Journal of Business Cycle Research
https://doi.org/10.1007/s41549-021-00063-5

RESEARCH PAPER

A Customized Machine Learning Algorithm for Discovering the Shapes of Recovery: Was the Global Financial Crisis Different?

Gonzalo Castañeda[1] · **Luis Castro Peñarrieta**[1,2]

Received: 7 April 2021 / Accepted: 23 December 2021

Abstract
In this paper, we modify a conventional machine learning technique to classify recession-and-recovery events emerging in the countries' business cycles. We do this by analyzing output dynamics in time windows of the same size for a large set of countries. We show with quarterly GDP series that, despite the simplicity of the method, it is possible to describe analytically the shapes of recovery ('shapelets') that can be considered representative in a sample of 95 events coming from 47 advanced and emerging economies. The proposed methodology allows to depurate the number of shapelets empirically relevant, and also to produce groupings with economic meaning that are strongly associated with recession features such as depth, duration, cumulative losses, and others. Furthermore, we find that the relative frequency of these clusters can vary with the type of crisis. In particular, in the recent global financial crisis, shapelets describing severe recession events were very likely but mild recessions were also common.

Keywords Recessions · Business cycles · Global financial crisis · Machine learning · Shapelets

JEL Classification E32 · F44 · C20 · C60

✉ Gonzalo Castañeda
sociomatica@hotmail.com

Luis Castro Peñarrieta
luis.castro@cide.edu; luiscastro@upb.edu

[1] Department of Economics, Centro de Investigación y Docencia Económicas, A.C. (CIDE), Mexico City, Mexico

[2] Centro de Investigaciones Económicas y Empresariales (CIEE), Universidad Privada Boliviana, La Paz, Bolivia

Published online: 27 January 2022

1 Introduction

Financial analysts and the media pay much attention to the different shapes that are likely to appear in the output dynamic of recessions and their subsequent recovery.[1] A ‘soup of letters’ is commonly mentioned: L-shaped to describe an initial sharp fall in GDP and then a prolonged (long-term) path below the trend level observed in the pre-recession peak; V-shaped to describe a snap-back dynamic in which a quick fall comes along with a symmetrical speedy upturn so that the initial trend path is recovered; J-shaped (Nike logo or checkmark) when the fall is limited to a few quarters and the economy recovers steadily; W-shaped to describe ‘double-dip’ recessions in which the initial recovery is interrupted by another drastic fall, perhaps due to the presence of additional complications in the economy’s performance; U-shaped to describe a steady fall and a sluggish recovery; K-shaped when a general output drop in the economy is followed by some sectors recovering and others stagnating or falling even further.[2]

It seems to us that some of these letters are part of the folklore and that their frequent reference is due to media hype. Therefore, in this paper, we use a customized technique to find out whether these shapes or others are truly observed in the data of a sample of 47 advanced and emerging economies that have experienced significant (mild or severe) recessions in their business cycles. For this, we appeal to the concept of ‘shapelet’ from the machine learning literature. A shapelet is a subsequence of a time series that has been recognized as a representative segment of a class of series and can be used for classification purposes.[3] For instance, in Fig. 1, we compare the series of interest (X) to the shape of shapelet S through the normalized distance between the series and the shapelet (d); if the distance is below a certain threshold (δ), then that series belongs to shapelet S.

Consequently, in a first step (discovery), it is necessary to identify similitudes between different segments in a set of series. The seminal papers on shapelets are Ye and Keogh (2009, 2011), which have produced a flourishing line of research with applications in several disciplines: geology, biology, meteorology, etc. This classification method has the advantage that the discovered shapes can provide information for the analysts to interpret. This concern is highly relevant when the objective is to identify shapelets in subsequences that coincide with recession events in the GDP series.[4]

[1] There is even a Wikipedia entry describing recession shapes (https://en.wikipedia.org/wiki/Recession_shapes). When looking for the words ‘recession and recovery shapes’ in Google, we found 148 links to newspaper columns, magazine articles, financial and economic reports (accessed: February 17, 2021).

[2] There are also references to other symbols: ‘long checkmark’ when GDP falls slowly, but after many quarters exhibits an upturn, which could be slow or fast; or ‘square root’ to describe a symmetrical fall and recovery followed by a period of stagnation.

[3] Generally speaking, this vector of ordered observations does not need to be a time series, but only a collection of observations that are measured sequentially (e.g., the angle that forms between the stem of a leaf and each point in its periphery).

[4] From now on, we use the terms recession events, recession-and-recovery events, recovery from recession events indistinctively. That is to say, the empirical spells to be analyzed here describe the different phases observed when a crisis or shock hits an economy.

A classification procedure for recession events based on shapelets is ideal for the following reasons: it deals with sequential observations and, thus, it can identify shapes in the output dynamics; it does not require synchronicity for two events to be considered similar; it can deal with the noisy component of the series that might distort the identification process; it is a semi-automatized method and hence minimizes the use of subjective judgment for classifying events with specific shapes; it makes use of statistical metrics to determine if two shapes (letters or symbols) associated with empirical spells are substantially different; it allows the identification of several shapes of recovery within the same GDP series but in different historical periods; it reduces the possibility that two subsequences can be considered similar only for random reasons when a set of theoretical and interpretable shapelets is specified in advance.

Consequently, by studying recession events with shapelets, we can identify which shapes are more frequently observed in the data. Moreover, we can distinguish if a specific shape is more commonly observed in certain types of crises, economic or otherwise. For instance, in this paper, we analyze whether the global financial crisis of 2007–2008 produced a distribution of shapes, across countries, different from those observed in other recession spells included in the database. Besides, we show through descriptive statistics and a multinomial model that the clusters of empirical spells based on our shapelet classification scheme are strongly associated with different features (e.g., depth, output loss, mean drop, duration) of the recession-and-recovery events. With this association, we can establish a ranking of the severity of the recession (a comprehensive measure of the recessions' features), which helps provide economic interpretability to the empirically relevant shapes of recovery.

Since the study of recession events is constrained by relatively short-time series, we decided to modify a traditional unsupervised machine learning technique (nearest-neighbors) for the classification of these events. Another reason for preferring this approach is that we have additional information on the quarters in which recession spells start. Hence, we do not need a procedure to discover the position of the shapelets, as done in the traditional ML techniques. Therefore, in the initial stage of our methodology, we define *ex-ante* (i.e., from the visual inspection of empirical spells) ten theoretical shapelets to be considered as potentially representative shapes. Then, we automatize the method for the classification of spells and the depuration of clusters through different tools: the customized ML algorithm, a Relative Adjustment Score (RAS) that measures the separability of such clusters, and a multinomial model the measures the out-of-the sample predictability of clusters using recession features.

Through the RAS method and the multinomial model, we conclude that a grouping with four shapelets is capable of differentiating the nature of the recession events, in terms of their economic meaning and the specification of clear (non-blurred) borders between clusters. The method has a better performance than other classification techniques based on features (scaled KMeans, transformed KMeans through dimension contraction) for the following reasons: (i) it can produce a larger number of clusters (four *versus* two), which allows distinguishing more dynamics in recession events; (ii) the four clusters discovered in the shapelets approach exhibit

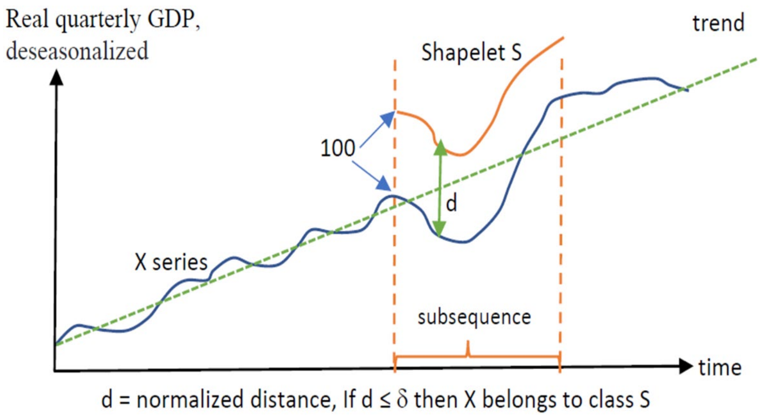

Fig. 1 Shapelet discovery in a subsequence of a GDP time series

good predictability when using recession features; (iii) the shapelets-clusters have economic interpretability, which scaled KMeans do not provide.

The rest of the paper is structured with eight more sections, besides the online supplementary material. In the following section, we present a literature review on other methods for identifying the shapes of recovery. In the third section, we extend on the concept of shapelet and explain how it can be associated with the output dynamic of recessions and their recovery within a prespecified time window. In the fourth section, we describe the database used for the study and define empirically a recession event. In the section on methods, we formulate analytically the configuration of theoretical shapelets and elaborate a metric for determining the quality of the empirical clusters. In the sixth section, we present our results concerning three classification schemes which vary in terms of the number of clusters considered. In the seventh section, we present some descriptive statistics for recession features and how they are related to the shapes of recovery. In the eighth section, we present the estimation of a multinomial model that validates the separability between clusters. In the ninth section, we review the relative frequency of shapelets in events associated with specific crises, like the Global financial crisis, and make comparisons between them. In the section of conclusions, we synthesize our main results and make suggestions on a possible research venue where the concept of shapelets could be handy.

2 A Review of the Empirical Work On Letters' Recognition

In the scholarly literature, there are two main procedures to identify the shapes of recovery through the letters' jargon. The most common one is a graphical procedure known as butterfly charts (Howard et al., 2011). The horizontal axis specifies several years (or quarters with deseasonalized series) before and after the recession trough, while the vertical axis measures a normalized GDP using the pre-recession peak (or the trough); although, occasionally, the growth rate with respect to this trough is specified in the y-axis. Then, to the naked eye, the analyst decides whether a recession event has a specific shape and how it should be classified. This manual method is highly subjective, prone to misjudgments because of the presence of noise, and not reliable to separate clusters with certain affinity.

