COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1027

CHRISTMAS, SPRING AND THE DAWNING OF ECONOMIC RECOVERY

Eric Ghysels

May 1992

CHRISTMAS, SPRING AND THE DAWNING OF ECONOMIC RECOVERY

Eric Ghysels
Centre de recherche et développement en économique (C.R.D.E.)
Université de Montréal

First draft : April 1992
This version : May 1992

* I would like to thank Chris Sims for invaluable discussions and comments. I would also like to acknowledge financial support from the Cowles Foundation, where part of this work was completed while I was Visiting Research Fellow. This research was also supported by the Social Sciences and Humanities Research Council of Canada, the Natural Sciences and Engineering Research Council of Canada and the Fonds FCAR of the province of Québec. As always, I bear full responsibility for any remaining errors.

1. INTRODUCTION

We know relatively little about what makes recessions come and go. In practice, economists often refer to changes in the economic environment such as housing starts or the relatively vague notion of consumer confidence as key elements which help jump start a recovery. Many of these changes in economic environment are not uniformly distributed throughout the year. Housing starts slow down because of the weather, and Christmas shopping is only a once-a-year ritual. Presidential elections are held in November every four years, and most incumbant politicians do not enjoy the idea of running a campaign, say six to nine months ahead of the election in the midst of a recession. All this may suggest subtle and nontrivial interactions between intra-year changes in the economic environment and the timing of business cycle phases, and leads to the question whether or not business cycle turning points are clustered around certain times of the year.

The analysis in this paper is based on a very simple and revealing classification of business cycle turning points and durations to show the interdependencies between cyclical and seasonal movements.[1] We pick six months of the year where the economic environment is potentially different from the other half of the year and test whether this two-sample classification has any impact on business cycle features. The first six months are March, April, May, June, July and December, while the remainder is classified in a second sample. As there is a fair amount of uncertainty about the timing of turning points, we prefer to keep a loose classification. The months of March through July cover most post-winter and pre-summer activity (which we call spring for convenience) while December covers the end-of-year effects (which we call Christmas for convenience as well). All other months include winter, (late) summer and early fall. The description of the economic environment will be kept deliberately vague, in the sense that "spring" and "Christmas" will stand for a multitude of factors including demand shocks, technology shocks as well as institutional arrangements which we will not explicitly identify. The statistical evidence reported in this paper is based on the NBER business cycle chronology as well as alternative chronologies. The latter are used to investigate the robustness of our results, since the NBER chronology has been

---

1 In Ghysels (1991), a quarterly classification was used and k-sample nonparametric tests were applied to test the hypothesis of uniformity. With a quarterly classification, the evidence is less compelling as presented here due to a combination of two factors; namely, the fact that k-sample tests have less desirable properties (with a given sample size), and the fact that the quarterly classification doesn't exactly match the more "economically based" classification used here.

criticized for possible inaccuracies, particularly for the dating of cycles before WWI. The statistical techniques are standard ones such as goodness-of-fit tests applied to the frequency of turning points and distribution-free statistical procedures making use of the qualitative information contained in the duration of cycles.[2]

This paper is surely about seasonality, yet it highlights aspects which have been unexplored. Namely, it focuses on dependencies between cyclical and seasonal fluctuations beyond the linear unobserved component paradigm almost exclusively used to discuss the separation of business cycle fluctuations and seasonals.[3] The interdependencies uncovered can be viewed as periodic variations throughout the year in regime switching probabilities or hazard rates which are intrinsically nonlinear and will not be detected by the usual seasonal adjustment filters such as X-11.[4] Hence, it does not really matter that much whether or not seasonal adjusted data were used to benchmark business cycle phases.

There are, of course, two very different interpretations to the results obtained in this paper. Either the turning points are unevenly distributed because the economy goes through fundamentally different phases throughout the year or else the outcome is purely the result of errors of type II. The NBER business cycle chronology is indeed a man-made series, and the committee members face the daunting task of deciding, for instance, whether or not the economy's turnaround coincided with the usual end-of-winter bust. Anticipating some of the discussions, it may be said that the results suggest that the periodicity is probably more likely the consequence of causes intrinsic to the economic structure. In addition, the issues being discussed in this paper relate to long-term (nonlinear) phenomena, while we retrieve our empirical evidence

---

2 See, for instance, Diebold and Rudebusch (1990b) where similar tests were applied to investigate the postwar stabilization question.

3 The unobserved component model has been used for the design of seasonal adjustment filters as well as for the documentation of the empirical evidence on seasonal fluctuations and their importance [see, for instance, Barsky and Miron (1989), Ghysels (1990) and Miron (1990) on documenting stylized facts of seasonal fluctuation via seasonal dummies]. The model has had its critics, however, as several researchers noted there was little economic rationale for the linear orthogonal decomposition [for various aspects, see, for instance, Sargent (1978), Ghysels (1986, 1988) and Singleton (1988), among others].

4 See Ghysels (1991) for a discussion of periodic Markov switching-regime models. The stochastic process theory and spectral representation of such models are discussed in Ghysels (1992). The latter provides a technical discussion on the question of adjustment filters and periodic stochastic switching.

from a relatively small sample (though approximately 150 years it *only* contains 30 business cycles). We will therefore also pay explicit attention to the small sample issues and their ramification.

The remainder of the paper is organized as follows. Section 2 discusses the economic environment as well as the data being used. Section 3 presents the empirical results. Section 4 concludes.

