mathematics

Article

Addressing Concerns about Single Path Analysis in Business Cycle Turning Points: The Case of Learning Vector Quantization

David Enck [1], Mario Beruvides [1], Víctor G. Tercero-Gómez [2] and Alvaro E. Cordero-Franco [3,*]

[1] Department of Industrial Manufacturing & Systems Engineering, Texas Tech University, Lubbock, TX 79409, USA; david.enck@ttu.edu (D.E.); mario.beruvides@ttu.edu (M.B.)
[2] School of Engineering and Sciences, Tecnologico de Monterrey, Monterrey 64849, Mexico; victor.tercero@tec.mx
[3] Facultad de Ciencias Físico Matemáticas, Universidad Autónoma de Nuevo León, San Nicolás de los Garza 66451, Mexico
* Correspondence: alvaro.corderofr@uanl.edu.mx

Abstract: Data-driven approaches in machine learning are increasingly applied in economic analysis, particularly for identifying business cycle (BC) turning points. However, temporal dependence in BCs is often overlooked, leading to what we term single path analysis (SPA). SPA neglects the diverse potential routes of a temporal data structure. It hinders the evaluation and calibration of algorithms. This study emphasizes the significance of acknowledging temporal dependence in BC analysis and illustrates the problem of SPA using learning vector quantization (LVQ) as a case study. LVQ was previously adapted to use economic indicators to determine the current BC phase, exhibiting flexibility in adapting to evolving patterns. To address temporal complexities, we employed a multivariate Monte Carlo simulation incorporating a specified number of change-points, autocorrelation, and cross-correlations, from a second-order vector autoregressive model. Calibrated with varying levels of observed economic leading indicators, our approach offers a deeper understanding of LVQ's uncertainties. Our results demonstrate the inadequacy of SPA, unveiling diverse risks and worst-case protection strategies. By encouraging researchers to consider temporal dependence, this study contributes to enhancing the robustness of data-driven approaches in financial and economic analyses, offering a comprehensive framework for addressing SPA concerns.

Keywords: data-driven methods; temporal dependence; Monte Carlo simulation; robustness; multivariate analysis; economic indicators

MSC: 62M10; 62H30; 9110

check for updates

Citation: Enck, D.; Beruvides, M.; Tercero-Gómez, V.G.; Cordero-Franco, A.E. Addressing Concerns about Single Path Analysis in Business Cycle Turning Points: The Case of Learning Vector Quantization. *Mathematics* **2024**, *12*, 678. https://doi.org/10.3390/math12050678

Academic Editors: Arne Johannssen and Nataliya Chukhrova

Received: 29 December 2023
Revised: 5 February 2024
Accepted: 6 February 2024
Published: 26 February 2024

1. Introduction

Data driven approaches in machine learning offer powerful alternatives for statistical classification, optimal decision-making, and pattern recognition. Its increasing use in economic analysis does not come as a surprise, where the application to identify business cycle (BC) turning points (TPs) has gained great popularity. A BC TP refers to the transition between different phases of an economic system. The recurring pattern of economic expansion and contraction consists of four main phases: expansion, peak, contraction (or recession), and trough. A TP marks the shift from one phase to another (NBER [1]). Identification of TPs allows economists to determine the impact of mitigating policies that were implemented, and timely identification helps countries, businesses, and citizens make financial decisions (Hamilton [2]). As BCs are related to time series data, they often exhibit autocorrelation, meaning that the observations are correlated with their past values. In such cases, the temporal structure of the data becomes crucial, and the performance of machine learning algorithms needs to address this reality. Related literature on BC analysis often fails to acknowledge this situation, and algorithms are evaluated with a given historical series that represents a single path of all possible potential routes a data structure can take. Approaches that fail to acknowledge this situation fall into the trap of what we call single path analysis (SPA). We encourage researchers doing BC analysis to be mindful of the temporal dependence when evaluating the performance of an algorithm to shed light on what to expect of an approach when applied to a new sequence of observations.

To illustrate the issue, we focus our attention on the learning vector quantization (LVQ) algorithm. LVQ is particularly useful for classification tasks where the decision boundaries between classes are well-defined. It has been applied in various domains, including pattern recognition, image classification, and the monitoring of BCs (Giusto and Piger [3]). In the context of BC monitoring, LVQ is used to classify economic indicators and determine the current phase of the BC based on historical indicators. The algorithm's ability to adapt to changing patterns and different levels of information makes it suitable for tasks where the characteristics of different cycles may evolve over time.

To address the temporal structure of the data, we use a multivariate Monte Carlo simulation with a structure with a specific number of change-points, autocorrelation and cross-correlations. A second-order vector autoregressive (VAR) model, calibrated with different levels of economic performance observed in the economic leading indicators (LIs) was used in the analysis. Results show the uncertainties of the LVQ algorithm, revealing different levels of risks and strategies for worst-case protections that were not possible to assess with single path analysis.

A common practice when evaluating BC TP algorithms is to use a single historical time series to select an algorithm's design parameters and to compare the ability of different algorithms to detect BC TPs. Given the defined design parameters, such as smoothing time periods and probability thresholds, an algorithm decides whether a time point is in a recession or growth period. Authors frequently evaluate the time to signal (TTS) for an algorithm's decisions against the economic status, and TTS is defined by the National Bureau of Economic Research (NBER) Business Cycle Dating Committee (BCDC) in Cambridge, MA, USA.

The NBER declares the TPs for peaks and troughs that are used by economists to define BCs. The NBER looks for multiple economic indicators to show a significantly broad decrease or increase in economic activity that lasts more than a few months to declare a BC TP. Declarations of peaks and troughs are important, as they impact behaviors of financial institutions, governments, and citizens (Hamilton [2]). There is no stated model for the analysis; however, the NBER offers guidance on variables and patterns they look for when making decisions (NBER [1,4]; The White House [5]). Due to the lack of a reproducible numerical approach, authors have developed algorithms to try and estimate the declared BC TPs while providing transparency and a shorter TTS. An important part of the method selection and algorithm optimization involves making decisions on the design parameters that adjust the decisions made by the algorithms.

Hamilton [6] wrote a foundational paper on how to effectively predict BC TPs using Markov Switching Models. Hamilton based his decisions regarding the smoothing periods and probability threshold on observed performance of the algorithm for the dates of study. Goodwin [7] explored Hamilton's model on eight developed market economies and made smoothing period and probability threshold decisions based only on the data being evaluated in his study. Some authors explored extensions or alternatives to Hamilton's Markov switching approach, including Birchenhall et al. [8], Chauvet and Piger [9], Giusto and Piger [3], Soybilgen [10], and others. These authors may use different criteria for selecting design parameters; however, they all based their selections on the single set of data they used in their study.