A more formal alternative for characterizing asymmetric dynamics over the business cycle, and for describing letters' types in recession events, is conducted by a non-linear estimation procedure known as the Markov-switching model. In a seminal paper, Hamilton (1989) studies expansion and recession regimes, in terms of an autoregressive framework with a state-dependent output mean, whose transitions are determined stochastically. Through the years, these models have become more complicated but also more informative. Three of these adaptations stand out since they have increased the model's potential to identify different shapes of recoveries. Incorporating three or more regimes in the state variable (e.g., Sichel, 1994; Ducker, 2006); for instance: recession, slow growth, ordinary growth, and snap-back growth. Introducing a bounce-back function in the expression of the state-dependent mean (e.g., Kim et al., 2005; Morley & Piger, 2012) that makes possible the modeling of the recovery phase in terms of the duration, magnitude, and depth of the previous recession. Adding a delayed bounced-back effect so that the rebound does not necessarily happen immediately after the economy has reached its trough (e.g., Bec et al, 2011; Teimouri, 2012, Chap. 4).

The combination of bounce-back functions and delayed effects allows the data to be highly informative regarding the shape of the event. This improves the possibility of selecting the appropriate shape of recovery in so far V, U, Long U, L, W shapes, as well as recoveries proportional to the recession's depth, become feasible options. The selection of a specific function is made either through the use of traditional information criteria (AIC, SIC) or by using statistical tests on parameter constraints for the bounce-back functions. However, this econometric method presents three important drawbacks: (i) it requires long time series where structural breaks have to be dealt with; (ii) it is computationally intensive since the previous model selection process has to be applied to each of the countries included in the database; (iii) it produces typical (average) shapes as a result of the estimated parameters (i.e., a family of bounce-back effects is selected) to describe all recession events found in the same time series. This average shape is not consistent with the empirical evidence since different historical periods of the same country tend to produce quite different recession events.

3 Identification of Theoretical Shapelets in Empirical Spells

In the machine learning literature, an empirical shapelet is a subsequence of observations that appears recurrently –with slight modifications– in several of the time series composing a data set, hence it can be considered one of the series' traits that make their classification possible. The form, extension, and position of these shapelets are not predetermined, hence a first algorithm is needed to discover shapelets, and with a second algorithm, the time series included in a database can be classified according to these shapelets. Some of these methods are computationally intensive, and they do not necessarily produce shapes that are interpretable. We show, in the top panel of Fig. 2, two empirical shapelets (green and red) that have some economic meaning when studying output dynamics, while in the bottom panels we detect its presence in two different time series.

However, in the problem at hand, we firstly identify when a recession event occurs. This helps us to avoid scanning the whole time series when searching for shapelets, reducing computational time considerably. Likewise, we start with a pre-determined set of theoretical shapelets, instead of discovering empirical ones algorithmically, since a wide array of tentative forms can be detected from visual inspection by only looking at the charts of empirical spells. Then, we proceed to lump together the spells of the database into different clusters using a classification algorithm and an adjustment metric to minimize the ambiguity of personal judgments and to refine the number of shapelet-clusters that exhibit clear boundaries. Finally, we apply a multinomial model to check the validity of the classification scheme in terms of the out-of-the-sample predictability of clusters in terms of recession features. An additional advantage in setting theoretical shapelets, instead of dealing with empirical shapelets directly, has to do with the fact that their economic meaning becomes more evident since they are configured as noise-free objects.

The reader needs to have in mind that the shapelets used in the paper describe the output dynamic that goes from the initial output drop to the end of the recession event –which can coincide or not with the pre-recession trend path. This implies that these shapelets can include additional drops, a transition phase, and a recovery phase as indicated in Fig. 3.[5] Because a recovery phase may last many quarters (or even years), we allow in the analysis the inclusion of censored spells in the database. Although this decision leads us to lose some information, we do so to simplify the analytical configuration of the shapelets, as explained in a section below. Since we are not directly interested in predicting the duration of some of these phases (e.g., the time that elapses between the trough and the trend) the omission of this information is not critical for our analysis.

Furthermore, in our analysis for classifying spells, we only consider shapelets with a predefined extension. Doing otherwise would be misleading since two GDP sequences with a similar shape but different sizes do not necessarily convey the same information with respect to the nature of the recession events. Ideally, a shapelet should have a distinctive economic meaning irrespectively of the historical period and time series in which it appears. This argument is diagrammatically exposed in Fig. 4, where two shapelets with similar forms but different extensions are associated with remarkably different features of the corresponding recession events.

When looking at the time window $[t_0, t_1]$ in the top panel and at the time window $[t_0, t_2]$ in the bottom panel one can notice that the output dynamic in both series can be described as U-shaped, hence the same theoretical shapelet would be assigned to the two events if different windows were allowed in the analysis. This is misleading since the top panel exhibits a checkmark shapelet in the time window $[t_0, t_2]$ and

[5] When yearly data is used, the transition phases tend to be short or non-existent. Yet with quarterly data, this phase can be relevant in some empirical spells.

describes a milder recession scenario (i.e., a lower quarterly mean drop and a lower cumulated output loss) in comparison with the one presented in the bottom panel. Therefore, the time window for clustering recession events in our subsequence-based classification algorithm has to be the same. Thus, for the configuration of theoretical shapelets, a decision has to be made about their extension. A larger size can accommodate more shapelet types to look for and fewer events will be censored, but it has the cost of increasing the possibility of mixing the dynamic of two or more shocks in the analysis.

4 Data

For building the empirical spells we use an unbalanced panel of quarterly GDP, expressed as a volume index, for 47 advanced and emerging economies. For few countries, the data available goes back to the '50 s or '60 s of the previous century, yet most of them have information since the '80 s or '90 s. Because the last date analyzed here is from the first quarter of 2020, we are mainly covering four decades of recession events. It is important to be aware that the real GDP indicator has also been seasonally adjusted in the source, and this is helpful for a classification exercise since it avoids unnecessary noise when comparing recession events in different quarters of the year.[6] In Appendix B of the online supplementary material, we include the full list of countries, periods, and recession events covered for this study.

4.1 Definition of a Recession Event

Using this information, we establish a recession event when, after a peak in a GDP series, we observe a cumulative output drop (i.e., a GDP reduction with respect to the pre-recession peak) higher than 5%. A peak occurs in the series if there are two consecutive drops in output (y_t); that is $y_t > y_{t+1}$ and $y_{t+1} > y_{t+2}$.[7] In this sense, we are discarding events with a light (non-significative) recession from the analysis. Following Hausmann et al. (2008), these empirical spells do not overlap; that is, a spell cannot start if the country is already in a recession event. To work with comparable shapelets, all spells need to have the same extension. We opt for a length of 11 quarters (the peak + 10 quarters ahead) because we want to minimize the possibility of including two crises affecting the economy in different quarters of the same recession event. Notice that the length of the empirical spell is different from the duration of a recession, a concept commonly referred to in the literature. Duration is usually understood as the number of periods (quarters or years) that the economy lasts between the first period after the peak (or after the trough) and the period where the GDP returns to its pre-recession level (or to its pre-recession trend).

[6] For details on the data, refer to the OECD website https://doi.org/10.1787/b86d1fc8-en

[7] This is a stricter condition than the Bry-Broschan procedure described in Harding and Pagan (2002).

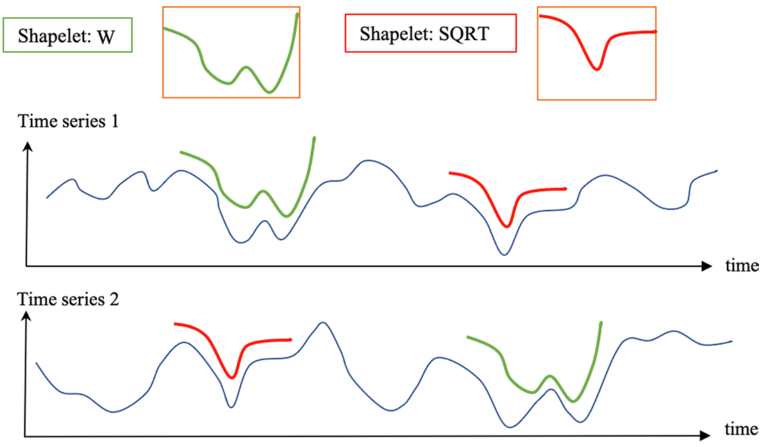

Fig. 2 Identification of empirical shapelets

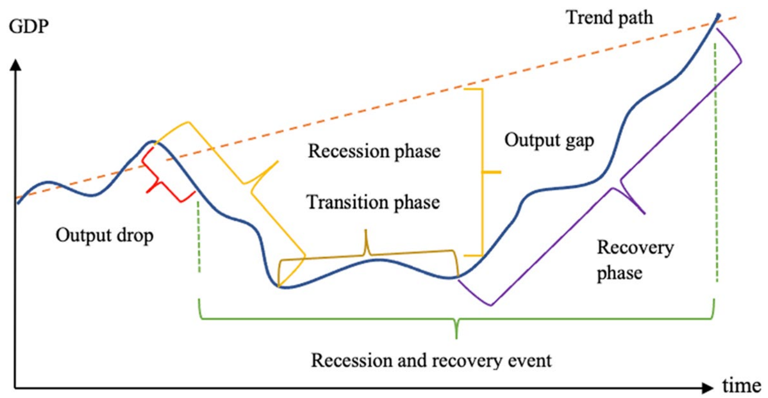

Fig. 3 The phases of recession events

Using this definition, we construct a database that contains a total of 95 empirical spells.[8] Then, for discovering shapelets, we need to make the spells comparable among themselves and, for that, the output indicator should be normalized. Therefore, we rescale (normalize) the data to generate an index of GDP such that its value is equal to 100 for each of the pre-recession peaks. This analytical trick makes it possible to compare recession events not only across time but also across countries,

[8] As a robustness check, instead of a 5% cumulative output drop, we also calculated the events with a threshold of 7% and 10%, resulting in 85 and 73 events, respectively. Table 3 of the Appendix, in the online supplementary material, presents the comparison of the number of events by country. Notice that the alternative thresholds reduce the number of events in several countries as the limit is set higher. With a 10% threshold, SC loses 7 out of 8 events; SP loses 10 out of 21 events; C loses 4 out of 22 events; P loses 1 out of 44 events. This implies that events with a large cumulative output drop are more likely to be associated with prolonged recessions shapes like P.