2. ECONOMIC ENVIRONMENT

We will be deliberately vague about the description of the environment and this for the following reasons : (1) as the analysis covers almost a span of one century and a half, it would be difficult to maintain a stable explicitly formulated seasonal pattern, as many institutional and other changes have occurred, (2) though our analysis will be highly critical of various aspects of the Burns and Mitchell methodology, we will adopt their simple notion that the economy goes through different "phases" without justifying the economic structure causing them and, finally, (3) a more or less full description would require a coherent data set of key macroeconomic variables over the past 150 years, preferably at the monthly frequency. In the remainder of the section, we will spell out in detail the three aforementioned main arguments for choosing an informal description of the environment.

It is clear that many institutional changes took place over the past 150 years which shaped and reshaped the intra-year dynamics of the economy. For instance, until 1975, fiscal years ended in March, but since then they end with the month of September. After 1914, the Federal Reserve System was in operation and may have had an impact on the seasonal pattern of interest rates.[5] On the other hand, some institutional aspects have remained stable, such as presidential elections always taking place in November. Hence, any political business cycle effect, like an incumbent president facing re-election and a recession at the same time, would be coherently present throughout the sample. While we will split the sample in several subsamples to account for certain heterogeneity effects, it is clear that by designating just six months

---

5 Much has been written about the foundation of the Federal Reserve, in particular about the impact it may have had on interest rates. The empirical evidence is fairly inconclusive, however, as many profound changes took place at about the same time. See, for instance, Miron (1986) and Canova (1991) emphasizing different interpretations of the empirical evidence.

of the year as being potentially different from the other six, we avoid many institutional and other details which would make the empirical investigation cumbersome.

Like Burns and Mitchell, we describe the economy as going through different phases, but instead of assuming only two or possibly three states of the world, namely expansion, recession and possibly recovery, we expand the notion of states of the world to include aspects related to the yearly repetitive character of the seasons.[6] Hence, it is not only important to know whether the economy is in a recession, it also matters whether the recession covers the summer, fall, etc. While we adopt certain aspects of the Burns and Mitchell reference cycle methodology, we must emphasize some profound differences, in fact, some fundamental criticism of their approach.[7] Indeed, during the thirties when their methodology took shape, it was assumed that seasonal adjustment filters, which were simple linear moving average filters, would take care of the problems surrounding seasonal fluctuations. Once the data were adjusted for seasonality, assumed to be an unobserved additive spurious source of fluctuations, they were ready to be used for business cycle analysis described via its traditional phases. The criticism, of course, is that the now widely used seasonal adjustment filters simply don't solve the problem since all time series movements are fundamentally intertwined in nontrivial (nonlinear) ways.

By adopting the notion that there are more than two states of the world, we circumvent some major issues regarding sources of shocks and their propagation mechanism. This is particularly appealing at this stage. Any more detailed description of the economic environment would force us to pin down the sources of shocks which make the six months of Christmas and spring distinctly different from the other six months. It is sometimes, perhaps somewhat naively assumed that "seasonal shocks" are exogenous. Yet, in any equilibrium model, one should expect seasonal phenomena to be the outcome of a complex interaction between changes in preferences throughout the year, changes in technology, the flow of information and possibly endogenous policy

---

6 This is exactly what happens if we were to formulate a model with Markov-switching probabilities for regime switching. Instead of having only two states, like in Neftçi (1984) or Hamilton (1989), one would augment the states with a seasonal classification. These issues are discussed in detail in Ghysels (1991).

7 Barsky and Miron (1989) also criticized the Burns and Mitchell methodology, yet for very different reasons. They argued that by ignoring seasonality, researchers would forego an opportunity to retrieve information about business cycle phenomena. In practice, many researchers have chosen to ignore seasonals, for good reasons or not. The question whether to use unadjusted data is a difficult one with many unresolved aspects. For further discussions of these issues, see, for instance, Hansen and Sargent (1991) and Sims (1991).

responses.[8] While there have been recent attempts to model such equilibrium interactions explicitly, see, for instance, Hansen and Sargent (1990), Braun and Evans (1991), Chatterjee and Ravikumar (1992), we have chosen in this paper to precisely follow a line of research which is to a large extent model-free and distribution-free by deliberately adopting a vague description of the economic environment and its dynamics. What we call Christmas and spring is a combination of many factors including intra-year demand shifts and technological changes together with an environment created by institutional factors which have changed throughout the last 150 years.

The third and final reason for not adopting a very detailed and explicit description of the economic environment is the lack of a coherent data set over such a long period that would allow us to describe explicitly the information set of agents. We sympathize with the arguments brought forward by, for instance, Diebold and Rudebush (1990b) choosing to rely only on the business cycle chronology to settle the postwar economic stabilization question because conventional measures of GNP, industrial production, employment, etc. suffer arguably from some serious data collection and construction incoherencies before and after WWII. Of course, the NBER chronology provides far less information, yet enough to answer some of the key questions via relatively robust statistical techniques. Moreover, the NBER chronology itself is constructed from a less than perfect data set of heterogeneous quality through time. Alternative chronologies will therefore also be used.

3. THE DATA, TESTS AND EMPIRICAL RESULTS

The business cycle chronology of the NBER is well known and has been the subject of much criticism and scrutiny.[9] Several papers critically reviewed the dating procedure and suggested alternative chronologies.[10] In particular, the dating of cycles before WWI is most controversial mainly because of a much more restrained data base and also because of some apparent inconsistencies between early dating rules and the more recent procedures. Recently, Romer (1991) set out to reexamine thoroughly the early chronology and carried out a consistent and coherent review of the dates. These

---

8 This is exactly the reason why an economic rationale for the unobserved component model is hard to find, as mentioned earlier.

