Give me a U, Give me a V, Give me an L!
How Effective are Countercyclical Policies in Shaping the Output Dynamic during Recessions

by

Gonzalo Castañeda and Luis Castro Peñarrieta

(This version: August, 11, 2021)

Abstract

In this paper, we argue that studying recession events through the shapes observed in subsequences of GDP data can help to avoid methodological complications and offer new insights to traditional inquiries. With a set of 147 recession events from 77 countries, we analyze whether the shape of the output dynamic might be affected by the application of countercyclical policies. Firstly, we apply a machine learning technique to discover and cluster the shapes (or 'shapelets') that prevail in the set of empirical spells. Secondly, we use a multinomial model to study fiscal and monetary interventions, in which we specify the categorical variable with a set of statistically distinguishable shapelets. Not only do we find strong empirical evidence that it is possible to overcome a recession by means of countercyclical policies, but also that there are non-linear effects that make it more likely when the strength of these policies crosses certain thresholds.

Key Words: Recessions, Countercyclical policies, Machine learning, Shapelets, Multinomial models

JEL Codes: E32, F44, C20, C60

1. Introduction

Nowadays, it is widely recognized that for having a good understanding of the countries' output dynamic, it is necessary to go beyond the statistical analysis of their average growth. Therefore, macroeconomists analyze recessions as perturbations in autoregressive growth models or employ events studies with different features characterizing such phenomena (e.g., depth, output loss, duration). However, an alternative way of looking at the 'canyons and cliffs' in GDP data –rephrasing Pritchett (2000)– is through the identification of shapes in short time series of output. As a matter of fact, financial analysts have done precisely that when describing the dynamic of recession events and their posterior recovery with a set of letters and symbols (e.g., L, U, V, W, checkmarks, square root).

The analysts' perspective is perhaps the result of their concern with firms, investors, and households preoccupied with the short-term prospects of their businesses, portfolios, or daily life once a recession occurs. Will the firm's inventories start to move quickly? Does the possibility of a double dip in the economy explain the volatility observed in financial markets? Is one's job at risk due to a prolonged recession? In contrast, the view of macroeconomic scholars can be related to their interest on mainly two topics: (i) the repercussions that sudden and deep output drops can have on the countries' potential GDP and, thus, on the long-term path of these economies;[1] (ii) the effectiveness of countercyclical policies for ameliorating the adverse effects of recessions.

In this paper, we argue that studying recession events through the shapes observed in the subsequence of GDP data can help to avoid methodological complications and offer new insights to traditional macroeconomic inquiries.[2] In particular, we analyze whether the shape of the output dynamic might be affected by the application of countercyclical policies. Since fiscal and monetary policies are frequently implemented in the aftermath of an output drop, we attempt to disentangle the magnitude of these macro-interventions and to explore their non-linear effects on the GDP path. However, for moving from a features-based description of recession events to a sequence-based description, some analytical

[1] Because of hysteresis and depending on the duration and deepness of the recession events, the pre-recession paths can be dislodged causing a long-term impact on cumulative growth.

[2] Throughout the text. we use the terms recession events and recession-and-recovery events, indistinctively. That is, the events under analysis describe the different phases that take place when GDP collapses, either from endogenous or exogenous factors. Likewise, as it will be explained below, the duration of a recession event does not necessarily coincide with the length of the empirical spells analyzed here; nonetheless, we use these two terms to convey the same idea: an empirical episode describing the dynamic of a recession over several quarters.

challenges need to be addressed. In other words, we need to elaborate a formal technique to discover and cluster the shapes that prevail in the empirical spells describing recession episodes.[3]

For doing that, we appeal to the concept of 'shapelet' developed in the machine learning (ML) literature. Formally speaking, a shapelet is a segment from a vector of observations measured sequentially (e.g., the angle that forms between the stem of a leaf and each point in its periphery). Since economic series are vectors of time-ordered data, this concept can be applied when analyzing recession events. In general, a shapelet is a representative form for a class of segments in a collection of ordered observations; hence it is useful to identify similitudes between series in any database, economic or otherwise. These objects have the additional benefit that their shape can convey information interpretable by analysts, such as the nature of the output dynamic in recession events. The seminal papers on shapelets are Ye and Keogh (2009, 2011). Since then, the methods for the discovery and classification of shapelets have reduced computational time substantially, and consequently, there is now a wide variety of applications in different fields.

Although this classification method is not the only one available for sequential observations, it has the advantage that does not require the synchronicity of events for them to be considered similar. When making event studies with economic time series, it is indispensable to leave aside the synchronicity requirement to pool spells from different historical periods. In an accompanying paper (Castañeda and Castro, 2021), we explain this ML tool in detail and produce the first application for classifying empirical spells associated with different recession varieties. In the said paper, we use a reduced sample of 95 events from an unbalanced panel of quarterly real GDP covering 47 advanced and emerging economies. Here, instead, we expand the database to 195 events coming from 95 countries, including several developing economies. This improved coverage allows us to apply the classification scheme in the analysis of fiscal and monetary interventions. This is done through a multinomial econometric model, in which we specify the categorical variable with a set of statistically distinguishable shapelets that characterize the empirical spells in the database.

This novel form of event study offers several methodological advantages, in so far as it makes the statistical analysis of recessions easier and permits exploring additional hypotheses: (i) we do not have to establish *ex-ante* an specific dynamic for the GDP downturn, such as the dummy variables used in autoregressive models that describe the sudden fall of output for one or few periods from the original

[3] In cluster analysis, the idea is to identify groups of objects through an algorithm that maximizes within-cluster similarities and between-cluster differences in the data.

trend; (ii) with the shapelets as the categorical variable, we can study a phenomenon whose main traits evolve through time but avoiding the econometric complications that arise when estimating with panel data; (iii) we can undertake a comprehensive analysis of the different features characterizing recessions events, instead of studying several dependent variables separately, as in traditional event studies; (iv) we can integrate the recession and the recovery phases in the statistical analysis and, in this manner, improve the study of the GDP dynamic produced in the aftermath of the initial shock; (v) we can analyze whether there are non-linear responses in the output dynamic when varying the strength of the countercyclical policies.[4]

These advantages allow us to assess, through a shapelet-multinomial model, how likely it is to push the economy into the recovery phase of a recession event (i.e., to reverse the downward trend) employing countercyclical policies. But most importantly, while traditional econometric models provide estimations of the fixed marginal effects on a specific feature of recessions due to changes in any explanatory variable, with our approach it is possible to quantify variable marginal effects and to infer if transition phases can occur. In this paper, we refer to a transition phase in a countercyclical policy when, around a threshold of a variable measuring said intervention cross-sectionally, the probabilities of events describing recessions of different kinds cross each other.[5] In other words, around this threshold, the likelihood of observing a mild *versus* a severe recession (i.e., a type of shapelet *versus* another) is reversed.

The rest of the paper is structured in eight more sections. We, firstly, present a brief literature review on econometric models analyzing recession events. Then, we explain which are the phases of a recession event and how the data should be processed to make these events comparable. In the next two sections, we describe the ML methods for the configuration, discovery, and classification of shapelets. We, then, proceed to show results for the fitting of shapelets to empirical spells, and the refinement of clusters with delimited boundaries. Afterward, we validate the selected clusters and their representative shapelets through a separability test. With the suggested classification scheme, we estimate the multinomial model and analyze the relevance of non-linear responses to intervention policies. We end up the paper summarizing its main findings.

[4] If it is considered convenient, we can also combine quarterly and yearly data, the former allows for the proper characterization of GDP subsequences since many of these spells usually do not last more than two years, even if the initial output drop is large; the latter might be useful when some tentative exploratory variables are not measured for a shorter frequency.

[5] Our concept of threshold is different from that of the complexity perspective. Under the latter, these transitions are sudden disruptions on a property of the system when some critical variable describing its operation crosses a threshold. Because of the nature of our database, composed of short subsequences of GDP, we are restricted to detect such thresholds when making cross-sectional comparisons of different recessions events.

2. On the statistical studies of recessions

In general, the complex dynamic of long-term growth is an outcome of a sequence of short-term spells of different magnitude and signs (Castañeda et al, 2021). Accordingly, a thorough empirical study of the difference in the countries' economic growth must go beyond the analysis of 5-10 years averages, like in Barro (1991), and instead to study in more detail the different features observed in the short-term dynamic of business cycles, as indicated in Pritchett (2000), Ben-David and Papell (1998) and Hausmann et al (2008). Along this vein, Becker and Mauro (2006) mention that GDP drops of a certain magnitude should not be conceived like mere fluctuations around the growth trend, but like factors contributing to the future decline of said trend. This view is consistent with macroeconomic studies concluding that output volatility is negatively associated with its long-term growth (Ramey and Ramey, 1991; Fatás 2002, Loayza and Hnatkovska, 2004), especially when there is a high frequency of episodes in the negative tail of the GDP distribution (i.e. recessions spells with large output drops);[6] or that severe downturns in the economy are not always offset by rapid recoveries and, hence they are capable of producing negative side effects in their long-run growth performance.[7]

Several theoretical arguments are put forward to explain the persistent negative consequences of recession episodes. The causal channels are grounded in the functioning of the labor market, the process of capital accumulation, modifications in the institutional architecture in response to the crisis, mechanisms of learning by doing, knowledge accumulation, R&D, and affectations to total factor productivity. Instead of a detailed description of all these channels, for illustration purposes, we specify three examples related to the capital accumulation channel and the problem of hysteresis that emerges in the aftermath of negative shocks.[8] Firstly, the productive capacity of the economy can be handicapped with the presence of deep or sustained contractions causing widespread bankruptcies.[9] Secondly, a sudden output drop affecting population consumption severely can produce political turmoil with a subsequent impact on the rule of law and on the confidence of investors which, in turn, dampens physical capital

[6] For seminal papers on the characterization of deep recessions in emerging economies caused by sudden stops of capital flows see Calvo et al (2005, 2006).

[7] The studies showing the persistent negative effect of recessions are plenty. Examples of these are Aguiar and Gopinath, 2007; Cerra and Saxena, 2008; Reinhart and Rogoff, 2009; Oulton and Sebastiá-Barriel, 2013; Furceri and Mourougane, 2012; Haltmaier 2013; Ball, 2014; Martin et al, 2015; Blanchard et al 2015; Cerra et al, 2020.

[8] By hysteresis, most economists understand that the current level of GDP is associated with the history of shocks disturbing the economy. This path dependency argument implies that certain contingencies exert permanent, or at least persistent, effects on the future evolution of GDP; this is so because the consequences of shocks are not easy to revert. For a literature review on alternative mechanisms of hysteresis see Cerra, et al (2020).

[9] This phenomenon can be especially harmful in emerging economies that lack a well-developed financial system and, hence, cannot smooth out firms' liquidity problems.

accumulation. Thirdly, the uncertainty prevailing in periods of prolonged recessions can generate erroneous investment decisions that are difficult to revert and, thus, the inefficient use of resources affects adversely the economy's potential output.

2.1 Four lines of research

In the literature, there are four main types of statistical analyses for studying recessions econometrically, either in terms of their determinants or consequences. In event studies, authors identify these kinds of episodes as major downward disruptions of growth. Then, for recessions to be defined as such, certain criteria have to be met, such as the output loss observed once GDP starts to fall, or the number of periods with continuous negative growth. Then, the presence, severity (output losses), depth (amplitude), steepness, or duration of these events are analyzed using regressions (least squares, Bayesian model averaging), logit models, and duration (survival) analyses, in which each dependent variable is considered in an isolated manner. Examples of these studies are found in Mora and Siotis (2005), Becker and Mauro (2006), Hausmann et al (2008), Kannan et al (2014), Francis et al (2018), Abiad et al (2014), and Chen et al (2019). In general, their main findings indicate that recession events are heterogeneous and that their nature, in the countries' business cycles, can be associated with different types of shocks (e.g., synchronized globally; financial, political),[10] stage of development, productive structure, initial macroeconomic conditions (e.g., investment ratios, trade openness, fiscal position at the start of the crisis), and the strength of the implemented countercyclical policies.[11]

In dynamic growth regressions, linear-autoregressive models in the spirit of Romer and Romer (1989) are used to study how recessions affect GDP dynamics. Some of these econometric models introduce lagged dummy variables or a summation of them for specifying the timing of the shock's impact. Examples of this approach can be found in Cerra and Saxena (2008), Furceri and Zdzienicka (2012), Teimouri (2012, Chap. 2), Teulings and Zubanov (2014), Morales-Zumaquero and Sosvilla-River (2015), Hamida (2018), and Terrones (2020). This type of statistical modeling has been used to simulate how the different shocks impinge on the economy through the use of impulse-response functions.[12] In general,

[10] For instance, recessions during financial or global crises are deeper and longer in advanced economies, while their recovery is more sluggish, in comparison with other economic shocks.
[11] For example, according to Kannan et al (2014), expansionary fiscal and monetary policies reduce the duration of recession in financial crises, although only the former policies have a statistical significance, perhaps because many economic agents face stringent liquidity constraints.
[12] Cerra and Saxena (2008) use an analytical impulse-response function whose estimated parameters come from the assumed data generating process. Instead, Teulings and Zubanov (2014) use forecast equations -local projections estimators, where new parameters are estimated for each year after the crisis; these equations are corrected by including intermediate shocks occurring

these studies show that short-term fluctuations around the trend are not always a measure of capacity utilization along the business cycle. Rather, severe downturns are frequently an outcome of adverse shocks affecting the trend and, thus, connecting the short with the long-term in a negative fashion. However, a criticism of these models is that the use of binary variables for describing recession episodes misrepresents the nature of recoveries and creates a bias to produce dynamics that have difficulties for bouncing back to the original trend, as in L-shaped events.[13]

In static growth regressions, different features measuring the severity of the recessions (e.g., depth, duration) or a dummy indicating the presence of a recession trough are considered as independent variables, while GDP growth in subsequent periods is the dependent variable to be analyzed.[14] In this manner, these models estimate how the presence of certain features of the recession events impacts the rate of growth in the recovery phase. For instance, Howard et al (2011) use recession features to find that GDP growth, in the aftermath of a banking or financial crisis, recovers faster the deeper is the recession, but slower the longer is such event.[15] Then, Cerra et al (2013) use a dummy for the presence of a trough to find, among other things, that monetary policy is effective in advanced economies for fostering growth in recovery phases, while the fiscal policy effectiveness is estimated to be significant across countries with different stages of development. The strength of the recovery (cumulative output growth one year after the trough) is also analyzed in Kannan et al (2014) with a static model. These authors find, anew, that fiscal and monetary expansions during recession phases are associated with stronger recoveries in advanced economies but, most importantly, that the degree of public indebtedness hampers the effectiveness of fiscal policies.

Another approach for analyzing recessions, which is closer to the idea of letters describing the shape of the output dynamic, is the non-linear growth framework known as the Markov-switching model.[16] Hamilton (1989) started to use this approach for studying expansion and recession regimes in an

---

within the forecast horizon. These authors show, theoretically and through Monte Carlo simulations, that their method is robust to misspecifications of the underlying dynamic of GDP, at least within a family of linear-autoregressive models.