Depending on the algorithm, the design parameters can include the number of time periods to smooth over, number of time periods in a forecast, the probability decision threshold, and/or the number of months the threshold should be exceeded. Design parameters that provide the best statistics for TTS and which incur the fewest false positives are selected, for that algorithm for that historical dataset. As an example, Chauvet and

Piger [9] evaluated a dynamic factor Markov-switching (DFMS) model of Chauvet [11] using a two-step approach to convert recession probability estimates into a claim on whether the economy is in recession or expansion. The two-step approach used 0.8 and 0.5 as probability thresholds with a 2-period lag. The parameters worked well, but the peak in 2001 was declared after the NBER announcement and declared to be 2 months prior to the NBER date of March 2001. Two other peaks found matched the NBER announcements, and one peak was only one month ahead. The DFMS model also identified the October 1982 trough 38 days ahead of the NBER announcement when the other three trough announcements studied had a range from 189 to 420 days ahead. Recessions and growth periods were clearly behaving differently relative to the specific changes in the economic indicators being monitored. This in-sample evaluation represents what we call the SPA problem. Algorithms calibrated using SPA usually show great performance in terms of false positives and TTS; however, evaluations cannot extend beyond the observed data for both the depth of the recession (growth) period and the distributional pattern of the observations.

This paper re-assesses the LVQ implementation by Giusto and Piger [3] with a Monte Carlo approach, allowing researchers to simulate specific changes in a data generating process (DGP) and evaluate the probability distribution for the TTS. The approach can be adapted to approach any algorithm used to evaluate BC TPs and avoid the SPA problem. Our proposed simulation allows researchers to develop a better understanding of performance of a set of definable design parameters for a particular TP.

Section 2 provides a comprehensive review of the LVQ algorithm, while Section 3 proposes a multi-path analysis to counteract SPA through Monte Carlo simulation. In Section 4, we apply the Monte Carlo approach to evaluate BC TPs, re-assessing the work of Giusto and Piger [3]. Section 5 presents a discussion on the obtained results, and finally, Section 6 provides conclusions and suggestions for future research.

2. Learning Vector Quantization in Business Cycle Analysis

LVQ is a family of machine-learning algorithms used for real-time statistical classification based on the use of codebook vectors, which are used to represent a particular class or category in a dataset. These vectors are used to classify new data points based on their similarity to the codebook vectors. In the LVQ algorithm, the codebook vectors are adjusted based on their proximity to new data points, allowing them to adapt to the characteristics of the data and improve the classification process (see Kohonen [12]). LVQ classifier presents some advantages when comparing multilayers perceptron and support vector machine algorithms, in terms of computational cost and interpretability (Nova and Estévez [13]). LVQ classifiers have been studied by several authors, either in theoretical studies looking for improvement terms of convergence or robustness, as it is seen in Sato and Yamada [14] in their General LVQ, and Seo and Obermayer [15] with their Soft LVQ, among others; or using the LVQ, or its adaptations, in different applications in image and signal processing (see Nanopoulos et al. [16]), medicine (see Pesu et al. [17]), or finance (see Giusto and Piger [3]).

Let $x_i \in \mathbb{R}^n,\ i = 1,\ldots,N$ a vector associated to a class y_i, where $y_i \in \{c_1, c_2, \ldots, c_k\}$. Algorithm LVQ trains over the pairs $\{x_i, y_i\}$ to create a model M that predicts the class of a new vector x. The LVQ algorithm is a supervised learning algorithm based on codebook vectors, which are vectors used to represent the different classes. Let $M = \left\{m_j^1\right\}_{j=1}^{N}$ be a set of initial codevectors, where $m_j^1 \in \mathbb{R}^n$ and $N \in \mathbb{Z}, N \in [k, n]$. It should be noted that one class can have more than one codevector. The algorithm uses two parameters, α to iteratively adjust the codevectors, and G as the number of iterations. The codevectors are updated in a linear fashion according to the distances of these codevectors to the closets datapoint. Let m_j^g be the codevector at step g, $g = 1,\ldots,G$. The description of the algorithm is given next:

1. Initialize $g \leftarrow 1,\ i \leftarrow 1$.
2. Identify the codevector m_c^g closest to x_i, using a preferred metric distance:

$$c = \underset{j \in \{1,2,\ldots,N\}}{\operatorname{argmin}} \left\{ \| x_i - m_j^g \| \right\}$$

3. Update the codevector set:
   - If x_i and its closest codevectors are in the same class:

$$m_c^{g+1} = m_c^g + \alpha^g \left(x_i - m_c^g \right)$$

   - If x_i and its closest codevectors are not in the same class:

$$m_c^{g+1} = m_c^g - \alpha^g \left(x_i - m_c^g \right)$$

4. If $i + 1 \leq N$, let $i \leftarrow i + 1$ and return to step 2. If not, go to step 5.
5. If $g \leq G$, let $i \leftarrow 1,\ g \leftarrow g + 1$, and return to step 2. Otherwise stop.

The final placement of the codebook vectors in an LVQ algorithm is not invariant to the initialization, so practitioners need to run multi-start procedures to initialize the codebook vectors. The class prediction of a new data point x consists of the class of the closest codevector:

$$c = \underset{j \in \{1,2,\ldots,N\}}{\operatorname{argmin}} \left\{ \| x - m_j \| \right\}$$

Giusto and Piger [3] proposed using the LVQ algorithm to identify United States BC TPs using the available NBER chronology and the four leading indicators (LIs) favored by the NBER Business Cycle Dating Committee (BCDC): non-farm payroll employment (E), industrial production (I), real personal income excluding transfer receipts (P), and real manufacturing and trade sales (M), over a period from November 1976 to July 2013. Recession and expansion phases are persistent according to their estimators from this period of time. The LVQ algorithm added this persistency by classifying a vector with their current values and a lag 1 of the series of data. In their research, at time T + 1, an analyst is looking to determine if a TP has occurred in the recent past, $T - j$ months, where the NBER classification is known. The time-window j varies according to some assumptions: (1) the date of a new peak or trough is known; (2) 12 months is the maximum time between a new peak and its announcement by the NBER; and (3) 6 months is the minimum amount of time of each BC phase. The determination of a new TP in a prediction period occurs if three consecutive months are classified in a different phase than the prior one, which is consistent with the NBER definition of a TP and the assumption of persistent periods. This process is executed 100 times looking to avoid a false identification of new TPs; if the LVQ identified a new TP in at least 80% of the runs, then this is assumed as true. This threshold presents better results in terms of accuracy and avoids false positives and negatives when comparing values from 50% to 90%.