9 For the sake of convenience, we reproduced the chronology in Table A.1 of the appendix.

10 A partial list includes Ayres (1939), Cloos (1963a, b), Fels (1959), Moore (1961), Moore and Zarnowitz (1986), Persons (1931), Trueblood (1961), Zarnowitz (1963a, b), for various periods.

efforts yielded an alternative chronology which we will also be using in this paper. In addition, we also relied on the algorithm advanced by Bry and Boschan (1971) and recently adopted in the analysis of the duration of the business cycle by Watson (1992).[11]

This section is divided into four parts. Evidence drawn from the NBER chronology is discussed first in section 3.1. Next, in section 3.2, we investigate the robustness of our results by looking at alternative chronologies. In section 3.3, we further scrutinize the evidence by relying on randomization arguments. The duration of business cycle contractions are used to finalize the empirical investigation in section 3.4.

3.1 The distribution of peaks and troughs throughout the year

The frequency distribution of peaks and troughs, based on the NBER chronology, appears for the entire sample as well as a set of subsamples in Table 3.1. Among the subsamples, we considered the pre-WWII and post-WWII eras separately. In addition, we also covered the period after 1885 to make comparisons with alternative chronologies discussed in later sections. Including the 1885-WWII sample yields four subsamples besides the entire sample. While Table 3.1 also lists results obtained from the alternative chronology suggested by Romer (1991), we will only discuss the NBER chronology in this section and defer the analysis of the alternative chronology to the next section.

The distribution of peaks throughout the year appears fairly regular in comparison to that of troughs over the entire sample as well as the subsamples. In the case of troughs, there is indeed more evidence of clustering around the spring and the latter part of the year. Of course, for the post-WWII era, there are relatively few observations; therefore, little information can be drawn from this particular subsample alone.[12] The clustering can simply be a purely random event, but on the other hand, it may also be the result of some fundamental unequal turning point distribution. Before we address this question, let us perform a few simple tests.

---

11 We will only report the results obtained from the NBER chronology and the alternative chronology suggested by Romer. The Bry and Boschan algorithm produced turning points essentially similar to those of the NBER and are therefore not reported.

12 There are other ways of looking at the post-WWII evidence, but they require parametric models [see Ghysels (1991, 1992)].

Table 3.1 : Distribution of Peaks and Troughs Throughout the Year from NBER and Alternative Chronologies

	NBER Chronology					Alternative Chronology		
Months	1855-1990	1885-1990	1855-WWII	1885-WWII	Post-WWII	1885-1990	1885-WWII	Post-WWII
				Peaks				
January	5	4	4	3	1	3	3	0
February	1	1	1	1	0	1	1	0
March	2	1	2	1	0	2	1	1
April	2	1	1	0	1	1	1	0
May	3	3	3	3	0	3	2	1
June	3	1	3	1	0	1	1	0
July	4	4	1	1	3	5	3	2
August	3	3	2	2	1	3	1	2
September	1	1	1	1	0	1	1	0
October	3	1	3	1	0	2	0	2
November	2	2	0	0	2	1	0	1
December	2	2	1	1	1	2	2	0
				Troughs				
January	1	1	1	1	0	2	2	0
February	1	1	0	0	1	2	1	1
March	4	3	3	2	1	3	3	0
April	2	2	1	1	1	2	0	2
May	3	3	2	2	1	1	1	0
June	5	4	5	4	0	2	2	0
July	3	3	2	2	1	6	4	2
August	1	1	1	1	0	1	0	1
September	0	0	0	0	0	0	0	0
October	2	2	0	0	2	1	0	1
November	3	3	1	1	2	2	1	1
December	5	2	5	2	0	2	2	0

Notes : Data and sources are described in the appendix, Tables A.1 and A.2.

As the troughs seem to cluster around March-June/July and November/December, we shall choose the March through July and December classification. We could have substituted November for July, at least for the entire NBER chronology, since they have the same number of troughs, but as further analysis will show, there is some clear indication that July carries more of the fallout of the spring activity than November does for Christmas. Hence, the six months of March-July and December represent one category of turning points, while the rest of the year represents another, both potentially distinct if any of the arguments layed out in the previous section have an impact.

At this stage, there is no clear reason why we should also apply the classification of spring and Christmas to peaks. We have used this classification for peaks as it will become useful in the analysis of duration data discussed later. It should be noted though that other classifications for business cycle peaks yielded essentially the same results.[13] In Table 3.2, the Christmas/spring classification is used to test the hypothesis of uniformity.

Over the entire sample, there are a total of 22 troughs versus 8 during the off-season months which is a ratio of almost three-to-one. Is this in any way statistically significant? A first answer, which will not be the only answer, however, is to use a relatively simple test known as the χ^2 goodness-of-fit test. The null hypothesis is that there is a 50-50 distribution of troughs (or peaks, as we shall discuss later) over the two six-month samples covering the year. The test is an asymptotic test and simply measures the difference between the expected frequencies under the null versus the empirically observed frequencies, which in this case equal, for instance, 22 and 8 out of 30 observations for the entire sample. Formally stated :

$$T = \sum_{j=1}^{2} \frac{(O_j - 0.50N)^2}{(0.50)N} \overset{a}{\sim} \chi^2(1)$$

where N is the sample size and O_j are the observed number of turning points over the two samples j = 1, 2. Table 3.1 reports for all the samples considered, covering peaks and troughs, the goodness-of-fit tests as well as their p-value. Some reservation

---

13 As already noted, a k-sample general classification was used in Ghysels (1991) where no evidence against the null hypothesis of uniformity of peaks was found.