[13] This criticism is partially circumvented by introducing dummy variables for episodes of recession and recovery, see for instance Terrones (2020).

[14] The observation period can be one or a few years after the trough, and the data can include only recession events or be restricted to expansion years (i.e., with a positive growth).

[15] A similar analysis is presented in Haltmaier (2013), although the paper´s main hypothesis is related to the consequences of recessions on the output trend (potential GDP).

[16] There are other empirical studies emphasizing letters' recognition, for instance: Challet et al (2013) and Gregory et al (2020). The former is an econophysics approach, which estimates a three-parameter model to produce response functions that can generate different shapes during episodes of recession and recovery. The latter is a search-theoretic approach, that studies the aggregate dynamic of the labor market during and after the application of lockdown measures to curtail the rate of contagions in a pandemic. Depending on the conditions of the labor markets, and the value of the model's parameters, these authors can

autoregressive model with a state-dependent output mean, whose transitions are determined stochastically. Since then, several adaptations have been implemented to identify different shapes of recoveries: (i) incorporating more than two regimes in the state variable (e.g., Sichel, 1994; Ducker, 2006); (ii) introducing a bounce-back function in the expression of the state-dependent mean (e.g., Kim et al 2005; Morley and Piger, 2012); (iii) adding a delayed bounced-back effect to reflect the possibility that an economy can take more time to recover (e.g., Bec et al, 2011; Teimouri 2012, Chap. 4). The combination of bounce-back functions and delayed effects allow a better characterization of the recovery phases given that the V, U, Long U, L, W shapes, as well as recoveries proportional to the recession's depth, become feasible options to be fitted into the data.[17]

Undoubtedly, these studies have increased the economists' knowledge on recessions, however, there are still several gaps in the literature that need to be filled in. One of these gaps has to do with integrating different features of a recession and its ulterior recovery (e.g., depth, duration) in an informative item so that the entire output dynamic can be studied with traditional econometric techniques. This challenge is tackled in this paper by discovering and classifying the shapes of recession events through a sequence-based ML scheme. This approach is analytically convenient since the recession and recovery phases are part of the same phenomenon and, hence, require to be studied jointly. Clustering these empirical spells into shapelets not only allows to classify different nonlinear GDP subsequences but also to analyze nonlinear responses to intervention policies.

3. The identification of empirical spells in the GDP data

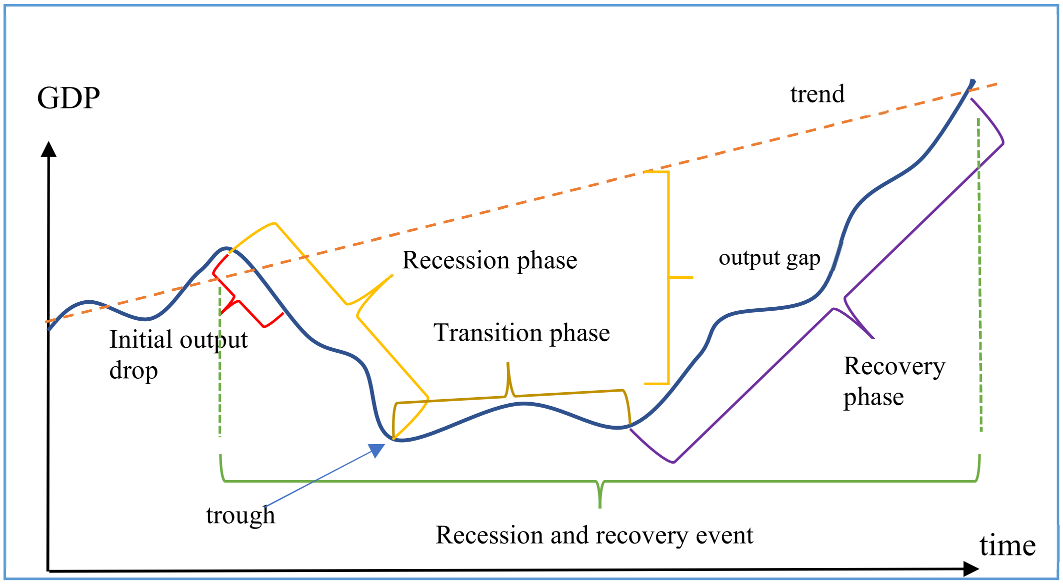
Generally speaking, a recession event is composed of three phases that start once GDP is positioned at a peak: recession, transition, and recovery. As seen in Figure 1, the recession phase begins with the initial output drop and lasts until the plummeting economy attains its trough. Then, during the transition phase, the economy stagnates or moves erratically during some quarters. Finally, in the recovery phase, GDP bounces back and increases steadily until the pre-recession trend is reached. At the end of this event, a new period of expansion ensues, and the economy keeps cycling with different degrees of oscillation while

explain two scenarios for the evolution of employment. A V-shaped dynamic in which the employment level quickly returns to normality once the restrictions on mobility are lifted, and an L-shaped recession in which employment does not recover for several years.

[17] The Markov-switching models present two important limitations for studying recessions: (i) it uses long time-series and, thus, structural breaks have to be considered, making the analysis more cumbersome; (ii) with the estimated parameters a family of bounce-back effects is selected, which implies that all recession events found in the time series have to belong to the same family.

the trend might or not shift. The vertical distance between the potential GDP (i.e., the trend) and the observed GDP is known as the output gap and, when summed, it reflects the quarterly (or yearly) losses that accumulate during a recession event. The maximum output gap is attained at the trough of the GDP fall and it is defined as the recession depth.

Figure 1: The different phases of recession events



For this event study, we define the empirical spells as the period of eleven quarters starting with the pre-recession peak. This implies that for some of these spells we are able to capture the whole recession and recovery event but for others the event will be censored. Although, this strategy implies losing some valuable information concerning the nature of the recessions, it is used here since a good characterization of the output dynamic in terms of shapelets requires empirical spells of the same length; doing otherwise would produce shapelets that are not comparable. For instance, a U-shaped empirical spell in a time-window of 20 quarters is quite different to the same type of shapelet but constructed in a time-window of just 7 quarters. Presumably, the two events have quite different recession features despite having the same shape, being the most evident the differences observed in the duration of the recession from peak to trough. Likewise, the loss of information with censored spells is not that relevant if we are not concerned in getting precise estimates of specific features of recessions, such as the duration of the transition-recovery phases (i.e., the time that elapses between the trough and the pre-recession trend), or the magnitude of the cumulative losses.

To identify these empirical spells in the GDP time series, we define a recession event when, after a peak, we observe a cumulative GDP reduction 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 $y_{t+1} > y_{t+2}$.[18] In this sense, we are leaving aside events exhibiting light recessions. Like in Hausmann et al. (2006), we avoid overlapping events so that an empirical spell cannot start if the country is already in one. As mentioned above, the length of the empirical spell considered here is only 11 quarters (the peak + the initial output drop + 9 quarters ahead); hence, with a relatively short spell, we abate the possibility of including two different crises. Accordingly, the length of the empirical spell does not necessarily coincide with the duration of a recession event that includes the three previously mentioned phases of the business cycle.

For this paper, we built a database with 195 empirical spells coming from the GDP time series of 95 countries that includes all sorts of economies (advanced, emerging, and developing). The database is an extension of the data described in Castañeda and Castro (2021), using the OECD database for 47 economies and data from The Global Economy for the remaining 48 countries.[19] For having comparable shapelets, we not only need spells of the same length but also to work with normalized GDP series. With this objective in mind, we rescale (normalize) the data to produce an index of GDP so that the value in each pre-recession peak is equal to 100. This procedure allows comparing recession events across time and space, no matter the size that the countries' economies had previously to the crisis. The reader should be aware that the output dynamic underlining the subsequences of GDP is not perturbed when rescaling the series. In other words, the recession features in the original series have a one-to-one correspondence with those observed in the transformed series; hence, the shapelets discovered when using the normalized units reflect the recession features observed in the original series.

4. The discovery of 'shapelets' and the classification of empirical spells

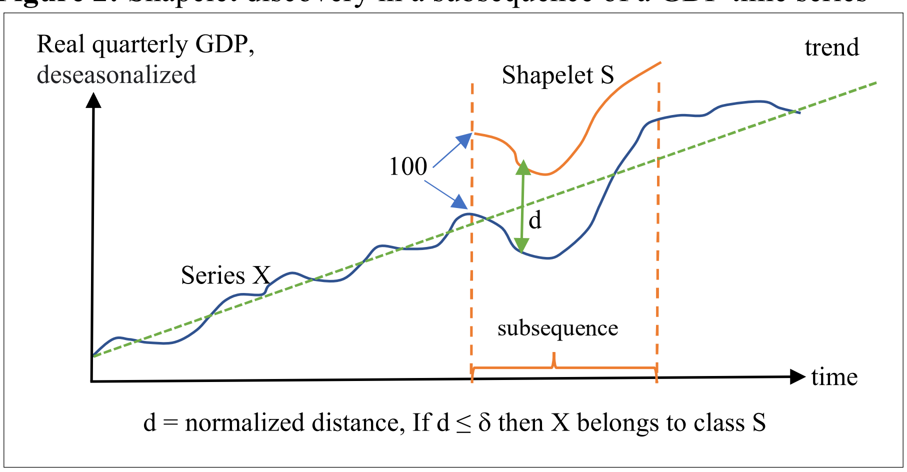
As mentioned in the introduction, an empirical shapelet is a subsequence of observations that, for some reason, appears with a certain frequency in a database of ordered series describing the phenomenon under study. Because of this, such pattern (or dynamic in the case of time series) can be considered one of the series' main traits. On some occasions, these patterns suggest interpretations that contribute to developing the analysts' insights but most importantly make the classification of the series possible. In general, the

[18] This condition is stricter than the one used in the Bry-Broschan procedure (Harding and Pagan, 2002).
[19] For details on the data, refer to the OECD website https://doi.org/10.1787/b86d1fc8-en and The global Economy website https://www.theglobaleconomy.com/

form and extension of these shapelets are not known in advance. Hence, two algorithms are required: one for the discovery of the shapelets and another for the classification of the events in the database.

In the study of recessions, we avoid the computational intensity that characterizes some of these ML methods by identifying, in a first step, the empirical spells in which this phenomenon occurs using the criteria indicated in the previous section. Therefore, we do not need to scan the whole series searching for shapelets of different sizes. In a second step, we use a tentative set of theoretical shapelets, instead of employing a complicated discovery algorithm, because several forms can be detected from a plain visual inspection when looking at the empirical spells. Afterward, we use a customized classification algorithm to cluster the spells of the database. This allows reducing the ambiguity that personal judgments produce. Finally, we implement a depuration method that can reduce the number of statistically relevant shapelets, so that each cluster has distinguishable boundaries. An additional advantage in starting the procedure with a set of predetermined theoretical shapelets consists in providing noise-free objects whose economic meaning in the analysis of recession events is straightforward.

Figure 2: Shapelet discovery in a subsequence of a GDP time series



The basic idea behind the classification process can be better explained with the support of Figure 2. Notice, firstly, that the GDP series has been deseasonalized for eliminating the noise created by seasonal conditions that are commonly observed in quarterly data. Then, we identify a theoretical shapelet (S) by visual inspection, either in the empirical spell of series X bounded by the dashed vertical lines or in the empirical spells of other time series. The classification algorithm starts once the theoretical shapelet is calibrated with the first-quarter output drop (see Figure 1) and with the quarterly average of the GDP changes in absolute value observed throughout the empirical spell. Both subsequences, the calibrated shapelet and the empirical spell of series X are rescaled such that their initial value is equal to 100. Then, we use a distance metric to determine how close is the empirical spell to the calibrated shapelet and repeat the same process for all the other theoretical shapelets initially considered. Therefore, with a variant of the k-nearest neighbor algorithm (with k = 1 and neighbor = shapelet), we establish which is the calibrated shapelet closets to the empirical spell. Once this procedure is done for all recession events in the database, each spell forms part of a cluster with a representative shapelet.

5. Machine learning methods

Next,[20] we explain with some detail the one-nearest-shapelet algorithm (1-NS) for detecting patterns in short sequences of GDP describing the recession spells included in the database.[21] The advantages of 1-NS are threefold: (i) its objects are not described by features but by a sequence of observations, and, thus, it is an appropriate tool to cluster recession events in terms of their output dynamic; (ii) it is designed to operate with sequences composed of a small number of observations (e.g., 10 quarters of a recession event); (iii) the economic meaning of these sequences is not only reflected in the shape of the output dynamic but also in terms of certain features, the latter allows identifying the severity of the recessions associated to each shapelet, as defined below. The interpretability of the cluster-shapelet is extremely important since traditional ML algorithms hardly offer analytical reasons why certain objects belong to specific clusters. Likewise, when using these shapelets as the categories of a multinomial model, we can formulate insightful hypotheses when analyzing countercyclical policies in recession events.

5.1 Configuration of shapelets

To come up with an analytical description of the GDP dynamics, through theoretical shapelets, we need to specify a precise configuration of the different patterns that might be observed.[22] We do this in Table 1

[20] This section is based on Castañeda and Castro (2021), where the reader can obtain a longer explanation of our ML methodology.

[21] This algorithm can be considered a quasi-supervised variant since we identify the spells with a collection of patterns (i.e. the proposed shapelets); although, in the training set there is not a one-to-one identification as in full-fledged supervised algorithms. In contrast, in unsupervised algorithms, such as 1-NN and KMeans, no identity is determined *ex-ante* for the objects in the database to be clustered.

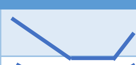
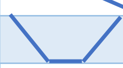
[22] The 'soup of letters' that is frequently mentioned in the media and financial reports is as follow: an L to characterize a sudden output drop in the GDP and then a prolonged path below the trend observed in the pre-recession peak; a V to describe a snap-back dynamic in which the quick fall is offset by a speedy recovery; a J (Nike logo or checkmark) in which the fall is limited to a few quarters and the economy recovers steadily; a W to represent a 'double-dip' recessions in which a tentative recovery is halted by the presence of additional complications in the economy; a U to describe a steady fall and a transition phase that makes the recovery sluggish; a K when a comprehensive output fall is followed by some sectors recovering while others stagnate or drop even further. Other symbols are also mentioned: a 'long-checkmark' when GDP falls slowly, and the

utilizing a binary representation of the shapelets with vectors of 20 bits describing events of 10 periods.[23] Because a binary pair corresponds to one quarter, the combination (1, 0) indicates a drop in the quarter, the pair (0, 0) describes the absence of change, and (0, 1) reflects a surge. Then, with these binary vectors, we can formulate mathematically how the values of the theoretical shapelets evolve in the different quarters that form part of an empirical spell:

$$S_t = S_{t-1} + S_{t-1}(-\tau \cdot \alpha_{t-1} + \tau \cdot \beta_{t-1}) \qquad \ldots(1)$$

where the parameters (α_t, β_t) establish whether in quarter *t* (2,…10), an upward, downward, or no-change movement is observed in the shapelet; τ is an average rate of change specifying the size of each step between periods; *S(0)* = 100 is the normalized GDP in the peak of the event; *S(1)* is the initial output drop.

