The Cooper Union
Albert Nerken School of Engineering

CLASSIFYING PHASES OF THE BUSINESS CYCLE: A MACHINE LEARNING APPROACH

A Thesis in
Electrical Engineering
by
Julia Astrauckas

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Engineering

September 2015

The thesis of Julia Astrauckas was reviewed and approved by the following:

Sam Keene
Professor of Electrical Engineering
Advisor

Richard Stock
Dean of Engineering

Abstract

In a capitalist economy, it has long been observed that the GDP of a country fluctuates about its longer-term trend. The nature and causes of this business cycle are a subject of interest among economists and econometricians. If patterns can be identified in economic indicators that are reported promptly and frequently, a classification system could serve as an early indicator for peaks and troughs in the business cycle. In recent years, the field of machine learning has provided new techniques for classifying and modeling data. This thesis seeks to apply some of these techniques to classify the current phase and predict the near-future phases of the business cycle at different points in time, using contraction and expansion periods hand-labeled by economists as ground truth. A novel classification method using wavelet features and hidden Markov models is proposed, and the accuracies and advantages of different methods are compared.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I owe enormous gratitude to the Cooper Union for allowing me to conduct this research with a full tuition scholarship, and to the inspiring faculty and student body for providing an environment in which technical and academic conversations flourish.

I would like to especially thank Dr. Sam Keene, my advisor, for monitoring both the technical and administrative aspects of my project, and Dr. John Sarich, for providing much-needed insight into the fields of Economics and Econometrics.

Additional thanks to David Rubinstein for the LaTeX assistance, Kathy Astrauckas and Nick Ober for voluntarily proofreading my drafts, and Daniël Reichman for answering my most existential math questions.

Chapter 1

Introduction

Hundreds of years after the origin of classical economics, economists and econometricians continue to debate the causes of fluctuations in capitalist economies, and whether anything should be done to intervene. Meanwhile, a greater breadth and depth of economic data are available than ever before. As the history of our financial bookkeeping grows longer and longer, the opportunities to find and identify recurring patterns increase. At the same time, recent innovations in mathematics, econometrics, and machine learning provide a new arsenal of techniques that may be used to process and make inferences about these data sets. Still, macroeconomic data is noisy and complicated, capturing a myriad of sudden, unpredictable changes, yet displaying reliable trends in the long run. Somewhere between the noise and the trend in frequency lies the business cycle.

The idea of a somewhat reliable, recurring business cycle is a relatively recent idea. Economic crises in capitalist economies were long considered to be the product of exogenous shocks, each unrelated to the others, in accordance with equilibrium theory [2]. While the existence of business cycles is now generally acknowledged, a debate continues on the extent to which they can be controlled by intervention. As more evidence emerges to suggest these cycles are a natural and inherent part of capitalism, more questions arise about whether they can or should be mitigated by public policy [3][4]. Applying new techniques or analyzing longer data series to demonstrate the existence of business cycles (or lack thereof), is therefore of some practical interest to economists. Additionally, if patterns can be found in the cyclical components of economic data, the phases of the business cycle could be classified automatically. This could provide an indicator of economic turning points more quickly than the labels assigned by economists. These labels are taken as ground truth, and rightfully so, since identifying peaks and troughs with complete certainty requires the consideration of many factors and some degree of hindsight [5]. However, the most recent National Bureau of Economic Research turning point, which marked the end of the contraction associated with the financial crisis of 2008, was announced over a year after it occurred [6]. An automatic classification system certainly wouldn't replace this kind of expert analysis, but it could provide a much faster estimation of turning points if it could operate on quarterly data available with a low latency. Since business cycles often last years, it would be beneficial and nontrivial to classify turning points between periods of contraction and expansion as they occur, without mistaking high-frequency noise for meaningful shifts [7]. A more difficult but arguably more useful task would be to use current economic data series to predict *future* turning points in the business cycle.

This thesis introduces a novel approach for classifying business cycle stages, with the aforementioned motivations in mind. Chapter 2 presents background information on the relevant economic theories and econometric analysis techniques, the machine learning algorithms applied to perform classification, and previous approaches to cycle classification. Chapter 3 details the data series that were considered as classification features, as well as the actual feature spaces and classifiers used, culminating in the combination of wavelet decomposition features and a hidden Markov model. Chapter 4 compares results of different classification systems and discusses evaluation metrics. Finally, Chapter 5 describes conclusions and future work.

Chapter 2

Background

2.1 The Business Cycle: Definitions, Theories, and Models

In economics, the business cycle describes alternating periods of growth and contraction about a long term trend. The existence and causes of the business cycle have been debated by economists for centuries. Many early economists believed equilibrium would occur in the long run, and cited individual, unrelated events to explain any fluctuation in the rate of economic growth [2]. As time passed and industrial capitalism became more mature, the recurring periods of expansion and contraction led economists to reconsider the classical idea of steady-state equilibrium [4]. Business cycle theories fall into two main categories: exogenous and endogenous. Exogenous business cycle theories posit that phases of the business cycle are caused by external events, which may themselves be recurring or periodic. Endogenous models, by contrast, claim fluctuations arise inherently from the natural course of economic activity and are caused by factors within the system. With the increasing duration of macroeconomic data series, it has become increasingly possible and interesting to examine the strength of these claims empirically.

2.1.1 The Business Cycle in Classical Economics

In the eyes of classical economists, capitalist economies unfettered by regulations and external crises were bound to head towards equilibrium. J.B. Say, famed for his eponymous Say's Law, stated in 1803 that "supply calls forth its own demand" [4]. Say, along with James Mill and David Ricardo, argued that the very act of production creates enough income and demand to absorb the produced supply. Under such a system, reliably repeating decreases in economic activity should not be possible in the absence of external shocks. They further reasoned that any money not spent would be invested (and thus, would be spent elsewhere), since investments would yield a positive return that savings would not. While it is true that the level of production equals the level of income, there is nothing in Say's Law to suggest that that level must be fixed at full employment (maximum production).

As capitalist economies began experiencing rapid growth in the early 19th century, several economists and other observers began to note fluctuations in the growth of the economy. One of the earliest was the Swiss writer Jean Charles Lonard de Sismondi. While not a formal economist himself, Sismondi built upon the theories presented in Adam Smith's magnum opus, *The Wealth of Nations*. He argued in *Nouveaux Principes d'conomie politique* that the unregulated activity of banks and investors would tend to lead to periodic economic crises, rather than the stable equilibrium imagined by Smith [8]. Though his writings were mainly loquacious musings rather than mathematical or empirical theories, his ideas provided one explanation for the Panic of 1825, an English stock market crash with global consequences. Unlike previous economic depressions, this crisis had not been caused by a war or any other obvious external force [9].

2.1.2 Karl Marx

Marx was one of the first and best-known economists to reject the idea that an economy with a monetary system and credit would be headed toward equilibrium. He argued that credit was speculative in nature, and could postpone a decline, but at the expense of the production of a surplus of goods resulting in an intensified recession [4]. He theorized that consumers would also use credit too liberally in anticipation of future income, resulting in increased debt and financial vulnerability. In this model, economic growth naturally leads to a period of decline and recession in which consumers cannot buy and producers cannot sell at the optimistic rates anticipated during the boom. At the trough of this cycle, the low demand for capital would depress interest rates to the point where they would encourage borrowing and an increase in economic activity, thus restarting the cycle [4].

2.1.3 John Maynard Keynes

With the Great Depression as a backdrop in the 1930s, John Maynard Keynes launched a formal and largely successful attack on the conventional wisdom attached to Say's Law. With unemployment rates at 25%, and an observable excess of supply that could not be absorbed by the existing demand, the atmosphere was ripe for this sort of challenge [4]. Rather than assuming that demand would rise to meet supply, Keynes envisioned the supply falling gradually to meet the reduced demand, which would negatively affect the employment rate. Furthermore, Keynes' work presented a scenario in which money could temporarily drop out of circulation, in direct opposition to the classical assumption of savings-investment equality [10]. Keynes argued that interest rates would stop falling below a certain point, since prospective lenders would speculatively keep their savings out of circulation in anticipation of higher interest rates [4]. The introduction of money into an economy, according to Keynes, creates this potential discrepancy between savings and investment due to speculation and hoarding. These kinds of activities both limit the fall of interest rates and the fall of prices during a period of underconsumption [4].

Post-Keynesian economists, following Keynes' lead, further developed the problems that could result from interest rate rigidity and investor speculation. In the event of an excess aggregate supply, an unchecked decline in prices would decrease the revenue resulting from the sales of goods, but this would also decrease the aggregate income. The same can happen with labor supply; if labor costs and thus wages fall dramatically, so does the capacity of workers' consumption [4]. In this Keynesian view of the economy, the supply-demand equilibrium cannot be reached in a case of excess supply without a decrease in the aggregate income, causing a recession.

2.1.4 Wesley Mitchell and the National Bureau of Economic Research

While the concept of periodic expansions and contractions of the economy were observed in the early 19th century, the first concise mathematical theories wouldn't be developed until the middle of the 20th century. Wesley Mitchell, who founded the National Bureau of Economic Research (NBER), created a quantitative method for measuring the business cycle in the 1940s. [4]. In a collaboration with Arthur Burns, Mitchell wrote that the business cycle is characterized by expansion in economic activities followed by a period of contraction. He claimed they could last from one year to as many as twelve years, and that they were not divisible into shorter cycles. Mitchell's business cycle is "recurrent but not periodic" [11], acknowledging that cycles occur routinely, but the duration of each will vary. Mitchell dated troughs and peaks of the business cycle, and this work was continued by NBER, resulting in the business cycle phase labels (expansion and contraction) used later in this work. These cycle labels are perhaps the most widely used, but certainly not flawless, as they do not differentiate between major depressions and minor recessions [4], and they do not account for the relationship between prices and output that is better described by the four phase cycle.

