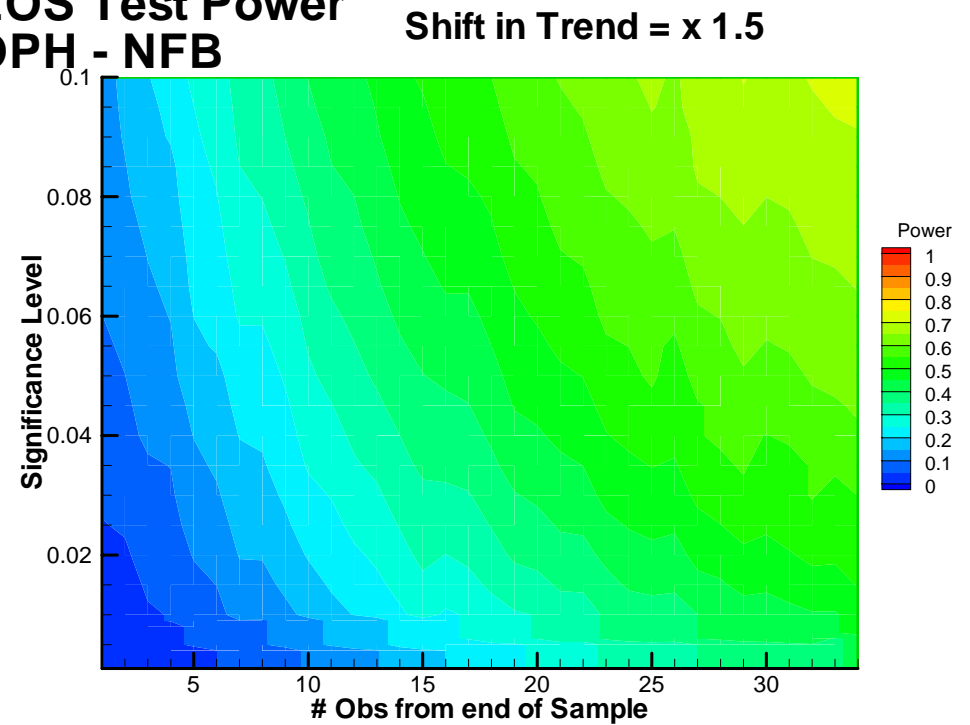
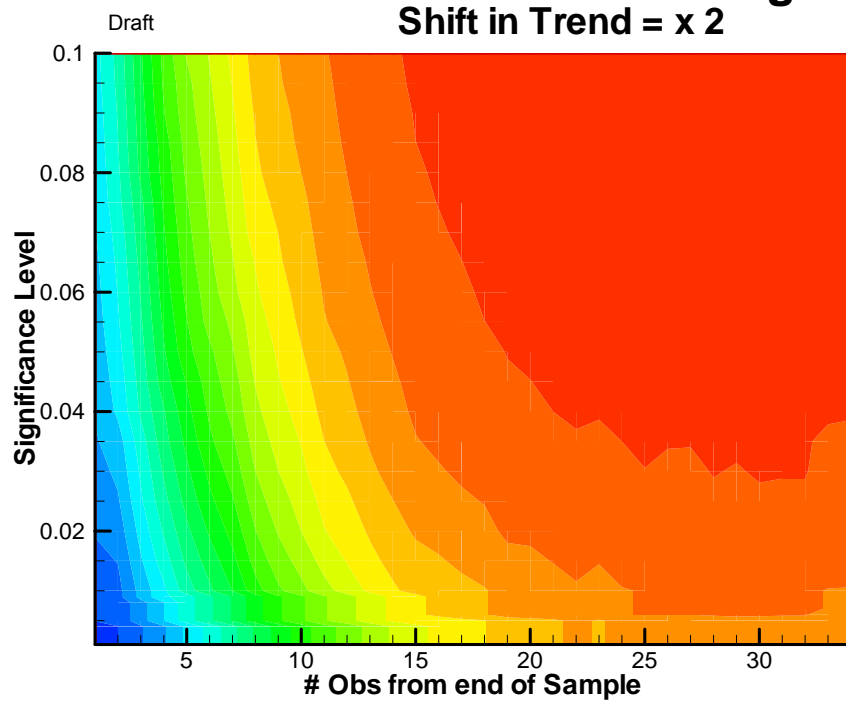


