UNIVERSITA' DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI
"M.FANNO"

CORSO DI LAUREA MAGISTRALE IN
ECONOMICS AND FINANCE

TESI DI LAUREA

DATING AND FORECASTING THE G7 BUSINESS CYCLE

RELATORE:

CH.MO PROF. Giovanni Caggiano

LAUREANDO/A: Kristian Racocha

MATRICOLA N. 2051110

ANNO ACCADEMICO 2022 – 2023

Il candidato dichiara che il presente lavoro è originale e non è già stato sottoposto, in tutto o in parte, per il conseguimento di un titolo accademico in altre Università italiane o straniere.
Il candidato dichiara altresì che tutti i materiali utilizzati durante la preparazione dell'elaborato sono stati indicati nel testo e nella sezione "Riferimenti bibliografici" e che le eventuali citazioni testuali sono individuabili attraverso l'esplicito richiamo alla pubblicazione originale.

The candidate declares that the present work is original and has not already been submitted, totally or in part, for the purposes of attaining an academic degree in other Italian or foreign universities. The candidate also declares that all the materials used during the preparation of the thesis have been explicitly indicated in the text and in the section "Bibliographical references" and that any textual citations can be identified through an explicit reference to the original publication.

Firma dello studente

Table of Contents

1 Introduction

The determination of business cycles has been at the forefront of economic research in recent years. This trend has only been exacerbated by the Global Financial Crisis in 2007 and the COVID-19 related recession in 2020. These two recessions were not constrained to a single country, rather they were global recessions. Moreover, the increased level of globalization puts a greater emphasis on the analysis of the global business cycle. But what constitutes a recession? And how is the business cycle measured?

Jorda & Berge (2011) summarize the four basic definitions of a recession. The first two rely solely on the growth rate of GDP and are considered to be the "rule-of-thumb" definitions. One views a recession as any period in which GDP growth is negative, which results in a very noisy series of recessions, lasting a very short time but occurring at a high frequency.

The other requires two consecutive quarters of negative GDP growth. This definition is generally the preferred method used by a majority of countries, politicians, and media across the world due to its simplicity. This rule is much more conservative but at times fails at identification. Consider a scenario where the economy is in a recession but experiences a temporary reversal, resulting in one quarter of positive growth. If only one negative quarter was observed prior to this period, the recession would be shifted by six months or missed completely as was the case in the 2001 recession in the United States. Moreover, quarterly GDP statistics are subject to large revisions. For example, when the official channels announced that the United States have entered into a recession at the start of 2008, quarterly GDP growth was still positive. Therefore, due to the overreliance on the GDP growth rate and the occasional misses, a more complex definition is preferred to identify recessions with greater confidence.

The second two definitions provided by Jorda & Berge (2011) rely on an indicator of economic activity, which is used to date peaks and troughs of the business cycle. The period between a peak and a trough is classified as a recession, and conversely, the period between a trough and a peak is deemed to be an expansion. The trough months are generally accepted to be included in the recession period. However, the inclusion of the peak months depends on the particular research. The peak-to-trough definition is used among most researchers and supranational organizations tasked with dating recessions.

The Business Cycle Dating Committee (BCDC) of the National Bureau of Economic Research (NBER)[1] is an organization formed in 1978[2] to establish the historical chronology of U.S. business cycle and is one of the main institutions cited in the relevant literature. Their definition of a recession is the following:

> *A significant decline in economic activity that is spread across the economy and that lasts more than a few months.*
>
> – BUSINESS CYCLE DATING. BUSINESS CYCLE DATING COMMITTEE OF THE NATIONAL BUREAU OF ECONOMIC RESEARCH.

The European equivalence of the BCDC is the Euro Area Business Cycle Dating Committee of the Centre for Economic Policy Research[3] (CEPR) founded in 2002. Their definition is very similar and also relies on indicators of economic activity. For CEPR, the task of dating turning points may be considered even more complex than that of the NBER since it must identify cyclical behavior in the multi-country context of the euro area.

As demonstrated by Kose et al. (2020), two approaches are employed to identify turning points of a business cycle: a judgmental method and a statistical method. As business cycles do not behave in a regular manner, despite historical efforts attempting to find periodicity (Juglar, 1862; Kitchin, 1923; Kondratiev, 1925; Mitchell, 1927; Schumpeter, 1939), the official dating committees rely on their judgement to better evaluate evidence presented by models and composite indicators. Since the dating committees are governmental organizations, there is great emphasis put on confidence and precision. In the case of NBER, this process generally takes between 4 and 21 months to eliminate any doubts about a turning point's existence. But once the recession is identified, the decision is not typically revised.

The statistical method has its origins in the work of Burns and Mitchell (1938), who developed a list of leading, coincident, and lagging indicators of economic activity in the United States. These indicators have since served an important role in dating and forecasting macroeconomic

[1] https://www.nber.org/

[2] NBER itself was founded in 1920.

[3] https://eabcn.org/

activity. Stock and Watson (1989) build on their work to develop monthly indexes of coincident economic indicators (CEI) and leading economic indicators. Moreover, they try to summarize them into a recession probability index (Stock & Watson, 1991, 1993). Hamilton (1989) also tackles the question as a formal statistical issue through a Markov-switching model and finds evidence of asymmetries in cyclical expansions and contractions and establishes the differences in the dynamics of business cycle phases.

Diebold and Rudebusch (1996) suggest a combination of the modeling framework and the coincident index. This task is then taken on by Chauvet (1998) who builds an alternative coincident index to identify business cycle turning points and automate the NBER's assessment process. After these initial efforts, the statistical method became subject to a wide range of research with the aim to create the optimal indicator. The indicators differ in their inputs, weighting structures, and coverage.

A particularly impactful paper was presented by Burns and Mitchell (1946), who identified turning points in a number of series and searched for a common date, i.e., an aggregate turning point. Stock and Watson (2014) automate this mechanism in the aim of dating the U.S. business cycle. They use a large, disaggregated dataset of the most common coincident monthly measurements of economic activity: production, sales, income, and employment. They use the Bry-Boschan (1971) procedure to identify recessions and individual series and determine an aggregate turning point conditional on an NBER date occurring. (See Section 3.1). They were able to match the NBER dates with great precision, which makes this methodology very intriguing for the purpose of studying global business cycle. It is well characterized by Fushing et al. (2010), who develop business cycle chronologies for 22 OECD nations without an official dating committee. Instead of finding common factors that explain fluctuations of economic activity across many countries, one must focus on finding business cycles. They follow by aggregating the chronologies to find the turning points in the global business cycle.

Although the total GDP of the G7 has declined from over 40% of the World's GDP in 2000 to approximately 30% in 2023 due to China's economic development, the country group remains a highly influential in terms of the global economy (Statista Research Department, 2023). Aruoba et al. (2011) implement a framework for characterizing and monitoring the global business cycle by focusing on the G7. They show that the common G7 real activity factor is able to track the major global cyclical events since the 1970s. Among other papers examining the dynamics of the G7 business cycle Bodman and Crosby (2005) study interconnectedness of the country-specific cycles and de Bondt and Vermeulen (2021) examine the effect of negative spillovers from one member to another. Enea et al. (2015) find a stronger degree of synchronization during recessions among the G7 members, with expansions following a continent-specific cycle.

There are some institutions, which provide indicators concerned with international business cycles. Private institutions, such as Economic Cycle Research Institute (ECRI) or Conference Board, offer many the business cycle indicators but they are not freely available and are not considered in this thesis. In terms of the G7, there is only one publicly available chronology, which is provided by the OECD. They base their methodology on their Composite Leading Indicators (CLI), an aggregate of economically significant variables. Using an approach based on the Bry-Boschan algorithm combined with a weighting scheme based on GDP, they are able to date business cycle chronologies for country groups, as well as individual countries.

In this thesis, a business cycle chronology for the G7 area is presented using a dataset of 91 monthly real economic activity indicators specific to member states. This provides a more nuanced approach to the chronology of the OECD. The individual turning points are obtained using the procedure of Stock and Watson (2014). The chronologies are evaluated based on an in-sample forecast of a machine learning algorithm know as boosting, which has been gaining popularity among economic researchers in recent years (Berge, 2015; Ng, 2014; Vrontos et al., 2021; Yousuf & Ng, 2020). Moreover, the model presents the importance of each predictor, which allows for a better understanding of the mechanics underlying the forecasts.