Table 3.2

Goodness-of-fit Tests for the Uniform Distribution
of NBER Business Cycle Chronology Peaks and Troughs

	1855-1990	1885-1990	1855-WWII	1885-WWII	Post-WWII
			Peaks		
Number of peaks Christmas/spring	16	12	11	7	5
Number of peaks Other months	15	12	11	8	4
Goodness-of-fit test	0.52	0.00	0.00	0.07	0.11
p-value	0.59	1.00	1.00	0.41	0.74
			Troughs		
Number of troughs Christmas/spring	22	17	18	13	4
Number of troughs Other months	8	8	3	3	5
Goodness-of-fit test	6.53	3.24	10.71	6.25	0.11
p-value	0.01	0.07	0.00	0.01	0.74

Note : The Christmas/spring classification corresponds to March, April, May, June, July and December. The data appear in Table 3.1 and are described in the appendix.

needs to be expressed at this point about the use of this test in some of the smaller samples.[14] The pattern of clustering of troughs present over the entire sample persists in all samples except the post-WWII sample where a relatively equal distribution emerges.

For business cycle peaks, there is no evidence of an uneven distribution according to the tests reported in Table 3.1. The contrast between troughs and peaks is sharp and quite significant. The striking difference between troughs and peaks brings us to several important issues. Namely, the fact that peaks show a uniform distribution and troughs don't, at least if we discount the evidence from the smaller samples including post-WWII, certainly suggests that it cannot simply be an error of type II factor produced by the various NBER dating committees. There are as many good reasons to have errors of type II in the dating of peaks as there are in the dating of troughs. It should be equally difficult to figure out whether or not an expansion ended in November, in the midst of the Christmas shopping season, as it would be to figure out whether a recession ended in the spring. Therefore, the asymmetry in our findings hints at the true cause being the intrinsic structure of the economy and the cycles it produces rather than the human error factor coming into play.[15]

3.2 Robustness

Does the uneven distribution of troughs critically depend on the uncertainty surrounding the location of turning points? Or, put differently : how robust are the results appearing in Table 3.2? We now turn to these questions, first, by making some general observations followed by inference drawn from an alternative chronology.

The uncertainty regarding turning points is well known and has been widely discussed. We did, of course, allow for a certain amount of heterogeneity, since we broke up the entire sample in smaller samples. But apart from that, there are some more important arguments in favor of the results in Table 3.2 stressing their relative robustness. Indeed, many of the alternative chronologies move the dating of turning

---

14 See Conover (1980) for a technical discussion of using χ^2 tests in small samples.

15 There is, of course, a very different way of distinguishing the error of type II factors from the intrinsic ones which is not covered here. It consists of estimating a stochastic switching-regime model from data like GNP following Hamilton (1989), allowing for intra-year heterogeneity using a periodic Markov chain model. This is done in Ghysels (1992) with the results favoring intrinsic periodicity.

points from only one up to three months relative to the official NBER chronology. This is particularly so for the more recent dates, while the early part of the chronology is often the subject of larger variations. Any movement of a trough from March to July or months in between, from August to November or months in between or any switch between January and February simply does not affect any of our results. In addition, any relocation from January/February to the summer or fall has no impact either, as does March/July to December. All such movements remain inside the two-sample classification. It is only those changes which relocate the turning point from the Christmas/spring sample to the other or vice versa that will have an impact. However, since the rejection rates are so high, at least for the larger samples, it would take a fair amount of reshuffling to turn over the results, keeping in mind that most of the reshuffling will not affect any of the results and that it might turn the results both ways, i.e., reinforcing or weakening them.

To make this argument more concrete, let us consider a recently suggested alternative to the NBER chronology to reinforce our results. It has already been noted that this chronology is the outcome of an effort to clear some of the apparent inconsistencies in the construction of the chronology through the decades. Moreover, it relies on a more coherent data base. The chronology, discussed in detail in Romer (1991), appears in Table A.2 of the appendix. It is clear from a comparison of Tables A.1 and A.2 and from the elaborate discussion in Romer (1991) that significant changes were made in the location of turning points. Let us first return to Table 3.1, where the frequency distribution of troughs and peaks throughout the year is listed. As the chronology starts in 1885, comparisons will be made with the NBER chronology for the corresponding sample sizes only.

Peaks show again a relatively more evenly spread distribution. Over the 1885-1990 period, the differences between the frequency distribution for the NBER and the alternative are not, in fact, that great, confirming the point made earlier that despite the fair amount of reshuffling, there is not necessarily much difference in the frequency distributions. For troughs, we note again a low during August through October, while March and April carry the same number of troughs in the 1885-1990 sample. May through July cover about the same number of troughs (10 for the NBER while only 9 for the alternative one), but it is clear that most of the troughs shifted to the month of July. This argument also holds for the smaller 1885-WWII subsample. We do not try to explain why such a shift occurred, keeping in mind that both chronologies are prone to errors and criticism. What is most important for us is the fact that one can again reject the uniformity of the distribution of troughs according to

Table 3.3 : Goodness-of-Fit Tests for the Uniform Distribution of Alternative Business Cycle Chronology Peaks and Troughs

	1885-1990	1885-WWII	Post-WWII
		Peaks	
Number of peaks Christmas/spring	14	10	4
Number of peaks Other months	11	6	5
Goodness-of-fit test	0.36	1.00	0.11
p-value	0.55	0.32	0.74
		Troughs	
Number of troughs Christmas/spring	16	12	4
Number of troughs Other months	8	4	4
Goodness-of-fit test	2.67	4.00	0.00
p-value	0.10	0.05	1.00

Note : The Christmas/spring classification corresponds to March, April, May, June, July and December. The data appear in Table 3.1 and are described in the appendix.

the results listed in Table 3.3. The evidence in the 1885-1990 period is only at 10 % while the pre-WWII sample yields a p-value of 5 %. Those results are perhaps a bit more borderline, yet they still represent fairly strong evidence of asymmetries in the empirical distribution. Peaks are again clearly uniformly distributed.