Figure 6
EOS Break Tests for US OPH

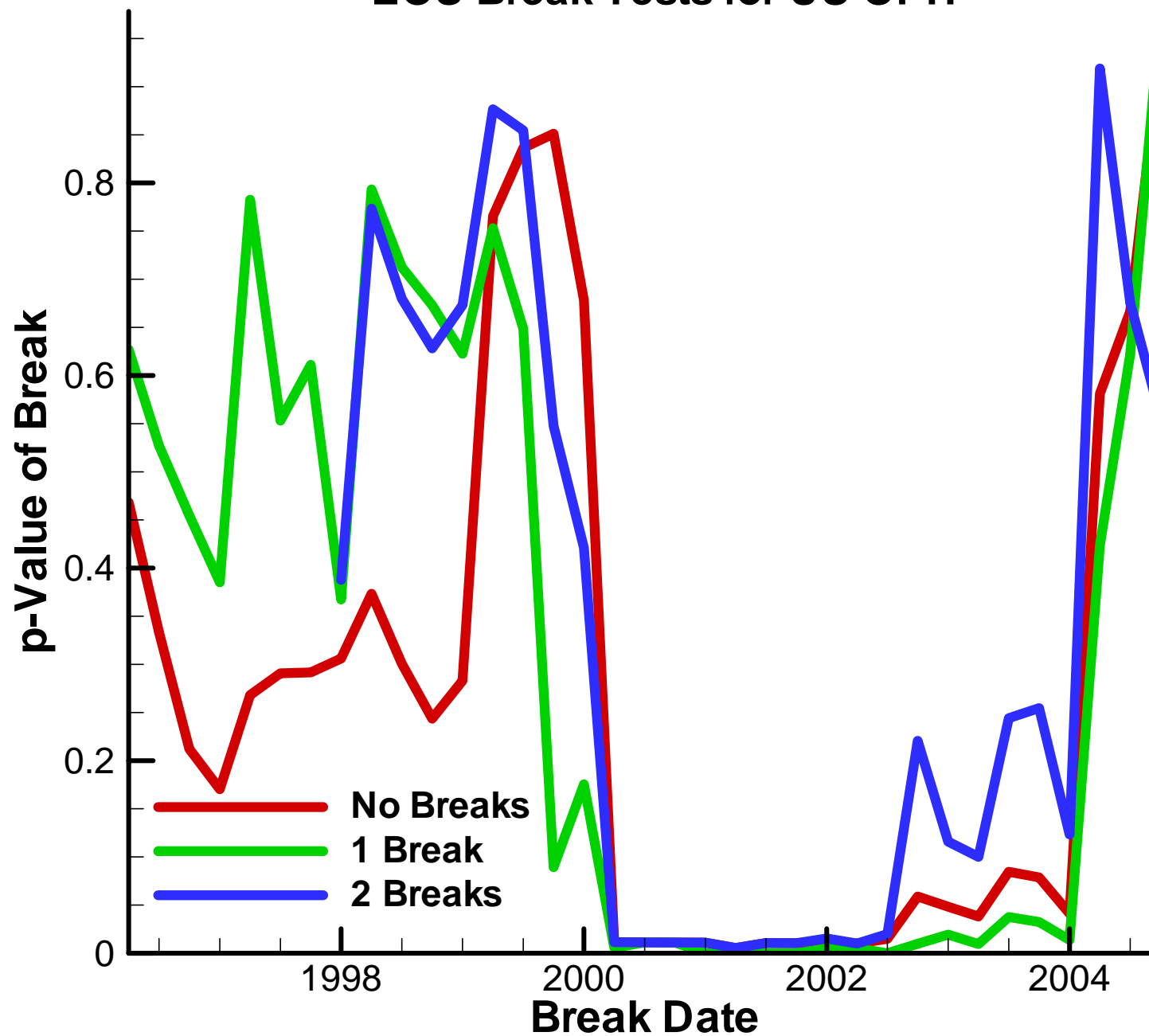


Figure 7 - End of Sample Breakpoint Tests
RealTime - US Output per Hour - Non Farm Business