Although Mitchell only labeled two phases of the business cycle, he defines other models that break the cycle into four and nine phases, respectively.

2.1.4.1 Two Phase Model

The two phase model of the business cycle was Mitchell's simplest model and the one that NBER currently provides dates for. This cycle model consists simply of *expansion*, or economic growth, and *contraction*, or economic decline. This is the easiest model to work with, as it contains the smallest number of phases and NBER provides dates hand-labeled by experts dating from the 1800s to the present.

2.1.4.2 Four Phase Model

In his four phase model, Mitchell begins to address varying rates of expansion and contraction. The four phases are [4]:

- *recovery*, the rapid increase in economic activity, occurring after the cycle reaches its trough
- *prosperity*, the slower period of expansion at a relatively high level, as the cycle approaches its peak
- *crisis*, the fast downturn occurring after the cycle has reached its peak
- *depression*, the period of low economic activity and slow decline leading to the trough of the cycle

The four phases that make up the business cycle in this model are clearly defined, but still difficult to label without a significant amount of economic expertise.

2.1.4.3 Nine Phase Model

Mitchell's nine phase model arises from a very detailed analysis, but the number of phases is somewhat arbitrary [4]. The first phase is the trough of the cycle and the fifth is the peak. The intermediate three phases describe the expansion in more detail, and the remaining four phases describe the contraction and the next trough [4]. This complicated model is perhaps more useful for qualitatively describing the progression of the cycle than quantitatively classifying it, as no analysis encountered in the course of this thesis has been able to find this many statistically distinct states of the cycle. The phases from the four phase model can also be described in terms of a sequence of phases from the nine phase model.

2.1.5 Richard M. Goodwin and Goodwin's Model

Goodwin's formulation of the business cycle is inspired by Marxian theories, with wages and the employment rate at the center. [12]. Goodwin's cycle model is endogenous, and the solutions to the differential equations he lays out are the same Lotka-Volterra equations used to describe predator-prey relationships [13].

2.2 Macroeconomic Data

The ability to draw conclusions or make predictions about the state of an economy depends upon the routine, accurate reporting of macroeconomic variables. This is challenging for a number of reasons. First, any measurement has some amount of error due to noise, and noise is exacerbated in data collection processes that require human reporting and estimation. The unemployment rate is one particularly problematic example of a biased statistic, since it fails to account for people who have given up on looking for work, or those who would like a full-time job but can only find part-time work [14]. Second, any change or improvement to a measurement process creates a discontinuity that prohibits long-term analyses of that particular variable. For example, the way states' gross domestic products were calculated changed in 1997 [15]. Nonetheless, many economic time series are available, some dating as far back as the nineteenth century. This includes leading and lagging indicators [8], meaning that variables are not necessarily synchronized and may precede or react to economic shocks. The following are some of the key economic indicators used in this and other analyses, along with a description of transformation commonly applied to these series.

2.2.1 Gross Domestic Product

The Gross Domestic Product, or GDP, is an aggregate measurement of production [16]. Loosely, it is a measure of the total output (goods and services) of a nation or other entity, excluding imports. It is often computed per capita so that economies can be compared among nations of dramatically different sizes.

2.2.2 Consumer Price Index

The consumer price index, or CPI, measures the changes in the prices of a fixed reference set of common consumer goods over time [16].

2.2.3 Unemployment Rate

The unemployment rate is formally a measure of the percentage of the population of potential workers who are seeking employment but are not employed [16]. Most computations of the unemployment rate do not take into account underemployment or partial employment (workers who would like full-time jobs but can only find part-time work) [14].

2.2.4 Federal Surplus

The surplus (or deficit, if it is negative) is the difference between the total revenue and expenditures of the federal government. It is one of the more contentious economic indicators, as not all economists agree that it matters significantly to the health of an economy [17].

2.2.5 Stock Market Composite Indicies

Stock market composite indices do not provide a measure of a specific economic variable, but are often used as a barometer of economic health. Because the data used to compute these indicies are available publicly and virtually instantaneously, they can be much more useful for timely observations and predictions than variables that must be computed annually or quarterly by government agencies [18][6]. However, stock market data is also much noisier than other series, as it contains many speculative upturns and downturns that are ultimately short-lived and not representative of actual business cycle turning points [6]. The National Bureau of Economic Research considers stock market composites to be leading business cycle indicators [6], so if the signal can be separated from the noise, these indicies can be very useful for cycle classification and prediction.

2.2.5.1 Dow Jones Industrial Average

The Dow Jones Industrial Average (DJIA) is a composite index which has been reported continuously since 1896, although its composition has changed many times since then. It was originally an average of 11 industrial companies' stocks, and has since evolved to include 30 modern companies [18]. The value of the index is computed as a price-weighted average of the stocks of the 30 designated companies, and normalized based on its value at some reference point.

2.2.5.2 NASDAQ Composite Index

The NASDAQ composite is an index measuring the average levels of stocks listed on the NASDAQ stock market. It only dates back to 1971, which makes it one of the shorter time series considered for classification [19].

2.2.5.3 S&P 500

As the name suggests, the Standard & Poor's 500 is a composite index computed from the stock prices of 500 large companies, both from the NASDAQ and the NYSE [18].

2.2.6 Common Transformations for Macroeconomic Time Series

While the series described in sections 2.2.1 through 2.2.5 contain useful information about the state of the economy, some processing is often required to create data that is well suited for examining specific properties. Using the raw values of economic time series is undesirable when searching for cycles and other repeated patterns, since these series all have a strong exponential trend due to inflation. The following are a number of simple transformations that can produce more useful series depending on the data and application.

2.2.6.1 Percent Change over Previous Period

Percent change is simply computed by taking the difference between two subsequent measurements and dividing by the earlier measurement. Since we expect an exponential increase between linearly spaced points in time, the percent change series should be stationary for a series with constant exponential growth, and thus highlights fluctuations rather than the predictable trend.

2.2.6.2 Year-Over-Year Percent Change

Year-Over-Year (YOY) percent change is computed in the same way as the percent change over the previous period, except the previous data point used as a comparison is from one year in the past instead of one measurement period. For example, monthly data points would be compared to the same month in the previous year. This adjusted percent change computation accounts for seasonality, such as consumer spending increases every December.

2.2.6.3 Seasonal Adjustment

Like YOY percent change, seasonal adjustment seeks to mitigate the effects of predictable, fixed-period variations in time series. A variety of techniques exist for performing this adjustment.

2.2.6.4 Natural Logarithm

Logarithms will remove any exponential trend, but can only be applied to data that is guaranteed to be nonzero. Luckily, that includes any prices or relative indices; it only excludes data such as federal deficit or series that have already been reported in terms of percent change.

2.2.6.5 Real Chained Dollars

An attempt to account for inflation directly, translating a series into "real dollars" adjusts a series by selecting a fixed reference point and normalizing all other dollar values in the series to their equivalent values at that point in time. Chained dollar adjustments allow the reference list of items and prices to vary annually to reflect changes in relative economic importance of goods and commodities, rather than fixing the list of items as some lose relevance.

2.3 Analysis Techniques for Economic Data

2.3.1 Stationarity Tests

A stationary time series is one whose statistical properties or structural composition does not vary over time. In the context of econometrics, two types of stationary series are useful to identify: *trend stationary* and *difference stationary* [20]. A *trend stationary* series can be represented as

$$x_t = \alpha + \beta t + u_t, \tag{2.1}$$

where x_t is stationary except for the deterministic trend βt, and u_t is a stationary stochastic process (its distribution is not time-varying). A *difference stationary* series is described by

$$x_t = \gamma + x_{t-1} + u_t, \tag{2.2}$$

where γ is the drift parameter and u_t is, once again, stationary. A *difference stationary* series can also be described as a random walk with drift [20]. Often, if an economic time series is not stationary, the difference will be tested for stationarity and used in place of the original series.

The ADF-GLS test is a unit-root hypothesis test [21], used in this thesis to test the stationarity of time series, once again to compare results to existing econometric work. A unit root test checks whether any of the roots of a time series' characteristic equation is close to 1, which means the series is nonstationary [22].

2.3.2 Hodrick-Prescott filter (HPF)

The Hodrick-Prescott filter is another common technique in econometrics used to create a stationary time series when using analysis or regression methods that require stationary data [23]. It is a filter that is commonly used to extract the long-term trend from cyclical components in nonstationary data. However, the HPF can actually introduce or exaggerate cyclical components when the input data is not stationary [23][24]. As such, it is used in this thesis as a means of understanding and comparing results with existing econometric research, but the resulting decomposition is not considered the most accurate representation of cyclical components in the data.

The Hodrick-Prescott filter models a time series as a trend with a cyclical component and stationary noise, as described by equation 2.3.

$$y_t = \tau_t + c_t + \epsilon_t \tag{2.3}$$

The cost expression to be minimized is given by equation 2.4, which is a balance between the squared error and the jumps between adjacent points in the trend component.

$$min\left(\sum_{t=1}^{T} (y_t - \tau_t)^2 + \lambda \sum_{t=2}^{T-1} [(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1})]^2 \right) \tag{2.4}$$

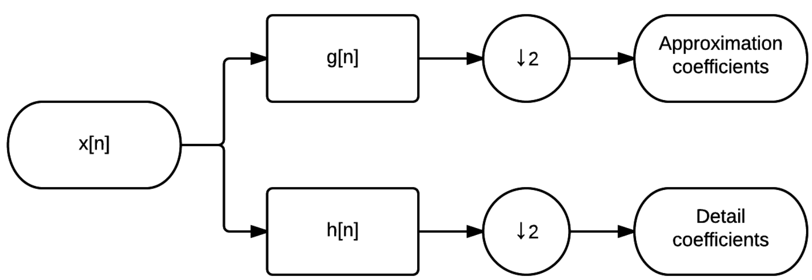
2.3.3 Vector Autoregression (VAR)

Vector autoregression is a technique from econometrics used to model linear interdependencies in multivariate data [12]. It describes the evolution of each time series in terms of the lags (previous discrete points in time) of the others. It can only be applied to stationary series, so the data may be filtered or the differences may be computed beforehand. See [12] for the application of VAR in an econometric context.