Table 1: Binary vectors for the configuration of shapelets

ID	Name	Shape	1	2	3	4	5	6	7	8	9	10
P	Pan		(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(0,0)	(0.0)	(0,0)	(0,1)	(0,1)
V	Letter V		(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(0,1)	(0.1)	(0,1)	(0,1)	(0,1)
B	Bowl		(1,0)	(1,0)	(1,0)	(0,0)	(0.0)	(0,0)	(0,0)	(0,1)	(0,1)	(0, 1)
W	Letter W		(1, 0)	(1, 0)	(0, 1)	(0, 1)	(0, 1)	(1, 0)	(1, 0)	(1, 0)	(0, 1)	(0, 1)
SL	Slash		(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)
SC	Short checkmark		(1, 0)	(1, 0)	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)
LC	Long checkmark		(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(0, 1)	(0, 1)
S	Stair		(1, 0)	(1, 0)	(1, 0)	(1, 0)	(0, 0)	(0, 0)	(0, 0)	(1, 0)	(1, 0)	(1, 0)
C	Cup		(1, 0)	(1, 0)	(1, 0)	(1, 0)	(0, 0)	(0, 0)	(0, 1)	(0, 1)	(0, 1)	(0, 1)
SP	Spoon		(1, 0)	(1, 0)	(0, 0)	(0, 0)	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)

upturn takes place well advanced the recession event; a 'square root' to describe a rapid fall and a recovery phase interrupted by a period of stagnation.

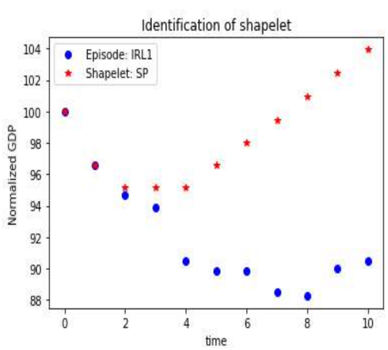
[23] Notice in Table 1 that the letter L –in our definition– is not included in the initial set of shapelets because, after a visual inspection, we opted for a symbol that looks like a pan's silhouette. While the letter U is substituted by the silhouette of a cup exhibiting a flatter transition stage. 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.

In the NS-algorithm, the empirical spells cannot be compared directly with the theoretical shapelets since the parameters τ and $S(1)$ in expression (1) have to be calibrated with the data corresponding to each recession event. Accordingly, for the same theoretical shapelet there are many calibrated shapelets, one for each empirical spell. Notice that, with the binary selection of αs and βs, it is possible to create a large number of configurations, yet many of these cases will not reflect a pattern but only plain noise, while others will not be empirically different to those presented in the third column of Table 1. This happens in a configuration describing a square root that, in a setting of only 10 quarters, looks like a W-shaped shapelet. Therefore, we opted for limiting the type of shapelets to a number equal or lower than the length of the empirical spells.

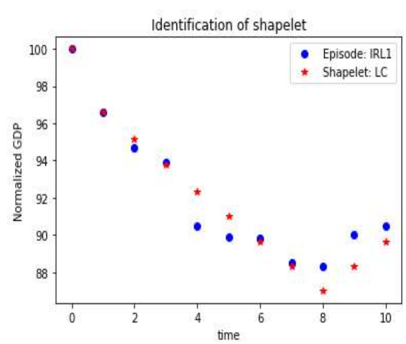
As an example of calibrated shapelets and empirical spells, we illustrate in Figure 3 the case of a recession event in Ireland starting in the 2007-IV quarter. While the small red stars identify the sequential values of the calibrated shapelets, the blue dots identify the rescaled GDP subsequences for this event. From these plots, one can see that the shapelets are noise-free objects describing different dynamics. Therefore, Plot 3.a shows a bad fitting scenario in which the output fall is much more prolonged than that suggested by the calibrated shapelet. In contrast, Plot 3.b shows a scenario of good fitting in which the vertical distance between stars and dot is rather small, being the discrepancy between the two a reflection of the noise observed in the empirical spell. Notice also that the two initial values of the shapelet and the empirical spell coincide; the first one corresponds to the rescaled pre-recession peak (100), while the second one corresponds to the initial output drop.

Figure 3: Calibrated shapelet *vis-a-vis* empirical spell

(a) Bad fitting



(b) Good fitting



5.2 The classification method

The classification method can be decomposed in two stages: a clustering algorithm that groups empirical spells into a predetermined number of theoretical shapelets, and a depuration algorithm that tries to reduce the number of categories by a sequential elimination of clusters with blurred boundaries. We describe the different steps of the clustering procedure in Algorithm 1. This pseudocode explains that the one-nearest shapelet algorithm works by comparing the empirical spells with all available theoretical shapelets once these have been calibrated with the data of the recession event. *Per se*, the algorithm does not guarantee that the fitting has to be good since that depends on the quality of the input; that is to say, on the theoretical shapletes obtained from visual inspection and educated conjectures. Neither that the assumed clusters correspond to statistically different shapelets; for that, it is necessary to improve the classification scheme with an analytical tool that indicates which shapelets of the initial set should be discarded.

Algorithm 1: Clustering pseudocode

```
Input: (a) GDP subsequences of empirical spells (E_i)
       (b) Set of X tentative theoretical shapelets (S_j)  (start with Table 1)
Output: empirical spells grouped in X clusters-shapelets

foreach empirical spell do:
    foreach theoretical shapelet do:
        calibrate S_j with E_i → C_j,i
        calculate the Euclidean distances → d(E_i, C_j,i)
    if S_N is such that min d(E_i, C_k,i)   k = 1,…,X then
        assign empirical spell (E_i) to cluster-shapelet N → SC_N
```

Precisely, Algorithm 2 presents a tool for analyzing whether the initial set of shapelets generates clusters with diffuse borders. This is done by comparing the fit of the empirical spells belonging to a specific cluster-shapelet with the fit of the same spells with their second-nearest shapelets (i.e., the degree of adjustment obtained if these spells were grouped into their closest clusters). As the pseudocode indicates, we calculate a Relative Adjustment Score (RAS_j) for each cluster-shapelet. The adjustment score, at the cluster level, is appropriate when the average distance of its members to the second closest neighbor is much higher than the average distance obtained with the 1-NS algorithm. Therefore, a positive but close to zero value indicates that this cluster has blurred borders.[24] When this occurs, we discard some shapelets corresponding to the lowest-quality clusters that, presumably, need to be merged into categories with the closest shape. Then, we keep iteratively removing shapelets until all clusters have had a score above a certain threshold ($RAS_i \geq 0.4$) in at least one previous round.[25]

Algorithm 2: Depuration pseudocode

Input: (a) GDP subsequences of empirical spells (E_i)
(b) 10 clusters-shapelets
(c) 1st nearest-shapelet of each E_i
Output: Z statistically distinguishable clusters ($Z \leq X$)

foreach empirical spell **do**:
calculate 2nd nearest-shapelet
(i.e. use Algorithm 1 discarding 1st nearest shapelet)

foreach cluster **do**:
calculate average distances of 1st nearest-shapelest $\rightarrow$ SI_i
calculate average distances 2nd nearest-shapelet $\rightarrow$ DI_i
calculate Relative Adjustment Score:

$$RAS_i = \frac{DI_i - SI_i}{max\{SI_i, DI_i\}}$$

if $RAS_i \leq 0.4$ for at least one cluster in all previous iterations **do;**
eliminate one or two shapelets with similar shapes and low RAS_i
apply Algorithm 1 and 2 again with X – J clusters (J = 1 or 2)

6. Clustering results

In this section we show, with 195 recession events, the capability of the clustering algorithm to find a good shapelet fitting, and how the depuration algorithm works in reducing the number of representative shapelets, so that those remaining can be distinguishable from one another. When implementing the initial 10-clusters configuration, we find that there is a positive number of empirical spells assigned to each shapelet {P : 25, V: 7, B: 27, W: 20, SL: 9, SC: 22, LC: 21, S: 9, C: 7, SP: 48}. Although we cannot

[24] A negative value is technically possible, for instance, when selecting clusters at random for each empirical spell. However, when using the 1-NS algorithm in the first stage of the classification procedure $RAS_i \geq 0$ by construction. In other situations, a $RAS_i < 0$ means that, in average, the empirical spells are wrongly positioned in such cluster.

[25] The reason why the halting criterion in the algorithm involves all previous iterations, and not only the most recent one, has to do with the possibility that a shapelet with a RAS > 0.4 can reduce its value below this threshold, in the following round. This occurs when the associated cluster receives empirical spells from the removed clusters, making the boundaries of the former a little bit more blurred.

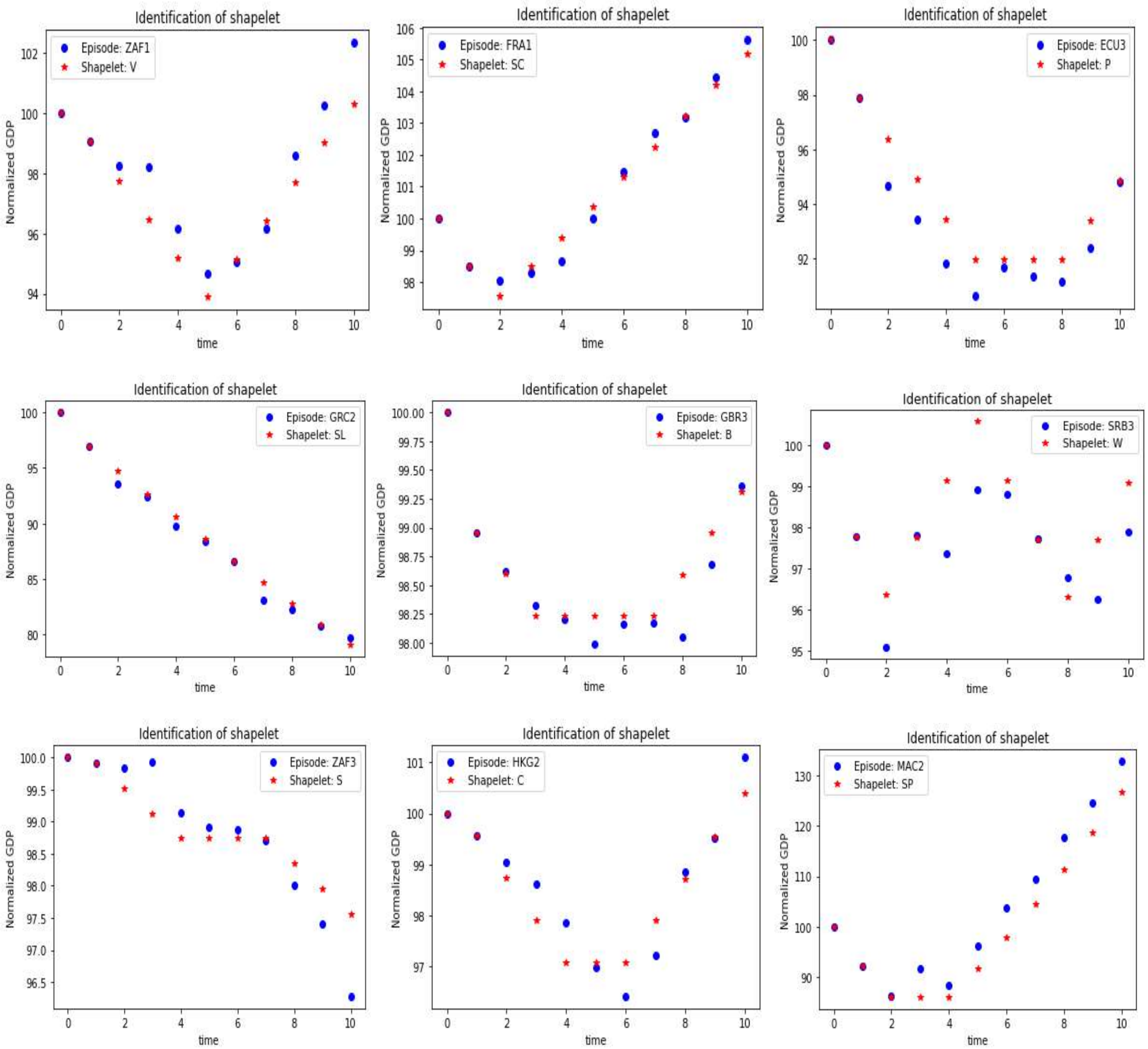
sustain that this set contains all possible shapes, we can argue that each of the available options can be identified with several empirical spell in the database. However, as stated above, the theoretical shapelets obtained from visual inspections is just the beginning of the analysis. The results presented here indicate that some of the associated clusters do not have clear boundaries. Indeed, the depuration algorithm suggests that some refinement is needed, in which only 4 or even 3 interpretable clusters-shapelets remain.

6.1 Fitting shapelets to the recession events

When applying the 1-NS algorithm to the events included in the database, we observe that the procedure performs relatively well in the sense that the method produces a relatively good fitting for most of these spells. Precisely, as an illustration, we show in Figure 4 the capacity of the algorithm to fit nine of the theoretical shapelets (the other is already presented in Figure 3.b) to empirical spells of different countries. Despite the noise in the empirical spells and the fact that we only calibrate the shapelets with the initial output drop and the step size (i.e., the average magnitude of changes in GDP across quarters), we find that these shapes can capture the severity of the recession reflected, among other features, in the cumulated losses and the mean gap (explained in detail below) that these events produce between the output trend and the subsequences of observed GDP.

This figure also shows that, for some noisy empirical spell, the first and second closest neighbors might not produce significantly different Euclidean distances. Thus, in practice, it is difficult to find delimited borders between some clusters when we consider many theoretical shapelets in the classification scheme. Although all the examples that we present here correspond to good-fitting cases, a careful look indicates that there are potential complications for the triad (V, C, and B) and the pair (S and SL) in terms of experiencing a poor separability. For addressing this problem, we require further statistical analyses to refine the number of representative shapelets. We do this either through the depuration algorithm or using a separability analysis that test the predictability of the clusters in terms of recession features, as explained below.