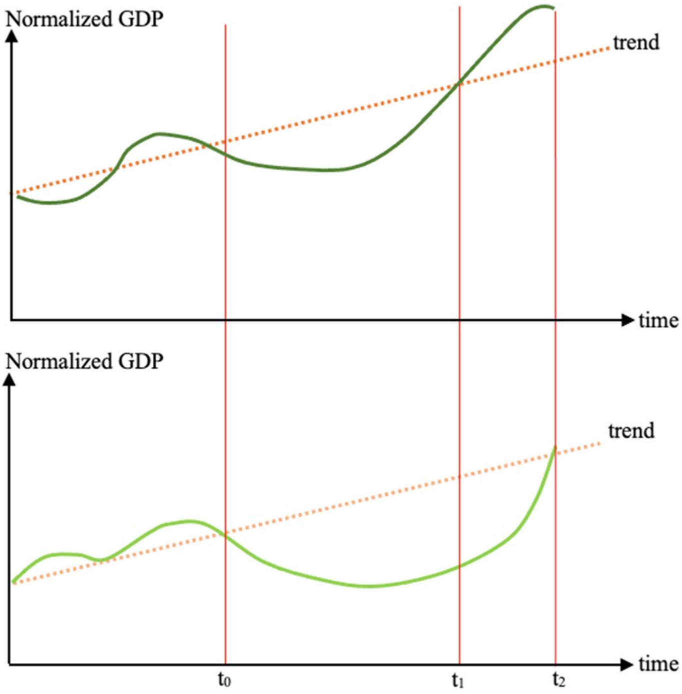

Fig. 4 Shapelets extension and their interpretation

irrespectively of the size of their economies. Note that rescaling the series does not hamper the nature of the output dynamic underlining the subsequences of GDP; hence, we can maintain the wide dispersion of shapes thar are observed in the data when expressed in the original units.

4.2 The Features of Recession Events

Besides studying, directly, the output dynamics of empirical spells, we want to analyze a collection of features characterizing these recession events. For that reason, we need to calculate the difference between the observed GDP and the potential GDP (i.e., the output gap, see Fig. 3) for each spell. Hence, we apply the Hodrick-Prescott (HP) filter to the GDP time series of each country using all the observations before a recession event. This procedure allows us to estimate a pre-recession trend by removing the cyclical component of the time series.[9] Because the estimated trend

[9] It is well known in the empirical literature of business cycles that the HP filter and others can produce an upwardly biased trend when including in its calculation periods close to a peak if the economy is experiencing a boom. This is likely to be the case in financial crises preceded by an extraordinary credit expansion. Therefore, regular practice is to calculate a trend removing several years (or quarters) before the pre-recession peaks (e.g., Blanchard, et al, 2015). Because this alternative method requires long time-series, we decided not to do so in order to maintain the largest possible number of recession events. We think that this is a preferable option since obtaining a precise estimation of the output gap is not the main purpose of the paper.

changes as we use more observations for its calculation, we can have trends that vary in each pre-recession peak along with the GDP series. Once the HP filter has been applied, we calculate the potential GDP for each recession event using the last growth rate observed in the filtered series and, then assume a constant growth rate during all the quarters that compose the empirical spells.[10] Finally, to make the scale of the trend comparable to the GDP transformed series, we rescale the HP filter too using a similar transformation so that the trend is also equal to 100 at the peak.

In the next step, we calculate the quarterly output gap as the distance between the GDP index and the potential GDP –in their rescaled versions– until the period that these two indicators intersect, or the recession event ends if the previous condition does not hold.[11] With the output gap, we can calculate the following features:

a) Duration: time to recover the GDP level that was reached at the pre-recession peak. This means that each spell has a maximum duration of 11 quarters, but it could be lower if GDP recovers sooner.[12]

b) Depth: maximum loss of GDP relative to potential GDP during the recession event.[13]

c) Cumulative output losses: cumulative GDP drop relative to potential GDP during all periods required for the economy to recover within the spell.

d) Statistical measures for describing the output gap: mean, variance, kurtosis, and skewness of the empirical distribution of losses reflected by the GDP gap. Thus, the mean measures the empirical average of the quarterly discrepancies between the observed and potential GDP; the variance estimates how diverse are these gaps during the number of quarters it takes the economy to reach its pre-recession peak (or during 11 quarters when that scenario did not occur during the empirical spell); the kurtosis measures if the quarterly output gaps are relatively uniform in size (i.e., there is a leptokurtic -flatter- empirical distribution), or if there are several gaps ostensibly larger than the remaining ones (i.e., there is a platykurtic -thinner tails- empirical distribution); the skewness measures if the quarterly output gaps are equally divided around the average or a biased towards one side (i.e. if most gaps are relatively small and few tend to be large, or the other way around).

Figure 5 shows an example of how we process the data to calculate the recession events observed in Argentina, which happens to be the country with the most spells in the sample. In the top-left panel, we depict the data from the source (OECD); the dashed lines represent the starting dates of each of the empirical spells for

[10] This approach was adopted since the estimated series for potential GDP with the HP filter may produce spurious cycles and miss nonlinearities (Hamilton, 2018).

[11] When the trend calculated with the HP filter is negative, then we define the output gap as 100—normalized GDP. This situation occurs for only three events: BGR2, HUN2, and IRL2.

[12] We preferred not to calculate the duration as the number of periods up to reaching the pre-recession trend since in a methodology permitting censored spells, like ours, this definition would produce many events with a duration of 11 quarters; consequently, the duration would be a meaningless concept since it could not be used as a feature capable of discriminating between recession events.

[13] An alternative definition would be the output drop between the peak and the trough; in general, these two definitions produce the same result, although, in some empirical settings they could give different values.

that country. In the top-right panel, we establish the empirical shapelets using the rescaled GDP for each recession event, so that these shapelets can be comparable across countries and periods. These empirical shapelets have to be identified, using the machine learning algorithm, with calibrated shapelets derived from the theoretical shapes of recovery.

Then, the lower-left panel shows the observed GDP and the potential GDP calculated with the HP filter; each colored line corresponds to the HP filter forecast for each one of the events. Notice that the forecasted trend along the GDP series, using the HP filter, changes over time, and thus, there is one trend for each peak detected in the series. The lower-right panel shows the distance between GDP and potential GDP, with the same color as the HP forecast, for each empirical spell. Note that the output gap is only calculated up to the quarter where the economy recovers, which can be up to 11 quarters when the pre-recession peak is not reached. Each line in this plot represents the spell with the same color as in the lower-left panel. In the Argentinian case, there are six recession events in the time range covered in the OECD database; those events are depicted in the top-right panel with their corresponding shapelet. Likewise, the colored curves in the lower-right panel define some sort of "empirical distribution" that we use to calculate additional features of the recession events (mean, variance, kurtosis, skewness).

To the naked eye, we detect in the top-right panel three types of shapelets: two checkmark shaped events (1994-IV and 2008-III), two V-shaped events (2001-I and 2015-II), and two W-shaped events (1998-II and 2011-IV), although the shape of the 2011-V spell could be easily confused with an SQRT-shaped event. However, preconceptions and noise can be misleading and produce erroneous inferences. Indeed, as it will be seen below, these preconceptions resulted to be wrong for several reasons: the presence of noise can make an event be identified with an incorrect letter (e.g., W in event 1998-II); one theoretical shapelet might not be distinguished from another (e.g., V in event 2015-II) in a statistical sense; the economic content behind an empirical shapelet might make more convenient to classify it in a different cluster (e.g., W in event 2011-IV). In fact, in our final classification scheme, the Argentinian recession events are identified with three theoretical shapelets but with groupings different from those produced by our subjective judgment.

5 Machine Learning Methods

In this section, we present in detail our customized machine learning (ML) algorithm for detecting patterns in short subsequences of GDP describing the recession spells identified in the data.[14] The procedure has five main steps, two for the definition of

[14] After the seminal papers, the literature of shapelets opted for separating the problem of extraction from the problem of classification. This decision helped to reduce computer time in a significant manner. Some of these methods are 'Shapelet Transform' (ST) developed in Davis et al (2011) and Hills et al (2014); 'Learning time-series shapelets' (LTS) developed in Grabocka (2014); 'Genetic Discovery of Shapelets' (GINDIS) developed in Vandewiele et al. (2021); 'Localized Random Shapelets' (LRS) developed in Mael (2020). None of these methods is convenient to the problem at hand since we do not need to explore all the subsequences of the GDP series. In the 1-NS (one-nearest-shapelet) procedure described below, the extraction problem is trivial since the empirical shapelets are identified once the

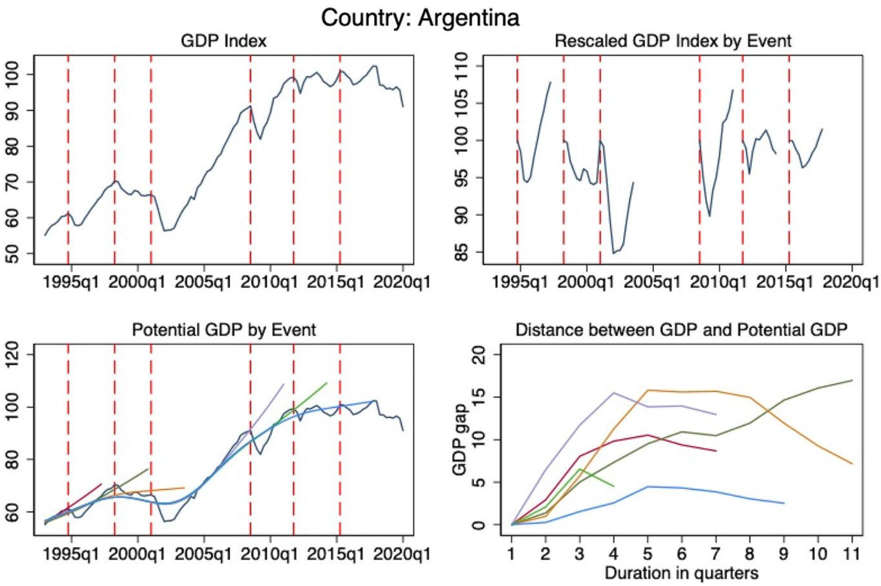

Fig. 5 Recession events, empirical shapelets, and spell's features. *Note:* Vertical dashed lines indicate the starting quarter of a recession event, the six curvilinear shapes in the upper-right panel are the empirical shapelets, the colored straight lines in the lower-left panel are our estimations of potential GDP during the period that each of these events last, and the colored curves in the lower-right panel are our estimations of the quarterly GDP gaps for each recession event

the theoretical and calibrated shapelets, one for the clustering algorithm, and two more for selecting and validating the number of clusters. (i) A comprehensive set of shapelets is proposed (e.g., 10 forms) from a visual inspection of the spells; (ii) each shapelet is calibrated with the initial output drop of the event and its average quarterly change rate in absolute terms across all periods of the sequence; (iii) each spell is compared with the calibrated shapelets and, then, the closest one is selected, this is done with a nearest-neighbor approach and the Euclidean distance metric; (iv) once each spell is associated with a shapelet, the set of clusters is refined by selecting the number of shapelets configurations so that the borders between clusters are not blurred (as described with more precision below); (v) a multinomial model is estimated with the tentative clusters, as the categorical variable, and the recession features as explanatory variables, in order to check the separability of the chosen classification in terms of its capability to produce clusters with economic meaning.