2.3.4 Wavelet Decomposition

While the business cycle's very name implies periodicity, a cursory examination of the expansions and contractions in the twentieth century shows that duration of cycles can vary dramatically. This is, in part, why the business cycle pioneers defined their theorized cycle durations in relatively broad ranges. Unlike the sine waves used in the conventional Fourier basis, wavelets are localized in both time and frequency [1], so they are particularly useful for an application in which the cycle "frequency" actually varies from cycle to cycle.

Figure 2.1. Diagram of single-level wavelet decomposition



2.3.4.1 The Discrete Wavelet Transform (DWT)

The wavelet transform decomposes a signal into components represented by a basis of dilated and translated wavelets. The individual wavelet functions are referred to as time-frequency atoms, or simply atoms [25]. A single level of the discrete wavelet transform is shown diagrammatically in figure 2.1. Mathematically, the discrete wavelet transform of a signal $x[n]$ is given by

$$y[n] = (x * g)[n] \tag{2.5}$$

In equation 2.5, $g[n]$ is a specific wavelet function, usually corresponding to the lowpass approximation coefficients. The output $y[n]$ is the convolution of the input signal $x[n]$ and the wavelet filter $g[n]$. To achieve the decomposition shown in figure 2.1, two filtering and decimation operations must be applied to $x[n]$.

$$y_{approx}[n] = \sum_{k=-\infty}^{\infty} x[k]h[2n-k] \tag{2.6}$$

$$y_{detail}[n] = \sum_{k=-\infty}^{\infty} x[k]g[2n-k] \tag{2.7}$$

Equations 2.5 through 2.7 are simply convolutions; the real point of interest when using the DWT is the selection of the mother wavelet and the dictionary of basis functions [25][26]. The mother wavelet is denoted $\psi(t)$, and the discrete dictionary of child wavelet functions is created by translating and dilating the mother wavelet by powers of two. The child wavelet is given by

$$\psi_{j,k}[t] = \frac{1}{\sqrt{2^j}}\psi\left(\frac{n-k2^j}{2^j}\right) \tag{2.8}$$

where the dilation parameter j and translation parameter k are integers. For a particular scaling j, the continuous time wavelet filter can be written as

$$h(t) = \frac{1}{\sqrt{2^j}}\psi\left(\frac{-t}{2^j}\right) \tag{2.9}$$

There are many different choices of wavelets, with varying shapes, support lengths, and vanishing moments that make them well-suited for different applications.

2.4 Machine Learning Techniques

In classifying phases of the business cycle, the choices of data series and preprocessing techniques are only half the challenge. The other half is choosing an appropriate classifier to robustly identify meaningful patterns out of the noisy, multidimensional data. The following are classifiers that were used in the systems proposed in this thesis.

2.4.1 Linear Discriminant Analysis (LDA)

A discriminant function is any function that takes a feature vector $\boldsymbol{x}$ and assigns it to one of K classes, C_k [1]. If the discriminant function is linear, then it can be called a linear discriminant and its decision surface is a hyperplane. In the simplest case of two classes, the relationship between the feature vector $\boldsymbol{x}$ and the linear discriminant is

$$y(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x} + w_0 \tag{2.10}$$

where $\boldsymbol{w}$ is the vector of weights that define the discriminant. Since there are only two classes, the decision threshold is at $-w_0$, where w_0 is the bias weight. If a vector $\boldsymbol{x}$ generates a value of $y \geq 0$, the input vector will be classified as class C_1. Otherwise, it will be placed in C_2. The task of linear discriminant analysis is to learn the values of the weight vector that optimize some desirable criterion.

2.4.1.1 Fisher's Linear Discriminant

One way of selecting the weights for a linear discriminant is to maximize the separation between the two classes. To do so, the distance between the centroid of each class and a given decision line or plane $\boldsymbol{w}$ must be computed. If there are N_1 points in C_1 and N_2 points in C_2, then the mean of each class is given by

$$\mu_1 = \frac{1}{N_1}\sum_{n\in C_1}\boldsymbol{x}_n, \tag{2.11}$$

$$\mu_2 = \frac{1}{N_2}\sum_{n\in C_2}\boldsymbol{x}_n. \tag{2.12}$$

The distance between the means of the two classes is

$$m2 - m1 = \boldsymbol{w}^T(\mu_2 - \mu_1) = \boldsymbol{w}^T\mu_2 - \boldsymbol{w}^T\mu_1. \tag{2.13}$$

However, this quantity should not be maximized directly, since this would result in making $\boldsymbol{w}$ arbitrarily large. Instead, the goal is to maximize the variance between classes while minimizing the variances within each class. This promotes a large separation between the classes while adding a penalty for a large variance within each class. The intra-class variance for C_K is

$$s_k^2 = \sum_{n \in C_k} (y_n - \mu_k)^2. \tag{2.14}$$

The Fisher Criterion, denoted $J(\boldsymbol{w})$, is the inter-class variance divided by the sum of the intra-class variances.

$$J(\boldsymbol{w}) = \frac{(m2 - m1)^2}{s_1^2 + s_2^2} \tag{2.15}$$

Using the preceding equations, the $\boldsymbol{w}$ that optimizes the Fisher Criterion is found to be

$$\boldsymbol{w} \propto S_W^{-1}(\mu_2 - \mu_1), \tag{2.16}$$

where S_W is the within-class (intra-class) variance

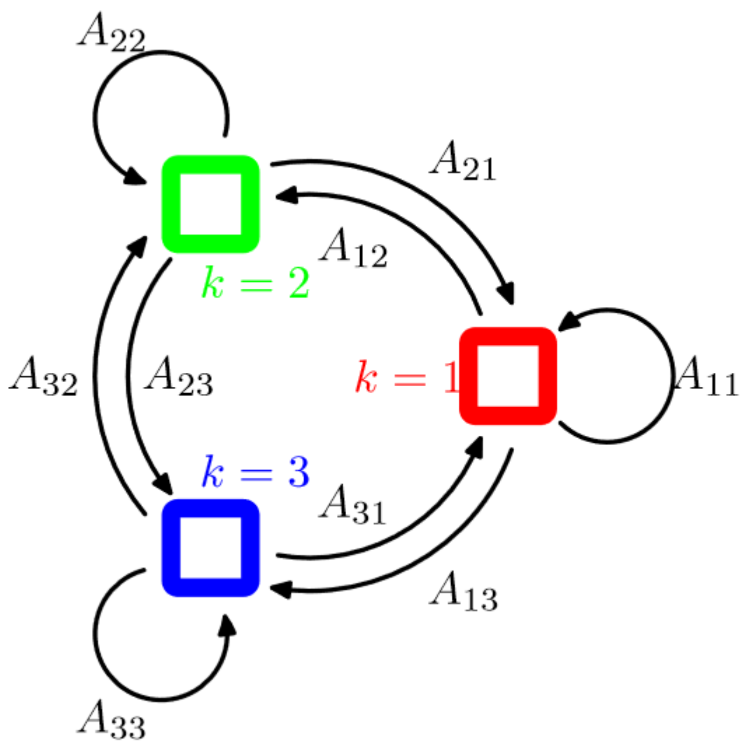
$$S_W = \sum_{n \in C_1} (\boldsymbol{x}_n - \boldsymbol{\mu}_1)(\boldsymbol{x}_n - \boldsymbol{\mu}_1)^T + \sum_{n \in C_2} (\boldsymbol{x}_n - \boldsymbol{\mu}_2)(\boldsymbol{x}_n - \boldsymbol{\mu}_2)^T. \tag{2.17}$$

A full derivation of the weights that optimize $J(\boldsymbol{w})$ is given in [1]. Once the weight vector has been learned from a labeled set of training data, it can be used to classify new, unlabeled feature vectors.

2.4.2 Hidden Markov Models (HMM)

Hidden Markov Models interpret a time series as a sequence of emissions generated by some underlying sequence of unobservable latent states. The latent variables are discrete, corresponding to some set of possible states. In a classification problem, the states are the possible classes, and the goal is to find sequence of states that is most likely responsible for the observed emissions. The emissions at time t depend on the latent state x_t, which in turn depends on the previous state x_{t-1}[1]. Figure 2.4.2 shows an HMM with three states and their respective transition probabilities A_{ij}.

Figure 2.2. Hidden Markov Model State Transitions



An HMM with three possible states. The black lines indicate the transition [1].

Let the parameters of the HMM be defined as follows:

- X: state space
- Y: observation space
- a: state transition probabilities
- A: transition matrix
- b: emission probabilities

- B: emission matrix

Then, given a fully observed training data set that includes both $\boldsymbol{x}$ and $\boldsymbol{y}$, the transition matrix A and emission matrix B can be estimated. For future data (test data) fed into the trained HMM, the Viterbi algorithm can be used to efficiently compute the most likely values of the unknown state sequence $\boldsymbol{x'}$, given a test sequence of observations $\boldsymbol{y'}$ [1].

2.4.2.1 Mathematical Model with Discrete Emissions

Using the parameters enumerated above, the following describes the way these parameters interact in a hidden Markov model. The state variable, X, can take on N discrete values, with a possible transition at every discrete time index t, where $t \in 1 : T$. The transition matrix A is an NxN matrix, where the set of transition probabilities from each state must sum to 1. The diagonal represents the probability of remaining in each state at time $t + 1$.

For each state $x \in X$, there is a different probability distribution governing the value of the emission variable Y. If the emission values are discrete, then $y \in Y$ is a discrete random variable and the emission matrix B contains the probabilities of each possible emission value given the current state (b_{ij} is the probability of observing emission y_i in state x_j. Thus, if there are M discrete values which Y can take on, B is an MxN matrix.