The thesis is organized as follows. Section 2 covers relevant literature on dating business cycle chronologies and forecasting recessions. Section 3 covers the methodology used to date turning points of the G7 business cycle and the model used in the forecasting exercise. Section 4 describes the data chosen for the purpose of this research. In the fifth section, the adjusted chronologies are presented, and the model is tested for in-sample performance and economic relevance of the findings is discussed. The sixth section, concludes, and suggests directions for future research.

2 Literature Review

A range of studies have been conducted in the search for a robust tool that improves upon or completely bypasses the NBER and is capable of identifying turning points and recessions. Essentially, research aims to create an algorithm observing similarities in changes of economic indicators conditional on a NBER recession happening and to provide a judgement on a purely objective basis, avoiding the long subjective assessment period preceding an official announcement. Hamilton (2011) stresses that the automation relieves political pressure on the business committees and sheds light on the drivers of economic downturns. He follows by summarizing the efforts to automate dating of business cycle turning points.

There are two main approaches to turning points dating. Stock and Watson (2014) define them as "average-then-date" and "date-then-average" methods. The average-the-date approach is based on one or a few highly aggregated series, usually the GDP, from which the turning points are extracted to identify the business cycle chronology. On the other hand, the date-then-average approach takes a large number of disaggregated series, finds the turning points in each one, essentially creating a chronology for each series, and aggregates them into a final business cycle chronology.

2.1 Average-then-Date

Most of the methods showcased in this section utilize Markov-switching time series models which, at the time of their publication, were the most established in out-of-sample real-time performance. In simple terms, the Markov-switching models are able to handle change in regime depending on the given period (high/low volatility or bull/bear market) and can accordingly adjust the estimation.

Chauvet and Hamilton (2006) investigate the U.S. GDP growth rate in 45 quarters between 1947 and 2004 that NBER declared as part of an economic recession. They were able to match the NBER dates with a reasonable accuracy and turning points would be announced at about the same time as the NBER announcements. They created recession probabilities,[4] which were regularly updated after the report was published and the performance was reviewed by Hamilton (2011). The Global Financial Crisis was dated consistently with the NBER, and the

[4] https://fred.stlouisfed.org/series/JHGDPBRINDX

announcement would have been made with a similar timing. Nalewaik (2007) uses GDI for this purpose and, although the announcement was timelier than the NBER for the 2007 crisis, there were significant inconsistencies in preceding episodes. Hamilton (2005) attempts to model recession probabilities using unemployment rate and offered very promising results. Hamilton (2011) argues that the cyclical behavior of unemployment rate exhibits changes over time, due to demographics and other variables evolving significantly. Similarly to GDI, the unemployment rate would be better suited as a supplement to the inference from GDP.

An alternative to the use of a single indicator for dating business cycle turning points it's to incorporate multiple monthly series. The challenge is not only to choose a set of series that best reflect the business cycle chronology, but also to create a suitable framework that can extract the desired information.

Stock and Watson (1989) aim to clarify the state of the economy in real time by revising the coincident index that was originally developed by Mitchell and Burns (1938). They identify four crucial indicators — growth rates of industrial production, personal income, sales, and employment — and develop a model where business cycles are measured by comovements in various components of economic activity. The paper was extended to produce probabilities for a range forecast horizons (Stock & Watson, 1991, 1993). However, after out-of-sample testing on real recessions, the indexes were deemed to be better-suited for historical inferences rather that predictions. Their model has since been modified by, for example, adding regime shifts to capture the asymmetry of the business cycle because expansions are gradual and display a high mean duration while recessions are shorter and steeper (Chauvet, 1998).[5]

Despite the modifications to the model, the main indicators remain relevant. Chauvet and Piger (2008) have conducted simulated real-time exercises on such models suggesting that they could perform well in real time. They have been periodically posting the probabilities on a publicly available web page.[6] The announcement of the Global Financial Crisis would have come three months before the one of Chauvet and Hamilton (2006) which relied solely on GDP and five months sooner than the official announcement by NBER. This performance confirms that there

[5] Recessions only take up approximately 20% of the time, while expansions span across 80% of a given time period (Chauvet & Hamilton, 2006).

[6] https://fred.stlouisfed.org/series/RECPROUSM156N

is useful additional information in monthly indicators. However, the adoption of monthly indicator models requires greater caution and tolerance of instability over time.

2.2 Date-then-Average

Burns and Mitchell (1946) are the first to implement a date-then-average method. They assessed the turning points in individual series and aggregated them to arrive at the desired date. The first modern attempt to automate such approach was carried out Harding and Pagan (2006), who use develop an algorithm to aggregate the specific cycle turning points in industrial production, employment, sales, and income. They are able to track the NBER troughs with great precision but missed the peaks of some recessions by a large margin. Chauvet and Piger (2008) adopt their approach to determine recessions in real time with a non-parametric algorithm and a Markov-switching model as in Chauvet (1998). Again, they observed much better performance for trough dates compared to peaks.

Stock and Watson (2014) propose a more complex dating mechanism, specifically the date-then-average method mentioned previously, which they compare to a set of average-then-date methodologies. They use the Bry-Boschan procedure to identify recessions in individual series conditional on an NBER date occurring (See Section 3.1.1 for a in dept explanation). The turning points are then aggregated to find the actual peak or trough. The methodology is applied to a set of 270 disaggregated monthly series covering the most common coincident monthly measurements of economic activity: production, sales, income, and employment.

The distribution of turning points in not normal in all cases and, especially when dating global business cycles, it takes on undesirable forms. A potential solution is provided by Burns and Mitchell (1946, pp. 77-80): “In many cases the turning points of different series were bunched so closely that we could not go far astray. But there were cases in which the turning points were widely scattered, and others in which they were concentrated around two separate dates. If there was little else to guide us, we placed the reference turn toward the close of the transition period.” Stock and Watson (2014) address this by focusing on local measures of central tendency. The mode is preferred to the mean and median as it is a non-parametric measure and is not easily influenced by outliers.

Turning points are identified with relatively high precision and with small standard errors relative to the NBER chronology. The empirical results suggest that the date-then-average procedures provide useful information and can potentially supplement the process of determining reference cycle chronologies.

Pacella (2021) uses the algorithm the algorithm developed by Stock and Watson (2014) to study the business cycle dating formulated by the CEPR committee for the euro area. Her dataset includes over 100 macroeconomic variables, but the historical coverage is limited to 1995 for most of them. Nevertheless, the Bry-Boschan peaks and troughs align very well with the turning points presented by CEPR. She rejects a reliance on other dating rules built on GDP dynamics and stresses that the CEPR considers a wide range of variables when making their announcements.

2.3 Dating Turning Points in the Global Business Cycle

Most research studying the chronology of business cycles is focused on the United States. This is an understandable decision as the United States is the world's leading economy with the highest GDP and its political and economic climate largely influences global affairs. However, this thesis is focused on determining the drivers of global (understood as G7) recessions, and thus focusing on the United States is not sufficient. Majority of research studying the U.S. business cycle relies on the NBER, however, no such reference chronology exists for the global business cycle and a sufficient placeholder must be found to transition the country-specific research to an international scale.

An obvious alternative to the reference chronology of the NBER is the one of OECD, who is the only organization publicly publishing international chronologies. They utilize their composite leading indicator system, which is based on a "growth cycle" approach. Business cycles and turning points are measured by observing the deviations from trends in a series, with country-specific GDP serving as a reference.

The CLIs are comprised of time series which exhibit a leading relationship with the GDP (the reference series) at turning points. For each country, the components are selected depending on their economic significance, cyclical behavior, data quality, timeliness, or availability. For zones aggregates, such as the G7, the composite leading indicators and the reference series are calculated as weighted averages of the corresponding zone member series. The identification scheme is a simplified version of the original Bry-Boschan algorithm without the correction for outliers, which is implemented at an earlier stage of the filtering process by the OECD.

.