Figure 4: Some examples of the shapelets fitting to the GDP subsequences



6.2 Adjustment scores and depuration of classification schemes

Before applying the depuration algorithm, we established that all theoretical shapelets in the original configuration are nearest-neighbors of several empirical spells, being the V and C clusters-shapelets the smallest ones (7 spells each). Although there are, indeed, some cases with a bad fitting, in which the appropriate shape of a recession event has not been properly identified, we can argue that, in general, these shapelets describe the most notorious dynamics observed in the database. However, the results of the clustering algorithm generate poor relative Adjustment Scores ($RAS_i << 0.4$) for seven clusters-shapelets (V, B, W, SL, LC, S, and C). This outcome does not mean that all of them should be removed from the tentative set of candidates, but only that some mergers are needed. For instance, when looking at the shapes in the third column of Table 1, it is clear that three shapelets (V, B, and C) have closer forms. Hence, because C can be considered the intermediate case, we decided to preserve only this shapelet in the following iteration.[26]

Table 2: Relative Adjustment Scores for each cluster in different classification schemes

No. Clusters	P	V	B	W	SL	SC	LC	S	C	SP
10	0.297	0.065	0.160	0.243	0.264	0.465	0.161	0.180	0.049	0.463
7	0.325			0.314	0.261	0.465	0.174		0.418	0.464
5	0.415			0.314		0.465			0.418	0.464
4	0.415					0.465			0.383	0.435
3	0.415								0.383	0.577
3	0.621					0.465				0.427

In the third round (second row in the table), we removed S which is a member of the triad (S, SL, LC) exhibiting low RAS and sharing a shape with a relatively long recession phase. Notice, in the 7-cluster classification scheme, that the adjustment score for C is now above the 0.4 threshold, as expected, and that SL and LC still present low scores, indicating that a merge with P seems a convenient step. This decision is implemented in the fourth round, where we have 5 clusters, and its outcome indicated that all but one shapelet (W) meets the halting criterion. When the associated cluster is removed, in the fifth round, we find that the merging of empirical spells reduces the RAS for C, blurring its boundaries. In any case, this outcome is good enough for stopping the depuration algorithm. Nonetheless, as a robustness check, we analyzed two schemes with three clusters each: (P, C, SP) with a mean adjustment score of 0.458 and (P, SC, SP) with a mean score of 0.504. Notice that both of these triads are composed of a theoretical shapelet with a prolonged recession phase (P) and two shapelets with a short one.

Accordingly, in Tables 3 and 4, we present the cluster membership for the 4-shapelets scheme and one of the 3-shapelets schemes. For reasons explained below, we opted for the triad (P, C, SP) despite the more erratic behavior of C's RAS and its slightly lower mean score. The shapes of the short checkmark

[26] An alternative would be to maintain the shapelet with the largest membership in the classification with 10 clusters, in this case, B. However, the separability test presented below shows that, in terms of predictability, it is better to preserve the intermediate case (C).

(SC) and the cup (C), described in Table 1, indicate that the corresponding recession events are more sluggish in the latter shapelet. It is important to clarify that a classification scheme with more clusters produces empirical spells with better fittings (i.e. empirical spells with clearer output dynamics) but, in contrast, its clusters' boundaries are more diffuse. The latter implies that a depurated scheme might be beneficial in detecting clusters-shapelets whose recessions events are more distinguishable and easier to analyze.

Table 3: Classification scheme with 4 clusters-shapelets (P, SC, C, SP)

P	SC	C	SP
ARG2 (LC), ARG3 (P), AUT1 (P), BGR1 (P), BLR4 (LC), BRA2 (LC), CAN1 (P), CHE2 (S), CYP1 (P), CYP2 (S), CZE1 (P), DEU1 (P), DNK1 (LC), ECU3 (P), EST2 (SL), ESP1 (SL), ESP2 (SL), FIN1 (LC), FIN2 (LC), FIN3 (S), FRA2 (P), GBR1 (S), GBR2 (P), GBR4 (LC), GRC1 (S), GRC2 (SL), HKG3 (P), HKG4 (V), HRV2 (LC), HRV3 (SL), HUN1 (LC), IDN1 (**P**), IRL1 (LC), ISL1 (S), ISR1 (P), ITA2 (LC), ITA3 (SL), JAM3 (LC), JPN3 (P), KGZ3 (**LC**), LTU2 (LC), LUX2 (P), LVA1 (SL), MAC3 (**LC**), NLD1 (P), PER1 (S), PHL1 (P), PRT2 (P), PRT3 (SL), QAT2 (LC), ROU1 (SL), ROU2 (P), RUS1 (P), RUS2 (P), SRB1 (**P**), SRB2 (P), SVN1 (LC), SVN2 (LC), ZAF3 (S), SWE2 (S), SWE3 (P), THA1 (LC), UKR2 (LC), URY1 (LC), URY2 (P)	BRA1 (SC), CRC1 (SC), EGY2 (**SC**), FRA1 (SC), GEO2 (SC), GUY2 (SC), HKG1 (SC), LUX1 (SC), MAC1 (**SC**), MDA1 (SC), NZL2 (SC), OMN1 (**SC**), PER3 (SC), PHL2 (SC), PHL3 (SC), POL1 (SC), QAT1 (**SC**), RWA1 (SC), THA2 (SC), URY3 (SC), USA2 (SC), USA3 (SC),	ALB1 (W), ALB2 (B), ARG6 (V), ARM2 (B), AUS1 (B), AZE1 (**C**), AZE2 (B), AZE3 (B), BEL1 (B), BGR2 (B), BGR3 (B), BWA1 (B), CAN2 (B), CHE1 (V), CPV1 (W), CZE2 (C), CZE3 (B), EGY3 (B), GBR3 (B), GEO1 (**W**), GUY1 (**W**), HKG2 (C), IRN1 (V), IRN2 (W), ITA1 (C), JAM2 (B), JAM4 (W), MEX2 (B), MNE1 (W), MNE2 (V), MYS1 (C), NLD2 (B), NOR3 (W), NZL3 (B), OMN2 (B), PRT1 (V), SGP2 (B), SGP3 (V), SLV2 (C), SRB3 (W), SVK1 (B), SWE1 (B), TUR2 (B), UKR1 (B), USA4 (V), ZAF1 (C), ZAF2 (B)	ARG1 (SP), ARG4 (SP), ARG5 (W), ARM1 (**W**), ARM3 (SP), BIH1 (W), BLR1 (**SP**), BLR2 (**SP**), BLR3 (SP), BWA2 (SP), BWA3 (SP), CHE3 (SP), CHL1 (SP), CMR1 (W), CMR2 (W), DOM1 (SP), ECU1 (SP), ECU2 (SP), EGY1 (W), EST1 (SP), GTM1 (SP), HRV1 (SP), HUN2 (SP), IRL2 (W), IRN3 (SP), JAM1 (SP), JOR1 (SP), JPN1 (SP), JPN2 (B), KGZ1 (SP), KGZ2 (**SP**), KGZ4 (SP), KOR1 (SP), KOR2 (SP), LTU1 (SP), MAC2 (**SP**), MEX1 (SP), MKD1 (W), MKD2 (SP), MLT1 (SP), MLT2 (SP), MNG1 (**SP**), MNG2 (**W**), MUS1 (B), NIC1 (SP), NOR1 (W), NOR2 (SP), NZL1 (W), PER2 (SP), PRY1 (SP), PRY2 (SP), SAU1 (SP), SGP1 (SP), SGP4 (SP), SLV1 (SP), THA3 (SP), TJK1 (SP), TUN1 (SP), TUR1 (SP), USA1 (SP), ZAF4 (SP).
65	22	47	61

With four clusters-shapelets, as in Table 3, the largest membership corresponds to cluster P with 65 empirical spells, followed by cluster SP with 61, then by cluster C with 47, and finally by cluster SC with 22. According to the shapelets' id, indicated in the parentheses, when moving from a 10-clusters configuration to a 4-clusters configuration, all but one empirical spell (HGK4: V → P) that have to be reassigned is merged in the expected shapelet-cluster. The SC-cluster remains with the same empirical spells; the P-cluster absorbs spells previously classified in LC, SL, and S; the C-cluster includes spells in V, B, and W; the SP-cluster receives spells in W and B, although just two in the latter case. In other words, when a theoretical shapelet is removed from the classification scheme, the associated spells are reassigned to categories with similar recession phases. When we type the id of the shapelet in boldface, this means that the 1-NS algorithm produces a bad fitting (i.e. an error above the mean plus a standard deviation), hence we can consider these events as undefined. There are 18 empirical spells with a poor-fitting; four of them assigned originally to the SC, W, and SP shapelets; 2 assigned to P and LC, and only one to C.

Table 4: Classification scheme with 3 clusters-shapelets (P, C, SP).

P	C	SP
ARG2, ARG3, AUT1, BLR4, BRA2, BGR1, CAN1, CHE2, CYP1, CYP2, CZE1, DEU1, DNK1, ECU3, EST2, ESP1, ESP2, FIN1, FIN2, FIN3, FRA2, GBR1, GBR2, GBR4, GRC1, GRC2, HKG3, HKG4, HRV2, HRV3, HUN1, IDN1, IRL1, ISR1, ISL1, IA2, ITA3, JAM3, JPN3, KGZ3, LTU2, LUX2, LVA1, MAC3, NLD1, PER1, PHL1, PRT2, PRT3, QAT2, ROU1, ROU2, RUS1, RUS2, SRB1, SRB2, SVN1, SVN2, SWE2, SWE3, THA1, UKR2, URY1, URY2, ZAF3	ALB1, ALB2, ARG6, ARM2, AUS1, AZE1, AZE2, AZE3, BEL1, BGR2, BGR3, BWA1, CAN2, CPV1, CZE2, CZE3, EGY3, GBR3, GEO1, GUY1, HKG2, IRN1, IRN2, ITA1, JAM2, JAM4, MEX2, MNE1, MNE2, MYS1, NLD2, NZL3, NOR3, OMN2, PRT1, SGP2, SGP3, SRB3, SVK1, SLV2, SWE1, CHE1, TUR2, UKR1, USA4 , ZAF1, ZAF2	ARG1, ARG4, ARG5, ARM1, ARM3, BIH1, BLR1, BLR2, BLR3, BWA2, BWA3, BRA1, CHE3, CHL1, CMR1, CMR2, CRC1, DOM1, ECU1, ECU2, EGY1, EGY2, EST1, FRA1, GEO2, GTM1, GUY2, HKG1, HRV1, HUN2, IRL2, IRN3, JAM1, JPN1, JPN2, JOR1, KGZ1, KGZ2, KGZ4, KOR1, KOR2, LTU1, LUX1, MAC1, MAC2, MEX1, MDA1 MKD1, MKD2, MLT1, MLT2, MNG1, MNG2, MUS1, NIC1, NOR1, NOR2, NZL1, NZL2, OMN1, PER2, PER3, PHL2, PHL3, POL1, PRY1, PRY2, QAT1, RWA1, SAU1, SGP1, SGP4, SLV1, THA2, THA3, TJK1, TUN1, TUR1, URY3, USA1, USA2, USA3, ZAF4
65	47	83

When the classification scheme is formed with only 3 shapelets, as in Table 4, the 22 empirical spells that are grouped in SC in the 4-clusters scheme, now integrate cluster SP, while P and C keep the same empirical spells. These three categories present boundaries relatively separated, although the fitting quality of the representative shapelet for the reassigned spells is lower, by construction, than the one obtained in a 10-cluster scheme. Despite the broad level of aggregation of this 3-clusters scheme, there is an extremely large number of countries whose events are classified in different schemes. For instance, Argentina with four recession events presents spells in the three categories (two in SP, one in C, and three in P). This outcome shows that the Markov-switching model is mistaken by assuming that the same family of recession dynamics applies through the entire time series of each country.

7. Separability tests

As a way to validate the selection of shapelets suggested by the RAS criterion, we test the separability of clusters by checking the predictability of the different classification schemes. This is done through a multinomial model, in which the categorical variable is composed of the clusters included in the scheme under analysis, while the explanatory variables are the recession features associated with the empirical spells. This method has the added benefit of providing additional economic meaning to the clusters, beyond the shape of the output dynamic. When a combination of these features impinges on the predictability, we can argue that these shapelets can be categorized as cases of mild or severe recessions.

As traditionally done in the ML literature, we divide the events in the database into a training subset and a testing subset. Hence, we can analyze the out-of-the-sample predictability of the clusters suggested by the classification scheme through recession features. Therefore, when the Accuracy Score is relatively high in a certain scheme, we can argue that the 1-NS algorithm does not produce blurred borders, in the sense that recession features make a reasonably good job in predicting specific output dynamics. In contrast, when the number of errors is large, we can state that the classification scheme, in a particular iteration of the depuration algorithm, fails to assign a significant number of empirical spells into clusters-shapelets in a meaningful way, precluding the interpretability of the scheme.

In the Table 10A of the online Appendix, we present some descriptive statistics related to seven recession features and their formal definitions: duration, depth, cumulative loss, mean, variance, kurtosis, and skewness. To calculate the last six features, we implement the Hodrick-Prescott filter to establish a trend in the business cycle that varies in each pre-recession peak. This trend is a proxy of the potential output that allows estimating these features once the output gap is calculated: the quarterly differences between the potential and observed GDP; for instance, depth is the maximum output gap observed during the empirical spell. While duration is the number of quarters that the empirical spell takes to reach its pre-recession peak, or 11 quarters if the previous condition is not met during the time the spell lasts.

In the following tables, we present the so-called confusion matrixes that show the predictability of the multinomial models at the shapelet level. First of all, after a series of estimations, we conclude that the highest Accuracy Scores are obtained with a combination of just three features (cumulative loss, mean, and kurtosis); hence, the results presented here correspond to multinomials using these explanatory variables.[27] Table 5 illustrates the inadequacy of the 7-cluster classification scheme, as suggested by the corresponding RAS presented in the previous section. The off-diagonal numbers of the confusion matrix indicate the presence of 30 errors out of 59 events, which implies a poor Accuracy Score (0.491).[28] In particular, four of the shapelets (SC, W, LC, and SL) have more errors than successes, although for the latter shapelet the sample only includes three spells of this kind. This matrix is highly informative since it makes salient the theoretical shapelets that might be excluded from the final classification scheme.

Table 5: Confusion Matrix for the 7-clusters scheme

	Pred-SC	Pred-W	Pred-C	Pred-SP	Pred-LC	Pred-P	Pred-SL
SC	3	0	0	6	0	0	0
W	3	1	2	1	0	0	0
C	0	0	5	1	0	0	0
SP	3	1	0	11	0	0	0
LC	0	0	1	0	0	7	1
P	0	0	2	0	0	8	0
SL	0	0	0	0	0	2	1

According to this information and the Relative Adjustment Scores (Table 2) generated with the depuration algorithm, we decided to analyze the confusion matrix of 4 clusters-shapelets (SC, C, SP, and P). By construction, the Accuracy Score should exhibit some increase by the mere fact that there are fewer categories to predict. However, the Accuracy Score with 4 candidates shows that the increase is far from being marginal (0.746).[29] It is important to highlight that several shapelets classified in the SC cluster are wrongly predicted with the explanatory variables of the multinomial model (Table 6), an outcome that questions the validity of this classification scheme, at least if interpretability is a desired trait in the final scheme.