2.4.2.2 Multivariate Gaussian Emissions

If Y is a continuous random variable, the emission probabilities cannot directly be described by a matrix, since there is an uncountably infinite number of possible emission values. However, the parameters of a multivariate Gaussian may be used in lieu of the discrete emissions matrix. Not only does this allow for continuous emissions variables, but also a multivariate emission space. Instead of a single observation variable Y, there may be a vector of observations at each time t, denoted $Y = [Y_1, Y_2, ...Y_D]$ for a D-dimensional observation space. In this model, there are D parameters describing the means of the multivariate Gaussian distributions for each state, and a DxD covariance matrix. The means and covariance can be computed directly for a particular state from the samples in the observation sequence

$\boldsymbol{y}$ corresponding to that state.

2.4.2.3 The Viterbi Algorithm

Once the transition and emission probabilities of a hidden Markov model have been learned, estimating the transition and emissions probabilities from a known sequence of observations and latent states, that model can be used to compute the most likely state sequence responsible for new observations [1].

2.5 Related Work

2.5.1 Business Cycle Analysis

Although there have been relatively few attempts at automatic classification of business cycle phases and turning points, many economists have examined data for quantitative evidence and characteristics of business cycles. In [12], Artur Tarassow analyzes quarterly employment and wage data in search of evidence of Goodwin's endogenous business cycle. Using bivariate vector autoregression (VAR) on cyclical components extracted both by the Hodrick-Prescott filter and the Baxter-King bandpass filter, he finds that employment dynamics lead and influence real wages, but there is not much evidence of the converse.

Recently, John Sarich conducted an empirical investigation of business cycles of different durations in U.S. corporate profit data [27]. He used an unobserved components method to extract trends and cycles from annual profit and equity price series. Unlike most of the other research mentioned, he checks for cycles of dramatically different lengths, including the 50-year Kondratieff long wave. He is successful in identifying two cycles of this long wave in corporate profit data, as well finding statistically significant structural time series model for the more conventional shorter cycles.

The preceding papers are not the full scope of econometric business cycle research, but they are two recent and successful attempts at extracting business cycle information from economic data. While econometric analyses of the business cycle do not necessarily provide features or models that are directly suitable for an automatic classification task, they do provide evidence that business cycle information can be extracted from real data series, which lends credibility to the creation of an automatic classification system.

2.5.2 Cycle Classification

The earliest attempt to classify phases of the US business cycle encountered in this research used linear discriminant analysis (LDA) to distinguish among phases in the US business cycle [28]. Using less than 30 years worth of economic data, Meyer and Weinberg attempted to identify four phases of the business cycle using an ensemble of indicator variables. They defined the four phases as recession, recovery, demand-pull, and stagflation, and they correspond to Mitchell's four phase model. The authors hand-labeled ground truth for their four phase model in order to go beyond the two phase labels provided by the National Bureau of Economic Research. They generate two canonical functions to be used for classification, and tabulate both the a-priori phase labels and the labels assigned by the LDA. 13 variables were used in total, including real GNP, the consumer price index, unemployment rate, and the NYSE composite index, among others. It was demonstrated that the linear discriminant could assign, with roughly accurate placement, cycles with three to four phases in post-war data.

More recently, Heilemann and Muench applied a similar technique to West German data [29]. They used a similar four phase model as well as a linear discriminant for classification. Their overall classification error rate was a very respectable 13.7% using leave-one-out cross validation, but rose sharply to 43.8% when an entire cycle was withheld during the training process.

Continuing the task of cycle phase classification with West German data, Ralf Klinkenberg applied his novel concept drift algorithm to the same data used by Heilemann and Muench [30]. They achieved a quarterly classification accuracy of 79.45% using a binary cycle phase model.

In the US, the task of automatic turning point classification was revisited in 2003 in an effort funded by the Federal Reserve Bank of Atlanta. [31]. Using Markov-switching multifractals with GDP and employment data, they provide automatic binary phase classifications and compare them to the NBER labels. Rather than evaluating their system in terms of percent accuracy, they compute the mean and standard deviation of their distance from the ground-truth NBER labels. The system was fairly successful at matching NBER's turning point dates; some of their results are tabulated for comparison in Chapter 4.

2.5.3 Wavelet Analysis of GDP Data

Motohiro Yogo's 2003 research explored the use of wavelet domain filtering for economic time series [32]. In particular, Yogo created a filter bank designed to extract cyclical components from high-frequency noise and low-frequency trend. This filter bank was applied to quarterly GDP and inflation time series. This filter bank was proposed as an alternative to the conventional bandpass filter used to extract cyclical components from economic time series [24] or the Hodrick-Prescott filter [23]. Because each decomposition stage of the wavelet filter bank results in an effective decimation by two, it was convenient to define the business cycle components as any component with a frequency between four and thirty-two quarters; higher-frequency components would be considered noise, and lower-frequency components would be considered long-term trend (although the long-wave Kondratief cycle would fall into this trend component as well) [4].

Chapter 3

Business Cycle Data and Models

3.1 Statement of Problem

The main goal of this thesis is to classify each quarter in quarterly economic time series as one of two categories: *expansion* or *contraction*. The reasons for selecting this binary classification problem rather than a more-complex four class model such as the one used in [28] include the availability of ground-truth labels from NBER, as well as research that suggests no more than two phases can be reliably identified with statistical significance [30].

In order to perform this classification task, a meaningful *feature space* must be derived from the raw data, through variable selection and potentially manipulations to those variables, and a suitable *classifier* must be selected, trained, and tuned. Econometric techniques will be employed in order to establish the relevance and statistical characteristics of the data series considered for the feature space. The feature vectors derived from economic time series will then be fed into a classifier to obtain cycle phase classifications of individual points in time.

The resulting models will be evaluated based on their percent accuracy in the classification of individual quarters, as well as their similarity to the NBER dates in estimated cycle turning points (the peaks and troughs separating phases of the cycle). The classifiers will also be evaluated to investigate whether it is possible to predict *future* turning points using current economic data.

3.2 Data Sources and Series

While different data series were more useful for different techniques, a common set of features was used across many of these approaches. In most cases, percent changes are computed from the raw levels of each indicator variable, since these differences tend to be more stationary over longer periods of time. Table 3.1 contains data that were examined and considered for classification purposes and brief descriptions of each.

Index	Name	Description
1	GNP Deflator %	GNP price deflator percent change
2	Real GNP %	Percent change over previous period
3	GNP %	Percent change over previous period
4	GNP	Gross national product
5	Real GNP	billions of chained 2009 dollars
6	GNP Deflator	GNP price deflator
7	CPI	Percent change over previous period
8	Real CPI	2010 = 1
9	Surplus	Millions of dollars
10	Surplus %	Percent change over previous period
11	GCE	Govt consumption & expenditures
12	GCEI	Real consumption & expenditures
13	M2 US	M2 supply
14	M2 US	M2 supply, seasonally adjusted
15	M1 US	M1 supply, seasonally adjusted
16	M1 %	Percent change over previous period
17	Unemployment	Unemployment rate
18	Unemployment %	Percent change over previous period
19	NASDAQ %	Percent change over previous period
20	DJIA %	From Central Bank of Brazil Database

Table 3.1. Series used for cycle classification

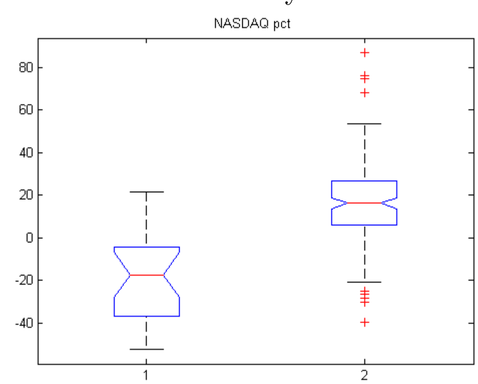
3.3 Data Analysis and Feature Selection

In examining raw data series for inclusion in a classification system, there are two main concerns: which raw series to select, and which transformations and pre-processing techniques to use. These two considerations are not truly independent. For example, a particularly noisy time series might be useless for classification as is, but a filtered version of the same series might contain an abundance of useful information. Furthermore, two series which are individually useless might contain valuable information when combined. Therefore, it would be a mistake to choose economic variables using *only* the predictive power of a single raw time series. However, due to the abundance of economic data available, examining the class-separability of individual series can be a good way to start narrowing down potential variables.

3.3.1 Kruskal-Wallis

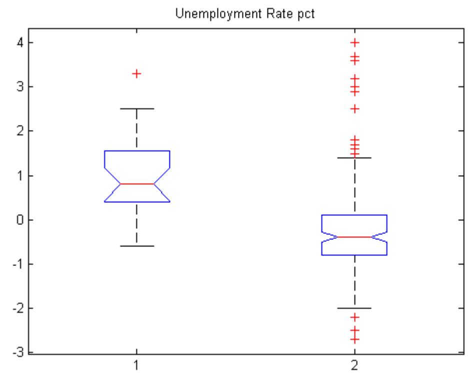
While it's not the last word in feature selection, the Kruskal-Wallis one-way analysis of variance is useful for testing whether the value of a certain time series varies across classes of interest. It tests whether samples originate from the same distribution [33]. In this instance, the two samples tested were the values of a single economic variable from periods of contraction and expansion.

Figure 3.1. Kruskal-Wallis analysis of NASDAQ composite



The null hypothesis is that both series (or all, in the case of more than two) come from the same distribution. The test statistic is computed using the ordered ranks of the data, rather than actual values. The p-value is the probability of observing the provided data under the null hypothesis. The null hypothesis is typically rejected for p-values less than 0.05. Figures 3.1 and 3.2 show the results of the Kruskal-Wallis test performed on the NASDAQ composite and unemployment rate, respectively. The red line is the median of the values from each series. The blue box shows the $25th$ amd $75th$ percentiles, and the whiskers ecompass the

Figure 3.2. Kruskal-Wallis analysis of unemployment rate



outermost values that are not considered outliers [33]. Indicator variables for which the null hypothesis was rejected at at least the 5% level were retained for future optimizations, while variables with poor separability based on this test were mostly ignored.