The one-nearest-shapelet algorithm (1-NS) used in step (iii) can be considered a quasi-supervised variant since the spells are expected to have patterns like the proposed theoretical shapelets, despite that the algorithm does not operate with a

Footnote 14 (continued)

recession events have been established with a remarkably simple procedure. Accordingly, our method tackles mainly the classification and depuration problems.

database whose objects (items) have a fully specified identity. In full-fledged supervised algorithms, the objects composing the training and testing datasets present features and a one-to-one classification identity. In contrast, in unsupervised algorithms, such as I-NN and KMeans, no identity is determined *ex-ante* for the objects in the database to be clustered. The advantages of 1-NS are threefold: it is a clustering device whose objects are not described by features but by subsequences of observations like in the recession events; it works well when these subsequences are composed by a small number of observations (i.e., the number of quarters of a recession event); as it will be shown below, these subsequences describe shapes in the output dynamic that have economic meaning and describe the severity of recessions produced by shocks to the economy and policy interventions. The latter point is extremely important since traditional ML algorithms hardly offer analytical reasons why certain objects belong to specific clusters.

5.1 Shapelets Configurations

For the patterns of GDP dynamics, in terms of letters or other distinctive symbols, to be analytically tractable, it is convenient to specify theoretical shapelets with precise configurations. Therefore, in Table 1, we describe three stages of different duration for characterizing possible shapes of recession spells. These heterogeneous stages can be composed of quarters of steady drops (downs), quarters of transition (repose or erratic ups and downs), or quarters of steady surge (ups) so that each of these profiles gives form to a specific shapelet. Notice that our definitions of shapelets do not necessarily coincide with the descriptions presented in macroeconomic reports using butterfly charts.[15] Moreover, instead of referring to a U-shaped shapelet, we consider two subcategories: a 'bowl' (long U) and a 'cup' (short U).

With these shapes, we attempt to identify particular dynamics of GDP in recession events defined with time windows of ten quarters. For instance, a 'short checkmark' can be associated with a brief recession phase; a 'slash' with a deep and prolonged recession; a W-shaped event with an erratic transient phase; a 'long checkmark' with a deep recession that bounces-back at the end of the spell; a 'cup' and a 'bowl' with events whose recovery phase is symmetrical to the recession phase but exhibiting different transition phases.[16] According to

[15] In some of these reports (see, for instance, Morley, 2009), the letters' shapes are established having the pre-recession trend as an axis, while in our definition the shapes of the letters are established with respect to the horizontal axis. Because of this, our 'short checkmark' does not correspond to the traditional L-shaped dynamic where the economy recovers its growth rate but moves across a lower-level trend. The definitions presented here are preferred since for 10 quarters it is common to find empirical spells with transition phases reflecting a stagnated economy and others more with no recovery phase whatsoever.

[16] Notice in Table 1 that the letter L –in our definition– is not included in the initial set of shapelets because, after the visual inspection, we opted for a symbol that looks like a pan's profile. While the letter U is substituted by the profile of a cup exhibiting a flatter transition stage. In addition, the square root is not considered relevant since it is difficult to differentiate it from the letter W in empirical spells given the reduced number of quarters composing a shapelet. Moreover, in two of the shapelets (SL and S), there is no upward trend, this is so because the data reflect some recession spells that do not exhibit a recovery phase during the 10 quarters following the peak.

the data, 45 out of 95 empirical spells recover the GDP level observed in their pre-recession peak, and 44 reached their trough before the 11^{th} quarter. These numbers indicate that in this time window it is possible to find shapelets that exhibit a recovery phase, either short or long and others where the recession phase prevails in all or most of the quarters.

In Table 2, we present binary representations of the shapelets' configurations with vectors of 20 bits describing 10-quarters events. Because each pair of bits corresponds to one quarter, the combination (1, 0) indicates a drop in the quarter, a (0, 0) describes the absence of change, and (0, 1) reflects a surge. These binary vectors help to formulate mathematically the value that a theoretical shapelet attains in each quarter, as indicated in the following expression:

$$S_t = S_{t-1} + S_{t-1}\left(-\tau \cdot \alpha_{t-1} + \tau \cdot \beta_{t-1}\right) \tag{1}$$

where the parameters (α_t, β_t) correspond to the binary pair describing whether in quarter t of the shapelet an 'up', 'down', or 'none' is observed; τ is the average rate of change that determines the size of each step between periods; $S(0) = 100$ is the normalized GDP in the pre-recession peak; $S(1)$ is the initial output drop observed in the empirical spell to be analyzed.

Because the sequence of observations in a recession event has to be compared with the different theoretical shapelets, through a Euclidean distance metric in the 1-NS algorithm, it is necessary to calibrate the parameters τ and $S(1)$ using the empirical data of the corresponding event. Therefore, the calibrated shapelet can have different output drops and step sizes depending on the spell to be analyzed. With the binary selection of α and β, it is theoretically possible to create a large number of configurations, however many of them would only reflect random noise, and others will not be different enough from those already presented in Table 1. This is the case of a 'square-root' that tends to resemble a W-shaped event when the sequence is limited to only 10 periods after the peak. Therefore, we limit the number of shapelets to be analyzed to a number equal or lower than the number of periods composing the empirical spells.

As an example, we present two diagrams in Fig. 6 describing with blue dots the GDP subsequence of a recession event in Ireland that started in the 2007-IV quarter. In the left panel, the empirical spell is overlapped with the calibrated shapelet of type SP in which the red stars correspond to its quarter values. Notice that their first two observations, corresponding to the peak and the initial output drop, coincide because of the calibration procedure. Therefore, with these shapelets, we attempt to explore the output dynamic from the initial output drop onwards, which includes the GDP remaining fall until reaching the trough as well as the transition and recovery phases (see Fig. 2) taking place within the 10 quarters after the peak. In the right panel, the calibrated shapelet is of type LC which exhibits a much better fit than SP. Since this is the best fit when compared with the other eight shapelets, this event is classified in the 1-NS algorithm as a member of the LC-cluster.

Table 1 Theoretical shapelets' profiles using sequences of ups and downs

ID	Name	Shape	1st Stage: (downs)	2nd Stage: transition	3rd Stage (ups)
1	Pan: P		5 quarters	3 flat quarters	2 quarters
2	Letter: V		5 quarters		5 quarters
3	Bowl: B		3 quarters	4 flat quarters	3 quarters
4	Letter: W		2 quarters	3 ups & 3 downs	2 quarters
5	Slash: SL		10 quarters		
6	Short-check: SC		2 quarters		8 quarters
7	Long-check: LC		8 quarters		2 quarters
8	Stair: S		4 quarters	3 flat quarters	3 quarters (downs)
9	Cup: C		4 quarters	2 flat quarters	4 quarters
10	Spoon: SP		2 quarters	2 flat quarters	6 quarters

Table 2 Shapelets' binary configurations

ID/periods	1	2	3	4	5	6	7	8	9	10
P	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(0, 0)	(0.0)	(0, 0)	(0, 1)	(0, 1)
V	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(0, 1)	(0.1)	(0, 1)	(0, 1)	(0, 1)
B	(1, 0)	(1, 0)	(1, 0)	(0, 0)	(0.0)	(0, 0)	(0, 0)	(0, 1)	(0, 1)	(0, 1)
W	(1, 0)	(1, 0)	(0, 1)	(0, 1)	(0, 1)	(1, 0)	(1, 0)	(1, 0)	(0, 1)	(0, 1)
SL	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)
SC	(1, 0)	(1, 0)	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)
LC	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(0, 1)	(0, 1)
S	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(0, 0)	(0, 0)	(0, 0)	(1, 0)	(1, 0)	(1, 0)
C	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(0, 0)	(0, 0)	(0, 1)	(0, 1)	(0, 1)	(0, 1)
SP	(1, 0)	(1, 0)	(0, 0)	(0, 0)	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)