Giusto and Piger [3] presented excellent results in the accuracy of their proposed methodology. With respect to the TPs identified by the NBER, the average lag to a new peak and a new trough improves in 90 and 212 days, respectively. Also, the LVQ algorithm never produces any false positives and negatives with respect to the ones predicted by the NBER. Because the NBER corrects their prediction results historically, the LVQ algorithm was also compared with other statistical methods. For the period of the great recession, 2007–2009, Hamilton [2] reported that the methods in use during that time did not identify a BC peak, whereas the LVQ algorithm detected one peak on 6 June 2008. He also compared the LVQ algorithm with the results presented by Chauvet and Piger [9] of the dynamic factor Markov-switching (SDMF) model of Chauvet [11] for identifying TPs over the period of November 1976 to June 2006. LVQ improves the results of SDMF in the identification of both peaks and troughs. Economic models had such trouble with the great recession that Congress held hearings and invited five renowned economists to testify regarding the identification of BC TPs (U.S. Congress, [18]).

3. Using Monte Carlo Simulation for Multi-Path Analysis

3.1. A Change-Point Model to Operationalize Business Cycles

The algorithms that model BC TPs typically assume a bivariate response for economic performance and model the probability of a recession. The probability is translated into a prediction on whether the current time segment will be classified as a BC TP, which is a transition from recession to growth or growth to recession periods. Rather than relying just on the single data path observed during the period of the study, we propose that measures of economic performance, such as TTS, should account for stochastic variation of the LIs under study.

The distribution of the LIs can be used to reflect historical or future behaviors of interest, thereby allowing investigators the ability to explore an algorithm's performance across a broad range of simulated scenarios from historically observable recessions to potential future scenarios. This section describes a method for using Monte Carlo Simulation to select design parameters and compare algorithms by comparing distributions for the TTS across a range of changes in economic LIs. The description of the methodology assumes that multiple LIs were used to make the BC TP decision. The algorithm is provided at the end of this section.

Our proposed Monte Carlo simulation method uses a population model or DGP for a multi-variate change point process for a series of LIs. The performance of a range of algorithm design parameters is evaluated for changes in the DGP of varying severities. The change-point model:

$$\underline{Y}_{ij} \sim \begin{cases} MV\left(\underline{\mu}_1, \Sigma_1\right), & i = 1, \ldots, \tau_1 - 1;\ j = 1, \ldots, m \\ MV\left(\underline{\mu}_2, \Sigma_2\right), & i = \tau_1, \ldots, \tau_2 - 1;\ j = 1,\ \ldots,\ m \\ & \vdots \\ MV\left(\underline{\mu}_k, \Sigma_k\right), & i = \tau_{k-1}, \ldots, \tau_k - 1;\ j = 1, \ldots, m \end{cases} \tag{1}$$

represents the DGP from time point 1 to τ_k, where $\underline{Y}_{ij}$ follows a multivariate distribution for a set of LIs of interest with a mean vector of $\underline{\mu}_i$ and a variance–covariance matrix Σ_i. Parameter τ_n represents the nth change point, and m the number of LIs that make up $\underline{Y}_{ij}$.

The model provides the flexibility to test changes in the population mean and variance–covariance structure. Not all changes specified in Equation (1) might be large enough to be declared as a BC TP. Equation (1) represents a model of economic behavior, while the objective of the NBER BCDC is to identify the changes in economic performance significant enough to declare a BC TP. An algorithm's sensitivity should be adjustable to detect this range of changes based on what is important for the user. Detection of smaller changes may be of interest to an analyst while deeper changes would be the focus for researchers predicting BC TPs. This population model allows for the assessment of a spectrum of TPs, both historically observed and unobserved. Researchers can select design parameters for their algorithm such that the distribution for the TTS shows a high probability of detecting the shallowest recession the BCDC has ever declared and a low probability of declaring a BC TP when a change point in the DGP is shallower than the shallowest recession—a false alarm. Our reassessment of the LVQ algorithm in the following section provides an example of this type of analysis.

3.2. Evaluating Practical Significance

The idea of monitoring a time series and declaring a change for fluctuations that are considered significant is not new. The concept of identifying significant changes in a system over time has been addressed in the process monitoring literature. Ewan and Kemp [19], Freund [20], Woodall [21,22], Box et al. [23,24] and Yashchin [25] partitioned a response over time into: good, undefined, and bad zones.

A basic application of a practical significance analysis is the 2-zone approach by Woodall and Faltin [26], which combines the good and indifferent zones into a "Dead-Zone" where an algorithm should have a low likelihood of detecting a change and an out of control or "Bad-Zone" with a high likelihood of detecting a change. For the BC TP problem, the TP would be the transition between zones. The 2-zone model gives context to language and measures in BC literature. When authors describe results that include false alarms, they are referring to a situation where the LIs are in the Dead-Zone, which means no changes occurred or the changes in the mean vector of the LIs are not large enough for the NBER to declare a BC TP recession (growth), and yet the algorithm declares a change. The Bad-Zone is represented by the space of LI distributional characteristics where the NBER flagged BC TPs. The simulation approach that we propose allows researchers to account for how the TTS distribution changes across the Dead-Zone into the Bad-Zone. It also allows for the tunning of an algorithm's design parameters based on these probabilities.

3.3. A Fitting a Model for Monte Carlo Simulation

The Monte Carlo simulation approach has the potential of helping researchers decide on design parameters by adding a simulated multivariate time series to the end of a historical series while evaluating the TTS distribution. This approach creates the opportunity to explore the impact of (i) typical variation within a growth or recession period based on the DGP; as well as (ii) changes in the DGP or distributions of the LIs on the TTS distribution for different design parameters. Researchers can study any construction of linear and non-linear distributional changes between growth and recession periods. Since the point of the paper is to propose the assessment of multi-path behaviors, we will keep the assumptions simple and use a Multivariate Normal distribution for the LIs, with a defined shift in the mean and a constant variance-covariance matrix.

The first step of the algorithm is to fit the recessionary distribution that is to be detected. Figure 1 shows the recession behavior for two important LIs between 1965 and 2013. Mfg. and Trade Sales is in a state of dynamic equilibrium across the 7 recessions displayed, while the Non-Farm Payroll Employment can change in a more unpredictable pattern over a recession. Because recessions are typically declared within 7 months of a TP (Giusto and Piger [9], p. 180), researchers should consider estimating the mean using the first 7 months of the LIs they plan to use for BC TP identification.

The simulated means for the LIs for a recession, in the case study, are estimated by the average of the months available from the start of the identified BC TP to the month the TP was flagged.

The Dead- and Bad-Zones should be determined through evaluation of the means of the 4 LIs. The means span a range from the deepest recession period to the shallowest. The deepest recession would be the recession where the average growth for each of the LIs was furthest from 0. "Shallow" recessions have average growth for each of the LIs which are closest to 0. The radar chart in Figure 2 shows the shallowest and deepest recessions analyzed by Giusto and Piger [3]. The shallowest recession, which started in March 2001 took the longest to detect at 7 months. The deepest recession started in January 1980 and was detected in 3 months by the LVQ algorithm.

Given the defined recessions of interest, the second step requires the selection of the final growth period. The simulated recession added to this final period of growth. Multiple final growth periods could be used to evaluate the TTS for given changes between the population means for the final growth and recession periods per the specified design parameters.