Variable	p-value
Real GNP	0.230
Real GNP % change	0.000
Unemployment Rate % change	0.000
DJIA % change	0.002
NASDAQ Composite % change	0.000
CPI % change	0.001
GCE	0.200
Surplus	0.059
Surplus % change	0.829

Table 3.2. p-values of select variables from Kruskal-Wallis test

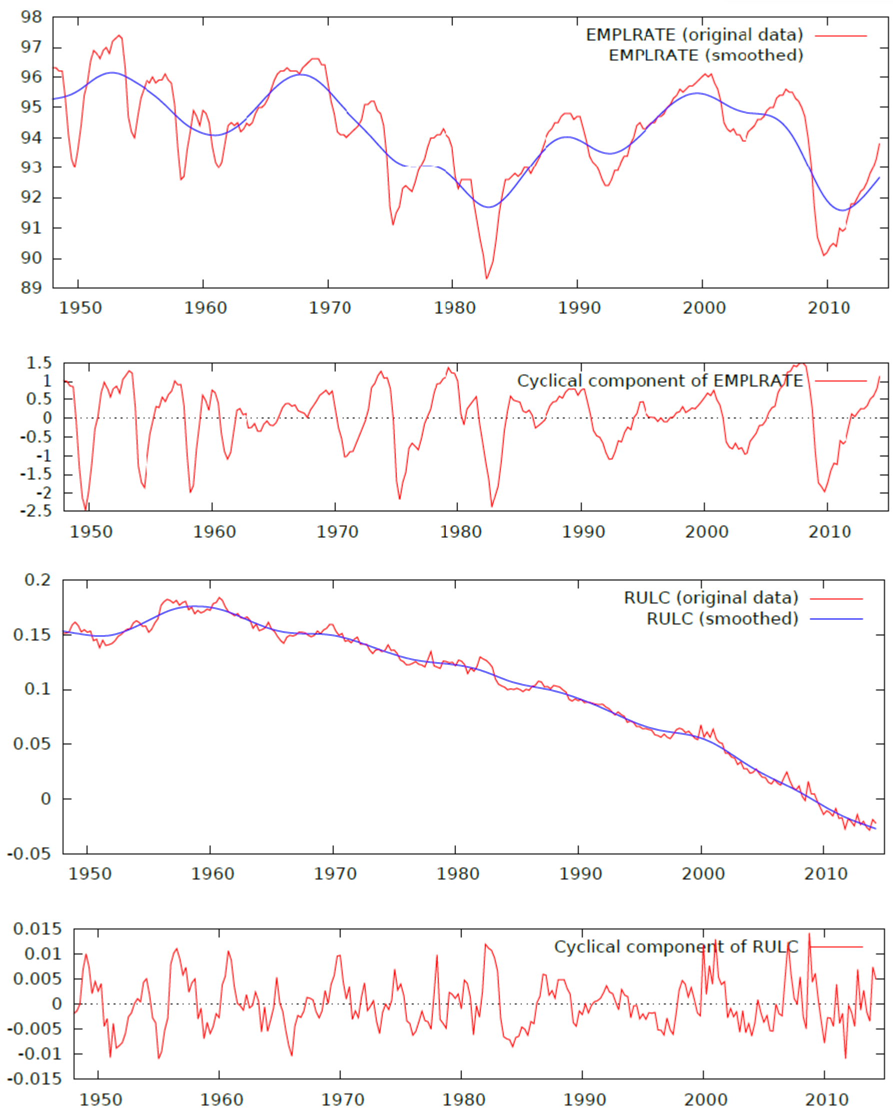
3.4 Econometrics Analysis

While the Hodrick-Prescott filter was not expected to be the most useful technique for pre-processing data for a classifier (see section 2.3.2), it is still useful to investigate which series are stationary after a moving-average filter has been applied, since structural discontinuities would render the filtered series nonstationary even with the trend component removed [12].

3.4.1 ADF-GLS Test

The ADF-GLS test was used to test the stationarity of several time series. From there, vector autoregression (VAR) could be applied if desired, as it requires that time series are either stationary or the same order of integration. Results similar to [12] were replicated.

Figure 3.3. Two time series before and after filtering



3.4.2 Hodrick-Prescott filter (HPF)

the Hodrick-Prescott filter is a low pass filter (LPF) commonly used to remove noise and reveal the trend in economic time series, as described in section 2.3.2. It is frequently used on business cycle data to remove cyclical components and examine long-term trends. Here, the standard value of $\lambda = 1600$ was used. Figure 3.3 shows two time series before and after HPF filtering.

3.5 Feature Space

The previous econometric analysis and Kruskal-Wallis tests served primarily to gain an understanding of the data set described in section 3.2 and to motivate the feature selection for a classifier down the line. The class-separability of individual features does not necessarily translate to the optimal features for a given classifier, but it does provide some insight into which features contain the most useful information for distinguishing between phases of the business cycle. The most promising series appeared to be the Real GDP, CPI, unemployment rate, and stock market composites. It should be noted that since most of these series are either lagging or leading indicators of the same economic trends, they are all somewhat correlated and should certainly not be considered independent. However, they are not necessarily totally redundant, either.

3.5.1 Wavelet Decomposition

In order to isolate cyclical frequency components of interest, a wavelet decomposition was performed using an multiresolution wavelet filter bank. The filter structure suggested in [32] was used, since the wavelet type and support length were specially tailored to quarterly economic time series, and the removal of noise and trend components is easily achieved. The effect is similar to a conventional bandpass filter, but with much better temporal localization. Figure 3.4 shows the decomposition filter bank that was used, and figures 3.5 and 3.6 show the results of this four-level decomposition for the quarterly GNP time series. Contraction periods are highlighted on the background of the plots for reference.

Figure 3.4. Wavelet decomposition subband tree

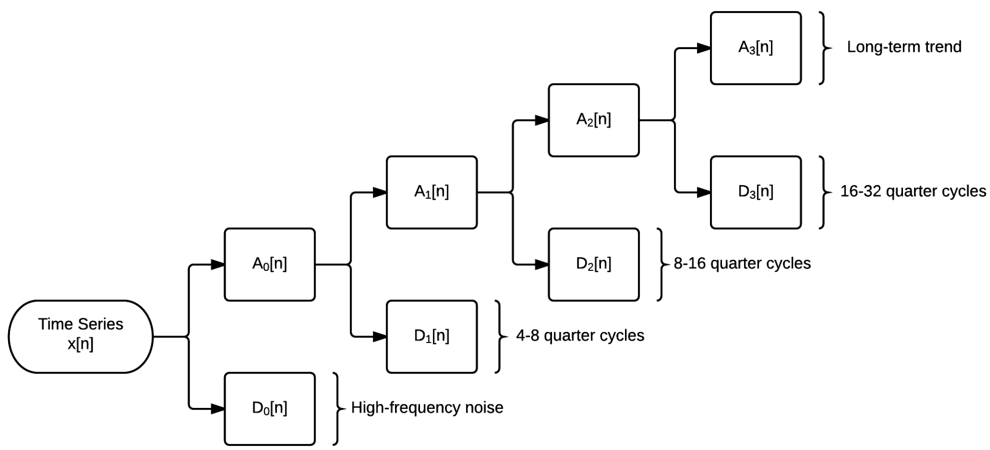
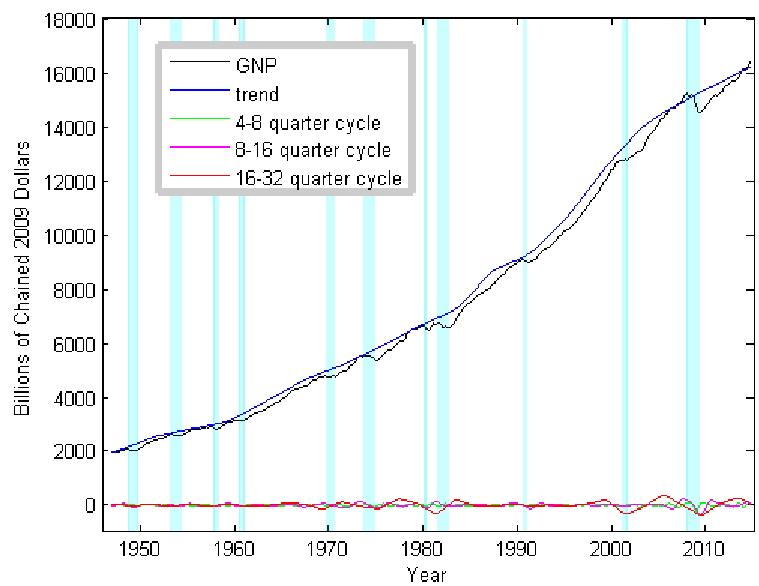


Figure 3.5. Wavelet decomposition of quarterly GNP data

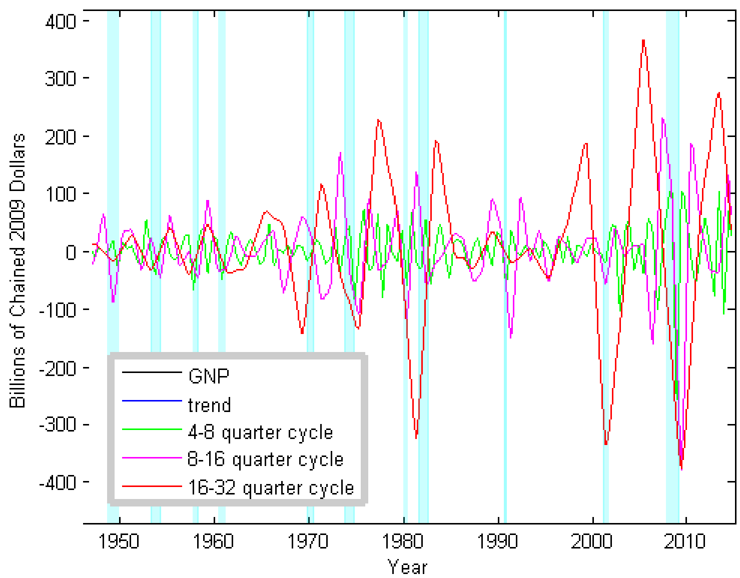


3.6 Binary Business Cycle Phase Classification

3.6.1 Phase Classification with Fisher's Linear Discriminant

Linear discriminant analysis was performed using a similar approach to [28], but with many of the less helpful variables excluded. The Pattern Recognition toolbox

Figure 3.6. Cyclical components of quarterly GNP data



was used to learn the weights which define the discriminant function, and to classify test data [34].