Table 6: Confusion Matrix for the 4-clusters scheme

	Pred-SC	Pred-C	Pred-SP	Pred-P
SC	2	0	7	0
C	1	6	1	1
SP	4	1	14	0
P	0	0	0	22

[27] For instance, with a classification scheme of 3 clusters-shapelets (C, SP, and P) we obtain an Adjustment Score of 0.932 when using three explanatory features (cumulative losses, mean, and kurtosis). In contrast, we get a lower score (0.881) when the seven features are considered.
[28] For this and the remaining out-of-the-sample predictions, we consider a testing subset with 30% of the 195 events in the database.
[29] Moreover, this score increased even further (0.81) when we removed from the sample 12 spells with extremely poor fitting to the 10 original theoretical shapelets.

Therefore, in Table 7 we show the confusion matrix with 3 clusters-shapelets (C, SP, and P). Anew, the jump in the Accuracy Score is substantial (0.932) and the predictability of the SP and P clusters is excellent with one and zero mistakes, respectively.[30] The performance of this three-clusters scheme is superb if we compare it with the score obtained with a manifold learning algorithm (T-sne) that contracts a 7 features space to only two dimensions before clustering with a traditional procedure (KMeans).[31] Despite that in T-sne only two categories are considered, its Accuracy Score is lower (0.864).[32] Consequently, the 1-NS algorithm produces a classification scheme with a remarkable accuracy that fully validates the separability of the three final clusters, with the additional advantage that the empirical spells associated with these clusters exhibit recessions with different degrees of severity.[33]

Table 7: Confusion Matrix for the 3-clusters scheme

	Pred-C	Pred-SP	Pred-P
C	6	2	1
SP	1	27	0
P	0	0	22

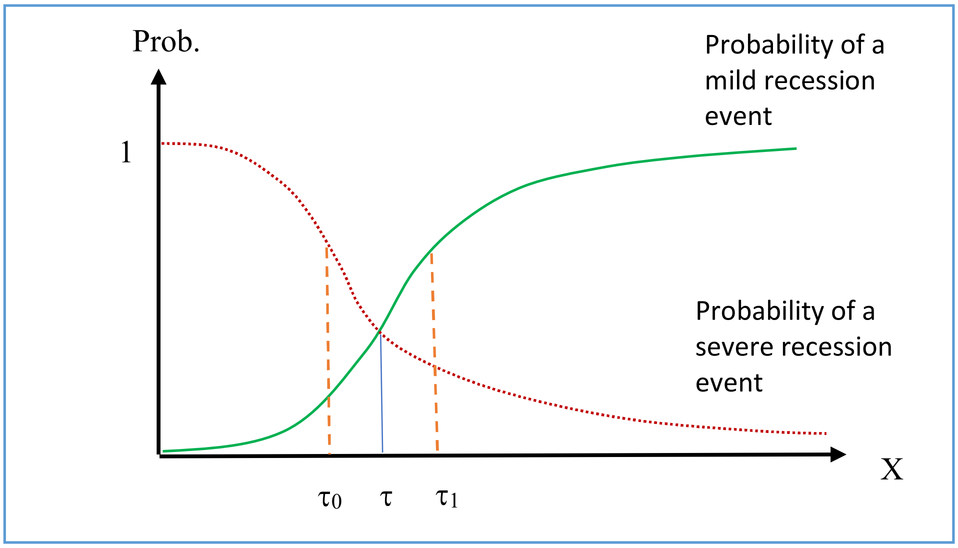
8. A multinomial analysis of output dynamics in recession events

In this section, we study several factors that can impinge on the shape of GDP subsequences during recession events through a multinomial logistic model (Verbeek, 2012, Chap. 7). This econometric technique allows estimating the likelihood of non-ordered outcomes,[34] as those described by the classification schemes obtained with the 1-NS algorithm. This modeling approach is used frequently for analyzing microeconomic choices in decision problems, but not as much in the macroeconomic literature. Three macro-application of this method are the study of countries selecting exchange rate regimes (Álvarez et al 2011; Dubas et al, 2010; Papaioannou, 2003), the country risk analysis of foreign direct investments (McGowan and Moeller, 2005), and the study of economic growth and structural change in

[30] The Accuracy Score for the triad (P, SC, SP) is lower (0.800), which validates our preference for the triad (P, C. SP).
[31] For additional details on this method and the performance of other algorithms consult the supplementary material in Castañeda and Castro (2021).
[32] As expected, when only two shapelet-clusters are considered (SP and P) in the 1-NS algorithm, the score increases even further (0.966), yet this improvement does not justify reducing the number of categories in the analysis of recession events.
[33] The reader should be aware that when reducing the number of theoretical shapelets, from the initial configuration with 10 candidates to only 3, some valuable information can be lost in terms of a poorer description of the different shapes of the output dynamic. This scenario happens when discarding the good fitting of some empirical spells (e.g., when changing a V-shaped spell for a C-shaped spell), yet the final classification has the advantage that the representative shapelets produce well-separated clusters which are also economically interpretable in terms of their recession severity (i.e. a comprehensive measure of different recession features).
[34] The unordered requirement holds for the cluster categories considered here since they do not identify a ranking of preferences or valuations.

the medium term (Castañeda and Romero, 2021).[35] Consequently, this paper is the first application of multinomial regressions for studying output dynamics in episodes of recession and their subsequent recovery.

Figure 5: Reversion of probabilities

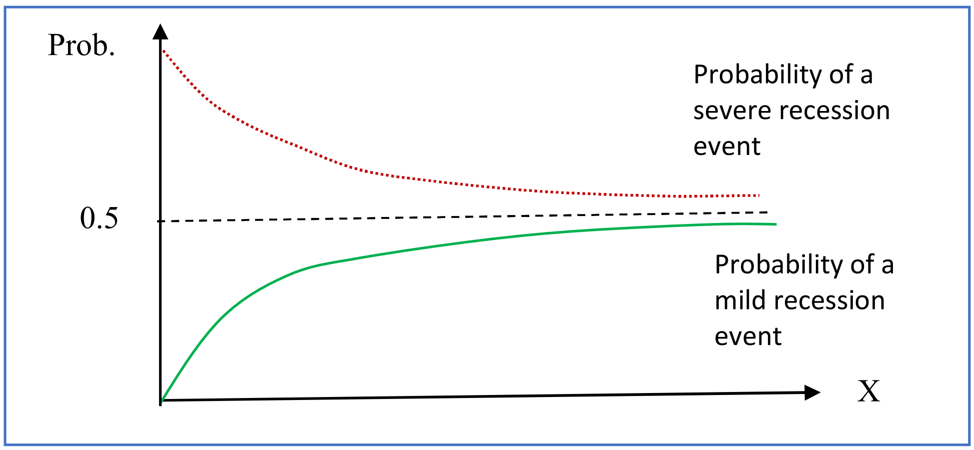


More importantly, this type of study allows inferring the existence of a threshold in an explanatory variable that when crossed produces a reversal in the likelihood of observing a certain type of recession event vis-à-vis other categories. For instance, the crossing of the threshold could imply that a significant change in the severity of the recession can be expected. By the nature of the econometric framework, it is possible to estimate different thresholds for each pair of categories. Therefore, if the explanatory variable under analysis is the strength of a fiscal or monetary policy, then this crossing point indicates the level where such policy can produce, in a statistical sense, significant changes in the kind of recession to be observed. In other words, while traditional methodologies for analyzing recessions provides estimations of the fixed marginal effects of an explanatory variable on a specific feature of recessions, the multinomial approach can generate transition phases (i.e., nonlinear effects produced by changes in the probability distribution around a threshold) of the same variable on the nature of a recession.

[35] Binomial models (logit or probit) are more commonly used in macroeconomic event studies. For instance, when analyzing the determinants of financial crises (Jemovic and Marinkovic, 2019; Comelli, 2014; Schularick and Taylor 2012; Cerra and Saxena, 2008; Hausmann et al 2008; Falcetti and Tudela, 2006), current account reversals (Cavdar and Aydin, 2015; Liesenfeld et al, 2010; Milesi-Ferretti and Razin, 1998). and credit-less recoveries (Bijsterbosch and Dahlhaus, 2015).

In Figure 5, we describe a scenario of probability reversion in which severe recessions become the least likely outcome once the level of a variable X crosses a threshold τ when moving from τ_0 to τ_1. Because of this, the estimated marginal effects are not only variable –moving along the distribution function– but also in some circumstances can entail transition phases –crossing thresholds. Consequently, if X measures the strength of the monetary policy at the onset of the recession, this setting indicates that an accommodative policy is unlikely to have a significant positive effect unless its strength is high enough ($X > \tau$). In contrast, in Figure 6, we show an alternative scenario in which changes in X, across the whole domain of the function, do not produce a reversion of probabilities. In this case, a severe recession event is always the most likely outcome, irrespectively of the level of X.

Figure 6: Unresponsive accommodative policy



8.1 Nonlinearities of countercyclical policies

The macroeconomic literature offers several theoretical explanations for the nonlinear effects of fiscal policies (e.g., Bertola and Drazen 1993, Blanchard 1990). These nonlinearities have to do with the size, persistence, and direction (expansion or contraction) of the fiscal impulses, with their composition (changes in taxes, transfers, or other government expenses), and with the existing level and growth of the debt-to-income ratio.[36] The theoretical underpinnings of these studies focus on how these policies affect consumer behavior, hence the horizon of decision-making, expectations, and the distortionary consequences of taxation are critical channels to determine their consequences on aggregate consumption, national savings, or growth. For instance, as suggested by Feldstein (1982) and Drazen (1990), a fiscal

[36] For a succinct literature review on this issue see Giavazzi et al (2000).

package with a large increase in public spending can modify consumers' expectations when the policy is seen as a sign of a new practice of high government spending. If this change is expected to be permanent, higher permanent taxes create a reduction in private consumption and, thus, national saving remains unaffected.[37] In contrast, when the increase in government spending is relatively small, no expectation of a regime switch occurs, hence private consumption remains unaltered, and national savings falls.

There are several variants of econometric analyzes involving nonlinearities in fiscal or monetary policies. Most of these studies indicate, for different sets of countries and periods, that there is a sound empirical basis to argue that the effectiveness of macroeconomic policies depends on the policies' characteristics and on the countries' economic context. One of these variants consists of reduced-form equation models with national savings as the dependent variable, in which nonlinearities are tested using dummy variables interacting with net taxes and government expenditures ratios. These dummies describe economic conditions in the business cycle (e.g., financial crises, expansions, recessions) as in Bui (2018), or characteristics of the fiscal variables (e.g., thresholds in the size of the fiscal impulse, or in the level and growth of the debt-to-income ratio) as in Giavazzi et al (2000).

Another approach is the use of nonlinear vector autoregressive models that attempt to capture, through impulse-response functions, asymmetric responses of policy shocks on different macroeconomic variables (e.g., output growth, inflation) depending on the economy's regime. In Markov-switching VAR models, it is assumed that lagged output growth affects the probability of the state of the economy. For instance, Jackson et al (2016) use a model of this type to evaluate whether the impact of the monetary policy depends on the current state of the economy, its history, and the size of current and future monetary shocks. Then, in threshold VAR models structural changes take place when a prespecified transition variable crosses a critical value. For example, Saldias (2017) studies the transmission of the monetary policy depending on whether the economy is in a situation of low or high financial stress.[38]

Although it is not the purpose of this paper to develop a theoretical model to explain why nonlinearities in countercyclical policies can be relevant in the context of recession events, we present an analytical narrative to illustrate why this might be the case. Emphasizing the supply side of the economy, we can argue that the production system is composed of a network of activities connecting producers of all sorts of goods (consumption, intermediate, capital, and raw-material extraction). In the course of a

[37] For this outcome to happen, it is necessary to assume an infinite-horizon setting and no tax distortions.

[38] Other applications of the TVAR model are the following: Avdjiev and Zeng (2014), which analyzes the interaction between credit market conditions, monetary policy, and economic activity; Fazzari et al (2015), which studies state-dependent effects of government spending conditioned on the economy's capacity utilization.

business cycle, continuous idiosyncratic shocks are affecting the demand for consumption goods across all economic sectors then, because of the need to produce in terms of batch-orders and the limited availability of suppliers, cash flow problems (e.g., failures to meet payments to suppliers) can produce cascade effects in the production network. These cascades can be of vastly different magnitude, some of them producing a small deviation from the output trend but others capable of generating sharp recessions endogenously.

Let us assume that at the start of a recession, the government decides to implement a fiscal accommodative policy. Say that this policy consists of tax rebates assigned to firms (i.e., the nodes in the network) and that its strength refers to the number of nodes covered. While the additional cash flow allows these firms to meet their accounts payable and keep operating, it might not be enough to preclude the recession if the number of nodes benefited from the rebates is not relatively large. In other words, to reduce the probability of a severe cascade of failures, these 'rescued' firms must be spread across different sectors of the economy and their production chains. For this to happen, the fiscal package requires being implemented with sufficient resources, otherwise, it will not be possible to avert firms stopping production, even those receiving rebates.

8.2. Empirical results

To check the effect of countercyclical policies on the shapes of the recessions, we construct two different variables: one for fiscal policy and one for monetary policy. For the fiscal policy variable, we use the quarterly government expenditure as a percentage of GDP; obtained from the Global Economy database,[39] while the monetary variable corresponds to the International Financial Statistics (IFS)'s quarterly Policy-Related Real Interest Rate (Percent per Annum) collected by the International Monetary Fund.[40] However, there is a data limitation since, when using jointly the fiscal and monetary policy variables described below, we only have 59 events with full information out of the 195 events, which are depicted by shapelets in Tables A1 through A3 in the Online Appendix.

For the econometric analysis, and since both variables are expressed in quarterly data, we compute their growth rate with respect to the previous year, to consider any seasonal pattern in both policies. [41] Yet,

[39] This database can be consulted in the following URL address: https://es.theglobaleconomy.com/

[40] For the monetary policy, we used the real interest rate calculated as $real = \{(1 + nominal)/(1 + inflation)\} - 1$. The inflation rate was calculated using the IFS quarterly Consumer Price Index. It is important to note that the model estimated with the nominal policy rate exhibits strikingly similar results.

[41] That is, for example, for the fiscal variable: $gfiscal_t = (fiscal_t - fiscal_{t-4})/fiscal_{t-4}$.

these growth rates do not reflect policy reactions *per se*. Since the government only realizes that there has been a recession some quarters after it started, we introduce the fiscal and monetary policy variables as changes in the growth rate after the peak in period (t). For instance, five and three quarters further down $(t+5, t+3)$, respectively:

$$Fiscal\ Policy_{t+5} = gfiscal_{t+5} - gfiscal_t$$
$$Monetary\ Policy_{t+3} = gmonetary_{t+3} - gmonetary_t$$

Notice that monetary policy reacts faster than fiscal policy since the latter is restricted by the annual budget, while monetary policy can react as soon as the government (central bank) learns of the recession. We obtain the response periods after the peak using the Akaike Information Criteria (AIC) and the Bayesian Information Criteria (BIC).[42] Also note that, for fiscal policy, a positive value of the variable means that the government is spending more after the crisis started than before; that is, it is acting countercyclically. While, for the monetary policy variable, a positive value means higher interest rates; thus, it is acting procyclically.