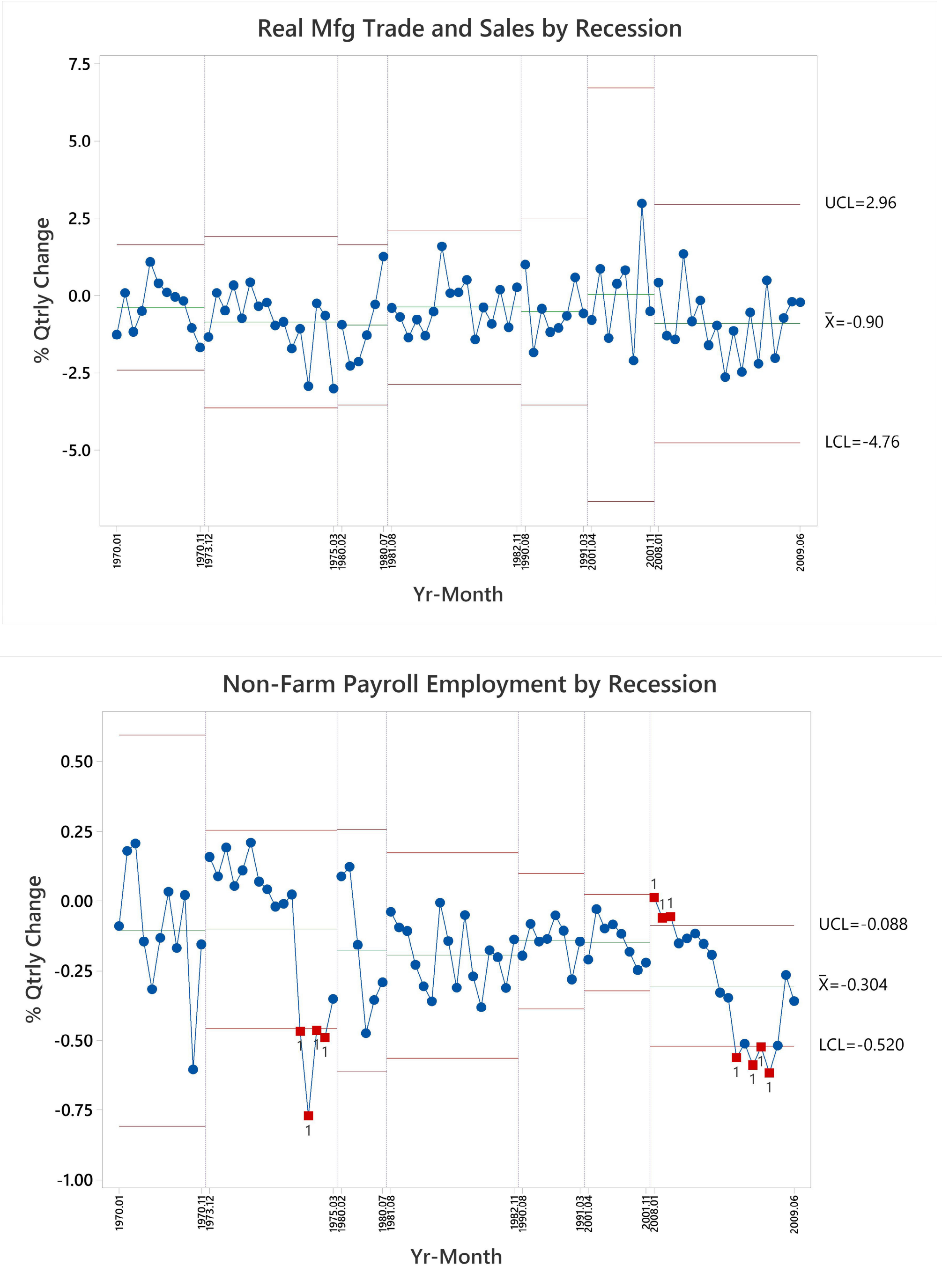

Figure 1. The graphs depict a concatenation of the last 7 recessions for real mfg trade sales and non-farm payroll employment. The upper and lower lines for each recession represent a rudimentary assessment of typical variation and the middle line represents the mean. The Red points represent values in a recession which are flagged as unusual. Understanding the type of variation a leading indicator (LI) experiences along with an understanding of when the variation occurs helps the researcher determine how best to model the behavior of the LI during the period that BC TP signals are typically observed, within 7 months. Performances of the two LIs depicted in Figure 1 reveal different behaviors between and within LIs for the 7 recessions. Real manufacturing trade and sales values less variation, however there are fewer values flagged as unusual than the non-farm payroll employment values. Given the flagged values tend to occur after 7 quarters we are less concerned about special patterns that occur within 2 years of the onset of a recession. This simplifies modeling recession behaviors for these two LIs.

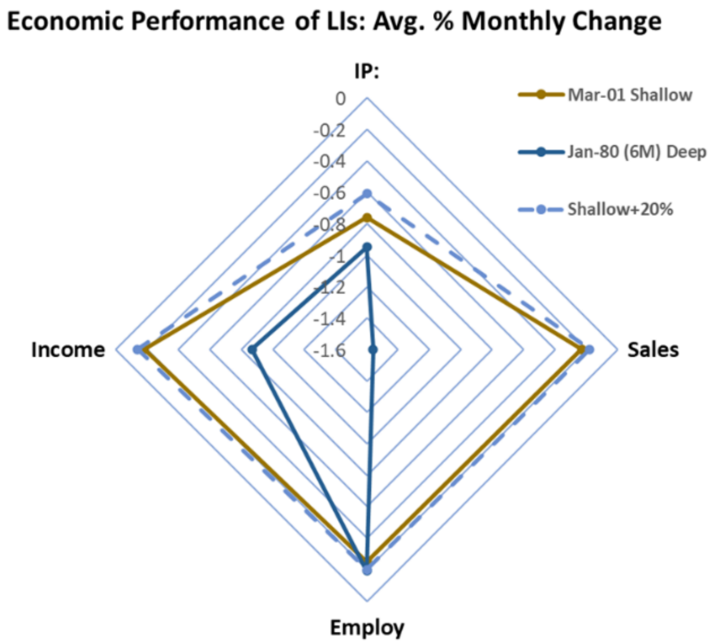

Figure 2. Depiction of 3 levels of economic performance in 2 zones: the Dead-Zone is represented by the Shallow Recession + 20%, and the Bad-Zone is represented by the Shallow Recession period starting in March 2001 and deep recession starting in January 1980.