3.6.2 Phase Classification with Hidden Markov Models

As discussed in section 2.4.2, hidden Markov models describe the relationship between a set of latent states and a sequence of observations. This model can be used for different purposes, but the classification task is the application of interest to this thesis.

Classification was performed using HMMs on the same time series used in the preceding linear discriminant analysis.

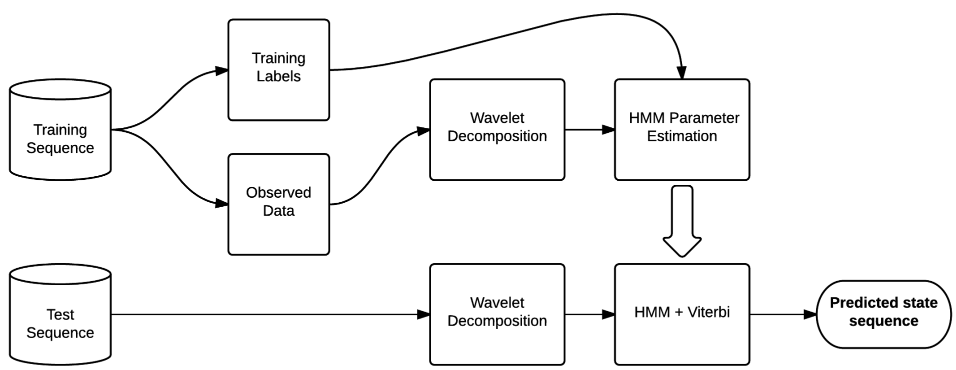
Two approaches were used in creating HMMs for this application. Initially, the continuous levels of the time series were binned into a small number of discrete intervals, allowing for discrete emission probabilities. This model only operated on a one-dimensional time series, so if multiple series were to be used, separate HMMs would have to be trained.

Next, HMMs were trained using multivariate Gaussians, which obviated the need for the binning and allowed multiple time series to be considered by the same model.

3.6.3 Combining HMMs and Wavelet Features

Since multiple variables can be used to train a hidden Markov model and predict state sequences, there is a good amount of flexibility in selecting a set of features to use for classification inside the HMM framework. In addition to multiple economic time series, decompositions or transformations of each of those time series may be used. It is worth considering, however, that the model complexity of the multivariate Gaussian HMM increases quadratically with the dimensionality of the feature space, since the size of the covariance matrix is D^2 for a D-dimensional feature [1].

Figure 3.7. HMM classification scheme with wavelet-based features



Once the transition and emission probabilities of an HMM have been estimated from training data, the Viterbi algorithm can be applied to the lattice of all possible state transitions to determine the most likely state sequence that resulted in the sequence of test features. This most probable state sequence is the set of class predictions that would be obtained from any other classifier. Figure 3.8 shows the lattice for a sequence of 4 observations in an HMM with 3 latent states.

3.6.4 Predicting the Future

While classifying the current phase of the business cycle based on quarterly data could still provide a more timely indicator of turning points than waiting for the release of NBER labels, it would be even more useful to predict what will happen next quarter, or even next year. Figure 3.9 shows the slightly modified classification system, in which data is shifted relative to the labels prior to training.

Figure 3.8. HMM trellis formed by all possible latent state sequences [1]

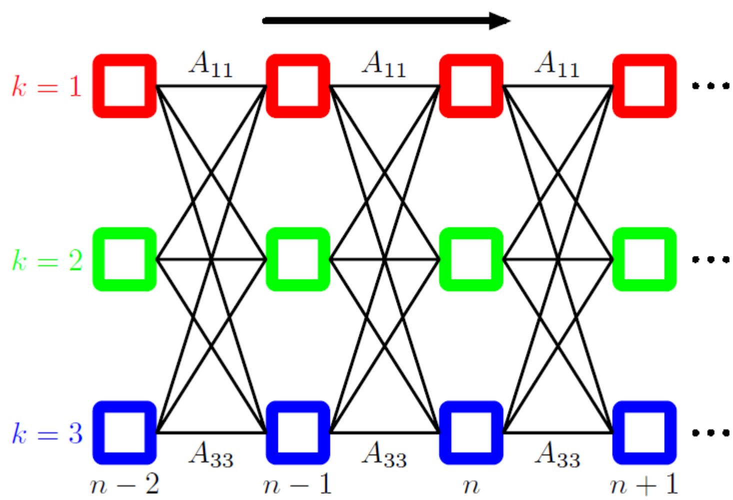
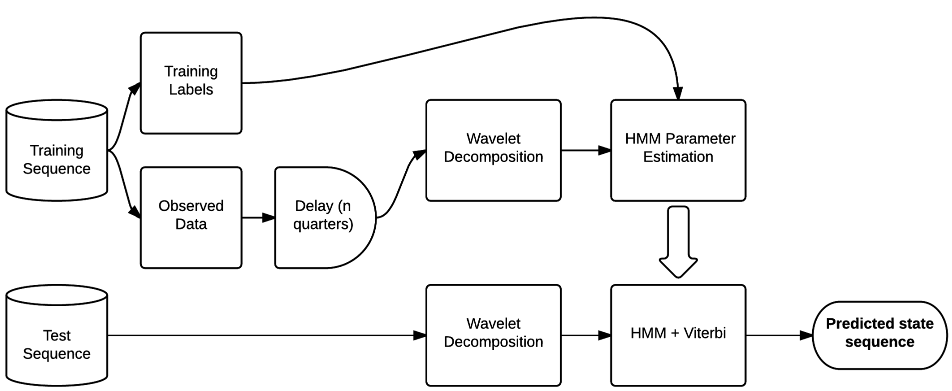


Figure 3.9. HMM classification with delayed wavelet features



Chapter 4

Results

4.1 Classifier Accuracy

The percentage of test data points classified correctly is one key metric used to evaluate a classification system. In this section, the success rates achieved by systems with different features and classifiers are compared. Table 3.1 briefly describes all of the economic time series considered for classification. All data are quarterly series from the Federal Reserve Economic Data (FRED) unless otherwise noted. Table 4.1 lists the classification success rates for each system. There are some caveats to note in comparing these numbers directly. First, depending on the variables used, the range of dates available for training and testing may vary, as some series were introduced earlier than others. To control for this, all variables used in the final tests range from 1949 to 2014, even if older or more recent data was available for some variables. Additionally, there are some fundamental differences in the way the HMM classifiers were trained and tested compared to the other classifiers.

4.1.1 Cross Validation

A common way to evaluate classifiers with a limited amount of data is k-fold cross validation. This entails dividing the available data into k sets of approximately equal sizes. Then, k iterations of classification and testing are performed, each time with a different subset held out for testing and all the others used for training. This way, a classifier can be tested on all available data. When data is in short supply, leave-one-out cross validation is ideal ($k = N$, where N is the total number of observations), since it maximizes the amount of training data in each iteration. However, this may not be a valid approach in business cycle classification. The underlying causes and properties of separate recessions and expansions are hotly debated in the field of economics; it should certainly not be assumed that one cycle is representative of others. For a business cycle classifier to be robust enough to be useful, it would need to be able to identify changes in the cycle *without training data from that cycle*. For this reason, cross validation was performed using a leave-one-cycle-out procedure, as suggested in [29], rather than withholding individual quarterly samples for testing. This way, in every iteration of testing, the points being classified were from a cycle that was not used to train the classifier.

4.1.2 HMM Training and Testing

Due to the temporal continuity inherent in HMMs, the leave-one-cycle-out technique used for other classifiers does not translate easily. Removing a cycle from the middle of the data set for testing would leave the HMM with two disjoint training sets. To keep things simple, the HMM classifiers were trained once on approximately the first two thirds of the data, and tested on the final third. This is probably a reasonable model for how such a classifier would actually be used; it would likely be trained on all available data up to the recent past and then applied to the most recent data points. The best sets of features were selected from the full list of 20 economic indicators based on the p-values of the Kruskal-Wallis tests, and their performance when tested in pairs with the linear discriminant classifier. The best performing features were consistently variations of Real GDP, unemployment rate, and the stock market indices. The classification accuracy of select combinations of classifiers and feature sets is shown in 4.1.

Variables	Feature Type	Classifier	Percent Correct
5,17,18,20	levels	LDA	79%
5,17,18,20	wavelet cycle components	LDA	86%
5,18	levels	HMM	90%
5,18	wavelet cycle components	HMM	92%

Table 4.1. Classification accuracy of different features and classifiers

Because we care mainly about identifying turning points in the cycle, it is somewhat more useful to compare the dates of the predicted and actual turning points directly. Table 4.2 compares the turning points predicted by an HMM with wavelet features. The HMM was trained using data from 1949 through 1990 and evaluated on the two business cycles from 1991 to the present. The HMM perfectly classified the contraction period in 2001, and was only off by a few quarters in the turning points for the 2008 contraction. However, the HMM also falsely identified a period of contraction in 1992 which did not exist, according to NBER.