Since we aim to explain the shapes of the recessions, we also include the GDP drop (cumulative area below potential GDP), and the GDP gap (distance between potential GDP and the GDP Index) as explanatory variables in the multinomial model:

$$\begin{aligned} Shapes3_{t+10,i} = \beta_0 + \beta_1 GDPdrop_{t+10,i} + \beta_2 GDPgap_{t+2,i} + \beta_3 FiscalPolicy_{t+5,i} \\ + \beta_4 MonetaryPolicy_{t+3,i} + \varepsilon_{t,i} \end{aligned} \quad \ldots(2)$$

where t is the period of the peak, t + 10 is the period in which the recession spell ends, i is the event, and $Shapes3$ refers to the three clusters scheme of shapelets discovered in the previous sections.[43] It is important to note that the GDP gap of enters into the equation in period $t+2$, since we use the gap of $t+1$ to calibrate the shapelet. Moreover, we take P shape as the base category in the multinomial model, since it is the one associated with "severe" recessions.

With this model formulation, we are controlling for GDP drop and GDP gap. Hence, we attempt to measure the impact of countercyclical policies on the shape of the events, and in particular on the nature

[42] The values for the AIC and BIC were calculated for different response periods for both fiscal and monetary policy. With both criteria, we selected the five periods ahead for the fiscal policy and the three periods ahead for the monetary policy, for both the 3 and 4 cluster schemes. These calculations are available in Table A4 and Table A5 of the online Appendix.

[43] In the results, we also present a model with the four different clusters (*Shapes4*). Both models have qualitatively similar results.

of their recovery, irrespectively of the magnitude of the recession. Since we derive the events' shapes from GDP subsequences, there is an implicit time component in the model. However, each observation of the dataset corresponds to a different spell, that is, to different countries and periods. Thus, in essence, the multinomial model is a cross-section analysis. This approach helps to avoid identification issues related to the time dependence of the variables, and to establish empirically a possible causal relationship since the strength of the countercyclical policy is measured ahead of the potential recovery phase.[44]

Table 8 depicts the estimation results which are in line with what is expected.[45] From these, we can note that: i) with a higher GDP drop, the log-odds for shapelets C and SP decrease with respect to the P shapelet, for both the 3 and the 4-cluster schemes; ii) with a bigger GDP gap, the log-odds for all the shapelets increase with respect to the P shapelet for both cluster schemes, this might be due to the fact that bigger gaps mean there is more room for a quicker recovery; iii) with a countercyclical fiscal policy, the log-odds for all the shapelets (i.e., C, SP, or SC) increases with respect to the base category; iv) finally, a procyclical monetary policy results in lower log-odds for all shapelets with respect to the P shapelet; this implies that with countercyclical monetary policy, it would be less likely to end up in a P type of scenario. All in all, we can argue that the empirical evidence shows it is possible to overcome a recession with the use of countercyclical policies.

To check if the limited number of observations is driving the results, we also estimate two additional models. One using only the fiscal policy (with a resulting set of 147 events), and the other one, including the fiscal policy variable and a monetary policy variable using the monthly Money Supply as a percentage of GDP, obtained from the Global Economy (with a resulting set of 103 events).[46] The results are qualitatively similar to the ones presented in Table 8, which we depict in Tables A6 through A8 of the online Appendix.[47] Moreover, when comparing the AIC and BIC, the model with fewer observations but with the policy-rate variable is preferable.

[44] Formally, the problem of causality is not entirely solved using a lagged independent variable. For instance, it could be the case that higher government expenditure, with respect to GDP, in t +5 can be a consequence of an expected recovery of the economy between t+5 and t + 10, and not in the other way around (i.e., from an increase in the government spending / GDP ratio to an observed increase in GDP). An explanation for this unlikely first scenario is that with a future rebound of the economy larger tax revenues can be expected which, in turn, encourages current government spending.

[45] From this table, we only interpret the statistical significance and the sign of the estimated parameters. We analyze below the estimated probability values for each shapelet category with respect to the strength of the fiscal/monetary policies.

[46] Money Supply is the total amount of currency and other liquid instruments circulating in the economy (broad money) divided by GDP. We used monthly data corresponding to the 3rd, 6th, 9th, and 12th months as quarterly data.

[47] Although, two out of five parameters of the Money Supply lose statistical significance.

Table 8: Multinomial results for the 3 and 4-cluster schemes

Variables	Shapes3		Shapes4		
	C	SP	C	SP	SC
GDP drop	-0.124**	-0.219***	-0.124**	-0.205***	-0.370***
	(0.0487)	(0.0514)	(0.0489)	(0.0485)	(0.108)
GDP gap (t+2)	2.384**	2.825***	2.392**	2.705***	3.615***
	(0.987)	(1.015)	(0.991)	(1.018)	(1.086)
Fiscal Policy	20.67**	30.98***	20.69**	31.43***	26.16**
	(8.525)	(10.52)	(8.548)	(10.67)	(11.10)
Monetary Policy	-5.332**	-6.737**	-5.329**	-6.452**	-8.326**
	(2.478)	(2.680)	(2.487)	(2.688)	(3.725)
Constant	-3.454*	-2.923	-3.484*	-2.974	-4.985*
	(1.998)	(2.518)	(2.010)	(2.508)	(2.880)
Observations	59		59		
Pseudo R-squared	0.675		0.635		
AIC	62.431		85.243		
BIC	83.375		116.658		

Notes: Robust standard errors in parentheses, *** *p<0.01,* ** *p<0.05,* * *p<0.1*
The base category is the P shapelet; GDP drop is the cumulative difference of the GDP index vs. the potential GDP calculated with a Hodrick-Prescott filter. GDP gap is the difference between the GDP index and the potential GDP two periods after the peak. *Fiscal Policy* is the difference in growth rates five periods after the peak, a positive value is expansionary. *Monetary Policy* is the difference in growth rates three periods after the peak, a positive value is contractionary. The pseudo-R^2 is McFadden's R^2.

Another robustness test consists in testing whether the 2008 international financial crisis can affect the likelihood of the different shapelets. Thus, we created a dummy equal to one if there was a crisis in either 2007, 2008, or 2009, this is so because the crisis hit different countries at different times. The Table A9 of the Appendix presents these alternative results. The crisis is not significant in the 3-cluster scheme; however, it is highly significant in one of the categories of the 4-cluster scheme. This is because all the SC shapelets in the 4-cluster scheme appeared around the 2008 crisis.[48]

As mentioned at the start of this section, the multinomial model allows the inference of threshold effects of different policies on the shapes of recessions. With the results of Table 8, it is possible to plot the probabilities of different shapelets and their reaction to fiscal/monetary policy. In Figure 7, we present

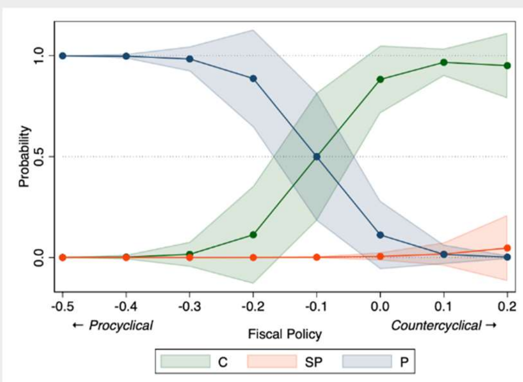
[48] Nevertheless, the 2008 crisis dummy is not significant in models with more events, either 103 or 147.

the predicted probabilities and check whether the probability reversion exists for both policies.[49] When analyzing the response of each shapelet to procyclical fiscal policy (negative values), in both the 3 and 4-cluster schemes, one can notice that there is a higher probability of observing a P event. Then, when moving along the x-axis, as fiscal policy becomes more countercyclical with respect to the peak, a C event increases its probability of occurring. In the case of a monetary policy, with a countercyclical policy (negative values), the C shapelet prevails over the others while if the monetary policy is procyclical, the P shapelet has a higher probability of occurring.[50]

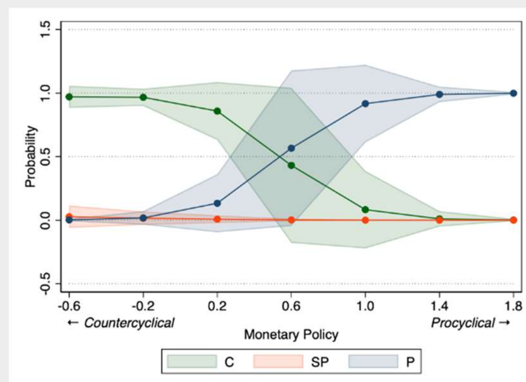
Figure 7: Predicted probabilities of shapelets and Fiscal/Monetary Policy

3-Cluster Scheme (calculated at mean values)

Fiscal Policy

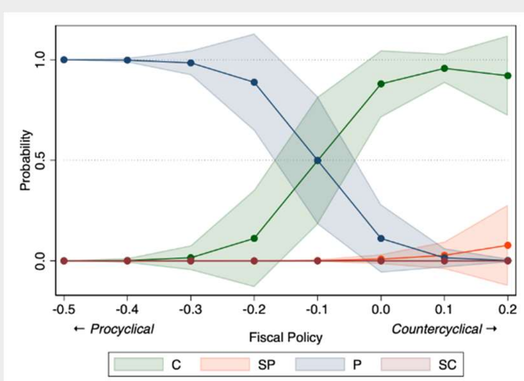


Monetary Policy

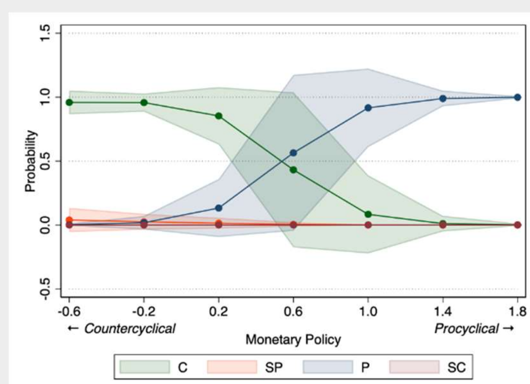


4-Cluster Scheme (calculated at mean values)

Fiscal Policy



Monetary Policy



[49] The shaded areas correspond to 95% confidence intervals.

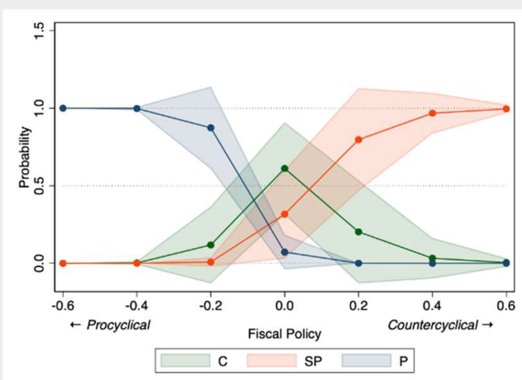
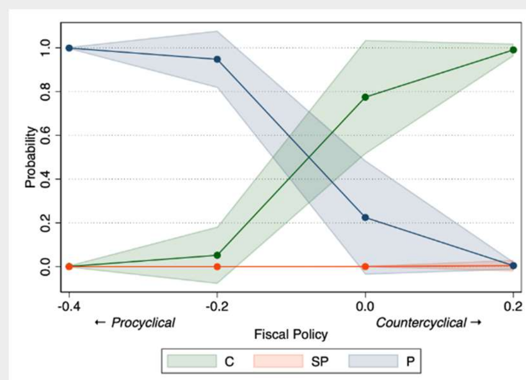
[50] Note that the range of values that is being presented for Figure 7 and Figure 8 is for values that are close to the probability reversion of shapelets. In the dataset, most of the values for fiscal and monetary policies show up within the range presented in both figures.

Notice that, in both policies, there is a threshold where the probabilities for C and P cross each other. However, the probability reversion does not happen at the zero value of the policy. This is due to two reasons, first, the zero value of the independent variable means that there is no change in the policy versus the peak, thus, it does not convey any information regarding the government's response to the crisis. Second, the probability reversion between shapelets does not have to happen at zero, since that value does not necessarily correspond to the "required force" of the policy to generate a different outcome. For example, if we take the top right panel of Figure 7, even if monetary policy is procyclical with respect to the peak at 0.2, it is more likely to result in a C shapelet than in a P shapelet. However, if the monetary policy becomes too procyclical (moving to the right), then the P shapelet is more likely.

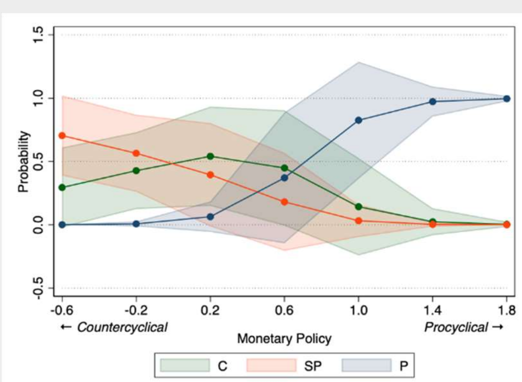
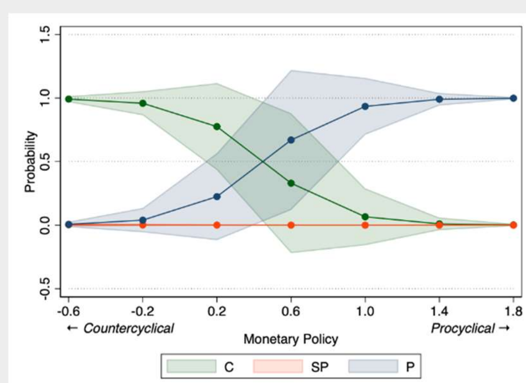
Figure 8: Predicted probabilities of shapelets and Fiscal/Monetary Policy

(Calculated at 1st and 3rd Quartile of GDP drop and GDP gap, 3-cluster scheme)

Fiscal Policy

At 1st Quartile of GDP drop and gap	**At 3rd Quartile of GDP drop and gap**
	

Monetary Policy

At 1st Quartile of GDP drop and gap	**At 3rd Quartile of GDP drop and gap**
	

Moreover, it is important to highlight that the SP and the SC shapelets are not likely to occur in these circumstances. To understand when such shapelets can occur, we explore in Figure 8 the predicted probabilities of shapelets at the first and third quartile values of GDP drop and GDP gap (mild vs severe crisis). In the top panel, we analyze the effect of fiscal policy, and in the bottom panel the effect of monetary policy when evaluated at these quartiles. In the case of fiscal policy, it is also possible to see that for mild recessions, the SP shapelet is more likely to occur when there is a countercyclical policy. Meanwhile, if the crisis is severe, the SP shapelet has virtually zero probability of happening, regardless of the level of fiscal policy response. In the case of monetary policy, the SP shapelet is likely to happen when there is a mild crisis and the policy is highly countercyclical, while with a severe crisis, monetary policy can only change between the C and P shapelets.[51]

9. Conclusions

In this paper, we take advantage of the shapes of GDP subsequences in the data of a large sample of countries to implement a new empirical strategy for testing the influence of countercyclical policies during recessions. Although financial analysts and the media commonly refer to letters and symbols to describe the shapes of recessions, these graphical descriptions have not been previously used to study how effective are monetary and fiscal policies in overcoming these events. We argue that this approach is convenient since it precludes methodological complications in the econometric analysis. Likewise, through the estimation of a multinomial model, we can discover the existence of nonlinear effects when applying intervention policies.