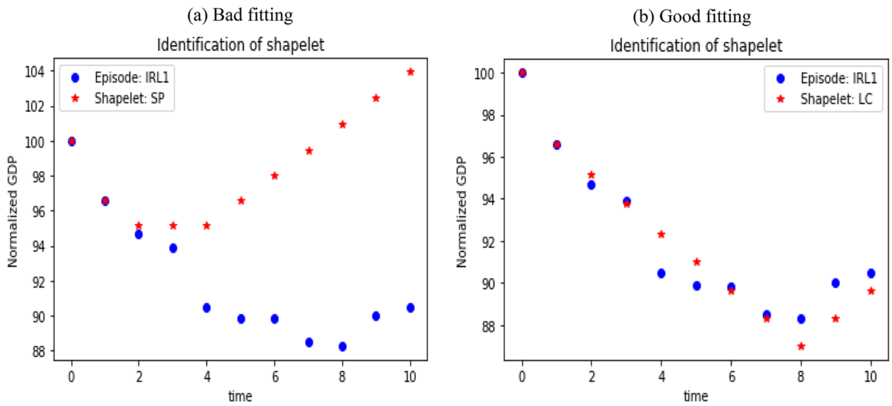

Fig. 6 Comparison between two calibrated shapelets and one empirical spell

5.2 Adjustment Scores and Depuration of Shapelets

For complementing the 1-NS algorithm, we analyze whether the initial set of configurations generates clusters with diffuse borders. With that aim, we compare the fit of the empirical spells that integrate a specific shapelet-cluster with the fit of the same spells with their second-nearest-shapelet; that is to say, with the adjustment attained if these spells were grouped in their closest clusters. With this scheme, we calculate a Relative Adjustment Score (RAS_j) for each shapelet-cluster, so that a negative value means that the empirical spells are on average wrongly positioned in such cluster, while a positive but close to zero value indicates that this cluster has blurred borders.[17] In the latter scenario, we remove the shapelet representative of the low-quality cluster from the set of tentative candidates. We proceed to remove shapelets in a piecewise fashion until all cluster scores are higher than a certain threshold.

$$RAS_i = \frac{DI_i - SI_i}{max\{SI_i, DI_i\}} \tag{2.a}$$

$$DI_i = \frac{\sum_{\varepsilon_k C_i} \min_{j \neq i} \{d(\varepsilon_k, S_j)\}}{|C_i|} \tag{2.b}$$

$$SI_i = \frac{\sum_{\varepsilon_k C_i} d(\varepsilon_k, S_i)}{|C_i|} \tag{2.c}$$

where $d(\varepsilon_k, S_i)$ indicates the Euclidean distance between the empirical spell ε_k and the shapelet of the *i-th* cluster (C_i) in which it was initially classified; $|C_i|$ is the size of cluster i; SI_i is a similarity index for the spells classified in cluster i so that a low index value means that, in average, the fit to this theoretical shapelet is strong; DI_i is a discrepancy index that measures, in average, how far the empirical spells grouped in cluster i are to their second-nearest shapelet representing other clusters. The division of the difference of these two indexes by $max\{SI_i, DI_i\}$ guarantees that the relative adjustment is bounded in the interval [− 1, 1].[18]

6 Fitting Theoretical Shapelets and Classifying Empirical Spells

In Fig. 7, we show with nine recession events –one for each theoretical shapelet (for LC see Fig. 6b)– the capacity of the 1-NS algorithm to fit the selected calibrated shapelets to the empirical spells. The method works relatively well, despite the existence of a few noisy spells and that it does not attempt to capture, directly, the different features characterizing recession events. However, as it will be shown below, the shapes of these shapelets do, indirectly, capture said features. Notice that the customized procedure can distinguish if a recession (or a recovery) phase is short or long, as well as the presence of an erratic or smooth transition phase. Nonetheless,

[17] The RAS is a variant of the Silhouette Coefficient used to calculate the goodness of fit in a clustering method (Kaufman and Rousseeuw 1990, p. 87). While in this metric two values are calculated (mean intra-cluster distance and the mean nearest-cluster distance) to get an overall score of adjustment, the RAS metric generates a score for each shapelet to identify which clusters present blurred borders. Moreover, instead of calculating an average distance between points (inside and outside clusters), RAS calculates the average distance of the events within a cluster with respect to calibrated shapelets (nearest and second-nearest).

[18] By construction, when the 1-NS algorithm is applied, there cannot be negative RAS since the nearest neighbor is always closer than the second-nearest neighbor; however, this metric can produce values near zero.

the application of the 1-NS algorithm jointly with the Relative Adjustment Scores shows that the presence of noise in some empirical spells can produce clusters with blurred borders.

Next, we apply an iterative depuration process, in which we gradually remove shapelet-clusters when according to the RAS we have evidence that some of these shapes do not provide distinctive borders between shapelets. In the first row of Table 3, we present the Adjustment Score for each shapelet-cluster included in the initial set of configurations, composed of 10 candidates. Then, in the following two rows, we exhibit the numerical calculations for the RAS when said set has 7 and 4 clusters. Notice that, in the first iteration, there are three Adjustment Scores close to zero (i.e., below 0.15) in the 10-clusters configuration (set 1), which indicates that the corresponding clusters are not distinctive. Because the letter C has a shape between letters V and B (see Table 1), we decided to keep it in the new configuration.

Then, all the clusters in the 7-clusters configuration (set 2) have a metric value above 0.15, including the C-cluster with a RAS of 0.438. However, not all of them are above the 0.4 threshold that we established as a cut-off point, which we consider to be high enough to obtain well-defined borders. Hence, we decided to make an additional cut by removing the theoretical shapelet below such threshold.[19] The final configuration presents 4 clusters, each with a score above the 0.4. The ML literature does not specify which score is high enough to declare a good fitting; hence, this threshold is somehow subjective. However. in Sect. 8, we show statistically that the last classification scheme is capable of producing a good separation between clusters, in so far as the predictability error of classifying clusters with recession features is pretty low when using this scheme.

For comparison purposes, we present in Table 4 the classification of the different empirical spells (country-period cases) into the clusters of the seven shapelets configuration set. The first thing to notice is that the smallest cluster is associated with the letter W and contains only six members, while the largest cluster is associated with P and has 22 members. The second thing to highlight is that the same country can be positioned in different clusters. For instance, Argentina presents episodes in five clusters: P, LC, W, C, and SP which indicates that within a country the output dynamics of recession events can be quite heterogenous, in contradiction to the assumption of Markov-switching models.[20] The third result to emphasize is that all clusters present countries in the two stages of economic development considered in the sample: emerging economies (EE) and advanced economies (AE).

However, as indicated by our cluster depuration procedure, a classification scheme with only four theoretical shapelets offers a better separation between clusters. Although, this scheme has the cost of producing less homogeneity in some

[19] In fact, we also analyzed an intermediate configuration in which only LC was removed for having the lowest RAS. The result of taking this step was that the spells previously classified in this shapelet-cluster were now grouped in the P cluster; likewise, the RAS values for the W and SL clusters remained low.

[20] A similar outcome occurs with South Africa that presents four recession events in the historical sample used here.

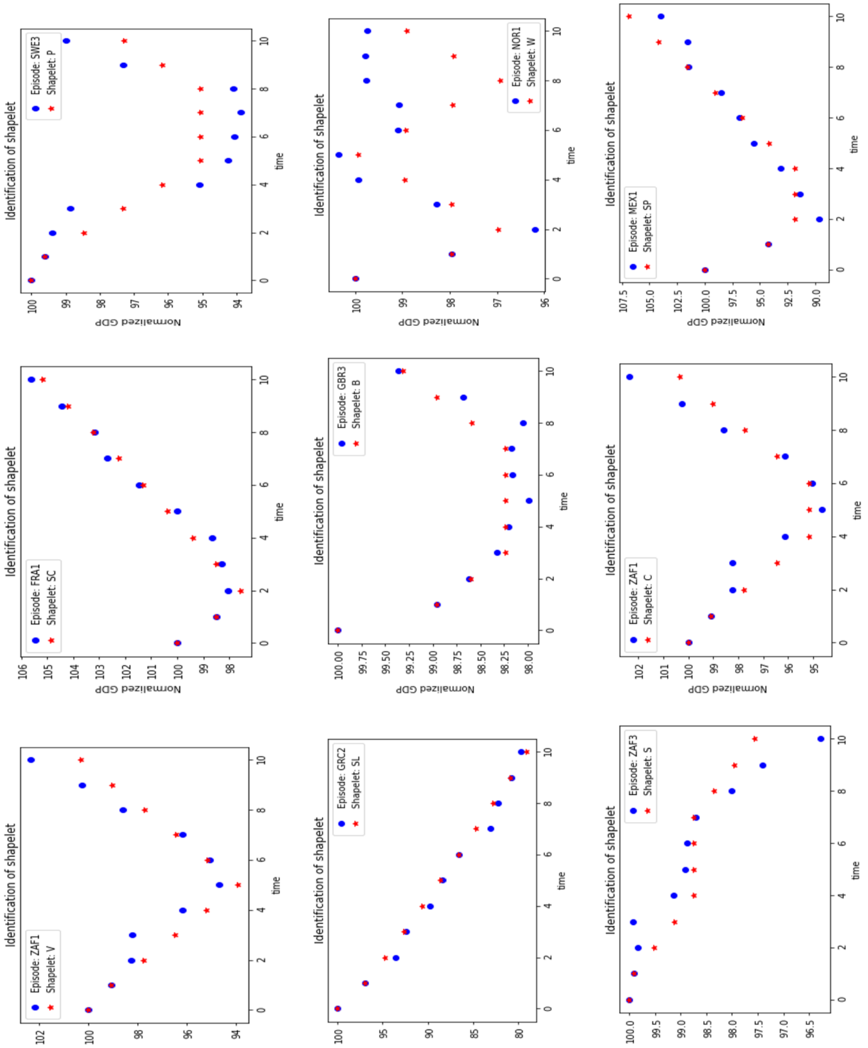

Fig. 7 Examples of calibrated shapelets fitted to different empirical spells

Table 3 RAS for each cluster in three different configuration sets

Clusters:	P	W	SL	SC	LC	C	SP	B	V	S
RAS in set 1	0.298	0.301	0.267	0.502	0.187	0.024	0.570	0.135	0.060	0.171
RAS in set 2	0.335	0.288	0.263	0.502	0.182	0.438	0.569			
RAS in set 3	0.404			0.502		0.440	0.519			

Table 4 Clustering of recession events in a classification scheme with 7 shapelets