The third step of the Monte Carlo simulation involves simulating the recession(s) of interest using the population model specified in Equation (1). Variables for the 4 LIs of a simulated recession stage are defined using the VAR model:

$$\underline{y}_{it} = \underline{b}_0 + B_t * \underline{y}_{it-1} + \underline{\varepsilon}_{it}, \tag{2}$$

where $\underline{y}_{it}$ is a $p \times 1$ vector, $\underline{\varepsilon}_{it} \sim N(\underline{0}, \Sigma)$, and Σ and B_t are $(p+1) \times (p+1)$ matrices with $i = 1, \ldots, 4$. We use lag 1 for simplicity in Equation (2). We use information from the case study to fit the model described, which results in:

$$\begin{aligned} \hat{y}_{I,t} = {} & -0.085 + 0.163I_{t-1} + 0.113E_{t-1} + 0.338P_{t-1} + 0.063M_{t-1} - 0.257I_{t-2} \\ & +0.050E_{t-2} + 0.095P_{t-2} + 0.033M_{t-2}, \end{aligned} \tag{3}$$

$$\begin{aligned} \hat{y}_{E,t} = {} & -0.001 + 0.126I_{t-1} + 0.067E_{t-1} - 0.101P_{t-1} - 0.050M_{t-1} - 0.156I_{t-2} \\ & -0.451E_{t-2} - 0.212P_{t-2} - 0.056M_{t-2}, \end{aligned} \tag{4}$$

$$\begin{aligned} \hat{y}_{P,t} = {} & 0.006 - 0.714I_{t-1} - 0.921E_{t-1} - 0.548P_{t-1} + 0.160Sales_{t-1} - 0.123I_{t-2} \\ & +1.957E_{t-2} - 0.239P_{t-2} + 0.175Sales_{t-2}, \end{aligned} \tag{5}$$

$$\begin{aligned} \hat{y}_{M,t} = {} & -0.094 + 1.557I_{t-1} - 8.748E_{t-1} - 0.113P_{t-1} - 1.235M_{t-1} - 0.586I_{t-2} \\ & -0.792E_{t-2} - 0.254P_{t-2} - 0.963M_{t-2}, \end{aligned} \tag{6}$$

for a second-order VAR model. The model was fitted with R (R Core Team, 2014) [27] using the function VAR from the R package vars (Pfaff, [28]). The equations show how the autocorrelation and cross-correlations are accounted for within a VAR model. The approach shown is used for the Monte Carlo simulation when reassessing the LVQ algorithm. The evaluation was set with lag 2 with a different subset of the data. The models in Equations (3)–(6) were generated using a combined dataset from the July 1990 and March 2001 recessions, starting the month after the peak and ending the month of the next trough. When reviewing the LIs for the recessions, we found that they showed qualitatively similar performance. The LIs within each recession were standardized to a zero mean.

The final step of the algorithm would be to append the simulated values onto the end of the defined dataset and repeat the process with a new simulated set of values each iteration.

3.4. Multi-Path Assessment Algorithm with Monte Carlo Simulation

The algorithm summarized here is used in Section 4 to show how a statistical assessment can be made regarding the design parameters for the LVQ approach utilized by Giusto and Piger [3]. It follows:

1. Fit the recessionary distribution of the LIs $N\left(\underline{\mu}_r, \Sigma_r\right)$ that you wish to evaluate:
   a. Where Σ_r is estimated from the last growth period or a combination of recession periods to provide a large enough sample size, and
   b. $\underline{\mu}_r$ represents the mean vector for the LIs of the recession under study.
2. Select the final growth period for the historical data under study.
3. Simulate N observations using an assumed distribution of the recession:
   a. Data is standardized to 0 growth,
   b. Desired mean vector, $\underline{\mu}_r$, is added to standardized data,
   c. Estimate a VAR model of order w,
   d. Select stopping point for N based on max. expected time to detect a signal. N should be $\leq$24, as TP should be detected within 24 months.
   e. Generate simulated values for the N time intervals using the VAR model.
4. Add the simulated values to the end of the historical dataset (ending in a growth period) and iteratively apply the test for whether a recession has occurred. Define test criteria:
   a. Decision level: Probability that the current time-period is part of a recession period,
   b. Decision threshold: the limit that a decision level must exceed a set number of times to declare a recession (growth) period,
   c. Define the number of time periods that the decision level must be greater than (less than) the decision threshold to declare a recession (growth) period.

4. Re-Assessment of Giusto and Piger's LVQ Implementation

We used the proposed Monte Carlo simulation method from Section 3.4 to provide statistical assessments of the design parameters for the LVQ algorithm as presented by Giusto and Piger [3]. Given the goal of the case study is to re-approach the LVQ in a BC TP context with a stochastic analysis, we provide a direct comparison with their published selection of design parameters. The TTS for shallow to deep recessions is evaluated using the 4 LIs described in Section 3. The LVQ algorithm of Giusto and Piger [3] is used to determine an estimated BC TP and TTS for the known change.

The four LIs favored by the BCDC are: (1) nonfarm payroll employment; (2) industrial production; (3) real manufacturing and trade sales; and (4) real personal income excluding transfer payments. We use monthly data as available to users from November 1976 to September 2000. The measure available for each LI at each month is the % monthly growth. For consistency, we used the vintage data available for download from Giusto and Piger [3]. Vintage datasets are also available from the Federal Reserve Bank of St. Louis (St. Louis Federal Reserve [29]).

Our Monte Carlo simulation measured the time, in months, required for the LVQ algorithm to signal a peak BC TP after a change was made to the underlying DGP. Key design elements of the simulation algorithm are provided below:

- Baseline period was March 1967 to September 2000.
- The simulated data were added after September 2000, which is 6 months prior to the March 2001 recession to avoid any end effects close to the start of a recession.
- A VAR process for the % monthly change of the 4 LIs was modeled using the March 1991 to March 2001 growth period. The Var–Cov matrix, Σ_g, was used when simulating the recession data, as the recession periods were short.

- The simulated shifts for data starting after the growth period spanned a range of means representing the shallowest and deepest recessions observed across the dataset studied by Giusto and Piger [3]. The shallowest recession was the recession with the longest TTS for the LVQ algorithm (started in March 2001), while the deepest recession (started in January 1980) had the shortest.
- The means for simulated recessions were estimated using the months from the recession start to the month the recession was flagged.
- The mean was calculated with the vintage data available the month the BC TP was flagged.
- The probability or decision thresholds studied ranged from 0.6 to 0.9, and the months required to be at or below the decision threshold were 2 and 3.
- The simulation was stopped after 24 months if the simulated change was not detected. To be useful, a peak BC TP should be flagged within 12 months. Using 24 months as the limit allowed more separation of the TTS distributions in the simulation.