3.3 Clustering and randomization

Empirical studies in macroeconomics typically fall short of very conclusive evidence because of the overall quality of data or the limited amount of data available to be retrieved from aggregate time series. Many empirical results can therefore easily be overturned either by choosing a different parameterization, a different sample period or a different assumption about the data generating process. The analysis in the two preceding sections is not immune to this either. In this section, we will indeed try to downplay what has been accomplished so far. This exercise will be useful because it will enable us through the criticism to provide further supporting evidence in the next section.

Although the NBER and alternative chronologies both represent over a century of data, it is clear that with only a maximum of 30 turning points we do not particularly enjoy the luxury of a large sample.

Suppose we were allowed to select in a sample of 30 turning points spread over 12 months those months which show the highest empirical frequency of turning points. Furthermore, suppose we would choose six such months with the highest frequencies, then the average number of peaks or troughs collected by those six months would be 21.99 with the other six months left with 8.01 if the frequency distribution were generated by a uniform distribution throughout the year.[16] Remember that we found exactly 22 troughs in a sample of 30. So, while the χ^2 goodness-of-fit test yielded a low p-value, it is not designed against a "scheme" of picking the highest frequency months and testing for uniformity.

Does this mean that the results in Tables 3.1 and 3.2 are not reliable? Not necessarily, and this for a number of reasons we will elaborate on in the remainder of this section. First of all, the scheme of picking the six most dense months yields most

16 All calculations reported in this section were either done analytically or via simulation. The simulation results are based on 10,000 replications using the SAS function RANTBL.

often an odd collection of months for which little economic rationale would be possible.[17] If turning points are uniformly distributed, how likely is it to have six high frequency months clustered in two sets of months similar to the classification used in the previous two sections? The probability of observing exactly this Christmas/spring classification with 22 troughs, under the assumption of uniformity, is only 0.0012. Moreover, we did not exactly choose the six months with the highest empirical frequency of troughs. Indeed, when we return to Table 3.1, that November has more troughs than April in the 1855-1990 or 1885-1990 samples with the NBER chronology, for instance. In the same table, we observe that with the alternative chronology, both January and February have more troughs than May. All this taken together makes the search scheme argument, credible, yet not all that damaging to the empirical results reported so far.

It should be noted though, that other clustering patterns could have emerged to which perhaps an economic rationale could be given. In fact, again under the assumption of a uniform distribution, the probability of observing the six highest frequency months neatly clustered into two bins is about 18 %, though in many of those cases the two bins would be located during, say, winter or summer or just separated by one month.

What these probability calculations yield is essentially a caveat and word of caution about the error of type I problem associated with the χ^2 goodness-of-fit test results reported in the two previous sections. Is there a way to reinforce the empirical results such that it would be immune to the criticism just raised? Fortunately, there is and it is essentially based on the principle of retrieving more information from the data we have.

3.4 Tests for distributional uniformity with duration data

So far, we focused exclusively on peaks and troughs but never linked the two together. Linking peaks and troughs amounts to analyzing the duration data of cycles, of course; i.e., the length measured in months of expansions and recessions. The idea to apply nonparametric methods to duration data has been advanced in a number of papers by Diebold and Rudebusch (1990a, b) either to investigate the question of

17 Selecting the six high frequency months could yield a collection of every other month starting with January or February, for instance.

duration dependence in economic cycles, or to study the stabilization hypothesis comparing the pre- and post-WWII era. It is in a similar vein that we shall use two-sample nonparametric tests applied to duration data classified along the lines of the seasonal benchmark. The fact, as it appears from the previous section, that troughs display an uneven distribution while peaks don't should translate into uneven distributions of contractions, depending on the time of the year the recession started. An alternative way of saying this is that the hazard rate varies throughout the year and, as a consequence, we should expect duration distributions to differ as well.[18] To test such hypothesis, we shall use nonparametric rank-based tests. Diebold and Rudebusch (1990b) explain in great detail much of the appeal for using such tests. Part of the motivation, of course, is again, as emphasized in section 2, that we do not need to say much about what constitutes the economic environment in order to actually perform those tests. We will first briefly explain the test that we will apply before reporting the empirical results.

Consider two samples of durations of size n_1 and n_2 with $n_1 + n_2 = N$. The durations may represent contraction lengths following peaks in either one of the two seasonal classification samples. The data of the two samples, denoted $\{X_1, .. X_{n_1}\}$ and $\{Y_1, ... , Y_{n_2}\}$ are assumed to be drawn from their corresponding population distributions F and G. If the classification of duration data along seasonal lines has no impact, then we should expect that the two population distributions F and G are identical. In contrast, if either Christmas or spring tend to promote economic recovery as the clustering of trough suggests, then we should expect contraction durations to be shorter or longer, depending on the season. The alternative that will be of particular interest is the one-sided alternative that Y is stochastically larger than X, namely $G(k) \leq F(k)$ for all k. It means that recessions are relatively longer when they follow, for instance, a peak in one of the two samples.

Two test statistics will be used, both rank-based, with known finite sample distribution. The first is the Wilcoxon, or rank-sum test, the second is the Savage test. Diebold and Rudebusch (1990b) used the former, while the latter was introduced in Ghysels (1991). We will first present the two tests and then discuss their power properties.

---

18 A more formal argument can be made, under somewhat more restrictive conditions, using a periodic Markov chain model, which is covered in detail in Ghysels (1991).