Cluster:	SP	LC	P	W	C	SC	SL
Events:	**ARG1**	**ARG2**	**ARG3**	**ARG5**	**ARG6**	**BRA1**	*ESP1*
	ARG4	**BRA2**	*AUT1*	**BGR3**	*AUS1*	**CRI1**	*ESP2*
	CHE3	*DNK1*	**BGR1**	*IRL2*	*BEL1*	*FRA1*	*EST2*
	CHL1	*FIN1*	*CAN1*	*NOR1*	**BGR2**	*LUX1*	*GRC2*
	EST1	*FIN2*	*CHE2*	*NOR3*	*CAN2*	*NZL2*	*ITA3*
	HUN2	*GBR4*	*CZE1*	*NZL1*	*CHE1*	**POL1**	*LVA1*
	JPN1	**HUN1**	*DEU1*		*CZE2*	*USA2*	*PRT3*
	JPN2	*IRL1*	*FRA2*		*CZE3*	*USA3*	**ROU1**
	KOR1	*ITA2*	*GBR1*		*GBR3*		**ZAF3**
	KOR2	*LTU2*	*GBR2*		*ITA1*		
	LTU1	*SVN1*	*GRC1*		**MEX2**		
	MEX1	*SVN2*	**IDN1**		*NLD2*		
	NOR2	*SWE2*	*ISL1*		*NZL3*		
	SAU1		*ISR1*		*PRT1*		
	TUR1		*JPN3*		*SVK1*		
	USA1		*LUX2*		*SWE1*		
	ZAF4		NLD1		**TUR2**		
			PRT2		*USA4*		
			ROU2		**ZAF1**		
			RUS1		**ZAF2**		
			RUS2				
			SWE3				
N. spells	17	13	22	6	20	8	9

Note: Categories of countries considered, EE in bold and AE in italic

clusters. In Table 5, we present the spells' membership using this criterion of classification. Notice that the P-cluster also has the largest membership and that its new members come from the LC- and SL-clusters, in the 7-clusters scheme, which have 13 and 9 spells, respectively. This outcome makes sense since these three shapelets were configured with a long recession-transition stage (see Table 1). In contrast, the W-shaped spells in the 7-clusters scheme are clustered now in the C- and SP-clusters because the three types coincide with a short recession phase. Finally, the SC-cluster did not capture any new member since its recovery phase is too large and the reclassified spells do not exhibit this attribute.

7 The Features of Recessions Events

Next in Table 6, we present some descriptive statistics of recession features that are helpful for the economic interpretation of the output dynamic prevailing in each cluster. When reviewing these calculations, we notice that our grouping method can discriminate the empirical spells in two main categories: mild and severe recessions, measured in terms of the following features: duration, depth, cumulative loss, mean, and variance. Therefore, the first five rows in Table 7 depict variables with an increasing ranking of the average severeness along the columns. That is to say, the first four shapelet-clusters (columns) are classified in the subset of mild recession spells (SC, W, C, SP), while the lasts three lie in the subset of severe recession spells (LC, SL, P). According to the information provided in Table 7, the division between these two subsets is not established with an arbitrary split in the ranking positions but based on significant quantitative differences in the border values of these metrics. Moreover, within each subset, there is also a pattern in the ranking position of the shapelet-clusters. For instance, the SC-cluster is always at the bottom (i.e., with the lowest value) of the 5 features characterizing the mild recession category, while LC-cluster has the lowest values in the shapelets observing severe recessions.

Another distinctive outcome is that the severity of the recession events is also, on average, associated with the shapes of the output gap distributions, as indicated by the metrics of kurtosis and skewness. Note that the flatter distributions are related to spells with mild recessions,[21] and while both distributions are asymmetrical (i.e., tend to have frequent gaps below its mean value) the small gaps are less common in severe recessions. We want to emphasize that the strong and positive relationships between features at the cluster average levels also exist at the empirical spell level for the first five features, as shown in Appendix A. However, kurtosis and skewness present a strong negative relationship at the spell level, but their association with the other features seems to be random.[22]

With the statistics of Table 8, using the 4-clusters classification, we confirm the division of spells into two categories depending on the severity of the recession. However, in this setting, the severe recession category is composed only of the P-cluster. In this cluster we observe long recession phases that also show a deep impact, high cumulative losses, and high quarterly deviations from the trend in average. Moreover, the distribution of the output gaps exhibits a large variance, which indicates that during these episodes there is a large disparity in quarterly

[21] When the kurtosis is lower than 3, it means the distribution of the output gaps has lighter tails than the normal distribution; in other words, the kurtosis decreases as the tails become lighter (i.e., less salient).

[22] In the same Appendix, we show that the KMeans clustering mechanism applied to the events' features produces two large clusters in which one of them presents only severe recessions, but all sorts of spells are combined in the second cluster; making the formation of these clusters difficult to interpret. This outcome improves when a manifold learning technique is used to contract the features in a 2-dimensional space. With this transformation we obtain, again, two large clusters, however, each of them is strikingly aligned with the two categories presented in Table 8. This result validates the good performance of the 1-NS algorithm, with the added benefit that it offers a ranking of clusters within each category and a straightforward economic interpretation of the GDP dynamics during recession events.

Table 5 Clustering of recession events in a classification scheme with 4 shapelets

Cluster:	SC	C	SP	P
Events:	BRA1, CRI1, FRA1, LUX1, NZL2, POL1, USA2, USA3	ARG6, AUS1, BEL1, BGR2, BGR3, CAN2, CHE1, CZE2, CZE3, GBR3, ITA1, MEX2, NLD2, NOR3, NZL3, PRT1, SVK1, SWE1, TUR2, USA4, ZAF1, ZAF2	ARG1, ARG4, ARG5, CHE3, CHL1, EST1, HUN2, IRL2, JPN1, JPN2, KOR1, KOR2, LTU1, MEX1, NOR1, NOR2, NZL1, SAU1, TUR1, USA1, ZAF4	ARG2, ARG3, AUT1, BGR1, BRA2, CAN1, CHE2, CZE1, DEU1, DNK1, ESP1, ESP2, EST2, FIN1, FIN2, FRA2, GBR1, GBR2, GBR4, GRC1, GRC2, HUN1, IDN1, IRL1, ISL1, ISR1, ITA2, ITA3, JPN3, LTU2, LUX2, LVA1, NLD1, PRT2, PRT3, ROU1, ROU2, RUS1, RUS2, SVN1, SVN2, SWE2, SWE3, ZAF3
N. spells	8	22	21	44

Table 6 Features' average values by type of cluster with 7 shapelets

Feature / Shapelet	Duration	Depth	Cumulative Loss	Mean	Variance	Kurtosis	Skewness
SC	5.250	6.038	19.055	3.713	6.139	2.053	– 0.611
W	7.333	6.886	35.359	4.235	7.977	2.594	– 0.546
C	9.500	7.998	49.679	5.042	9.169	2.437	– 0.652
SP	7.412	9.124	42.658	5.788	14.125	2.388	– 0.780
LC	11.000	13.995	95.707	8.701	33.033	2.006	– 0.574
P	11.000	16.460	106.228	9.657	56.836	1.953	– 0.527
SL	11.000	17.605	105.923	9.629	65.160	1.772	– 0.251

Table 7 Rankings of clusters' types according to their average features' values

Feature:	Category: Mild recession spells				Category: Severe recession spells		
Duration	SC	W	SP	C	LC=	P=	SL=
Depth	SC	W	C	SP	LC	P	SL
Cumulative Loss	SC	W	SP	C	LC	SL	P
Mean	SC	W	C	SP	LC	SL	P
Variance	SC	W	C	SP	LC	P	SL
	Flatter and biased toward small gaps				Thinner tails and less asymmetrical		
Kurtosis	W	C	SP	SC	LC	P	SL
Skewness	SP	C	SC	LC	W	P	SL

Note: the sign = indicates that the corresponding values for this feature are identical in the three clusters

losses. Again, the SC-cluster is at the bottom of the mild recession clusters for the first five features, while the C- and SP-clusters alternate their ranking in the mild category depending on which features are considered. The C-cluster exhibits empirical spells with longer recession phases and larger cumulative losses on average –when considering the output that could have been achieved if the economy had continued moving along the trend.

In Fig. 8, we show sets of stacked histograms for two of the recession's features (depth and mean) where the unit of observation is the empirical spell. Notice that, in both features, SC-shaped spells (red color) tend to have values that are similar or below those of C-shaped spells (green color), while the latter spells' values tend to be similar or below those of the P-shaped spells (orange color). When comparing the SP-shaped spells with the P-shaped spells, we can observe that the values of the former accumulate around its average while for the latter there is a larger dispersion with several extreme values. In brief, these plots make clear that, indeed, there is a distinctive pattern in the shapes of the recession events associated with their severity but also that it is possible to find some spells that do not conform to the norm (e.g., few P (orange) spells exhibiting mild recessions).

8 A multinomial Procedure for Testing the Separability of Shapelets

The previous results indicate that the shapelets, either included in classification schemes with 4 or 7 clusters, are strongly associated with the features of recession events. On the one hand, this provides economic meaning to the different shapelet-clusters and, in particular, permits to determine if the GDP subsequences described with these shapelets can be categorized as cases of mild or severe recessions. On the other hand, this strong relationship can be helpful to test the separability of the different classification schemes using a multinomial model, in which the categorical variable is composed of the different clusters and the explanatory variables are the recession features.

When the database with the 95 spells is divided into a training-subset and a testing-subset, as traditionally done in the ML literature, we can analyze the out-of-the-sample predictability of cluster through recession features. The idea is to check whether the model has the capability of predicting the cluster where a spell belongs by only looking at its features. Then, when the errors of predictability are relatively few, we can argue that the 1-NS algorithm plus the cluster depuration procedure have been able to produce meaningful clusters; that is to say, clusters well separated that do not present blurred borders. On the contrary, if the number of errors is substantial, we can argue that the classification scheme under consideration is not robust since many spells are classified in the wrong clusters when attending to the features observed in the recession events.

For this exercise, we consider 30% of the events in the testing-subset, that is, events that were not considered in the estimation stage. By splitting the database in this manner, we can produce out-of-the-sample predictions which allow having a powerful test. Because we have a limited number of observations and many clusters in the categorical variable, we need to reduce the number of explanatory variables to just three (cumulative loss, mean, and kurtosis), which were selected using the Akaike Information Criterion (i.e., a low AIC). With these explanatory variables and the 7-clusters classification scheme, we obtain a pseudo-R-square of only 0.43.