The performance of the OECD composite leading indicator was evaluated by Ojo et al. 2023) who utilized a wavelet transformation to track comovements of the CLI with industrial production, unemployment, and real GDP growth. This methodology has become very popular for economic research and was used to analyze topics, such as the synchronization of business cycles in Europe (Aguiar-Conraria & Joana Soares, 2011) or even the comovements of drought and commodity prices (Racocha, 2017). The evaluation showed that the CLI is a valid leading indicator for the industrial production index, slightly less effective with the unemployment rate, and consistently ineffective with GDP growth. On the other hand, it is most significantly coherent with the real GDP growth but with a lag, meaning the changes of GDP growth precede the changes in OECD'S CLI. Astolfi et al. (2016) suggest that the OECD CLIs were able to anticipate the Great Recession in G7 countries at an early stage but could not predict its severeness. Therefore, the OECD business cycle turning points are well suited to serve as a reference chronology in a global setting.

Kose et al. (2020) carried out an extensive analysis of the global business cycle. Firstly, they considered long annual dataset (1950-2019) of economic series and 180 economies across the world (36 advanced economies and 144 emerging market and developing economies, EMDEs). Secondly, they present an analysis of the phases of the global business cycle using quarterly output series of 106 countries over the period 1960:1-2019:3. The study employs global real GDP per capita as its primary variable in its analysis as it is the main indicator of well-being over the world.

They utilize two assessment methods for turning point identification: a statistical method and a judgmental method. The statistical approach defines recessions as declines in annual GDP per capita. It follows the methodology of Harding and Pagan (2002) who extend the algorithm developed by Bry and Boschan (1971) to identify the turning points in a logarithmic transformation of per capita GDP. The judgmental method is a subjective assessment, similar to the practices of the BCDC of the NBER or CEPR.

In their study, the researchers focus a wide range of real and financial variables. They argue against the use of a single threshold under which the world economy is deemed to be in a recession as the characteristics change throughout time. The uncovered turning points are displayed in Table 2. They find the average duration of a recession using quarterly data is slightly less than a year. This coincides with the finding of Jorda and Berge (2011) that a recession in the United States lasts on average 11 months. In addition to the global recessions, the researchers have identified four global economic downturns. These episodes occurred in the years 1958, 1998, 2001, and 2012, but they fall short of their recession definition.

Fushing et al. (2010) utilize a new empirical strategy to identify turning points across 22 OECD countries and subsequently to obtain a global business cycle chronology. The intuition is similar to the date-then-average methodology used by Stock and Watson (2014), but in the context of different countries, rather than sectors. They stress that instead of finding common factors that explain fluctuations of economic activity across many countries, one must focus on finding business cycles. The regional chronologies offer valuable information about the contagion of business cycles in addition to the propagation of shocks to different economies. Bodman and Crosby (2005) try a number of different dating methods to uncover relationships between business cycles in the G7 member states but find mixed evidence for most countries with Japan being the most independent. Through a set of regime-switching logit models de Bondt & Vermeulen (2021) find that the spillovers are more likely to happen at the beginning of the recession rather than at the start of an expansion.

Fushing et al. (2010) propose a non-parametric, parsimonious, and computationally simple procedure called Hierarchical Factor Segmentation (HFS) algorithm. This algorithm is applied to each indicator separately and network analysis is used to combine the obtained information to better determine the onset of cyclical phases. When applied to U.S. data, the algorithm produced a classification of recessions that is sensible when compared with the NBER's benchmark. Apart from creating 22 chronologies for countries without dating committees, they identify four global recession episodes since 1965 (starting point of the dataset in this thesis), which are presented in Table 2.

2.4 Predicting Recessions

Many indicators of global activity have been created in the bid to identify the latent state of the economy and to predict the next global recession. Miranda-Agrippino and Rey (2020) create a single global factor that explains an important share (over 20%) of the variation of risky asset prices around the world, which was used to study the impacts of U.S. monetary shocks (Miranda-Agrippino & Ricco, 2021). Cuba-Borda et al. (2018) create a Global Conditions Index (GCI), a real-time measure of the health of the global economy. They combine the GCI with the excess bond premium[7] series developed by Gilchrist and Zakrajšek (2012) to create a world recession indicator. Another important indicator of world economic activity is the nowcast of global growth by Ferrara and Marsilli (2019). They utilize a regression framework called Mixed-frequency models (MIDAS), which allows the dependent variable (e.g., quarterly GDP) to be related to a higher-frequency explanatory variable (e.g., monthly industrial production).

Baumeister et al. (2020) create a global economic conditions indicator (GECON). They stress that the inclusion of a large number of disaggregated series leads to a diminished forecasting performance due to possible cross-correlation of error terms of series in the same category. They illustrate very good forecasting performance of the global business cycle using the MIDAS models.

Other indexes include World Industrial Production by OECD (currently maintained by Baumeister and Hamilton (2019)), Global Steel Production Factor (Ravazzolo & Vespignani, 2020), the Kilian index (Kilian, 2009), or the Real Commodity Price Factor (Alquist et al., 2020). However, their forecasting performance is worse compared to the GECON (Baumeister and Guérin, 2020).

Many academic studies have suggested a simpler and possibly more robust procedure. The slope of the yield curve seems to be a very promising predictor of recessions in the United States. Chauvet and Potter (2005) find that although the yield curve is a statistically significant predictor of future activity, the predictive power of the spread is not stable over time, citing the 1990 recession in particular. Hamilton (2011) highlights the weak performance before the Great Recession, showing that a good in-sample fit does not guarantee a good out-of-sample performance as predicting the future is extremely difficult. Kumar et al. (2021) study the relationship of the yield curve and economic activity in the G7 and find the predictive power is dependent on the inclusion of other variables.

2.4.1 Boosting

A relatively new approach to predicting recession is the use of machine learning applied to economic indicators. Ng (2014) presents a useful methodology for giving warning signals of

---

[7] It is used to predict the probability that the U.S. economy will enter into a recession sometime during the next 12 months.

recessions and consequently identifying the best predictors of recessions. The technology known as boosting is a machine learning algorithm that was, among other use cases, originally developed as a classification rule to determine if a message is a spam, or if a tumor is cancerous given gene expression data. Boosting is a very useful tool, especially when the analysis uses a large dataset as it assigns an importance to each predictor. This allows to rank the indicators based on their usefulness for predictions.

There are many different ways to implement a boosting algorithm. Ng (2014) uses the package GBM in the software R. Specifically, as her goal to is to study the best predictors of a binary variable, i.e., whether a recession will happen or not, she uses a boosting technique based on the logistic model. In a standard logit model, the selection of predictors occurs prior to estimation and the fit is based on a model that considers multiple predictors jointly. In contrast, gradient boosting, originally developed by (Friedman, 2001, 2002), performs variable selection and estimation simultaneously and without interaction between variables, and the final model is built up from an ensemble of models. She finds that even if when the pool of predictors is very large, the predictor set with systematic and important predictive power consists of only approximately 10 variables. (Yousuf & Ng, 2020) use the ability of the boosting model to utilize many potential predictors for forecasting and study the effects of time varying parameters and economic instability.

Berge (2015) compares the use of a non-linear and logit models as weak learners in his boosting procedure for predicting turning points of the U.S. business cycle. He argues in favor of boosting as a forecasting tool and shows that non-linear boosting models outperform their linear counterparts. Delgado et al. (2022) study the yield curve and individual spreads with the help of gradient boosting to uncover the best indicators of recessions.

Giusto and Piger (2017) propose new approach to identifying business cycle in real time. They utilize the machine learning algorithm known as learning vector quantization. It identified all five recessions recognized by NBER since the 1980s without any false positives. Moreover, the algorithm would have identified the December 2007 peak by early June 2008, which is several months ahead of the procedures reviewed by Hamilton (2011).

Vrontos et al. (2021) uses machine learning techniques to model and predict U.S. economic activity and recessions. They compare many machine learning methods including regularization techniques, such as Ridge, Least Absolute Shrinkage and Selection Operator (LASSO), and

Elastic Net, Discriminant Analysis classifiers, Bayesian classifiers, and classification and regression trees (CART), such as Bagging, Random Forests, and Boosting. They argue for the more frequent utilization of machine learning and point out the benefits for making predictions of recessions. Each technique exhibits strengths in different use cases. Specifically, they find that performance of tree-based algorithms is reliant on the underlying forecast horizon, and they generate the most reliable predictions in short- to medium-term. Jain and Bhasin (2022) try to predict recessions using a set of financial and real economic variables in different models. They identified XGBoost as the best choice of modelling technique, through which they identified industrial production as the most important variable.