The Wilcoxon test is a fairly standard nonparametric distribution-free test. Its distribution-free properties come from the fact that the actual data are replaced by their ranks relative to the joint sample of observations $\{X_1, \dots X_{n_1}, Y_1, \dots, Y_{n_2}\}$. The ranks of the joint sample will be denoted as $\{R_1, \dots, R_N\}$.[19] Then, the Wilcoxon test consists of computing the sum of the ranks over the second sample, namely :

$$W = \sum_{i=n_1+1}^{N} R_i$$

and for samples of even modest size (like 8 observations or more in each sample), a standardized version with a normal distribution is quite accurate. The standardized Wilcoxon statistic is :

$$WS = \frac{W - 1/2 \ N(N+1)}{(n_1 \ n_2 \ (N+1)/12)^{1/2}} \overset{a}{\sim} N(0, 1).$$

The test is designed to be powerful against location shifts and is essentially a nonparametric distribution-free variation on the t-statistic for testing the difference in mean of two samples drawn from Gaussian population distributions. The Wilcoxon test is particularly, but not exclusively designed to detect location shifts between two symmetric distributions. Duration distributions are, as we know, not symmetric. A shift in the mean coincides with an increase in the variance in most cases of asymmetric distributions encountered in duration data. The simplest example, of course, would be the exponential distribution.[20] The fact that an increase in the mean is associated with an increase in variance adversely affects the power of the Wilcoxon rank test, as we shall discuss shortly.

---

19 In the event of tied ranks among the duration data, we used the mean rank of the tied observations. As there are a few tied ranks, we also computed a version of the Wilcoxon test which corrects for the presence of ties [see Conover (1980)]. The results were essentially the same.

20 While Diebold and Rudebusch (1990a) report evidence in favor of no duration dependence. However, as reported in Diebold, Rudebusch and Sichel (1990) that evidence is overturned if the pre-end post-WWII samples are reported. They tend to find strong and positive duration dependence in pre-WWII expansions and post-WWII contractions.

A second nonparametric test for distributional heterogeneity particularly designed for asymmetrically distributed data is the Savage (1956) test. It is specifically but exclusively designed to compare two exponentially distributed data samples and to test the null that they are drawn from the same distribution. The Savage test is formulated as follows :

$$S = \sum_{j=n_1+1}^{N} s(R_j) \overset{a}{\sim} \mathcal{N}(0, 1)$$

where $$s(R_j) = \sum_{\ell=N-j+1}^{N} \ell^{-1}.$$

Hence, instead of taking the actual ranks, like the Wilcoxon rank test, a function of the ranks is used to construct the test statistic. This function makes the test more suitable when the distributions are exponential or asymmetric.

Let us now briefly discuss, in nontechnical terms, the relative merits of both tests for the purpose of our application. Both tests have tabulated finite-sample distributions, and the approximation via the normal distribution is quite accurate even in modestly sized samples.[21] Hence, neither of the tests have any size problems and, consequently, only power is a matter of concern. In the case of duration data, one should expect the Savage test to be more powerful, since it is designed to expect the mean and variance to move in the same direction. A larger variance combined with a mean shift poses more problems with the Wilcoxon test, because it can easily be confused with a situation where simply the variance moves without a mean shift, which is something the Wilcoxon test is not designed to pick up as it only focuses on location shifts.[22] Hence, the Savage test overall is preferred, but that doesn't, of course, mean that the Wilcoxon test is totally inappropriate.[23]

We now turn to the empirical results reported in Table 3.4. The results we focus on are contraction lengths and, more specifically, contraction lengths following a peak

---

21 See Conover (1980) for a technical discussion of using χ^2 tests in small samples.

22 There is a formal discussion of this issue in Savage (1956), Hájek (1969) and Lehmann (1975).

23 In cases like the post-WWII stabilization issue, the relative power of the two tests is less important, because the mean shifts are quite large relative to the dispersion [see Diebold and Rudebusch (1990b)].

Table 3.4 : Contraction Durations After Peaks
NBER and Alternative Chronology
Descriptive and Nonparametric Statistics

	NBER Chronology				*Alternative Chronology*	
	1855-1990	1885-1990	1855-WWII	1885-WWII	1885-1990	1885-WWII
Peaks						
Christmas/spring						
Mean	16.80	13.27	18.64	14.14	9.31	8.70
Standard deviation	8.06	3.00	8.64	2.91	4.73	3.33
Other months						
Mean	19.33	16.69	22.64	19.56	14.64	16.00
Standard deviation	15.90	10.14	17.41	10.82	7.71	9.94
Wilcoxon	0.60	0.30	0.36	0.14	0.02	0.03
Savage	0.08	0.00	0.03	0.00	0.00	0.00

* Entries to the Wilcoxon and Savage tests are p-values for one-sided tests of Christmas/spring versus other month samples.

in either one of the two samples of the Christmas/spring and off-season classification.[24] The results in Table 3.4 list the average duration of contractions, their standard deviation and the p-values of the Wilcoxon and Savage tests. We do not report the post-WWII sample as it is very small. A first thing to note is that all contraction lengths following a peak in the off-season are longer than contraction lengths following a peak in Christmas/spring. The difference is at least three months, but sometimes also more than that. Also, note that every time the mean duration is higher, the standard deviation is higher as well, often significantly higher by a factor of two or three.

The empirical finding that contraction lengths are longer when they start in the off-season is supportive of the fact that troughs are clustered around Christmas/spring for reasons other than randomness. Indeed, when a recession starts in the off-season, its first encounter with a month or months where a turning point is more likely to happen is at an early stage of the recession, i.e., one to three months into the recession. It is partly because of censoring and partly because of more fundamental reasons that we do not have a recession end so early on. Hence, recessions following a peak in the off-season have at least one more hurdle to pass than recessions following a peak in, say, the spring which might easily recover by Christmas.