The predictability of the multinomial model with 7-clusters is somehow reduced since the Accuracy Score generates a low number (0.586). Moreover, Table 9 illustrates its deficiencies and, thus, the inadequacy of the proposed classification scheme. For instance, the confusion matrix indicates that there are 12 errors out of 29 events (good predictions correspond to the diagonal numbers in bold), and that the W and LC shapelets have serious shortcomings. This matrix is highly informative since it shows that the two W-shaped events were classified wrongly in SC- and SP-clusters, the 6 LC-shaped events in the P-clusters, and the single SL-shaped event in the P-cluster. In other words, with the off-diagonal numbers, we can discover the same kind of theoretical shapelets that the Relative Adjustment Scores suggested to remove previously (see Table 3).

Now, when the multinomial model is estimated for the 4-clusters classification scheme, using the same percentage of events in the testing-subset and the same set of explanatory variables, we get a higher pseudo-R-square: 0.734. Furthermore, on this occasion, the models' features produce a good out-of-the-sample classification

Table 8 Features' average values by type of cluster with 4 shapelets

Feature / Shapelet	Duration	Depth	Cumulative Loss	Mean	Variance	Kurtosis	Skewness
SC	5.250	6.038	19.055	3.713	6.139	2.053	– 0.611
C	9.364	8.196	51.140	5.180	9.675	2.419	– 0.657
SP	7.333	8.384	38.374	5.271	12.311	2.461	– 0.720
P	11.00	15.966	103.057	9.369	51.506	1.931	– 0.484

mark with an Accuracy Score of 0.86.[23] This generates a confusion matrix with just 4 errors out of the 29 events in the testing-subset. According to the results in Table 10, the predictability precision for shapelets P and SP is 100%. In brief, the multinomial separability test confirms the 4-clusters classification scheme as the most appropriate configuration set for grouping recessions.

9 Was the Recent Global Financial Crisis Different?

Another type of pattern that is important to study, from an economic point of view, has to do with the association between recession variants and the nature of the underlying crises. In this paper, we are mainly interested in the global financial crisis that hit the world economy in 2008–2009, although we also analyze three other types of crises: banking, currency, or fiscal. With the 4 and 7 clusters classification schemes, we explore whether a certain crisis is more likely to exhibit a specific output dynamic and, indirectly, if such crisis generates recession events with specific features. Readers should be aware that the statistical comparisons presented below do not control for other relevant variables, such as the policy interventions that are usually implemented when a crisis shocks the economy.

Firstly, we check if the empirical spells coincide with at least one of the four types of crises mentioned above; otherwise, we declare these events as unassigned. Hence, some spells in the database could have experienced more than one type of crisis at the beginning of the recession phase. Secondly, we make visual comparisons with overlapped histograms to assess whether some theoretical shapelets are more salient in the global crisis (or in any of the other three) vis-à-vis other recession categories. Thirdly, we apply the Mann–Whitney U test (Hollander et al., 2014) to explore statistically if pairs of recession categories tend to produce the same type of shapelet-clusters (i.e., if the two independent samples come from the same population). This popular non-parametric test is convenient because we do not have a guarantee that these samples come from a normal distribution and the number of observations is relatively small. Finally, we calculate odds ratios with the historical

[23] Moreover, when we increase the number explanatory variables to five (depth, cumulative loss, mean, kurtosis, and skewness), the Pseudo-R-square increases to 0.81 and the Accuracy Score to 0.9.

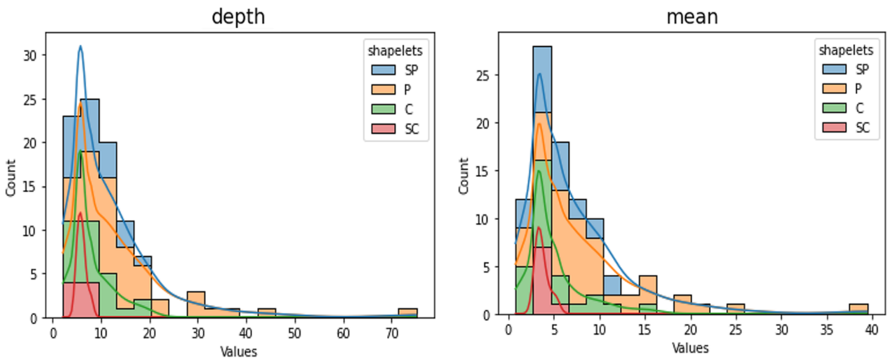

Fig. 8 Frequencies of features' values at the spell level

frequencies to identify which shapelets are the most and the least likely to appear in each type of crisis.

In Fig. 9, we present histogram plots with four pair-wise comparisons, in which the most distinctive implication is that in three types of crises (global, banking, and currency) the P-shaped spells are much more common (see blue bars) when comparing with the reference group (i.e., the other variants of recessions); that is, events with severe recessions are more likely to appear. In contrast, in fiscal crises (lower-right panel) the SP-shaped events increase considerably their frequency, which means that spells with mild recessions and short recession phases are empirically more relevant.

Other characteristics in these four sets of overlapped histograms can be highlighted. For instance, the scarce presence of spells associated with mild recessions, in any of their forms, is evident in the global, banking, and currency crises; inclusively, in the latter two there are no SC-shaped spells at all. In fiscal crises, the presence of P-shaped spells is also highly relevant, but they are not much more common than SP-shaped spells. Accordingly, the global financial crisis of 2008–2009, in advanced and emerging economies, is different from most of the crises observed in the last 40 decades when comparing their corresponding recession events in terms of their severity. As opposed to the banking and currency crises, which also tended to generate severe recessions, the global crisis produced relatively more mild recessions and even some with a quick recovery.

In order to apply the Mann–Whitney U test, we need to express the two samples (crisis and other recessions) in ordinal values since this non-parametric test works with ranked values. For doing this, we use the rankings of Table 7 establishing the severity of the recession events according to the following features: duration, depth, cumulative loss, mean, and variance. The ranking is uncontroversial for three of the shapelets (SC, W, LC) but not for the other four since their position change slightly depending on the feature considered. Thus, in a first approximation, we use a majority rule to establish the ranking (SC, W, C, SP, LC, P, SL) and assign a natural

Table 9 Confusion matrix with a 7-clusters classification

	Pred-SC	Pred-W	Pred-C	Pred-SP	Pred-LC	Pred-P	Pred-SL
SC	**3**	0	0	1	0	0	0
W	1	**0**	0	1	0	0	0
C	0	0	**4**	1	0	0	0
SP	0	0	0	**4**	0	0	0
LC	0	0	0	0	**0**	6	0
P	0	0	1	0	0	**6**	0
SL	0	0	0	0	0	1	**0**

Table 10 Confusion matrix with a 4-clusters classification

	Pred-SC	Pred-C	Pred-SP	Pred-P
SC	**3**	0	1	0
C	0	**3**	3	0
SP	0	0	**5**	0
P	0	0	0	**14**

Good predictions are given in bold

number to each of these shapelets in ascending order.[24] However, as a robustness check, we also reverse the ordering between C$\leftrightarrow$SP and P$\leftrightarrow$SL. Likewise, we apply this non-parametric test for both classification schemes: 4 and 7 clusters.

The p-values of the test, presented in Table 11, confirm statistically that, irrespectively of the number of clusters considered, the global financial crisis produced a different pattern of spells in comparison with the other types of recession combined (p-value < 0.1). Moreover, the same result holds for the two alternative shapelets' rankings considered here. In three of the tests, the level of significance is lower than 0.05, and in one of them inclusively lower than 0.01. In the case of the sample with banking crises, the pattern observed is undoubtedly different with a p-value of 0.000 in the four cases. However, for the currency crisis, the statistical results depend on how the ranking of severity is established, at least for the 4- cluster scheme. In three of these tests, the two samples do not seem to come from the same population. Finally, in the four tests presented in the table for fiscal crises, the null hypothesis is not rejected, hence statistically speaking this type of crisis does not generate different output dynamics in comparison with other recession events as a whole.

With the odds ratios of Table 12, calculated with the historical record, we can also infer which shapelets are more likely for certain types of crises. For instance, in the global financial crisis, the P-shaped spells are 1.64 times more frequent than other types of shapelets; although, these events are more prevalent in banking crises

[24] Because the test operates with rankings, the number assigned is irrelevant as long as the order is preserved (i.e., cardinality does not matter).

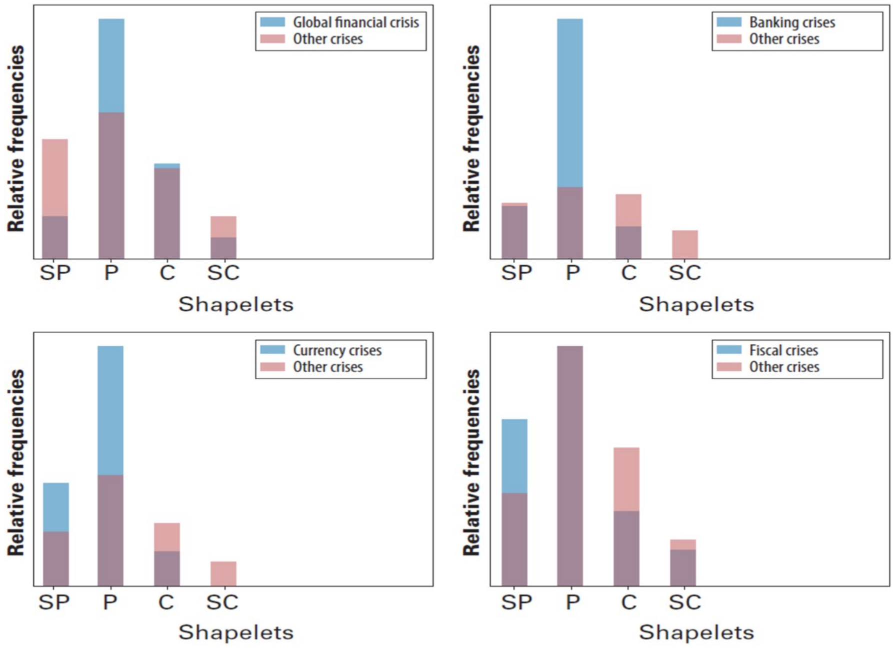

Fig. 9 Comparisons of clusters' frequencies for different types of crises

(2.26 times more), but less so in currency crises (1.455). Also, notice that the possibility of SC-shaped events is null in banking and currency crises, while the global financial crisis did produce this type of events but 0.5 times less than other shapelets. In contrast, SP-shaped events are 1.79 times more likely in fiscal crises. These ratios seem to indicate that the recession events are more benign in fiscal crises than in the other three crises analyzed here; in particular, strong recessions are less likely, and the recovery phase tends to start quickly in mild recessions (i.e., SC-shaped events are more likely).