3 Methodology

The methodology used to determine turning points in the in the global economy follows the one of Stock and Watson (2014). The problem of dating a reference cycle turning point (peak or trough) is conditional on the event that a single turning point occurred during a certain time span. If a recession happened during a certain period, there must have been a turning point in the economic activity, i.e., a switch from an expansion to a recession. All that remains is to date the peak within this interval.

3.1 Dating Turning Points

3.1.1 Bry-Boschan

The algorithm for identifying turning points in individual series is taken over from Bry and Boschan (1971). This method is widely used among business cycle researchers, only with some slight adjustments made since its conception (Otsu, 2017). One of the potential drawbacks is the fact that it was developed based on the dating rule heuristically found with the U.S. data in the first half of the 20th century and may need some adjustments to fit modern data. Moreover, some of the moving averages were chosen in a specific way to best identify peaks and troughs as closest possible to the selected by the staff at NBER. The procedure assumes all series are seasonally adjusted and is split into five steps.

In Step I, the algorithm addresses outliers. A Spencer curve is computed using a 15-month symmetric moving average with particular weights. The outliers, defined as values whose ratios to the 15-point Spencer curve are larger than 3.5 standard deviations, are replaced by values of the Spencer curve. Stock and Watson (2014) found that the 15-month moving average occasionally produced some anomalous results and they eliminated then by using a centered three-month moving average. The results presented in their paper and in this thesis use the three-month moving average in the first Bry-Boschan step.

Step II calculates a 12-month moving average (MA12) from the series previously filtered for outliers. The length of the moving average is chosen to best fit the Spencer curve and its fluctuations. Any date with the highest (lowest) value among the 6 preceding and the 6 following months is tentatively regarded as the date of a peak (trough). The peaks and troughs must be altering to ensure cyclicality. If peaks or troughs are contiguous, the highest value is chosen for a peak, and the lowest for a trough. If the values are same, the algorithm sets an earlier date for a peak, and a later date for a trough, respectively. The next steps are seen as refinements to the dating mechanism as most of the work has been done in the first two steps.

Step III compares the findings of MA12 and the Spencer curve to ensure peaks and troughs within ±6 months. The turns of the Spencer curve of the outlier-free series are heuristically closer to those of the original series than those of MA12. If there are any ties within ±6 data points on the Spencer curve, an earlier date is chosen for a peak, and a later date for a trough. The turning points are checked for alteration as in Step II and the durations of a full cycle (a peak to peak or a trough to trough are enforced to be at least 15-month period. Shorter cycles than 15 months result in an elimination of the lower (higher) of the two peaks (troughs). Same values are treated in the same way as in Step II.

In Step IV, a short-term moving average called MCD (Months of Cyclical Dominance) curve is used to further refine the procedure and is obtained in the following way. First, the Spencer curve of the original series is taken as the trend-cycle component. The irregular component is obtained by calculating the difference between the original series and the trend-cycle component. Next, a ratio is computed of the average change in the irregular component to that in the trend-cycle component. The MCD is the shortest period of months that it takes for the change in the trend-cycle component to dominate that in the irregular component, which corresponds to a ratio less than 1. The Bry-Boschan procedure confines the MCD between 3 to 6 months. Then, a short-term moving average is computed over the span of MCD to ensure peaks and troughs are within ±6 months as in Step III and alternation is checked as in Step II if modified.

The fifth and final Step V is a series of tests that are conducted to determine final turning points from the tentative ones obtained previously. First, the alteration of turning points, as in Step II, is checked. Second, the original series is used to ensure peaks and troughs within ±4 months or ±4 MCD, whichever is longer. Third, any turning points closer than 6 months from the ends are removed. Fourth, if the first or the last peak (or trough) takes a smaller (or greater) value than any value between it and the end of the original series, it is removed. The code for the algorithm written in the programing language Gauss by Watson (1994) compares the first and the last turns only with the initial and the last data points, respectively, not with all the values between them. It could be a cause of a nontrivial difference, but it does not change the results of the paper. The same is also the case for the algorithm utilized in Stock and Watson (2014) in this thesis. Fifth, if the duration of a full cycle is checked to be at least 15-month length, as in Step III. Sixth, the final test checks whether a phase (peak to trough or trough to peak) duration is at least 5 months. If it lasts less than 5 months, the two turning points are eliminated, and if the violation is found at the last turning point, only the last point is removed.

By completing these five steps, turning points are calculated for each individual series in the dataset. Considering this information, the following dating mechanism is implemented to aggregate the sample into a single turning point for each recession episode in the reference business cycle chronology.

3.1.2 Dating Using a Simple Random Sample of Disaggregated Series

A population of economic series is taken, each of which describes a different aspect of economic activity. This population is approximated to be infinitely large, and its members have turning points uncovered by the utilization of the Bry-Boschan procedure. For a given episode s (i.e., recession or expansion), which covers a known time interval, there exists a population distribution $g_s(\tau)$ of turning points τ of specific series in this population. The reference cycle turning point (estimand) is defined as a functional of the population distribution $g_s(\tau)$.

For such a population, a sample of turning points is given by $\{\tau_{is}\}, i = 1, 2, \dots, n_s$ where τ_{is} is the turning point date of series i in episode s and n_s is the number of turning points observed in episode s. The mean (D_s^{mean}), median (D_s^{med}), and mode (D_s^{mode}) of the distribution $g_s(\tau)$ can be estimated to using the sample $\{\tau_{is}\}$, resulting in the sample central tendencies $\widehat{D}_s^{mean}$, $\widehat{D}_s^{med}$, and $\widehat{D}_s^{mode}$ respectively. The mode is computed as the mode of a kernel density estimator $\hat{g}_s$ of g_s, with kernel K and bandwidth h.

The sample obtained by simple random sampling from the population of series ensures the turning points are i.i.d. and the asymptotic distributions of the three estimators are,

$$\sqrt{n_s}\left(\widehat{D}_s^{\text{mean}} - D_s^{\text{mean}}\right) \xrightarrow{d} \mathrm{N}\left(0, \sigma_{\tau,s}^2\right) \quad (1)$$

$$\sqrt{n_s}\left(\widehat{D}_s^{\text{med}} - D_s^{\text{med}}\right) \xrightarrow{d} N\left(0, \frac{1}{4g_s\left(D_s^{\text{med}}\right)^2}\right) \quad (2)$$

$$\sqrt{n_s h^3}\left(\widehat{D}_s^{\text{mode}} - D_s^{\text{mode}}\right) \xrightarrow{d} N\left(0, \frac{g_s\left(D_s^{\text{mode}}\right)\int [K'(z)]^2 dz}{\left[g_s''\left(D_s^{\text{mode}}\right)\right]^2}\right), \quad (3)$$

where $\sigma_{is}^2 = var(\tau_{is})$ in episode s and where the bandwidth sequence h_n satisfies $h_n \to 0$, $nh_n^3 \to \infty$. The variances in $(1) - (3)$ can be estimated using kernel estimators of g_s and its second derivative, ${g_s}^2$.

3.1.3 Adjusting for Weighted Random Sampling

If a certain class exhibits turning points that systematically lead the turning points in a different class, bias would be introduced in the simple random sample. It will tend to bias the estimator towards estimated reference cycle turning point that leads the population turning point.

Similarly, if a class of series is represented more heavily than another class, a simple estimation would also result in a bias. Stock and Watson (2014) use the four disaggregated coincident indicators, with some classes being represented to a greater extent than others (e.g., industrial production class contains 69 distinct component series in contrast to only 14 series for real personal income less transfers).

They model this problem of unequal representation of classes of series by utilizing the method of stratified sampling. The initial stratum is the class of series (e.g., industrial production). The subaggregate (e.g., industrial production of durable goods) is then randomly sampled within the class. The number of series (observations) differs from one class to the next. This leads to some classes of series receiving larger weight in the sample than in the population.

The dataset used in this research covers economic indicators across the G7 countries. It expands the number of sectors included to six, industrial production, employment, real manufacturing and wholesale-retail trade sales, real personal income, and the two added ones, housing starts and personal consumption (see Section 4 for further details).