The top four graphs of Figure 3 show actual data for the first 5 months of the LIs for the December 2007 recession added onto the growth period ending in September 2000. The observed data are overlaid on a single set of simulated observations to provide a relative reference for the location and spread of the two datasets. The specific simulated result shown provides a reasonable representation of the actual mean and standard deviation of the observed data. Simulating the 4 LIs 5000 times generates different paths with varying TTS results. The second set of four graphs in Figure 3 displays three data sequences appended onto the final growth period. Each of the three sequences represents a simulated result that could have occurred from the same DGP. The sequences were selected from 5000 multivariate Monte Carlo runs. All four LIs were considered together for a TTS determination. One of the sequences generated a TTS of 3 months. The 3-month sequence had the lowest values of the three sequences for Personal Income and Mfg/Trade Sales, while its values were similar to the other two sequences for Industrial Production and Payroll Employment. The differences were enough to cause a TTS of 3 months. The sequence of four LIs with a TTS of 7 months had averages for Personal Income and Mfg/Trade Sales that were higher than the averages for the sequence with a 3-month TTS and less than the averages of the sequence with a 24-month TTS. The averages were similar for Industrial Production and Payroll Employment. For the same DGP, these types of fluctuations can arise due to differences in small nuisance factors that affect how the LIs occur. A researcher can evaluate different amounts of variation caused by the nuisance factors using historical or expected future economic behavior.

For the same VAR model, the TTS can vary from 3 to 24 months depending on fluctuations in the cross and auto-correlated LI results. The important part of the simulation is the distribution of the TTS. This potential for variation becomes important when selecting design parameters such as the probability threshold and number of months at or above the threshold. Rather than using an average of a single path to select the best options for decision criteria, we select decision criteria that perform well probabilistically (i.e., over a collection of future possible paths).

Monte Carlo simulation provides a flexible approach to simulating behaviors of future recessions. Whether the mean is degrading over time or making a step change, the Monte Carlo simulation approach can provide a general prediction or understanding of the BC TP algorithm's performance with regards to TTS for a selection of design parameters.

Using the LVQ algorithm and SPA, Giusto and Piger [3] (p.181) found that for the 3-month criterion, the probability threshold of 0.9 and 0.8 required a similar amount of time to identify a peak, approximately 4.5 months, with no false alarms. The 0.8 threshold required 1 month more than the 0.7 threshold; however, the 0.7 threshold produced a false alarm. The 0.6 and 0.5 thresholds required the same average TTS, but the 0.6 threshold produced one false alarm and the 0.5 threshold produced three false alarms. The general implication of the SPA is that the 3-month, 0.8 threshold should be used to identify peak BC TPs. A limitation of this analysis is that there is no opportunity to study how adjusting the month and probability threshold criteria can be optimized for the severity of recession that the researcher may be interested in.

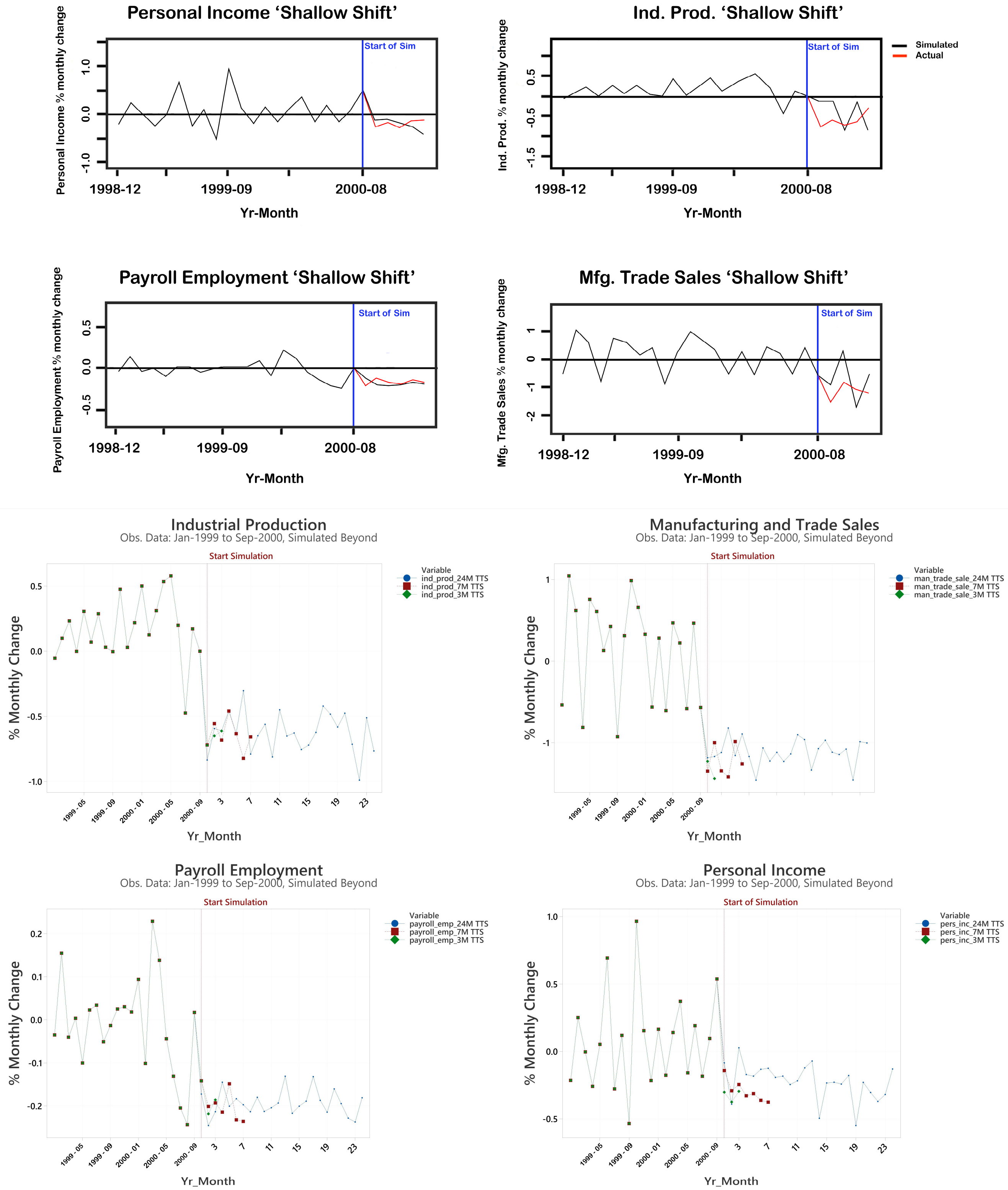

Figure 3. Top four graphs represent a simulation of December 2007 recession added on to the end of the March 1991 Growth period. Bottom four graphs show three simulated sequences, from the same Data Generating Process, that resulted in a 3, 7, and 24 month TTS. The TTS was determined through joint assessment of the four LIs.

We analyzed the data with the R program provided by Giusto and Piger [3] and did not find the false signal at the 0.7 threshold, which implies that the 3-month criterion with 0.7 threshold might be a good option as it identifies peak BC TPs 1 month earlier than the 0.8 threshold, and no false alarms occurred. However, this leaves the question of how much of a risk is there for a false alarm and how consistently can the 0.7 threshold identify peak BC TPs across a range of recession severities? And does the 0.8 threshold provide similar performance with less risk of a false alarm? These questions cannot be addressed by the SPA.