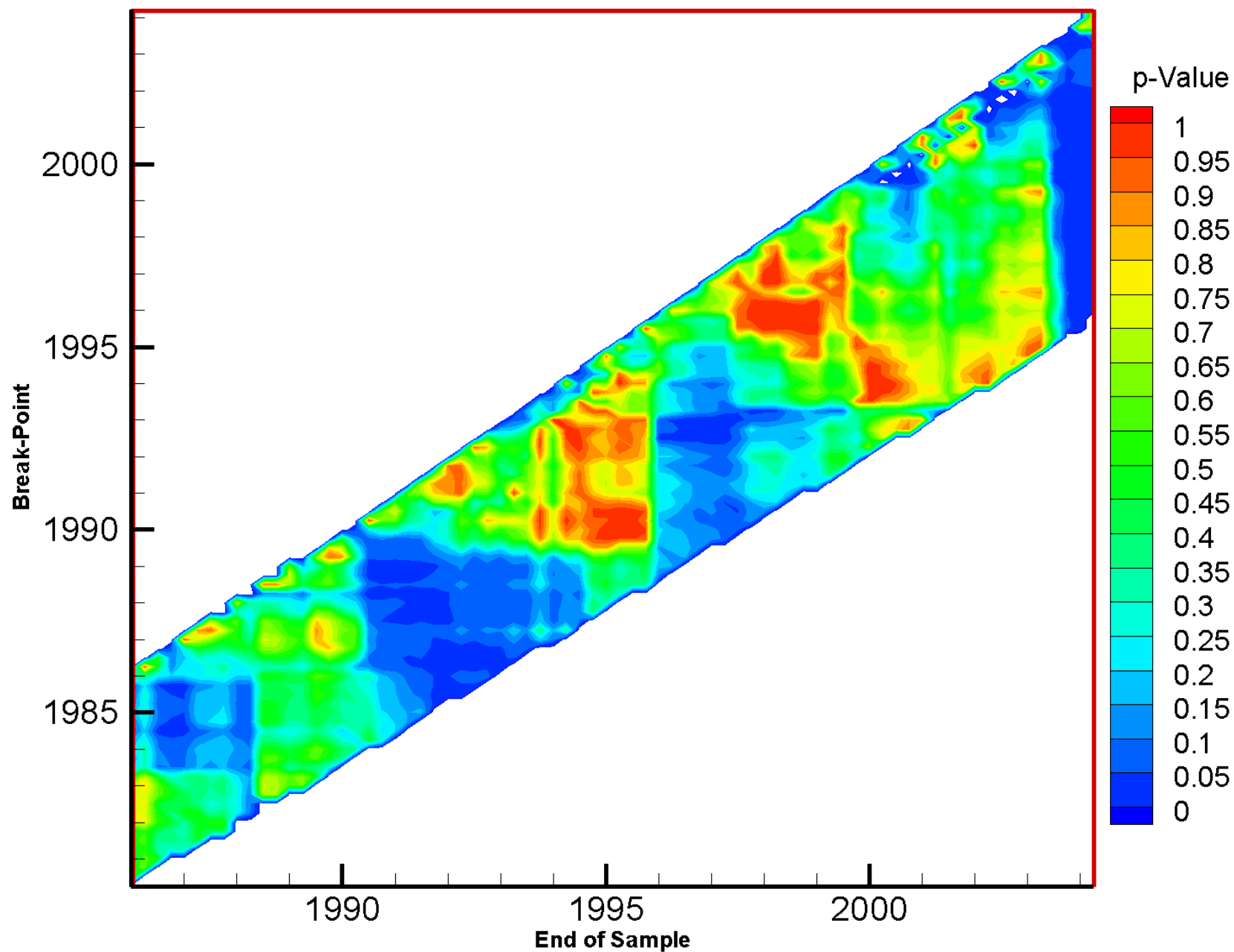


Figure 8 - End of Sample Breakpoint Tests
Final - US Output per Hour - Non Farm Business