Moreover, the series are categorized in two different ways. Firstly, the economic indicators are classified by the country origin which results in seven categories, each containing a different number of series. Secondly, the series are distributed among six classes depending on the economic sector of the specific indicator.

Therefore, instead of weighting only by an aggregated sector (e.g., Industrial production), the methodology in this thesis will also create weights using individual countries. Thus, in the classification scheme based on countries, the initial stratum would be, for example, the United States and the subaggregated series is employment in manufacturing.

Two different methods are used for adjusting for discrepancies between sample and population weights by series class: lag adjustment and weighted estimation. Both procedures are based on weighting by the ratio of population to sample probabilities. By classifying the series in different ways, it allows for a deeper insight of how individual countries (sectors) experience turning points.

Let m index classes of series, let M be the number of classes (finite), let m_i be the class containing series i, let π_m be the population probability assigned to class m, and let p_{ms} be the fraction of series of class m in the sample of turning points for episode s. Then the ratio w_{is} of population weights to sample weights for series i in episode s is,

$$w_{is} = \frac{\pi_{m_i}}{p_{m_i s}}. \tag{4}$$

3.1.4 Lag Adjustment

The process of lag adjustment estimates a mean lag for each series. If a series in class m has a population mean lag k_m, relative to the reference cycle date, then the turning point of the i^{th} series in episode s can be written as the sum of the population mean reference cycle turning point D_s^{mean}, the mean lag for its class, and a discrepancy η_{is}:

$$\tau_{is} = D_s^{mean} + k_{m_i} + \eta_{is}. \tag{5}$$

The reference cycle turning point can be identified as the mean, D_s^{mean} by using two assumptions. It is assumed that $E(\eta_{is}) = 0$ and k_{m_i} is normalized in such a way that the mean lag in the population is zero, which corresponds to $\sum_{m=1}^{M} \pi_m k_m = 0$. A general estimation of (5) by OLS will not satisfy the latter assumption and it will produce biased estimates of the class lags, unless $w_{is} = 1$ for all i and s. Therefore, the lag adjustment is carried out in two steps. Firstly, $\{k_m\}$ are estimated by restricted least squares (with the restriction $\sum_{m=1}^{M} \pi_m k_m = 0$), yielding an estimator $\{\hat{k}_m\}$. Secondly, the sample of lag-adjusted turning points is obtained by $\tilde{\tau}_{is} = \tau_{is} - \hat{k}_{m_i}$. The mean, median, and mode estimators are subsequently computed episode-by-episode using the lag-adjusted data, $\{\tilde{\tau}_{is}\}$.

3.1.5 Weighted Estimation

After adjusting for lags in different series, the sample is weighted in a way, where the sample weights on individual observations (series-specific turning points) match the population weights.

Let g_{ms} be the distribution of turning points among series of class m in episode s. Then the population distribution of turning points in episode s is given by $g_s = \sum_m \pi_m g_{ms}$, where the sum is over a finite number of classes of series, i.e., different countries or sectors.

Due to the nature of the Bry-Boschan algorithm, which produces integer-valued turning points, the output is a set of histograms of turning point dates for each class of series, by episode. These histograms are weighted by class, yielding a weighted histogram for episode s, given by $\hat{g}_s^{hist-wtd}(t) = \sum_{i=1}^{n_s} w_{is} 1(\tau_{is} = t) / \sum_{i=1}^{n_s} w_{is}$. The individual weights are the ratio of the population to sample weights. The mean and median are derived from the weighted histogram and the mode is obtained from the kernel density estimator computed by smoothing $\hat{g}_s^{hist-wtd}$. The variances of the weighted estimators for mean and median are,

$$\mathrm{var}\left(\hat{D}_s^{\text{mean,wtd}}\right) = \sum_m \left(\frac{\pi_m^2}{n_{ms}}\right) \sigma_{ms}^2 \tag{6}$$

$$\mathrm{var}\left(\hat{D}_s^{\text{med,wtd}}\right) = \sum_m \left(\frac{\pi_m^2}{n_{ms}}\right) \frac{G_{ms}\left(D_s^{\text{med}}\right)\left[1 - G_{ms}\left(D_s^{\text{med}}\right)\right]}{g_s\left(D_s^{\text{med}}\right)^2} \leq \frac{1}{n_s}\left(\frac{1}{4 g_s\left(D_s^{\text{med}}\right)^2}\right) \sum_m \frac{\pi_m^2}{p_m}, \tag{7}$$

where G_{ms} is the cumulative distribution function corresponding to g_{ms}. Stock and Watson (2014) use a bound in the expression (7) for the standard errors in the weighted median due to the poor estimation of the terms in the first summation.

The variance of the weighted mode is equivalent to the one in (3) for the mode under simple random sampling. There is a slight modification that g_s is reinterpreted as the weighed density. The corresponding standard errors are computed using the kernel smoother of the weighted histogram to estimate g_s.

3.2 Boosting

The methodology used for the boosting section of this thesis implements a machine learning algorithm under the gradient boosting framework. Specifically, the Python package XGBoost (Extreme Gradient Boosting) is used due to its ease of use and good performance. (Chen & Guestrin, 2016) describe the procedure and its benefits in great detail.

The XGBoost algorithm is used for supervised learning problems, where we use the training data x_i, consisting of multiple features, to predict a target variable y_i.[8] The supervised learning is elevated to ensemble rating to improve the predictive performance. In supervised learning, the model searches for a suitable hypothesis through a certain hypothesis space. The chosen hypothesis should make good predictions for a particular problem. Even if the hypothesis space is well-suited for the problem in question, it may be very difficult to find a good fit. A common example of supervised learning is a linear model where the prediction is in the form of a linear combination of weighted input features. As (Liu et al., 2014) show, ensemble learning combines multiple hypotheses to form a potentially better hypothesis. It follows that ensemble learning is a collection of methods for studying a target function by training several individual base learners and combining their predictions to form a classification.

Rather than using a classical decision tree, which makes sequence of decisions to arrive at a prediction or classification for a given input, XGBoost uses decision tree ensembles. The tree ensemble model is comprised of a set of classification and regression trees (CART). The XGBoost is an instance of Gradient Boosted Trees modeling (GBT). In a basic decision tree, a leaf only contains decision values. In CART, on the other hand, a real score is associated with each of the leaves, which gives richer interpretations that go beyond classification.

As for all supervised learning models, an objective function is defined and minimized through the learning process to obtain best possible results. The objective function (loss function and regularization) at iteration t that needs to be minimized is the following:

[8] For an introduction to Boosted Trees see: https://xgboost.readthedocs.io/en/latest/tutorials/model.html.

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \omega(f_t) + \textit{ constant}, \quad (8)$$

where l is the loss function of CART learners and ω is the regularization term. ω penalizes the complexity of the model which helps to smooth the final learnt weights to avoid over-fitting.

The first goal is to find the parameters of the individual trees. The functions f_i each contain the structure of the tree and the leaf scores. Learning tree structure is much more difficult than in a traditional optimization problem where one can simply take the gradient. Due to the intractability of learning all the trees at once, XGBoost uses an additive strategy, where the model saves what it has learnt and adds one new tree at a time. The prediction value at step t is written as $\hat{y}_i^{(t)}$ and the iterations are the following,

$$\begin{aligned} \hat{y}_i^{(0)} &= 0 \\ \hat{y}_i^{(1)} &= f_1(x_i) = \hat{y}_i^{(0)} + f_1(x_i) \\ \hat{y}_i^{(2)} &= f_1(x_i) + f_2(x_i) = \hat{y}_i^{(1)} + f_2(x_i) \\ &\cdots \\ \hat{y}_i^{(t)} &= \sum_{k=1}^{t} f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i), \qquad f_k \in \mathcal{F}, \end{aligned} \quad (9)$$

where $\mathcal{F} = \{f(\mathbf{x}) = w_{q(\mathbf{x})}\}(q: \mathbb{R}^m \to T, w \in \mathbb{R}^T)$ is the CART. A mean squared error (MSE) is a very straight-forward evaluation metric of the loss function. Generally, it is not easy to get a nice form as with the MSE and a second order Taylor expansion of the loss function is taken to approximate the objective. For binary classification tasks, such as the determination of recession required in this thesis, a logistic loss is used, which relies on the approximation procedure. This changes the objective function as follows,

$$\mathcal{L}^{(t)} \simeq \sum_{i=1}^{n} \left[l\left(y_i, \hat{y}^{(t-1)}\right) + g_i f_t(\mathbf{x}_i) + \frac{1}{2} h_i f_t^2(\mathbf{x}_i)\right] + \omega(f_t), \quad (10)$$

where g_i and h_i are defined as

$$g_i = \partial_{\hat{y}_i^{(t-1)}} l\left(y_i, \hat{y}_i^{(t-1)}\right), \quad (11)$$

$$h_i = \partial^2_{\hat{y}_i^{(t-1)}} l\left(y_i, \hat{y}_i^{(t-1)}\right). \tag{12}$$