10 Conclusions

Although, the economic literature has emphasized the analysis of features when studying recessions,[25] we think that using shapelets to describe the output dynamic in recession events can be an attractive alternative in further studies. On the one hand, the shapes of the subsequence of GDP seem to be more informative than recession features for generating clusters with economic meaning. On the other hand, our statistical tests indicate that a collection of these features are strongly associated with clusters based on theoretical shapelets; hence, in a synthesized manner, these shapes of recovery carry the features' information content.

[25] See for instance: Becker and Mauro, 2006; Hausmann et al, 2008; Kannan et al., 2014; Abiad et al, 2014; Francis et al, 2018; Chen et al, 2019.

Table 11 Statistical tests for comparing a specific crisis with other types of recessions

	Analysis with 4 clusters				Analysis with 7 clusters			
	Global	Banking	Currency	Fiscal	Global	Banking	Currency	Fiscal
	Ranking: SC, C, SP, P				**Ranking: SC, W, C, SP, LC, P, SL**			
U-stat	868	526	333	870	888	546	341	842
p-value	0.041	0.000	0.055	0.280	0.067	0.000	0.078	0.215
	Reversed ranking: (SP, C)				**Reversed ranking: (SP, C) and (SL, P)**			
U-stat	767	573	378	891	807	513	341.5	936.5
p-value	0.005	0.000	0.150	0.344	0.017	0.000	0.078	0.497
N. spells	38	31	11	28	38	31	11	28

We make an analytical argument in favor of employing the one-nearest-shapelet algorithm, developed here, to discriminate between recession events. The graphical approach, known as butterfly charts, is prone to judgment biases and it has problems differentiating between certain shapes due to the presence of noise. Having said that, our framework shows that some of the shapes frequently mentioned in the media and financial reports are less common than their popularity would suggest. This seems to be the case of the W-shaped events, SQRT-shaped events, or even L-shaped events when this letter is said to describe a prolonged stagnation period (i.e., if it is drawn with respect to the horizontal axis, instead of using as a reference the pre-recession trend). Moreover, in a statistical sense, it is extremely difficult to distinguish between certain types of shapes; for instance, between V-shaped spells and U-shaped spells.

When using theoretical shapelets, and once these are calibrated with GDP data coming from empirical spells, we find four clusters with well-defined borders and economic meaning: ‘prolonged recession’, ‘symmetric fall and recovery’, ‘small fall and quick recovery’, and ‘prolonged recovery’.[26] Then, we show that the distributions of these clusters have particular characteristics depending on the nature of the crisis hitting the economy. In particular, we find that the global financial crisis of 2007–2008 tended to produce severe recessions that hardly recovered in a time window of ten quarters, although mild recession also occurred in some advanced and emerging economies. The latter is an outcome less likely when these economies are being affected by a banking or currency crisis. Furthermore, in fiscal crises, severe recessions do occur, yet mild recessions are more prevalent –especially those with a quick recovery– in comparison with other types of crises.

Possible applications of the shapelet methodology (1-NS) are the following: (i) the identification of a business cycle synchronization in countries or regions sharing certain policies (e.g., euro-states, sub-national sates); (ii) to study whether or not

[26] These four categories correspond to the following collections of shapelets: (P, LC, SL, S), (C, V, B), (SC), and (SP), respectively. While some of the original W-shaped events are reclassified in SP and others in (C, V, B).

Table 12 Odds Ratios for shapelets in different types of crises

	Banking	Fiscal	Global	Currency
C	0.326	0.532	1.038	0.364
P	2.261	1.003	1.643	1.445
SC	0.0	0.798	0.500	0.0
SP	0.645	1.795	0.353	1.273

business cycles in industrialized economies are intrinsically asymmetric (Morley & Panovska, 2020); (iii) to analyze if fiscal and monetary accommodative policies can modify the shapes of the recession events; (iv) to provide a forecasting technique of the shapes of the recovery based on observable variables during the recession stage. We believe that the shapelet approach offers methodological advantages that simplify the statistical analysis and allow exploring alternative insights from the GDP data. For instance, using the shapelet-clusters as a categorical variable, we can study a phenomenon whose main traits evolve through time, and avoid the econometric complications that arise when estimating models with panel data. Likewise, we can analyze whether there are no-linear effects (i.e., transition phases) on the nature of a recession when intervention policies are implemented.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s41549-021-00063-5.

Funding The authors did not receive support from any organization for the submitted work.

Declarations

Conflicts of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

Abiad, A., Balakrisnan, R., Brooks, P. K., Leigh, D., & Tytell, I. (2014). What's the damage? Medium-term output dynamics after financial crises. In S. Claessens, M. Ayhan-Kose, L. Laeven, & F. Valencia (Eds.), *Financial Crises. Causes, Consequences, and Policy Responses* (pp. 277–308). International Monetary Fund.

Bec F, Bouabdallah, O., Ferrara, L. (2011). The possible shapes of recoveries in Markov-Switching models. *Banque De France Working Paper* No. 321.

Becker, T.I., Mauro, P. (2006). "Output drops and the shocks that matter", *IMF Working Paper*, WP/06/172.

Blanchard, O., Eugenio, C., Summers, L. (2015). "Inflation and activity—Two explorations and their monetary policy implications", *IMF Working Paper*, WP/15/230.

Chen, W., Mico, M., Malhar, N. (2019). "The global economic recovery 10 years after the 2008 financial crisis", *IMF Working Paper* No. WP/19/83

Davis, L.M., Hills, J., Bagnall, A. (2011). "A shapelet transform for time series classification." In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 289–297.

Dueker, M. J. (2006). *Federal Reserve Bank of St Louis Review, 88*(2), 145–153.

Francis, N., Jackson, L. E., & Owyang, M. T. (2018). Countercyclical policy and the speed of recovery after recessions. *Journal of Money, Credit, and Banking, 50*(4), 675–704.

Grabocka, J., Schilling, N., Wistuba, M., & Schmidt-Thieme, L. (2014). Learning time-series shapelets. *International Conference on Data Mining, 14*, 392–401.

Hamilton, J. D. (2018). Why you should never use the Hodrick-Prescott filter. *The Review of Economic and Statistic, 100*(5), 831–843.

Hamilton, J. D. (1989). A New approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica, 57*(2), 357–384.

Harding, D., & Pagan, A. (2002). Dissecting the cycle: A methodological investigation. *Journal of Monetary Economics, 49*(2), 365–381.

Hausmann, R., Rodriguez, F.R., Wagner, R.A. (2008). “Growth collapses.” In Carmen M. Reinhart, Carlos A. Vegh, and Andres Velasco (eds.), *Money, Crises, and Transition: Essays in Honor of Guillermo A. Calvo*. Chap. 15, MIT Press Scholarship Online.

Hills, J., Lines, J., Baranauskas, E., Mapp, J., & Bagnall, A. (2014). Classification of time series by shapelet transformation. *Data Mining and Knowledge Discovery, 28*(4), 851–881.

Hollander, M., Wolfe, D. A., & Chicken, E. (2014). *Nonparametric statistical methods*. John Wiley and Sons.

Howard, G., Martin, R., Wilson, B.A. (2011). “Are recoveries from banking and financial crises really so different?”, *Federal Reserve System, International Finance Discussion Papers*, Number 1037.

Kannan, P., Scott, A., Marco, E. (2014). “From recession to recovery: How soon and how strong?”, Stijn Claessens, M. Ayhan Kose, Luc Laeven, and Fabián Valencia (eds.), *Financial Crises. Causes, Consequences, and Policy Responses*, pp. 239–274, Washington: International Monetary Fund.

Kaufman, L., Rousseeuw, P.J. (1990). “Finding groups in data: An introduction to cluster analysis”. Hoboken, NJ: Wiley.

Kim, C.-J., Morley, J., & Piger, J. (2005). Nonlinearity and the permanent effects of recessions. *Journal of Applied Econometrics, 20*, 291–309.

Mael, G., Malinowski, S., Tavenard, R., Renard, X. (2020). “Localized random shapelets” V. In Lemaire et al. (eds.): *Advanced Analytics and Learning on Temporal Data*, LNAI 11986, pp. 85–97.

Morley, J. (2009). The shape of things to come. *Macroeconomic Advisers, 4*(6), 1–11.

Morley, J., & Panovska, I. B. (2020). Is business cycle asymmetry intrinsic in industrialized economies? *Macroeconomic Dynamics, 24*(6), 1403–1436.

Sichel, D. E. (1994). Inventories and the three phases of the business cycle. *Journal of Business & Economic Statistics, 12*(3), 269–277.

Morley, J., & Piger, J. (2012). The asymmetric business cycle. *The Review of Economics and Statistics, 94*(1), 208–221.

Teimouri S. (2012). “Macroeconomic recovery after a currency crisis”, Ph.D. Dissertation in Economics, West Virginia University.

Vandewiele, G., Ongenae, F., & De Turck, F. (2021). GENDIS: Genetic discovery of shapelets. *Sensors, 21*, 1059.

Ye, L., & Keogh, E. (2011). Time series shapelets: A novel technique that allows accurate, interpretable and fast classification. *Data Mining Knowledge Discovery, 22*(1–2), 149–182.

Ye, L., Keogh, E. (2009). “Time series shapelets: a new primitive for data mining”. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.