NBER Peak	Predicted Peak	NBER Trough	Predicted Trough
2001 Q1	2001 Q1	2001 Q4	2001 Q4
2007 Q4	2008 Q2	2009 Q2	2009 Q1

Table 4.2. NBER turning points compared to HMM classifications

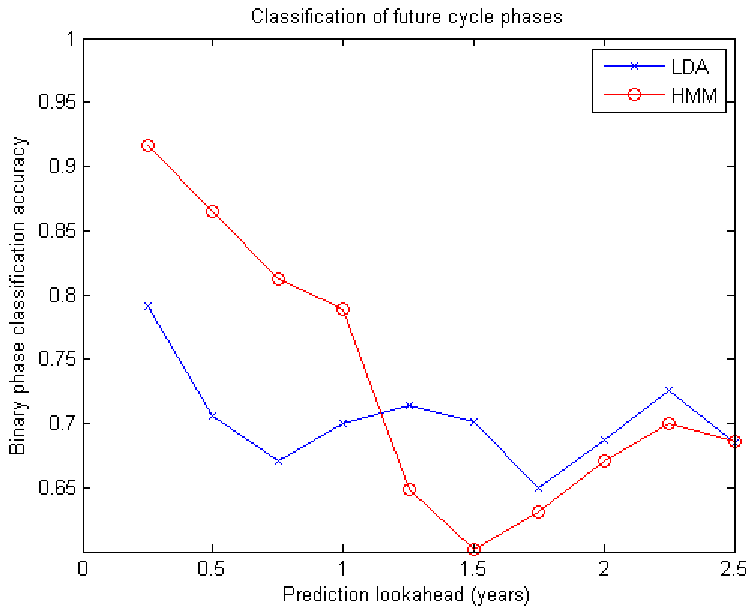
Compared to the linear discriminant analysis approach, the HMM provided fewer false positives (one extraneous contraction period instead of three short ones that showed up in the LDA classifications), and the predicted turning points were more closely aligned with the NBER turning points.

4.1.3 Classification Results with Lookahead

The ability to classify business cycle turning points using current economic data would provide turning point estimates much more promptly than the official NBER classifications. However, it would be even more useful to be able to predict what will happen *next* quarter, or maybe even several quarters in the future. Of course, this is only possible if the state that directly *leads to* peaks and troughs is already present in the data one or more quarters before the actual turning point occurs.

Figure 4.1 shows the quarterly classification accuracy as a function of how far in the future predictions are being made. Since this is a binary classification task, 0.5 is the lower bounds for how poorly a classifier can perform if it guesses at random. By about one year into the future, both the LDA and HMM classifiers have dropped off in accuracy to the point where they essentially provide no information. For the results shown, the same features were used as in the classification accuracy comparison from table 4.1, which are the wavelet components derived from real GNP and unemployment percent change. For classification of future quarters, it

Figure 4.1. Classifier accuracy as a function of lookahead time



might be more expedient to use series representing *leading* indicators, such as the stock market composites.

Chapter 5

Conclusions and Future Work

5.1 Conclusions

While more analysis is needed, feeding cyclical components from a wavelet decomposition into a hidden Markov model appears to be a promising scheme for automatically identifying turning points in economic activity.

5.2 Future Work

5.2.1 Other business cycle models

If successful and reliable, a four phase classification model would provide more information than the two phase model which has been the focus of this thesis. However, since these labels are not published by any agency, they would have to be hand-labeled and would rely on the economic expertise of the labeler.

5.2.2 Validate Model on Additional Data

One of the major disadvantages of working with a small data set is the inability to reserve a validation data set to be tested on only after a model has been tuned and optimized. Testing the models used in this thesis on similar data series from other capitalist countries could yield interesting insights into the robustness of the model, as well as the similarities and differences between patterns of economic activity in different countries. West Germany would be a good potential candidate over the time range utlized in [29].

Appendix A

Code

The code used to generate the results from section 4.1.3 is included in this appendix. A git repository containing the complete code used in this thesis is also available at `http://github.com/jastrauckas/thesis`.

A.1 HMM prediction with wavelet features

```
clc, clear all, close all
RELOAD = true;

%% INITIALIZE
addpath('C:\Users\j_ast_000\Documents\MATLAB\pmtk3-master');
initPmtk3
cd 'C:\cygwin\home\j_ast_000\Thesis\Code'

%% LOAD DATA and select features
% define what's happening in this csv
nLabels = 1; % how many columns at the end are class labels?
startDate = 1947.0;

if RELOAD
    rawData = csvread('../Data/masterData.csv', 2, 1);
    featureCount = size(rawData, 2)-1;
```

```
    observationCount = size(rawData, 1);
    featureNames = textread('../Data/masterData.csv', '%s', 'delimiter', ',');
    featureNames = featureNames(2:featureCount + 1);
    classLabels = rawData(:,featureCount + 1);
end

endDate = startDate + (observationCount-1)*0.25;
dates = linspace(startDate, endDate, observationCount);

%% WAVELET DECOMPOSITION
% goal is to separate long term trend, cyclical component(s), and high
% frequency noise using multiresolution wavelet analysis
% Baxter and King define the following:
% Long-term trend -- periodicity > 32 quarters
% Business cycle -- periodicity 4-32 quarters
% High frequency noise -- periodicity > 32 quarters

% reproduce wavelet filter bank on real GNP
gnp = rawData(:,5);
[y1, y2, i1, i2] = getValidDateRange(gnp, dates);
gnpDates = dates(i1:i2);
gnp = gnp(i1:i2);
gnpLabels = classLabels(i1:i2) - 1;
[conStarts, conEnds] = getContractionDates(gnpLabels);

% Perform subsequent single-level wavelet decompositions
% Yogo suggest 17/11 filter -> coiflet with N=2
wname = 'coif2';
[a0, d0] = dwt(gnp, wname);
[a1, d1] = dwt(a0, wname);
[a2, d2] = dwt(a1, wname);
[a3, d3] = dwt(a2, wname);

```

```
% since these filtered components are all critically sampled, interpolate
% to restore the correct magnitude and number of points
originalLength = size(gnp, 1);
cycle_4_8 = upcoef('d', d1, wname, 2, originalLength);
cycle_8_16 = upcoef('d', d2, wname, 3, originalLength);
cycle_16_32 = upcoef('d', d3, wname, 4, originalLength);
trend = upcoef('a', a3, wname, 4, originalLength);

figure
plot(gnpDates, gnp, 'k', 'linewidth', 1)
hold on
plot(gnpDates, trend, 'b')
plot(gnpDates, cycle_4_8, 'g')
plot(gnpDates, cycle_8_16, 'm')
plot(gnpDates, cycle_16_32, 'r')

legend('GNP', 'trend', '4-8 quarter cycle', '8-16 quarter cycle', '16-32 quarter
xlabel('Year')
ylabel('Billions of Chained 2009 Dollars')

ymin = -1000;
ymax = 18000;
axis([1946 2015 ymin ymax])
ymid = ((ymax-ymin)/2) + ymin;
yheight = (ymax-ymin);
for ind = 1:size(conStarts, 2)
    first = gnpDates(conStarts(ind));
    last = gnpDates(conEnds(ind));
    duration = last-first;
    center = (duration/2) + first;
    %rectangle('Position', [center, ymin, duration, yheight])
    p = patch([first last last first], [ymin ymin ymax ymax], 'c');
    set(p,'FaceAlpha',0.2);
```

```
    set(p,'EdgeAlpha',0.2);
    set(p, 'EdgeColor', 'c');
end
hold off

%% WAVELET CLASSIFICATION (LEAVE ONE CYCLE OUT)
% use the time-aligned output levels of the wavelet decomposition
% components to classify expansion/contraction
rowCount = size(gnpLabels,1);
cycleData = [cycle_4_8 cycle_8_16 cycle_16_32];
cycleStarts = getCycleStarts(gnpLabels);
weightedPctCorrect = 0;
for ind = 1:length(cycleStarts)
    if ind == length(cycleStarts)
        testInds = cycleStarts(ind):rowCount;
        trainingInds = 1:(cycleStarts(ind)-1);
    else
        testInds = cycleStarts(ind):(cycleStarts(ind+1)-1);
        trainingInds = [1:(cycleStarts(ind)-1) (cycleStarts(ind+1)):rowCount];
    end
    trainDS = prtDataSetClass(cycleData(trainingInds,:), ...
        gnpLabels(trainingInds));
    testDS = prtDataSetClass(cycleData(testInds,:), ...
        gnpLabels(testInds));
    classifier = prtClassFld + prtDecisionMap;
    classifier = classifier.train(trainDS);
    classified = run(classifier, testDS);
    pct = prtScorePercentCorrect(classified);
    weightedPctCorrect = weightedPctCorrect + length(testInds)*pct;
end
waveletPctCorrect = weightedPctCorrect / rowCount; % GNP ONLY

% Try using multiple time series that worked well before
```

```
% LEAVE ONE CYCLE OUT
% which features to use?
%selections = 1:19;
%selections = [2,7,18,19];  % NASDAQ
%selections = [2,7,18,20];   % DJIA
selections = [5,17,18,20]; % KEEP
%selections = [5,18,20];

selectedNames = featureNames(selections);
data = [];
for i=selections
%     figure
%     plot(dates, rawData(:,i));
%     title(featureNames{i});
    data = [data rawData(:,i)];
end

% see what range of dates we can use for this set of data
[d1, d2, i1, i2] = getValidDateRange(data, dates);
validData = data(i1:i2,:);
waveletData = [];
featureCount = size(validData,2);
for col=1:featureCount
    wd = getCycleComponents(validData(:,col), wname, 0);
    waveletData = [waveletData wd];
end
% ONLY FOR CONTROL COMPARISON:
%waveletData = validData;

validLabels = classLabels(i1:i2);
validDates = dates(i1:i2);
validLabels(validLabels==1) = 0;
validLabels(validLabels==2) = 1;
```

```
rowCount = size(validData,1);

iterations = 1:10;
pcts = zeros(1, size(iterations,2));
hmmPcts = pcts;
for ii = iterations
    labels = validLabels(ii:rowCount);
    data = waveletData(1:rowCount-(ii-1), :);
    rows = rowCount - (ii-1);
    cycleStarts = getCycleStarts(labels);
    weightedPctCorrect = 0;
    for ind = 1:length(cycleStarts)
        if ind == length(cycleStarts)
            testInds = cycleStarts(ind):rows;
            trainingInds = 1:(cycleStarts(ind)-1);
        else
            testInds = cycleStarts(ind):(cycleStarts(ind+1)-1);
            trainingInds = [1:(cycleStarts(ind)-1) (cycleStarts(ind+1)):rows];
        end
        trainDS = prtDataSetClass(data(trainingInds,:), ...
            labels(trainingInds));
        testDS = prtDataSetClass(data(testInds,:), ...
            labels(testInds));

        % normalize
        zmuv = prtPreProcZmuv;
        zmuv = zmuv.train(trainDS);
        trainDS = zmuv.run(trainDS);
        zmuv = prtPreProcZmuv;
        zmuv = zmuv.train(testDS);
        testDS = zmuv.run(testDS);

        % create classifier
```

```
        %classifier = prtClassFld + prtDecisionMap; % FISHER'S
        classifier = prtClassFld + prtDecisionBinaryMinPe;
        %classifier = prtClassLibSvm + prtDecisionMap;
        %classifier = prtClassKnn + prtDecisionMap;
        classifier = classifier.train(trainDS);
        classified = run(classifier, testDS);
        pct = prtScorePercentCorrect(classified);
        weightedPctCorrect = weightedPctCorrect + length(testInds)*pct;
    end
    pctCorrect = weightedPctCorrect / rows;
    pcts(ii) = pctCorrect;


    %% PMTK HMMs - can use multivariate Gaussians
    d = size(data, 2);
    nstates = 2;

    % now, without incest!
    targetIndex = 2 * round(size(labels, 1) / 3);
    while labels(targetIndex) ~= labels(1)
        targetIndex = targetIndex + 1;
    end
    split = targetIndex;

    Z = {labels(1:targetIndex)' + 1};
    Y = {data(1:targetIndex,:)'};
    X = data(targetIndex+1:end,:)';
    model3 = hmmFitFullyObs(Z, Y, 'gauss');

    % use Viterbi to predict state sequence
    path = hmmMap(model3, X) - 1;
    pctError_MVN_HMM = sum(path ~= labels(split+1:end)') / ...
        size(labels(141:end),1);
```

```
    hmmPcts(ii) = 1 - pctError_MVN_HMM;
end

%% Plot results
figure
years = 1:size(pcts,2);
years = years/4;
plot(years, pcts, 'bx-')
title('Classification of future cycle phases')
ylabel('Binary phase classification accuracy')
xlabel('Prediction lookahead (years)')
hold on
plot(years, hmmPcts, 'ro-');
hold off
legend('LDA','HMM')
```