After the removal of the constant parts, the objective to minimize at step t simplifies to

$$\tilde{\mathcal{L}}^{(t)} = \sum_{i=1}^{n} \left[g_i f_t(\mathbf{x}_i) + \frac{1}{2} h_i f_t^2(\mathbf{x}_i)\right] + \omega(f_t), \tag{13}$$

which is a sum of quadratic functions of one variable that can be minimize by the utilization of known techniques. One important advantage of this definition is that the value of the objective function only depends on g_i and h_i, which enables the support of custom loss functions.

Model complexity is defined through the regularization term $\omega(f_t)$. To do so, the definition of the tree $f(x)$ is refined as

$$f_t(x) = w_{q(x)}, w \in R^T, q{:}\, R^d \rightarrow \{1,2,\cdots,T\}, \tag{14}$$

where ω is the vector of scores on leaves, q is a function assigning each data point to the corresponding leaf, and T is the number of leaves. After the refinement of the tree, XGBoost defines the complexity as

$$\omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{15}$$

Using the reformulation of the tree model, the objective value with the t-th tree can be written as

$$\mathcal{L}^{(t)} \approx \sum_{i=1}^{n} \left[g_i w_{q(x_i)} + \frac{1}{2} h_i w_{q(x_i)}^2\right] + \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{16}$$

$$= \sum_{j=1}^{T} \left[\left(\sum_{i \in I_j} g_i\right) w_j + \frac{1}{2}\left(\sum_{i \in I_j} h_i + \lambda\right) w_j^2\right] + \gamma T, \tag{17}$$

where $I_j = \{i|q(x_i) = j\}$ is the set of indices of data points assigned to the j-th leaf. All data points in a single leaf have the same score. The expression is compressed by defining the terms $G_j = \sum_{i \in I_j} g_i$ and $H_j = \sum_{i \in I_j} h_i$ which yields.

$$\mathcal{L}^{(t)} = \sum_{j=1}^{T} \left[G_j w_j + \frac{1}{2} (H_j + \lambda) w_j^2 \right] + \gamma T. \tag{18}$$

Here, the w_j are independent with respect to each other. Given a structure $q(x)$, the best w_j and objective function reduction we can get is the following (the latter expression measures how good the tree structure $q(x)$ is),

$$w_j^* = -\frac{G_j}{H_j + \lambda} \tag{19}$$

$$\mathcal{L}^* = -\frac{1}{2} \sum_{j=1}^{T} \frac{G_j^2}{H_j + \lambda} + \gamma T. \tag{20}$$

Due to the nature of the research a binary classification with a log loss objective function is considered. This procedure is described well by Leventis (2018). The loss function takes the following form,

$$y \ln(p) + (1 - y) \ln(1 - p) \text{ where } p = \frac{1}{(1 + e^{-x})}, \tag{21}$$

where y is the real label in {0,1}, the indicator of a recession, and p is the probability score. The p (score or pseudo-probability) is calculated by applying the sigmoid function to the output of the GBT model x. The output x of the model is the sum across the CART tree learners. The 1st and 2nd derivatives (gradient and hessian) with respect to x must be found in order to minimize the log loss objective function. They are the following: gradient $= (p - y)$ and hessian $= p * (1 - p)$.

4 Data

4.1 Disaggregated Dataset

The disaggregated dataset consists of 91 time series of real economic indicators.[9] It covers similar sectors to the one in Stock and Watson (2014): industrial production (12 series), employment (27 series), manufacturing and trade sales (20 series), and personal income (18 series). They are widely used coincident indicators describing the current state of the economy. It adds two additional sectors representing real economic activity that are often present in business cycle research (Ng, 2014; J. Stock & Watson, 1989; Vrontos et al., 2021), personal expenditure (9 series) and housing starts (5 series). The latter two sectors are usually classified as leading indicators. This offers a valuable insight on whether they have more predictive power than the others in the second part of the empirical analysis.

Moreover, the data is collected across the G7 area in order to analyze global recessions. The number of series for each country is the following: Canada (11 series), France (11 series), Germany (10 series), Italy (9 series), Japan (15 series), the United Kingdom (14 series), the United States (20 series). All series have monthly frequency, with a maximum span of 1965:1 – 2023:4 (700 months). The algorithm allows for missing values, leading to more series being included as they become available. The series, their classes, mnemonics, and time spans for which they are available are listed in Table 1. The dataset stops at a higher level of aggregation then the one of Stock and Watson (2014), whose set contained 270 monthly real activity indicators (621 months) for the United States.

The individual series were chosen depending on the economic significance, data quality, coverage, and availability. It must be noted that not all the countries are represented equally as the number of series differs from one country to another. For some countries, not all sectors are available at a monthly frequency. For France, for example, there is no available data for earnings as they only report them at a quarterly or annual frequency. Moreover, the availability of different economic indicators is variable in time as they do not have the same starting point.

[9] The majority of the time series have been downloaded from Refinitiv Datastream. For more information, refer to: https://www.refinitiv.com/. Three indicators of earnings in the UK (indicated in Table 1) were taken from the Office for National Statistics in the UK.

This could be, among other reasons, due to a suspension of previously active series, a change of measuring methodology (e.g., from monthly reporting to quarterly or different model to calculate indexes), or a more recent beginning of reporting. A simple interpolation of quarterly data was not considered as valuable information about the specific month of the turning point is lost by approximating the two missing months.

Additionally, for some countries (such as Italy and France), business survey expectations or future tendencies for employment are the only available monthly time series covering a sufficient time period. Such data is not appropriate for the purposes of this research as it is subject to very significant period-to-period changes. In these cases, quarterly data has been interpolated to a monthly frequency incorporating the information provided in the employment expectation series. The procedure starts with a linear interpolation of quarterly data. Because of the very jittery nature of the business survey (or future tendency) data, they were considered only when the absolute value of the relative change surpassed 0.4. For the months where the threshold was surpassed, the interpolated value at time t was multiplied by $(1 + tendency_t * 10^{-4})$. The tendency value was multiplied by 10^{-4} to avoid extremely large jumps and maintain a relative smoothness of the data.[10] The additional step of including future tendencies, was included to obtain a more accurate estimate of the turning point in the series.

All series exhibiting seasonality were adjusted by extracting the seasonal factor in the time series decomposition procedure. Series, which were not available in real terms or in constant prices, were adjusted by the price level of the respective country.[11]

The monthly growth rates of the 91 series, divided by their standard deviation, are displayed as a heat chart in Figure 1. The vertical axis is series number as given in Table 1 and the horizontal axis is time in months. Blue represents periods of positive growth, yellow denotes moderately negative growth, and red denotes strongly negative growth.

The gray areas in Figure 1 represent missing data. The most relevant features of Figure 1 for the current purpose are the vertical yellow-red bands. Because the horizontal axis is calendar time, the vertical yellow-red bands show periods in which many of the component series

[10] Bot the threshold of 0.4 and the multiplier of 10^{-4} are arbitrarily chosen values to best fit the data.

[11] The inflation statistics are available at: https://data.oecd.org/price/inflation-cpi.htm.

experienced negative growth. In the context of Figure 1, the goal of the research is to date the beginning and end of the yellow-red band, which are respectively the cyclical peak and trough.

Figure 3 plots series-specific recession episodes, where a series-specific recession is defined to be the period from a Bry-Boschan peak to a Bry-Boschan trough. Black indicates a recession and gray indicates a period of expansion. Again, vertical clustering of black segments indicate that a large set of component series experienced a negative growth.