With a sample of 195 recession events in 95 countries, we use a customized machine learning technique to discover and classify these events in ten different clusters according to representative shapes of GDP subsequences of eleven quarters (or shapelets in terms of the ML semantics). Moreover, using formal techniques, we reduce the statistically distinguishable clusters in the database to only 3 or 4 shapelets. Three of these clusters tend to exhibit mild recessions (SC for describing a short checkmark shape, SP for describing a 'spoon' silhouette for a right-handed, and C for describing a 'cup' silhouette) in terms of several features: duration, depth (maximum loss), cumulative GDP drop, and mean drop, being

[51] As for the SC shapelet, its occurrence is quite similar to the SP shapelet, since all the occurrences in the 4-cluster scheme replace SP shapelets.

the last three concepts measured in relation to potential GDP; while the fourth one (P for describing a 'pan' silhouette for a left-handed) characterizes events with severe recessions.

Using these classifications, we estimate a multinomial model in which the clusters-shapelets are used as the categorical variable. Either with 3 or 4 shapes, we find strong evidence in favor of countercyclical policies moving the economy from a setting of severe recession to one of a mild recession, being fiscal policies more effective than monetary policies in statistical terms. This implies that once the fiscal accommodative policy is implemented (i.e., growth difference in the ratio of government expenditure to GDP), five quarters after the pre-recession peak, it is more likely to observe signs of recovery (i.e., a 'cup' than a 'pan' shapelet for some countries and a 'spoon' than a 'pan' for others). The response to monetary policies (i.e., growth difference in the policy rate) is swifter than the response to fiscal policies since the former interventions show an influence on the GDP sequences of only three quarters after the pre-recession peak. Because of data limitations for calculating the strength of the monetary intervention, we obtain the above results for a sample of 59 events and 41 countries. However, the efficacy of the fiscal policy prevails when we omit the 'policy rate' variable (147 events in 77 countries) or when we use "broad money" instead (103 events in 62 countries).

We calculate the strength of fiscal/monetary policies as the difference in their annual growth rate at the intervention period with respect to the rate observed at the pre-recession peak. Therefore, we can observe, in the data, negative or positive values in these differences. Accordingly, in some of the recession events, we find positive (values) for the fiscal intervention and negative values for the monetary interventions; that is, in many countries, countercyclical policies were implemented during the sampling period. However, there is also evidence of procicallity, in which some countries mismanaged the crisis and make it more severe. Finally, the model's estimates show that there are thresholds that need to be crossed for making a countercyclical policy more likely to exhibit signs of recovery in the GDP than further downfalls or a prolonged stagnation.

References

Abiad, A. G., Balakrishnan, R, Brooks, P. K., Leigh, D. & Tytell I. (2014). What's the damage? Medium-term output dynamics after banking crises. In *Financial crises: Causes, consequences, and policy responses*, edited by Stijn Claessens, S. Kose, M. A., Laeven, L. & Valencia, F. 277-308, Washington DC. International Monetary Fund.

Aguiar, M., & Gopinath, G. (2007). Emerging market business cycles: The cycle is the trend. *Journal of Political Economy*, *115*(1), 69–102.

Álvarez Ondina, P., Pérez Rivero, J. L., Vicente Queijeiro, S. de, & Vicente Cuervo, M. R. (2011). The determinants of the choice of exchange rate regimes in Latin America: A mixed multinomial logit approach. *Cuadernos de Economía*, *34*(95), 55–61. https://doi.org/10.1016/S0210-0266(11)70006-1

Avdjiev, S., & Zeng, Z. (2014). Credit growth, monetary policy and economic activity in a three-regime TVAR model. *Applied Economics*, *46*(24), 2936–2951. https://doi.org/10.1080/00036846.2014.916391

Ball, L. (2014). Long-term damage from the Great Recession in OECD countries. *European Journal of Economics and Economic Policies: Intervention*, *11*(2), 149–160.

Barro, R. (1991). Economic growth in a cross-section of countries. *Quarterly Journal of Economics*, 106(2), 407-443.

Bec, F., Bouabdallah, O., & Ferrara, L. (2011). The possible shapes of recoveries in markov-switching Models. *Banque de France Working Paper* No. 321. Available at SSRN: https://doi.org/10.2139/ssrn.1783585

Becker, T., & Mauro, P. (2006). Output drops and the shocks that matter. *IMF Working Paper* No. 06/172. Available at SSRN: https://ssrn.com/abstract=926227

Ben-David, D., & Papell, D. H. (1998). Slowdowns and meltdowns: Postwar growth evidence from 74 countries. *Review of Economics and Statistics*, *80*(4), 561–571.

Bertola, G., & Drazen, A. (1993). Trigger Points and Budget Cuts: Explaining the Effects of Fiscal Austerity. *The American Economic Review*, *83*(1), 11–26.

Bijsterbosch, M., & Dahlhaus, T. (2015). Key features and determinants of credit-less recoveries. *Empirical Economics*, *49*(4), 1245–1269. https://doi.org/10.1007/s00181-014-0910-y

Blanchard, O. J. (1990). *Suggestions for a New Set of Fiscal Indicators*. OECD Economics Department Working Papers No. 79, OECD Publishing, Paris. https://doi.org/10.1787/435618162862

Blanchard, O.J., Cerutti, E. & Summers. L. (2015). Inflation and activity – Two explorations
and their monetary policy implications. *IMF Working Paper*, WP/15/230, International Monetary Fund, Washington DC.

Bui, D.-T. (2018). Nonlinear effects of fiscal policy on national saving: Empirical evidence from emerging Asian economies. *Journal of Asian Business and Economic Studies*, *25*(1), 2–14. https://doi.org/10.1108/JABES-04-2018-0001

Calvo, G. A., Izquierdo, A., & Talvi, E. (2006). Sudden Stops and Phoenix Miracles in Emerging Markets. *American Economic Review*, *96*(2), 405–410. https://doi.org/10.1257/000282806777211856

Calvo, G. A., & Talvi, E. (2005). *Sudden Stop, Financial Factors and Economic Collpase in Latin America: Learning from Argentina and Chile* (Working Paper No. 11153; Working Paper Series). National Bureau of Economic Research. https://doi.org/10.3386/w11153

Castañeda, G., & Castro Peñarrieta, L. (2021). A Customized Machine Learning Algorithm for Discovering the Shapes of Recovery. Was the Global Financial Crisis Different?. *Working Paper CIDE,* Mexico City.

Castañeda, G., Pietronero, L., Romero-Padilla, J. & Zaccaria, A. (2021). The complex dynamic of growth: Fitness and the different patterns of economic activity in the medium and long terms. *Working Paper, CIDE*, Mexico City.

Cavdar, S. C., & Aydin, A. D. (2015). Understanding the Factors Behind Current Account Deficit Problem: A Panel Logit Approach on 16 OECD Member Countries. *Procedia Economics and Finance*, *30*, 187–194. https://doi.org/10.1016/S2212-5671(15)01282-4

Cerra, M. V., Fatás, A., & Saxena, M. S. C. (2020). *Hysteresis and business cycles*. International Monetary Fund.

Cerra, V., Panizza, U., & Saxena, S. C. (2013). International evidence on recovery from recessions. *Contemporary Economic Policy*, *31*(2), 424–439.

Cerra, V., & Saxena, S. C. (2008). Growth dynamics: The myth of economic recovery. *American Economic Review*, *98*(1), 439–457.

Challet, D., & Morton de Lachapelle, D. (2013). A robust measure of investor contrarian behaviour. In F. Abergel, B. K. Chakrabarti, A. Chakraborti, & A. Ghosh (Eds.), *Econophysics of Systemic Risk and Network Dynamics* (pp. 105–118). Springer Milan. https://doi.org/10.1007/978-88-470-2553-0_7

Chen, W., Mrkaic, M. & Nabar. M. (2019). The global economic recovery 10 years after the 2008 financial crisis. *IMF Working Paper*, WP/19/83, International Monetary Fund, Washington, DC.

Coenen, G., Erceg, C. J., Freedman, C., Furceri, D., Kumhof, M., Lalonde, R., Laxton, D., Lindé, J., Mourougane, A., & Muir, D. (2012). Effects of fiscal stimulus in structural models. *American Economic Journal: Macroeconomics*, *4*(1), 22–68.

Comelli, M. F. (2014). Comparing the performance of logit and probit early warning systems for currency crises in emerging market economies. *IMF Working Paper* WP/14/65. International Monetary Fund.

Drazen, A. 1990. Can severe fiscal contractions be expansionary? Tales of two small European countries: comment. *NBER Macroeconomics Annual*. 5(1), 117-122, https://doi.org/10.1086/654133

Dubas, J. M., Lee, B.-J., & Mark, N. C. (2010). A multinomial logit approach to exchange rate policy classification with an application to growth. *Journal of International Money and Finance*, *29*(7), 1438–1462. https://doi.org/10.1016/j.jimonfin.2010.03.010

Dueker, M. J. (2006). Using cyclical regimes of output growth to predict jobless recoveries. *Federal Reserve Bank of St. Louis Review*. 88(2), 145-153.

Falcetti, E., & Tudela, M. (2006). Modelling Currency Crises in Emerging Markets: A Dynamic Probit Model with Unobserved Heterogeneity and Autocorrelated Errors. *Oxford Bulletin of Economics and Statistics*, *68*(4), 445–471. https://doi.org/10.1111/j.1468-0084.2006.00172.x

Fatás, A. (2002). The effects of business cycles on growth. *Central Bank of Chile Working Papers*, *156*.

Fazzari, S. M., Morley, J., & Panovska, I. (2015). State-dependent effects of fiscal policy. *Studies in Nonlinear Dynamics & Econometrics*, *19*(3), 285–315. https://doi.org/10.1515/snde-2014-0022

Feldstein, M. (1982). Government deficits and aggregate demand. *Journal of Monetary Economics*, *9*(1), 1–20. https://doi.org/10.1016/0304-3932(82)90047-2

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.

Furceri, D., & Mourougane, A. (2012). The effect of financial crises on potential output: New empirical evidence from OECD countries. *Journal of Macroeconomics*, *34*(3), 822–832. https://doi.org/10.1016/j.jmacro.2012.05.010

Furceri, D., & Zdzienicka, A. (2012). How costly are debt crises? *Journal of International Money and Finance*, *31*(4), 726–742.

Giavazzi, F., Jappelli, T., & Pagano, M. (2000). Searching for non-linear effects of fiscal policy: Evidence from industrial and developing countries. *European Economic Review*, *44*(7), 1259–1289. https://doi.org/10.1016/S0014-2921(00)00038-6

Gregory, V., Menzio, G., & Wiczer, D. G. (2020). *Pandemic Recession: L or V-Shaped?* National Bureau of Economic Research (NBER) Working Paper No. 27105. https://www.nber.org/papers/w27105

Haltmaier, J. (2013). Do recessions affect potential output? *FRB International Finance Discussion Paper*, *1066*.

Hamida, A. (2018). Currency Crises and Growth. *International Journal of Economics and Finance*, *10*(3), 83–92.

Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, *57*(2), 357–384. https://doi.org/10.2307/1912559

Harding, D., & Pagan, A. (2002). Dissecting the cycle: A methodological investigation. *Journal of Monetary Economics*, *49*(2), 365–381. https://doi.org/10.1016/S0304-3932(01)00108-8

Hausmann, R., Rodriguez, F. R., & Wagner, R. A. (2006). *Growth Collapses*. KSG Working Paper No. RWP06-046, CID Working Paper No. 136. Available at SSRN: https://doi.org/10.2139/ssrn.902392

Hausmann, R., Rodrik, D., & Velasco, A. (2008). Growth Diagnostics. *The Washington Consensus Reconsidered: Towards a New Global Governance*, 324–355.

Howard, G., Martin, R., & Wilson, B. A. (2011). *Are Recoveries from Banking and Financial Crises Really so Different?* The Federal Reserve Board of Governors in Washington DC. International Finance Discussion Papers (IFDP).

Jackson, L. E., Owyang, M.T., & Soques. D. 2018. Nonlinearities, smoothing and countercyclical monetary policy, *Journal of Economic Dynamics & Control*, 95, 136-154 https://doi.org/10.1016/j.jedc.2018.08.007

Jemović, M., & Marinković, S. (2021). Determinants of financial crises—An early warning system based on panel logit regression. *International Journal of Finance & Economics*, *26*(1), 103–117. https://doi.org/10.1002/ijfe.1779

Kannan, P., Scott, A., & Terrones, M. E. (2014). From recession to recovery: How soon and how strong. *Financial Crises: Causes, Consequences, and Policy Responses*, *239*.

Kim, C.-J., Morley, J., & Piger, J. (2005). Nonlinearity and the permanent effects of recessions. *Journal of Applied Econometrics*, *20*(2), 291–309. https://doi.org/10.1002/jae.831

Liesenfeld, R., Moura, G. V., & Richard, J.-F. (2010). Determinants and Dynamics of Current Account Reversals: An Empirical Analysis. *Oxford Bulletin of Economics and Statistics*, *72*(4), 486–517. https://doi.org/10.1111/j.1468-0084.2010.00588.x

Loayza, N., & Hnatkovska, V. V. (2004). *Volatility and Growth*. Social Science Research Network (SSRN). https://papers.ssrn.com/abstract=636604

Maria Milesi-Ferretti, G., & Razin, A. (1998). Sharp reductions in current account deficits An empirical analysis. *European Economic Review*, *42*(3), 897–908. https://doi.org/10.1016/S0014-2921(97)00124-4

Martin, R., Sunley, P., & Tyler, P. (2015). *Local growth evolutions: Recession, resilience and recovery*. Oxford University Press UK.

McGowan, C. B., & Moeller, S. E. (2005). *Using multinomial logistic regression of political and economic risk variables for country risk analysis of foreign direct investment decisions*. Southwestern Economic Proceedings.

Milesi-Ferretti, G. M., & Razin, A. (1998). Sharp reductions in current account deficits An empirical analysis. *European Economic Review*, *42*(3), 897–908. https://doi.org/10.1016/S0014-2921(97)00124-4

Mora, R., & Siotis, G. (2005). External factors in emerging market recoveries: An empirical investigation. *European Economic Review*, *49*(3), 683–702.

Morales-Zumaquero, A., & Sosvilla-Rivero, S. (2015). Growth dynamics, financial crises and exchange rate regimes. *Applied Economics Letters*, *22*(10), 767–771.