Appendix B

Data

CSV files containing the full set of data used in this thesis can be obtained at `http://github.com/jastrauckas/thesis`. The data series that were used or considered for classification are listed below, along with their respective sources. Refer to the more verbose series names and descriptions provided in table 3.1.

Index	Name	Source	URL
1	GNP Deflator %	FRED	alfred.stlouisfed.org
2	Real GNP %	FRED	alfred.stlouisfed.org
3	GNP %	FRED	alfred.stlouisfed.org
4	GNP	FRED	alfred.stlouisfed.org
5	Real GNP	FRED	alfred.stlouisfed.org
6	GNP Deflator	FRED	alfred.stlouisfed.org
7	CPI	FRED	alfred.stlouisfed.org
8	Real CPI	FRED	alfred.stlouisfed.org
9	Surplus	FRED	alfred.stlouisfed.org
10	Surplus %	FRED	alfred.stlouisfed.org
11	GCE	FRED	alfred.stlouisfed.org
12	GCEI	FRED	alfred.stlouisfed.org
13	M2 US	FRED	alfred.stlouisfed.org
14	M2 US	FRED	alfred.stlouisfed.org
15	M1 US	FRED	alfred.stlouisfed.org
16	M1 %	FRED	alfred.stlouisfed.org
17	Unemployment	FRED	alfred.stlouisfed.org
18	Unemployment %	FRED	alfred.stlouisfed.org
19	NASDAQ %	FRED	alfred.stlouisfed.org
20	DJIA %	BCB	quandl.com/data/BCB/

Table B.1. Sources used for cycle classification

Bibliography

[1] Bishop, C. M. (2006) *Pattern recognition and machine learning*, springer.

[2] Belongia, M. T., M. R. Garfinkel, et al. (1992) "The business cycle: theories and evidence: proceedings of the sixteenth annual Economic Policy Conference held October 17-18, 1991," *Federal Reserve Bank of St. Louis Proceedings*.

[3] Piketty, T. (2014) "Capital in the twenty-first century," *Cambridge, MA, London*.

[4] Sherman, H. J. (1991) *The business cycle: growth and crisis under capitalism*, Princeton University Press.

[5] National Bureau of Economic Research (2008), "The NBER's Recession Dating Procedure," .
URL `http://www.nber.org/cycles/jan2003.html`

[6] ——— (2010), "US Business Cycle Expansions and Contractions," .
URL `http://www.nber.org/cycles.html`

[7] Stock, J. H. and M. W. Watson (2010) *Estimating turning points using large data sets*, *Tech. rep.*, National Bureau of Economic Research.

[8] de Sismondi, J. C. L. (1819) *Nouveaux Principes d'conomie politique*.

[9] Glasner, D. and T. Cooley (1997) *Business Cycles and Depressions: An Encyclopedia*, Garland reference library of social science, Garland Pub.
URL `https://books.google.com/books?id=e1ZEPd_pQnoC`

[10] Keynes, J. M. (1936) *General theory of employment, interest and money*.

[11] Burns, A. F. and W. C. Mitchell (1946) "Measuring business cycles," *NBER Books*.

[12] TARASSOW, A. (2010) "The empirical relevance of Goodwins business cycle model for the US economy," .

[13] GOODWIN, R. M. (1967) "A growth cycle," *Socialism, capitalism and economic growth*, **5458**.

[14] BUREAU OF LABOR STATISTICS, "How the Government Measures Unemployment," .
URL http://www.bls.gov/cps/cps_htgm.htm

[15] WYOMING DEPARTMENT OF ADMINISTRATION AND INFORMATION, "Income, Employment, and Gross Domestic Product by Industry," .
URL http://eadiv.state.wy.us/i&e/i&e.html

[16] ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, "Glossary of Statistical Terms," .
URL http://stats.oecd.org/glossary/

[17] BARTH, J. R., G. IDEN, and F. S. RUSSEK (1984) "Do federal deficits really matter?" *Contemporary Economic Policy*, **3**(1), pp. 79–95.

[18] (2013) "Dow Jones Industrial Average: Historical Components," *S&P Dow Jones Indices*.

[19] NASDAQ, "Stock Market Indices," .
URL http://www.nasdaq.com/markets/indices/major-indices.aspx

[20] BALTAGI, B. H. (1999) *Econometrics*, Springer Texts in Business and Economics.

[21] ELLIOTT, G., T. J. ROTHENBERG, and J. H. STOCK (1992), "Efficient tests for an autoregressive unit root," .

[22] BIERENS, H. J. (1997) "Testing the unit root with drift hypothesis against nonlinear trend stationarity, with an application to the US price level and interest rate," *Journal of Econometrics*, **81**(1), pp. 29–64.

[23] COGLEY, T. and J. M. NASON (1995) "Effects of the Hodrick-Prescott filter on trend and difference stationary time series Implications for business cycle research," *Journal of Economic Dynamics and control*, **19**(1), pp. 253–278.

[24] BAXTER, M. and R. G. KING (1999) "Measuring business cycles: approximate band-pass filters for economic time series," *Review of economics and statistics*, **81**(4), pp. 575–593.

[25] MALLAT, S. (1999) *A wavelet tour of signal processing*, Academic press.

[26] Crouse, M. S., R. D. Nowak, and R. G. Baraniuk (1998) "Wavelet-based statistical signal processing using hidden Markov models," *Signal Processing, IEEE Transactions on*, **46**(4), pp. 886–902.

[27] Sarich, J. (2007) "Is There a Long Wave in U.S. Corporate Profits?" *presented to the Eastern Economic Association*.

[28] Meyer, J. R. and D. H. Weinberg (1975) "On the classification of economic fluctuations," in *Explorations in Economic Research, Volume 2, number 2*, NBER, pp. 43–78.

[29] Heilemann, U. and H. J. Münch (1999) *Classification of West German business cycles*, *Tech. rep.*, Technical Report, SFB 475: Komplexitätsreduktion in Multivariaten Datenstrukturen, Universität Dortmund.

[30] Klinkenberg, R. (2005) "Meta-Learning, Model Selection, and Example Selection in Machine Learning Domains with Concept Drift." in *LWA*, vol. 2005, pp. 164–171.

[31] Chauvet, M. and J. M. Piger (2003) "Identifying Business Cycle Turning Points in Real Time (Digest Summary)," *Federal Reserve Bank of St. Louis Review*, **85**(2), pp. 47–61.

[32] Yogo, M. (2008) "Measuring business cycles: A wavelet analysis of economic time series," *Economics Letters*, **100**(2), pp. 208–212.

[33] MathWorks, "Kruskal-Wallis test," .
URL `http://www.mathworks.com/help/stats/kruskalwallis.html`

[34] Morton Jr, K. D., P. Torrione, L. Collins, and S. Keene (2014) "An Open Source Pattern Recognition Toolbox for MATLAB," *arXiv preprint arXiv:1406.5565*.