Moreover, as the series are organized by country, recessions specific to individual countries can be observed. For example, in the bottom 20 series, which correspond to France and Italy, a lot of series experience negative growth in the 2010s. This is a consequence of the sovereign debt crisis which has impacted many European economies. Many more such inferences can be made by studying these Figures, but the main goal is to aggregate the recessions through an algorithm to improve the business cycle chronology of the OECD.

As a sensitivity test of the Boosting framework (see Section 5.2), variables explaining the financial cycle are added to the set of predictors together with the disaggregated dataset. The first originates in a paper studying global economic conditions and their influence on energy markets by Baumeister et al. (2020). They develop a global economic conditions indicator (GECON) to track economic activity as a deviation from normal trend growth. A measure of uncertainty shocks is the one of Bloom (2007) who uses the Chicago Board of Options Exchange VXO index of percentage implied volatility, on a hypothetical at the money S&P100 option 30 days to expiration. The indicator of the global financial cycle (GFC) of Miranda-Agrippino and Rey (2020) is included. The single global factor explains an important share (over 20%) of the variation of risky asset prices around the world. It is constructed using a Dynamic Factor Model for a large panel of risky asset prices traded on all the major global markets, a collection of corporate bond indices, and commodities price series (excluding precious metals). The next variable is the Excess bond premium (EBP) presented by Favara et al. (2016). It is a widely recognized financial indicator and is used to produce probabilities whether the U.S. economy will enter a recession sometime during the next 12 months. And finally, the Global Conditions Index (GCI), a real-time measure of the health of the global economy constructed using a small set of world economic variables by Cuba-Borda et al. (2018).

4.2 Reference Chronology

An equally important task as choosing the correct economic indicators is choosing the reference chronology. As the methodology takes a moving average of 12 months around the reference cycle turning point and determines and adjustment, one must be sure that the actual turning point is present within the 25-month[12] sample. Although most available studies agree on the majority of recessions, the chronologies are not in unison on all dates.

The benchmark chronology chosen for the purpose of this research is the one capturing turning points of the major seven countries created by the OECD. They aggregate their set of composite leading indicators to form a reference chronology for the OECD members and country groups, like the G7 (see Section 2.3). This coincides with the disaggregated dataset covering the indicators in the G7 countries.

As a demonstration of the algorithm's abilities, turning points of three alternative papers studying global business cycles were taken as comparisons. The first of the three is the already discussed GECON indicator by Baumeister et al. (2020). Although the indicator tracks known episodes of expansions and contractions well, it must be noted, that the GECON was not created to serve as a reference chronology. The second chronology was taken from a paper by Kose et al. (2020). In the paper, they identify four global recessions and a set of economic slowdowns. Only the actual recessions were taken. Their turning points are reported in a quarterly frequency. For this reason, the first month of the given quarter was taken as a reference date. The third alternative chronology is the one of Fushing et al. (2010) which also identified only four global recessions. All turning points are shown in Table 2. An immediate observation is that the alternative business cycle chronologies do not recognize all of the OECD recessions. Moreover, even when the same recession is recognized, the studies define recessions in a different way, as there is no generally agreed upon recession definition, which leads to large differences in the dating. The analysis of the alternative business cycles implies interesting results (See Appendix) and serves as a cross-validation metric for the OECD chronology.

[12] 12 leads, 12 lags and the reference date.

5 Empirical Results

The creation of an improved chronology relies on the categorization by sector rather than by country. Although the results largely agree in both schemes the sector classification was chosen because of more stable performance, especially for the lag-adjustments. The research using the classification through countries was also carried out but serves as a source of supplemental information that brings more transparency into the process.

5.1 Estimated Chronologies

In Table 3, a set of three chronologies based on the 91 disaggregated series are reported. The OECD chronology is used as a reference. The proposed adjustments are reported alongside with their standard errors.[13] For each episode the mean, median and mode are calculated.

The first set of chronologies in Table 3 are the unadjusted estimates of turning points, in which the series are treated as if they were obtained from simple random sampling (Section 3.1.2). It must be noted that the U.S. has a greater representation within the dataset, resulting in the estimates relying more heavily on indicators from the United States. Therefore, there is already some weighting entering this unadjusted estimation.

The second set of chronologies in Table 3 are estimated based on the weighted lag adjustment procedure described in Section 3.1.4. They are adjusted by the class-specific lag as in (5). In this exercise, the estimated class-specific lags (k_m in the notation of (5)) are 0.04 for industrial production (IP), 0.57 for Employment (EMP), 0.55 for manufacturing and trade sales (MT), 1.35 for personal income (PIX), -1.67 for housing starts (HOU) and -2.57 for personal expenditures (EXP). This immediately points out the differences between the variables included in the dataset. Whereas the first four categories are an instance of coincident economic indicators and the classes do not undergo large lag-adjustment, HOU and EXP can be categorized as leading economic indicators (Stock & Watson, 1989). This lead is captured by the algorithm and is then adjusted accordingly.

[13] The kernel density estimator f_s was computed using the biweight kernel, $K(z) = (15/16)(1 - z^2)^2$, for which $\int [K'(z)]^2 dz = 2.1429$, with bandwidth h = 4 months.

The third set of chronologies are the obtained through a weighted estimation, computed as described Section 3.1.5. The lag-adjusted and weighted estimation methods require population weights $\pi_1, \ldots, \pi_6$ for the six classes of series. The results in Table 3 use population weights of 0.3 for EMP, 0.2 for IP and MT, and 0.1 for PIX, HOU, and EXP. They are similar to the ones of Stock and Watson (2014) who chose the weighting structure as 0.3 for IP, EMP, and MT, and 0.1 for PIX. In both cases, the PIX reported a low amount of turning points in some episodes and is underweighted to avoid influential outliers.

The weights used in this research are an approximation of the average importances obtained through the utilization of the Boosting framework (See Section 5.2). If these weights are applied, the mean the mean square error of the discrepancies relative to the OECD turning points is lower than when equal weights are used, indicating a better fit within the weighted chronology (See section 5.1.3).

A sensible choice of chronology would be the class-lag adjusted as it adjusts for systematic lead/lag relationships with the reference cycle and takes into account the weighting structure from the weighted estimation. However, the standard errors are generally higher than the ones observed in the random sampling framework and, as each recession is different in nature and the number of series is relatively small for some classes, the dates are more prone to be affected by outliers. Therefore, focus is set on the unadjusted mode as it relies on the actual distribution of turning points while the other measures of central tendency serve as a sensitivity check.

5.1.1 Comparison with U.S. Chronology

As the methodology for dating turning points was originally created to date the U.S. business cycle, the U.S. turning points can be used as a reference to evaluate the performance in dating the G7 chronology. When comparing Table 3 to the equivalent in Stock and Watson (2014) some important differences are observable.

First, in this research, the algorithm suggests larger adjustments to the reference chronology. This is due to the heterogeneity of the studied countries. With each country experiencing its own business cycle, the aggregation mechanism developed by the OECD may omit some nuance offered by the disaggregated dataset.

Second, the standard errors are significantly higher when using the G7 dataset. In most cases, the errors range from 0.5 to 1.5 (0.5 to 0.8 for the mode) so the typical confidence interval for a turning point is ±1 to ±3 months. The results Stock and Watson (2014) report lower standard errors, around 0.5 to 0.8 for all measures of local tendency. As recessions in different countries start and end at different points in time, the turning points are scattered across a longer time period. This necessarily leads to a higher standard error. It also gives more importance to the mode as the mean and median are largely influenced by this phenomenon, leading to higher standard errors. Additionally, the standard errors tend to be larger for earlier episodes. This is consistent with the increasing number of series over time.

Third, in most cases the mean and median suggest a correction which is closer to the reference date in comparison to the mode. Again, this occurs due to the broader dispersion of turning points across time as a consequence of including many different countries and sectors. Turning points on opposite sides of a reference date often cancel themselves out, resulting in the suggested turning point being close to the reference date. In a majority of the cases, the mean and median suggest an adjustment smaller or equal to one month. The mode is much less conservative in this regard offers valuable information as it is immune to outliers.