Morley, J., & Piger, J. (2012). The Asymmetric Business Cycle. *The Review of Economics and Statistics*, *94*(1), 208–221. https://doi.org/10.1162/REST_a_00169

Oulton, N., & Sebastia-Barriel, M. (2013). *Long and Short-Term Effects of the Financial Crisis on Labour Productivity, Capital and Output*. Bank of England Working Paper No. 470. Available at SSRN: https://doi.org/10.2139/ssrn.2206347

Papaioannou, M. M. G. (2003). *Determinants of the Choice of Exchange Rate Regimes in Six Central American Countries: An Empirical Analysis*. International Monetary Fund.

Pritchett, L. (2000). Understanding patterns of economic growth: Searching for hills among plateaus, mountains, and plains. *The World Bank Economic Review*, *14*(2), 221–250.

Ramey, G., & Ramey, V. A. (1991). Technology commitment and the cost of economic fluctuations. *NBER Working Paper*, *w3755*.

Reinhart, C. M., & Rogoff, K. S. (2009). The aftermath of financial crises. *American Economic Review*, *99*(2), 466–472.

Romer, C. D., & Romer, D. H. (1989). Does monetary policy matter? A new test in the spirit of Friedman and Schwartz. *NBER Macroeconomics Annual*, *4*, 121–170.

Saldías, M. (2017). *The Nonlinear Interaction Between Monetary Policy and Financial Stress*. International Monetary Fund.

Schularick, M., & Taylor, A. M. (2012). Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crises, 1870-2008. *American Economic Review*, *102*(2), 1029–1061. https://doi.org/10.1257/aer.102.2.1029

Sichel, D. E. (1994). Inventories and the Three Phases of the Business Cycle. *Journal of Business & Economic Statistics*, *12*(3), 269–277. https://doi.org/10.1080/07350015.1994.10524542

Teimouri, S. (2012). *Macroeconomic Recovery after a Currency Crisis* [Ph.D.]. West Virginia University.

Terrones, M. E. (2020). Do fixers perform worse than non-fixers during global recessions and recoveries? *Journal of International Money and Finance*, *104*, 102–160.

Teulings, C. N., & Zubanov, N. (2014). Is economic recovery a myth? Robust estimation of impulse responses. *Journal of Applied Econometrics*, *29*(3), 497–514.

Verbeek, M. (2012). *A Guide to Modern Econometrics* (4th edition). John Wiley & Sons.

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

Ye, L., & Keogh, E. (2009). Time series shapelets: A new primitive for data mining. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 947–956.

Online Supplementary Material

Appendix

Table A1: Classification scheme with 3 and 4 shapelet-clusters

(59 events, 41 countries)

Country	Time	Episode	Shapes3	Shapes4	Country	Time	Episode	Shapes3	Shapes4
Albania	2009q2	ALB1	C	C	Japan	2008q1	JPN3	P	P
Albania	2013q1	ALB2	C	C	Qatar	2014q2	QAT2	P	P
Armenia	2008q3	ARM2	C	C	Romania	1996q3	ROU1	P	P
Australia	1981q3	AUS1	C	C	Romania	2008q3	ROU2	P	P
Azerbaijan	2008q1	AZE1	C	C	Russia	2008q2	RUS1	P	P
Azerbaijan	2011q4	AZE2	C	C	Russia	2014q2	RUS2	P	P
Azerbaijan	2014q3	AZE3	C	C	Serbia	2008q1	SRB2	P	P
Bulgaria	2008q4	BGR3	C	C	Sweden	2007q4	SWE3	P	P
Canada	2008q3	CAN2	C	C	United Kingdom	1973q2	GBR1	P	P
Czech Republic	2008q3	CZE2	C	C	United Kingdom	1979q4	GBR2	P	P
Czech Republic	2011q4	CZE3	C	C	United Kingdom	2008q1	GBR4	P	P
Malaysia	2008q2	MYS1	C	C	Brazil	2008q3	BRA1	SP	SC
Mexico	2008q3	MEX2	C	C	Costa Rica	2008q3	CRC1	SP	SC
New Zealand	2007q4	NZL3	C	C	Peru	2008q3	PER3	SP	SC
Norway	2008q4	NOR3	C	C	Philippines	2008q3	PHL3	SP	SC
Serbia	2012q2	SRB3	C	C	Rwanda	2008q4	RWA1	SP	SC
Singapore	2000q4	SGP2	C	C	Thailand	2008q3	THA2	SP	SC
Singapore	2007q3	SGP3	C	C	Armenia	2013q1	ARM3	SP	SP
Turkey	2008q1	TUR2	C	C	Chile	2008q2	CHL1	SP	SP
United Kingdom	1990q2	GBR3	C	C	Guatemala	2012q4	GTM1	SP	SP
United States	2008q2	USA4	C	C	Hungary	2011q4	HUN2	SP	SP
Belarus	2014q3	BLR4	P	P	Kyrgyzstan	2014q4	KGZ4	SP	SP
Brazil	2014q4	BRA2	P	P	Mauritius	2008q1	MUS1	SP	SP
Denmark	2007q4	DNK1	P	P	Mongolia	2014q3	MNG2	SP	SP
Hong Kong	1997q2	HKG3	P	P	Norway	1987q4	NOR2	SP	SP
Hong Kong	2007q4	HKG4	P	P	Singapore	2012q1	SGP4	SP	SP
Hungary	2008q2	HUN1	P	P	South Africa	2008q3	ZAF4	SP	SP
Iceland	2007q4	ISL1	P	P	Switzerland	2008q3	CHE3	SP	SP
Indonesia	1997q3	IDN1	P	P	Thailand	2012q4	THA3	SP	SP
Israel	2000q3	ISR1	P	P					

Table A2: Classification scheme with 3 shapelet-clusters (P, C, SP)

(59 events, 41 countries)

P	C	SP
BLR4, BRA2, DNK1, HKG3, HKG4, HUN1, ISL1, IDN1, ISR1, JPN3, QAT2, ROU1, ROU2, RUS1, RUS2, SRB2, SWE3, GBR1, GBR2, GBR4	ALB1, ALB2, ARM2, AUS1, AZE1, AZE2, AZE3, BGR3, CAN2, CZE2, CZE3, MYS1, MEX2, NZL3, NOR3, SRB3, SGP2, SGP3, TUR2, GBR3, USA4	ARM3, BRA1, CHL1, CRC1, GTM1, HUN2, KGZ4, MUS1, MNG2, NOR2, PER3, PHL3, RWA1, SGP4, ZAF4, CHE3, THA2, THA3
20	21	18

Table A3: Classification scheme with 4 shapelet-clusters (P, SC, C, SP)

(59 events, 41 countries)

P	SC	C	SP
BLR4, BRA2, DNK1, HKG3, HKG4, HUN1, ISL1, IDN1, ISR1, JPN3, QAT2, ROU1, ROU2, RUS1, RUS2, SRB2, SWE3, GBR1, GBR2, GBR4	BRA1, CRC1, PER3, PHL3, RWA1, THA2	ALB1, ALB2, ARM2, AUS1, AZE1, AZE2, AZE3, BGR3, CAN2, CZE2, CZE3, MYS1, MEX2, NZL3, NOR3, SRB3, SGP2, SGP3, TUR2, GBR3, USA4	ARM3, CHL1, GTM1, HUN2, KGZ4, MUS1, MNG2, NOR2, SGP4, ZAF4, CHE3, THA3
20	6	21	12

Table A4: Tabulation of information criteria for different policy response periods (3 cluster – scheme)

Fiscal Policy	*Monetary Policy*					
	Criteria	*t+3*	*t+4*	*t+5*	*t+6*	*t+7*
t+3	AIC	73.044	75.583	76.561	78.022	77.504
	BIC	93.820	96.359	97.336	98.797	98.280
t+4	AIC	66.885	69.331	69.880	72.226	72.354
	BIC	87.661	90.106	90.656	93.002	93.129
t+5	AIC	62.110	64.938	65.841	69.397	69.804
	BIC	82.886	85.711	86.617	90.172	90.580
t+6	AIC	63.713	68.179	71.733	76.799	77.723
	BIC	84.489	88.955	92.508	97.574	98.498
t+7	AIC	67.048	73.553	77.251	80.793	80.983
	BIC	87.824	94.328	98.026	101.569	101.758

Table A5: Tabulation of information criteria for different policy response periods (4 cluster – scheme)

Fiscal Policy	*Monetary Policy*					
	Criteria	*t+3*	*t+4*	*t+5*	*t+6*	*t+7*
t+3	AIC	96.173	97.259	98.180	100.943	101.335
	BIC	127.336	128.422	129.343	132.106	132.498
t+4	AIC	90.309	90.895	91.351	94.906	95.892
	BIC	121.472	122.058	122.514	126.069	127.055
t+5	AIC	85.623	86.786	87.575	92.204	93.426
	BIC	116.786	117.949	118.738	123.367	124.589
t+6	AIC	86.133	88.787	92.017	97.950	99.522
	BIC	117.296	119.951	123.180	129.113	130.685
t+7	AIC	89.427	94.632	98.142	102.488	103.216
	BIC	120.590	125.795	129.306	133.651	134.379

Table A6: Tabulation of events in different models

Shapelet	*Using fiscal policy alone*		*Using fiscal policy and money supply*	
	Shapes3	***Shapes4***	***Shapes3***	***Shapes4***
C	32	32	24	24
SP	59	44	40	29
P	56	56	39	39
SC		15		11
Total	147	147	103	103

Table A7: Results using only fiscal policy (147 events)

Variables	**Shapes3**		**Shapes4**		
	C	SP	C	SP	SC
GDP drop	-0.0829**	-0.170***	-0.0830**	-0.165***	-0.311***
	(0.0397)	(0.0444)	(0.0396)	(0.0443)	(0.0567)
GDP gap (t+2)	1.529***	2.349***	1.532***	2.307***	2.857***
	(0.481)	(0.544)	(0.480)	(0.541)	(0.575)
Fiscal Policy	6.358***	5.567***	6.428***	5.710***	6.622***
	(1.957)	(1.942)	(2.025)	(2.178)	(1.825)
Constant	-0.862	-0.363	-0.873	-0.515	-0.852
	(1.157)	(0.850)	(1.164)	(0.867)	(1.114)
Observations	147	147	147	147	147
Pseudo R-squared	0.611	0.611	0.564	0.564	0.564
AIC	138.0409		189.8745		
BIC	161.9644		225.7597		

Notes: Robust standard errors in parentheses, *** *p<0.01,* ** *p<0.05,* * *p<0.1*
The base category is the "P" shapelet; GDP drop is the cumulative difference of the GDP index vs. the potential CDP calculated with a Hodrick-Prescott filter. GDP gap is the difference between the GDP index and the potential GDP two periods after the peak. *Fiscal Policy* is the difference in growth rates five periods after the peak, a positive value is expansionary. The pseudo-R^2 is McFadden's R^2.

Table A8: Results using fiscal policy and money supply (103 events)

Variables	Shapes3		Shapes4		
	C	SP	C	SP	SC
GDP drop	-0.236***	-0.342***	-0.236***	-0.334***	-0.528***
	(0.0771)	(0.0876)	(0.0773)	(0.0876)	(0.109)
GDP gap (t+2)	3.948***	5.047***	3.950***	4.985***	5.693***
	(1.212)	(1.283)	(1.214)	(1.284)	(1.308)
Fiscal Policy	17.30***	22.79***	17.27***	22.98***	18.15**
	(6.048)	(7.186)	(6.049)	(7.190)	(7.705)
Monetary Policy	-4.464	-12.40*	-4.511	-12.05*	-16.24**
	(6.192)	(7.073)	(6.183)	(7.050)	(7.641)
Constant	0.760	0.616	0.764	0.446	0.453
	(1.364)	(1.343)	(1.366)	(1.353)	(1.689)
Observations	103		103		
Pseudo R-squared	0.728		0.668		
AIC	80.30851		119.0812		
BIC	106.6558		158.6021		

Notes: Robust standard errors in parentheses, *** *p<0.01,* ** *p<0.05,* * *p<0.1*
The base category is the "P" shapelet; GDP drop is the cumulative difference of the GDP index vs. the potential CDP calculated with a Hodrick-Prescott filter. GDP gap is the difference between the GDP index and the potential GDP two periods after the peak. *Fiscal Policy* is the difference in growth rates five periods after the peak, a positive value is expansionary. *Monetary Policy* is the difference in growth rates three periods after the peak, a positive value is contractionary. The pseudo-R^2 is McFadden's R^2.

Table A9: Results using 2008 crisis dummy (59 events)

Variables	**Shapes3**		**Shapes4**		
	C	SP	C	SP	SC
GDP drop	-0.148**	-0.243***	-0.149**	-0.228***	-1.704***
	(0.0648)	(0.0666)	(0.0650)	(0.0645)	(0.0809)
GDP gap (t+2)	2.942**	3.460**	2.962**	3.266**	14.49***
	(1.313)	(1.384)	(1.316)	(1.363)	(1.395)
Fiscal Policy	23.77**	34.10***	23.90**	33.02***	98.49***
	(10.11)	(12.15)	(10.16)	(11.73)	(11.97)
Monetary Policy	-8.722*	-10.78**	-8.675*	-11.14**	-12.96**
	(4.909)	(5.205)	(4.945)	(5.262)	(5.408)
Crisis	-2.749	-3.416	-2.691	-3.805	82.47***
	(2.854)	(3.141)	(2.877)	(3.068)	(3.685)
Constant	-3.609**	-3.253	-3.673**	-2.903	-118.3***
	(1.625)	(2.293)	(1.658)	(2.356)	(3.853)
Observations	59		59		
Pseudo R-squared	0.693		0.748		
AIC	63.72909		74.31288		
BIC	88.65954		111.7086		

Notes: Robust standard errors in parentheses, *** *p<0.01,* ** *p<0.05,* * *p<0.1*
The base category is the "P" shapelet; GDP drop is the cumulative difference of the GDP index vs. the potential CDP calculated with a Hodrick-Prescott filter. GDP gap is the Difference between the GDP index and the potential GDP two periods after the peak. *Fiscal Policy* is the difference in growth rates five periods after the peak, a positive value is expansionary. *Monetary Policy* is the difference in growth rates three periods after the peak, a positive value is contractionary. The pseudo-R^2 is McFadden's R^2.

Table A10: Average values within cluster for the events' features in the 195 countries sample

Feature / Shapelet	Duration	Depth	Cumulative Loss	Mean	Variance	Kurtosis	Skewness
SC	4.727	8.690	24.287	5.074	19.434	2.040	-0.408
SP	6.639	10.290	40.714	6.059	20.119	2.302	-0.548
C	9.553	16.187	101.358	10.021	45.375	2.574	-0.647
P	11.000	21.424	136.494	12.408	111.654	1.944	-0.452

Notes: 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. b) Depth: maximum loss of GDP relative to potential GDP during the recession event. 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 (i.e., or 11 quarter 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 others (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).