Are the differences in mean duration reported in Table 3.4 just random or are they statistically significant? For the alternative chronology, there is simply no doubt, both the Wilcoxon and Savage tests strongly reject equal duration distributions in favor of one-sided (as well as two-sided) alternatives. For the NBER chronology, the test results of the Wilcoxon tests favor the null but the Savage test rejects again. Given what was said about the relative merits of both tests, we would tend to favor the Savage test, in particular, in light of the large differences in standard deviations across the samples.

Hence, to conclude, it is fairly clear that the additional information retrieved from the chronology via duration data clearly reinforces the results suggesting that Christmas and spring make economic recovery easier.

---

24 It is at this stage that the classification used throughout the paper is useful to the analysis of peaks, as we now turn to contraction lengths immediately following the peak.

4. CONCLUSIONS AND FURTHER RESEARCH

As explained in section 2, we deliberately chose not to get involved in any detailed description of the economic environment we tried to model. By simply dividing the year into two categories of months which might have an impact on business cycle properties, we did not take a stand on any well-defined and parametric model of sources of shocks and response functions at each stage of the business cycle throughout the year. Despite the abstractions and significant amount of information being lost by taking such a route, we were still able to test whether or not the propensity to leave a recession is uniformly distributed throughout the year.

Any paper in empirical macroeconomics has its flaws due to the shortcomings in data collection, model specification, etc. Despite these flaws, which we tried to minimize via a model-free and distrition-free approach as much as possible, it seems clear that the intra-year dynamics of the economy and the recurrent business cycle are intertwined. We haven't explicitly identified, of course, whether Christmas is a "demand shock" or spring is a "technology shock". In fact, as we observed, once we write down a fully specified equilibrium model, all shocks interact in nontrivial ways via the propagation mechanism.

There are reasons to believe, as we noted, that the uneven distribution is not solely the consequence of human error on the part of the NBER committee members. Other related research consisting of estimating a Markov switching regime model, in fact, confirms this. Hence, there is something intrinsic about the economic structure that is a genuine source of turnaround that is more present at certain times of the year than at others. It raises several fundamental questions, of course. The most obvious one, for instance, is that what many accept as a sound "seasonal adjustment" procedure probably needs some rethinking. But more importantly, we do want to understand some of these phenomena both for the purpose of model building, forecasting and above all economic policy - a task which is beyond the scope of this paper.

APPENDIX

Table A.1 : NBER Business Cycle Reference Dates and Durations

Trough	Peak	Contractions	Expansions
-	June 1857	18	-
December 1858	October 1860	8	22
June 1861	April 1865	32	46
December 1867	June 1869	18	18
December 1870	October 1873	65	34
March 1879	March 1882	38	36
May 1885	March 1887	13	22
April 1888	July 1890	10	27
May 1891	January 1893	17	20
June 1894	December 1895	18	18
June 1897	June 1899	18	24
December 1900	September 1902	23	21
August 1904	May 1907	13	33
June 1908	January 1910	24	19
January 1912	January 1913	23	12
December 1914	August 1918	7	44
March 1919	January 1920	18	10
July 1921	May 1923	14	22
July 1924	October 1926	13	27
November 1927	August 1929	43	21
March 1933	May 1937	13	50
June 1938	February 1945	8	80
October 1945	November 1948	11	37
October 1949	July 1953	10	45
May 1954	August 1957	8	39
April 1958	April 1960	10	24
February 1961	December 1969	11	106
November 1970	November 1973	16	36
March 1975	January 1980	6	58
July 1980	July 1981	16	12
November 1982	July 1990	-	92

Source : Diebold and Rudebusch (1990a).

Table A.2 : Alternative Business Cycle Reference Dates and Durations

Trough	Peak	Contractions	Expansions
–	February 1887	5	–
July 1887	January 1893	13	66
February 1894	December 1895	13	22
January 1897	April 1900	8	39
December 1900	July 1903	8	31
March 1904	July 1907	11	40
June 1908	January 1910	16	19
May 1911	June 1914	5	37
November 1914	May 1916	8	18
January 1917	July 1918	8	18
March 1919	January 1920	18	8
July 1921	May 1923	14	22
July 1924	March 1927	9	32
December 1927	September 1929	34	21
July 1932	August 1937	10	61
June 1938	December 1939	3	18
March 1940	October 1948	12	103
October 1949	August 1953	12	46
August 1954	August 1957	8	36
April 1958	May 1960	9	25
February 1961	October 1969	13	114
November 1970	November 1973	20	36
July 1975	March 1980	4	56
July 1980	July 1981	21	12
April 1983	July 1990	–	87

Source : See Romer (1991), Tables 2 and 3.