Monte Carlo simulation provides a more complete understanding of the decision criteria than the SPA. Following the Monte Carlo simulation approach in Section 3.4 and the key design elements from Section 3.3, we generated a distribution of 5000 TTS values for the decision thresholds of 0.6, 0.7, 0.8, and 0.9 for 2 months and 3 months with three severities of recession (Shallow + 20%, Shallow, Deep) using the 4 LIs previously mentioned. In Table 1, we show the TTS values for the Shallow + 20%, Shallow and Deep recessions with the 5th, 25th, 50th, 75th, and 95th percentiles for the simulated TTS values. The mean % monthly growth values for the Deep Recession (I, M, E, P) are: −0.95, −1.56, −0.192, −0.869, for the Shallow Recession: −0.76, −0.23, −0.25, −0.18, and for the Shallow Recession + 20% are: −0.61, −0.18, −0.20, −0.14.

Table 1. Percentiles (5th, 25th, 50th, 75th, and 95th) for varying decision level thresholds. The Dead-Zone represents a change that is real, albeit not significant enough for the BCDC to declare a recession. The means for the LIs in the Dead-Zone represent a 20% improvement over the Shallowest Recession and are still far from the means for the growth period (starting April 1991): 0.344, 0.359, 0.168, 0.326.

Monthly Criterion	Probability Threshold	Dead Zone: Shallow Rec + 20%	Shallow Recession (March 2001)	Deep Recession (January 1980)
$m = 2$	0.6	3, 5, 10, 24, 24	3, 3, 3, 3, 5	2, 2, 2, 2, 2
	0.7	3, 7, 19, 24, 24	3, 3, 3, 4, 6	2, 2, 2, 2, 2
	0.8	3, 9, 24, 24, 24	3, 3, 3, 4, 7	2, 2, 2, 2, 3
	0.9	4, 24, 24, 24, 24	3, 3, 3, 5, 9	2, 2, 2, 3, 3
$m = 3$	0.6	5, 24, 24, 24, 24	3, 4, 4, 5, 9	3, 3, 3, 3, 3
	0.7	7, 24, 24, 24, 24	4, 4, 4, 7, 11	3, 3, 3, 3, 3
	0.8	10, 24, 24, 24, 24	4, 4, 5, 8, 24	3, 3, 3, 3, 3
	0.9	12, 24, 24, 24, 24	4, 4, 7, 10, 24	3, 3, 3, 3, 3

There are several analyses that can be conducted. We compared the simulated results for the 3-month 0.8 and 0.7 thresholds with the SPA to show how the Monte Carlo simulation approach improves the understanding of performance. The analysis for the Monte Carlo simulation provides a more complete understanding of the impact on TTS for the design parameters due to the opportunity to understand the changes in distribution for different severities of recessions.

The analysis can be broken down into 3 steps that can be iterated: (1) select the recession depth; (2) pick the monthly criteria; and (3) compare the distribution for the TTS across different probability thresholds.

1. Select the shallow recession to start, as the deep recession is easily identified by all design parameter options.
2. Because the results of a SPA from Giusto and Piger [3] selected $m = 3$, start with 3 months for the number of months the probability threshold must be met.
3. Distribution of TTS for a peak:
    a. The 50th percentile of the TTS with probability threshold of 0.6 and 0.7 are both 4 months while 0.8 is 5 and 0.9 is 7 months. This aligns well with Giusto and Piger [3], who identified five peaks with a median of 4.1 and longest of

7.1 months. The months for the Monte Carlo simulation results are larger, as the SPA combines all recession severities and therefore biases the average TTS. The Shallow recession from Table 1 is the least severe recession within the Giusto and Piger [3] dataset, and the probability threshold of 0.7 provides a faster median TTS than 0.8.

b. The 75th percentile of the TTS with probability threshold of 0.6 is 5 months, 0.7 is 7 months, while 0.8 is 8 and 0.9 is 10 months. This makes sense, as increasing the proportion of detected peaks out of 5000 will result in an increase in the number of months to detect; however, there is still a 1-month shift between the 0.7 and 0.8 probability thresholds.

c. The 95th percentile of the TTS with the probability threshold of 0.6 is 9 months, 0.7 is 11 months, while 0.8 and 0.9 are greater than 24 months. This result was not possible to explain with the single path analysis, which addresses an average for all recessions. For a shallow recession, the selection of the probability threshold of 0.8 or 0.9 creates a 5% probability of taking longer than 24 months to detect a peak. The probability threshold of 0.7 provides much better protection against poor performance.

4. Iterate step 3 with a different severity of recession. The Monte Carlo simulation allows for the evaluation of changes in the LIs that have either not occurred or have not been declared as a recession by the BCDC. In our simulation, we created a Dead-Zone change in the LIs by adding 20% to the averages of the Shallowest recession declared by the BCDC for the dataset studied by Giusto and Piger [3]. This change is used to evaluate the probability of a false alarm.

   a. The 50th and 25th percentiles of the TTS, with probability threshold of 0.6, 0.7, 0.8, and 0.9, are all greater than 24 months (this is also true for the 75th and 95th percentiles).

   b. The 5th percentile of the TTS with probability threshold of 0.6 is 5 months and 0.7 is 7 months, while 0.8 is 10 and 0.9 is 12 months. There is a 5% probability of declaring this change a peak in 7 months if the probability threshold of 0.7 is used and in 10 months for 0.8. This translates to a low probability of a false alarm for 0.7 or 0.8.

5. The entire analysis can be repeated with the monthly threshold, m, set to 2. A review of Table 1 shows that the 0.7 and 0.8 probability thresholds improve the detection of a Shallow recession. The median TTS decreased from 4 and 5 months for $m = 3$ to 3 months for $m = 2$ while the 95th percentiles went from 11 and 24 to 6 and 7, respectively. However, when using $m = 2$, there is now a 25% probability of declaring the Dead-Zone change a BC TP in 7 months with a probability threshold of 0.7 and 9 months with a probability threshold of 0.8. This is an increase of 20% over the $m = 3$ probability of a false alarm. If a researcher is willing to be more aggressive, they have an option to use $m = 2$, and they would know the change in risk levels for different severities of recession.

5. Discussion

The ability to evaluate the probability of detecting differing severities of recessions provides a more detailed understanding of the impact of design parameter choices than SPA. The re-assessment of the LVQ showed that the options for design parameters were (i) $m = 3$, $p = 0.7$; (ii) $m = 3$, $p = 0.8$; and (iii) $m = 2$, $p = 0.7$ or 0.8 along with the risks and benefits of the different options. We also saw that $m = 3$, $p = 0.07$ provides a significant reduction in the worst-case probability of missing a BC TP. The worst-case protection is not possible to assess with the SPA.