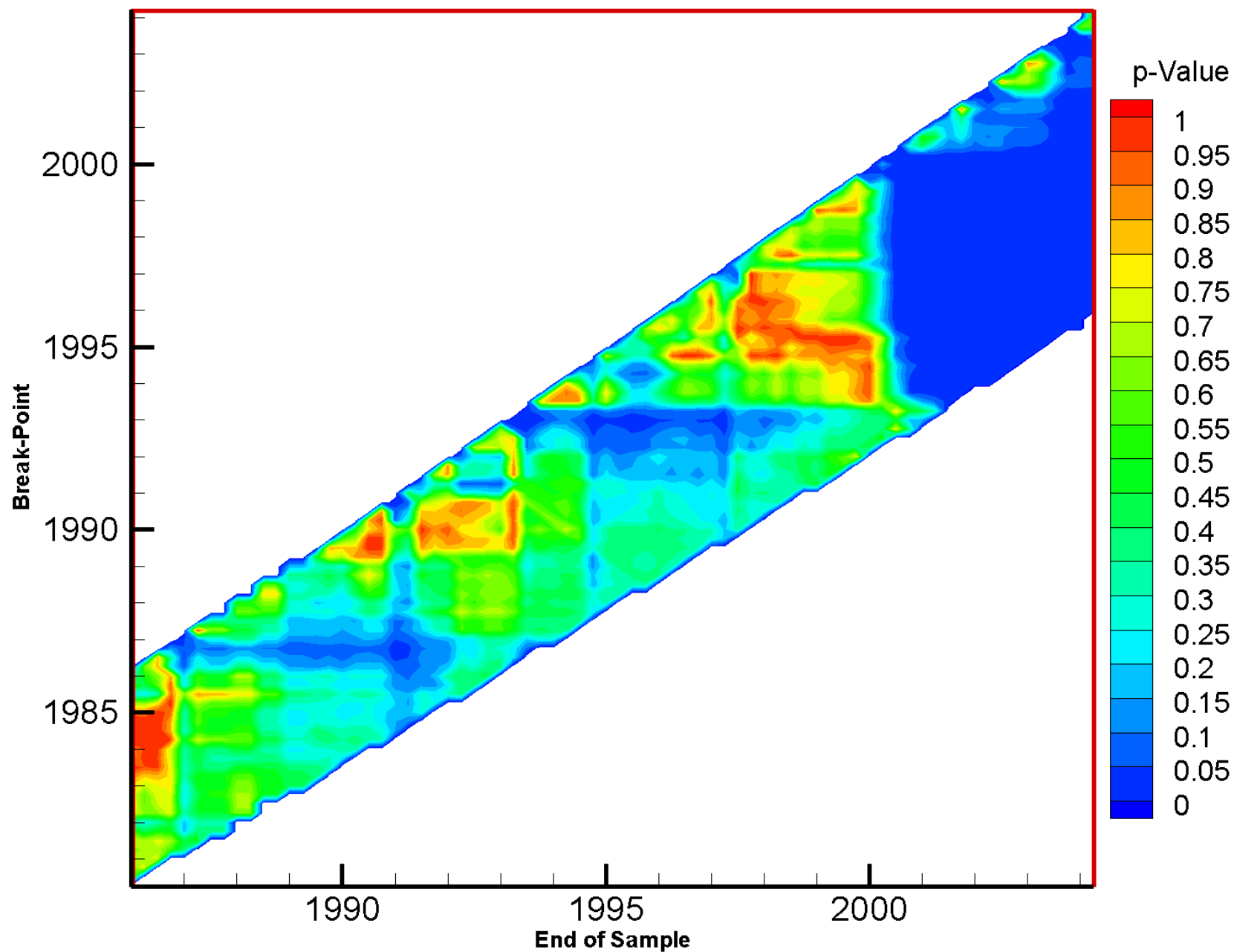


Figure 9 - End of Sample Breakpoint Tests
RealTime - Final p-values

Draft