REFERENCES

Ayres, L.P. (1939), "Turning Points in Business Cycles", New York, MacMillan.
Barsky, R.B. and J.A. Miron (1989), "The Seasonal Cycle and the Business Cycle", *Journal of Political Economy* 97, 503-534.
Braun, R. and C. Evans (1991), "Seasonality and Equilibrium Business Cycle Theories", Discussion Paper, The University of Virginia.
Burns, A.F. and W.C. Mitchell (1946), *Measuring Business Cycles*, New York, Columbia University Press (for NBER).
Canova, F. (1991), "The Sources of Financial Crises : Pre- and Post-Fed Evidence", *International Economic Review* 32, 689-713.
Chatterjee, S. and B. Ravikumar (1992), "A Neoclassical Model of Seasonal Fluctuations", *Journal of Monetary Economics* 29, 59-86.
Cloos, G.W. (1963a), "How Good Are the National Bureau's Reference Dates?", *Journal of Business* 36, 14-32.
Cloos, G.W. (1963b), "More on Reference Dates and Leading Indicators", *Journal of Business* 36, 352-364.
Conover, W.J. (1980), *Practical Nonparametric Statistics (2nd Ed.)*, John Wiley, New York.
Diebold, F.X., G.D. Rudebusch and D.E. Sichel (1990), "International Evidence on Business Cycle Duration Dependence", Manuscript, Department of Economics, University of Pennsylvania.
Diebold, F.X. and G.D. Rudebusch (1990a), "A Nonparametric Investigation of Duration Dependence in the American Business Cycle", *Journal of Political Economy* 98, 596-616.
Diebold, F.X. and G.D. Rudebusch (1990b), "Have Postwar Economic Fluctuations Been Stabilized?", Institute for Empirical Macroeconomics, Discussion Paper no. 33.
Fels, R. (1959), "American Business Cycles, 1865-1897", Univeristy of North Carolina Press, Chapel Hill.
Ghysels, E. (1986), *Seasonal Adjustment Without Too Much A Priori Economic Theory*, Proceedings of the American Statistical Association, Business and Economic Statistics Section, 150-153.
Ghysels, E. (1987), "Seasonal Extraction in the Presence of Feedback", *Journal of Business and Economic Statistics* 5, 191-194.
Ghysels, E. (1988), "A Study Toward a Dynamic Theory of Seasonality for Economic Time Series", *Journal of the American Statistical Association* 83, 168-172.
Ghysels, E. (1990), "On the Economics and Econometrics of Seasonality", Invited Symposia Paper presented at the World Congress of the Econometric Society, Barcelona.
Ghysels, E. (1991), "Are Business Cycle Turning Points Uniformly Distributed Throughout the Year?", Discussion Paper, C.R.D.E., Université de Montréal.
Ghysels, E. (1992), "A Time Series Model of Growth Cycles and Seasonals with Stochastic Regime Switching", Discussion Paper, C.R.D.E., Université de Montréal.
Hájek, J. (1969), "A Course in Nonparametric Statistics", Hidden-Day, San Francisco.
Hamilton, J.D. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle", *Econometrica* 57, 357-384.
Hansen, L.P. and T.J. Sargent (1990), "Recursive Linear Models of Dynamic Economies", Manuscript, Hoover Institution, Stanford University.
Hansen, L.P. and T.J. Sargent (1991), "Seasonality and Approximation Errors in Rational Expectations Models", *Journal of Econometrics* (forthcoming).
Lehmann, E.L. (1975), *Nonparametrics : Statistical Methods Based on Ranks*, Holden-Day, San Francisco.

Miron, J.A. (1986), "Financial Panics, the Seasonality of the Nominal Interest Rates and the Founding of the Fed", *American Economic Review* 76, 125-140.
Miron, J.A. (1990), "The Economics of Seasonal Cycles", Invited Symposia Paper presented at the World Congress of the Econometric Society, Barcelona.
Mitchell, W.C. and A.F. Burns (1938), *Statistical Indicators of Cyclical Revivals*, New York, National Bureau of Economic Research.
Moore, G.H. and V. Zarnowitz (1986), "The Development and Role of the National Bureau of Economic Research's Business Cycle Chronologies", in R.J. Gordon (ed.), *The American Business Cycle : Continuity and Change*, Chicago, University of Chicago Press (for NBER).
Neftçi, S.N. (1984), "Are Economic Time Series Asymmetric over the Business Cycle?", *Journal of Political Economy* 92, April, 307-328.
Persons, W.M. (1931), Forecasting Business Cycles", John Wiley, New York.
Romer, C.D. (1991), "Remeasuring Business Cycles : A Critique of the Prewar NBER Dates", Discussion Paper, University of California at Berkeley.
Sargent, T.J. (1978), Comments on "Seasonal Adjustment and Multiple Time Series Analysis", by K.F. Wallis, in Seasonal Analysis of Economic Time Series, A. Zellner (ed.), Washington, DC, U.S. Department of Commerce, Bureau of the Census, 361-364.
Savage, I.R. (1956), "Contributions to the Theory of Rank Order Statistics - the Two-Sample Case", *The Annals of Mathematical Statistics* 27, 590-615.
Sims, C.A. (1991), "Rational Expectations Modeling with Seasonally Adjusted Data", *Journal of Econometrics* (forthcoming).
Singleton, K.J. (1988), "Econometric Issues in the Analysis of Equilibrium Business Cycle Models", *Journal of Monetary Economics* 21, 361-386.

Stock, J.H. and M.W. Watson (1989), "New Indexes of Coincident and Leading Economic Indicators", in O. Blanchard and S. Fischer (eds.), *NBER Macroeconomics Annual, 1989*, Cambridge, Mass., MIT Press.

Trueblood, L.C. (1961), "The Dating of Postwar Business Cycles", *Proceedings of the American Statistical Association, Business and Economic Statistics Section, 1961*, 16-26, Washington, DC, American Statistical Association.

Zarnowitz, V. (1963a), "On the Dating of Business Cycles", *Journal of Business* 36, 179-199.

Zarnowitz, V. (1963b), "Cloos on Reference Dates and Leading Indicators : A Comment", *Journal of Business* 36, 461-463.

Watson, M.W. (1992), "Business Cycle Durations and Postwar Stabilization of the U.S. Economy", NBER Working Paper No. 4005.