Future research can evaluate the impact of changes in distributional forms, behaviors of the mean over time, and structure of the variance–covariance matrices. This paper focused on the identification of BC peaks which represent the onset of a recession. Similar analyses can be conducted for BC troughs to simulate capabilities for decision criteria in identifying the onset of growth periods.

6. Conclusions

We have shown that it is important to consider the statistical variation in the LIs when selecting design parameters for an algorithm that flags BC TPs to predict the state of the economy. Historical SPA provides an initial assessment; however, the results provide a narrowly focused analysis which is restricted to the observed data and is not capable of assessing all potential risks. The proposed Monte Carlo simulation methodology can be applied to any algorithm that predicts BC TPs. Monte Carlo simulation allows the researcher to avoid biases from specific single path sequences and provides a mechanism to make more informed decisions regarding the next BC TP flag. Future work needs to explore topics such as the impact of changing distributional forms, dynamic transition periods from peak to trough, and the structure of the variance–covariance matrices on the TTS distribution.

Author Contributions: D.E.: methodology, software, writing, investigation, validation, data curation, original draft preparation; M.B.: investigation, supervision, project administration, editing; V.G.T.-G.: conceptualization, supervision, review, editing; and A.E.C.-F.: review, formal analysis, editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. NBER. The NBER's Recession Dating Procedure. 7 January 2008. Available online: http://www.nber.org/cycles/jan08bcdc_memo.html (accessed on 5 April 2022).
2. Hamilton, J.D. Calling recessions in real time. *Int. J. Forecast.* **2011**, *27*, 1006–1026. [CrossRef]
3. Giusto, A.; Piger, J. Identifying Business Cycle Turning Points in Real Time with Vector Quantization. *Int. J. Forecast.* **2017**, *33*, 174–184. [CrossRef]
4. NBER. The National Bureau of Economic Research. September 2010. Available online: http://www.nber.org/cycles/recessions.html (accessed on 30 October 2022).
5. The White House. How Do Economists Determine Whether the Economy Is in a Recession. 21 July 2022. Available online: https://www.whitehouse.gov/cea/written-materials/2022/07/21/how-do-economists-determine-whether-the-economy-is-in-a-recession/#:~:text=The%20National%20Bureau%20of%20Economic,committee%20typically%20tracks%20include%20real (accessed on 7 February 2024).
6. Hamilton, J.D. A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica* **1989**, *57*, 357–384. [CrossRef]
7. Goodwin, T.H. Business-Cycle Analysis with a Markov-Switching Model. *J. Bus. Econ. Stat.* **1993**, *11*, 331–339.
8. Birchenhall, C.R.; Jessen, H.; Osborn, D.R.; Simpson, P. Predicting U.S. Business-Cycle Regimes. *J. Bus. Econ. Stat.* **1999**, *17*, 313–323.
9. Chauvet, M.; Piger, A. Comparison of the Real-Time Performance of Business Cycle Dating Methods. *J. Bus. Econ. Stat.* **2008**, *26*, 42–49. [CrossRef]
10. Soybilgen, B. Identifying US business cycle regimes using dynamic factors and neural network models. *J. Forecast.* **2020**, *39*, 827–840. [CrossRef]
11. Chauvet, M. An econometric characterization of business cycle dynamics with factor structure. *Int. Econ. Rev.* **1998**, *39*, 969–996. [CrossRef]
12. Kohonen, T. Learning vector quantization. In *Self-Organizing Maps*; Springer: Berlin/Heidelberg, Germany, 1997; pp. 175–189.
13. Nova, D.; Estévez, P. A review of learning vector quantization classifiers. *Neural Comput. Appl.* **2014**, *25*, 511–524. [CrossRef]
14. Sato, A.; Yamada, K. Generalized learning vector quantization. *Adv. Neural Inf. Process. Syst.* **1995**, *8*, 423–429.
15. Seo, S.; Obermayer, K. Soft learning vector quantization. *Neural Comput.* **2003**, *15*, 1589–1604. [CrossRef]
16. Nanopoulos, A.; Alcock, R.; Manolopoulos, Y. Feature-based classification of time-series data. *Int. J. Comput. Res.* **2001**, *10*, 49–61.
17. Pesu, L.; Helistö, P.; Ademovič, E.; Pesquet, J.; Saarinen, A.; Sovijärvi, A. Classification of respiratory sounds based on wavelet packet decomposition and learning vector quantization. *Technol. Health Care* **1998**, *6*, 65–74. [CrossRef] [PubMed]
18. US Congress; US House Committee on Science and Technology. *Building a Science of Economics for the Real World*; Serial Number 111-106; Government Printing Office: Washington, DC, USA, 2010. Available online: https://app.thestorygraph.com/books/8925845f-3a94-426f-aa55-c61747844938 (accessed on 10 February 2024).

19. Ewan, W.D.; Kemp, K.W. Sampling inspection of continuous processes with no autocorrelation between successive results. *Biometrika* **1960**, *47*, 363–380. [CrossRef]
20. Freund, J.E. Some Results on Recurrent Events. *Am. Math. Mon.* **1957**, *64*, 718–720. [CrossRef]
21. Woodall, W.H. Weaknesses of The Economic Design of Control Charts. *Technometrics* **1986**, *28*, 408–409. [CrossRef]
22. Woodall, W.H. The Statistical Design of Quality Control Charts. *J. R. Stat. Society. Ser. D* **1985**, *34*, 155–160. [CrossRef]
23. Box, G.; Bisgaard, S.; Graves, S.; Kulahci, M.; Marko, K.; James, J.; Van Gilder, J.; Ting, T.; Zatorski, H.; Wu, C. Performance Evaluation of dynamic monitoring systems: The waterfall chart. *Qual. Eng.* **2003**, *16*, 183–191. [CrossRef]
24. Box, G.; Graves, S.; Bisgaard, S.; Van Gilder, J.; Marko, K.; James, J.; Seifer, M.; Poublon, M.; Fodale, F. Detecting Malfunctions in Dynamic Systems. *SAE Trans.* **2000**, *109*, 469–479.
25. Yashchin, E. Statistical monitoring of multi-stage processes. In *Frontiers in Statistical Quality Control*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 185–209.
26. Woodall, W.H.; Faltin, F.W. Rethinking control chart design and evaluation. *Qual. Eng.* **2019**, *31*, 596–605. [CrossRef]
27. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2014.
28. R Core Team. vars: VAR Modelling. R Package Version 1.5-6. 2021. Available online: https://CRAN.R-project.org/package=vars (accessed on 10 February 2024).
29. St. Louis Federal Reserve. Archival Federal Reserve Economic Data (ALFRED). 12 April 2020. Available online: https://alfred.stlouisfed.org/ (accessed on 7 February 2024).