5.1.2 Shifts Using Different Adjustments

The mean, median, and mode estimates computed using different adjustment procedures (unadjusted, lag-adjusted, or weighted) generally agree with each other (±1 month). However, there are some episodes where the adjustments affect the chronologies in a significant way, as is the case in the 1993:8 and 2016:8 episodes.

In the 1993:8 trough, the weighted estimation dates the turning point significantly earlier compared to the other two methods. This occurs because highly weighed classes of industrial production and employment lead the turning point on average (i.e., suggesting it occurred earlier). In the 2016:8 trough, there are two instances of clustering of turning points, around -2 months and +6 months. The mode obtained through a weighted estimation identifies +6.2 months, which is a substantial shift from the unadjusted estimation of -1.9. Though the cluster around -2 months is larger, the shift is a consequence of many employment series with high weights (not specific to a single country) experiencing turning points with significant lags.

In the case of some episodes, e.g., 1967:07, 1987:02, or 2012:02, the weighted estimation of the mode introduces very large standard errors due to some highly influential outliers.

5.1.3 The Adjusted Chronology

In Table 4, the new *Adjusted chronology* is presented by rounding the turning points estimated by the unadjusted mode to the closest integer. An immediate observation is the small amount of dates which were left unaffected by the disaggregated series. This illustrates the large differences between the average-then-date approach of the Composite Leading Indicators used to date the OECD reference cycles and the date-then-average methodology. Moreover, the OECD dating mechanism identifies relatively long periods as recessions. This contrasts with a wide range of research identifying U.S. and global recessions to last approximately one year (Fushing et al., 2010; Jorda & Berge, 2011; Kose et al., 2020). By calculating the mean of the adjustment made for peaks and troughs independently, it is clearly visible that the disaggregated dataset suggests shorter recession periods on average. This suggests that the recessions provided by OECD last too long and should cover shorter periods.

The peaks have a much greater adjustment in absolute terms, this could be a similar phenomenon to that observed by Chauvet and Piger (2008) and Harding and Pagan (2006) who utilized the date-then-average approach with only four series. In their research, the troughs aligned much more closely to the NBER dates in comparison to the peaks.

In some episodes, e.g., 1980, 1990, and 2001 recessions[14], the reference peaks are far away from each other, and this adjustment is not observable. In Figures A.2, A.3, and A.4, the distribution of turning points around the OECD reference date is extended to encompass the alternative chronologies into the distribution. This view of the distribution raises a very interesting question. Are the OECD's reference cycles a realistic reflection of economic activity? As the OECD also dates the reference chronology of the U.S., it can be compared with the one provided by the NBER. The recession episodes provided by the OECD are generally longer in duration than the ones of the NBER, which last at maximum 18 months. For example, the OECD reports an almost three years long recession starting in 2000:05 and lasting until 2003:02, whereas the NBER only classifies it from 2001:03 – 2001:11. Moreover, the OECD presents recessions that are not recognized by the NBER, indicating an increased sensitivity of the procedure to classify a date as a recession.

[14] The recession period remains long even if the entire OECD area is considered. In the case of the 2020 recession the peak is reported to be 2018:04, extending the recession by more than one year compared to the Major Seven countries reference chronology. For a complete list of reference cycles see: https://www.oecd.org/sdd/leading-indicators/oecdcompositeleadingindicatorsreferenceturningpointsandcomponentseries.htm.

A similar comparison can be made with the chronologies of Fushing et al. (2010) who computed reference chronologies for the OECD nations. They also compared their U.S. chronology with the NBER, and they match up better than the one of OECD but not perfectly. Fushing et al. (2010) find recessions with shorter durations for individual countries than the OECD does. As the NBER and many researchers suggested a shorter recession time span, some dates may need to be adjusted by a large margin to reflect the movements in the global economy. This casts doubt on the ability of the OECD to accurately compute reference cycles and indicates the necessity for potential revisions to the automated dating mechanism.

Therefore, it must be stressed that not all recession episodes reported in Table 4 are necessarily instances of global (G7) recessions. They are, however, arguably a better approximation of the dates of economic downturns compared to the ones provided by the OECD. Pacella (2021) studied the CEPR chronology through the lens of the Stock and Watson (2014) using over 100 macroeconomic series and rejected a reliance solely on GDP. This serves as further evidence for a potential improvement to the OECD chronology by including additional series. An additional comparison of the turning points obtained using the alternative chronologies is available in the Appendix.

5.1.4 Significant Recession Episodes

The dating procedure by Stock and Watson (2014) presents the number of turning points triggered around the reference date together with the class the individual series belong to. In Table 4, the bold dates represent instances where more than half of the series reported a turning point. These episodes are considered to be significant. The dating procedure was carried out again but with the Adjusted chronology as a reference cycle, from which the significance measures are taken. It must be noted that if the classes were more equally represented, the procedure of selecting a significant recession would be more robust. However, it serves as a good indication of the severity of the specific economic downturn.

The significant recessions intersect with the chronologies of Fushing et al. (2010) and Kose et al. (2020) and add the 2020:04 trough. There are three instances of turning points where only one of the peak-trough pair is deemed significant.

1993:08 Trough. If the reference date is moved to 1991:01 as observable in Table A.2, the turning point suddenly becomes significant. This could be an indication that the actual turning point occurred more than two years prior to the OECD date. Fushing et al. (2010) identify the end of the episode to have occurred in the beginning of 1991 for the U.S., UK, and Canada. Only in Italy and France the episode extends to 1993. If the individual countries are examined, the 1991:01 date is driven mostly by the U.S. Chauvet and Yu (2006) find that the U.S. economy led the beginning and end of the global contractions in the early 1970s and early 1990s. On the other hand, the 1993:08 date was driven by other countries, with the turning points being spread out quite evenly among them. In the latter episode, U.S. did not report almost any series experiencing turning points. This is confirmed by the announcements of the dating committees for U.S. and Europe. NBER dates the trough to be 1991:03, whereas CERP identifies 1993:Q3 as the trough.

2003:03 Trough. The same phenomenon repeats as a shift of the reference date to 2001:11 (Table A.3) makes the turning point significant. Again, the findings of Fushing et al. (2010) align closer with the new proposed turning point. The same trend of a quicker recovery in the U.S. is present also in this episode, but the evidence of the U.S. trough occurring in 2001:11 is even stronger as the NBER also dates the end of the episode to have occurred in 2001:11. Interestingly, the CEPR does not recognize the early 2000s downturn as a recession, which would suggest relying more on the U.S. recession dates. On the other hand, the unemployment rate in the U.S. peaked in 2003:06, two months before the turning point found through the OECD chronology, indicating the implications of the recessions were still present at that time. GECON of Baumeister et al. (2020) dates the trough of economic activity to be in 2003:04.

2019:07 Peak. The downturn preceding the COVID-19 pandemic appears to be much more gradual, as many sectors experienced turning points from 2018 until 2020. The GECON indicator of Baumeister et al. (2020) indicates a decline in economic activity in starting in the second half of 2018. The OECD dates the peak for all its members combined to be 2018:04, with different countries also experiencing turning points in the two years preceding 2020. The CEPR and NBER report the trough to have happened in 2019:Q4 and 2020:02 respectively.

Essentially, all three mentioned episodes are an instance of turning points not occurring at the same time for all G7 countries but rather being more dispersed over time. This does not downplay the severeness of the recessions themselves but offers a more nuanced view of the economic climate during the time of the episodes. A further discussion is available in the Appendix.

5.1.5 Sensitivity Analysis

Three alternative chronologies were estimated as sensitivity checks. First, Table 2 was recomputed using equal population weights, i.e., $\pi_1 = \pi_2 = \cdots = \pi_6 = \frac{1}{6}$. This affects the class-specific lag and weighted estimations. The class-lag adjusted results remain within ±0.5 of the original chronologies in most cases. The weighted chronologies often differ by a larger margin, and in some episodes the difference is very significant. This occurs due to outliers in less significant classes receiving a larger weight. For example, in the episode 1990:04 the weighted mode moves from -1.1 to 7.8 after the introduction of equal weights. The personal income class reports only one series with a turning point, which occurred eight months after the reference date. Naturally, this leads to an introduction of a large bias to the mode of the distribution, but the other measures were unaffected. Same is observable in the 1997:10 episode, where only two series of personal expenditure with mean -11.5 influenced the turning point estimates in a very significant way.

Second, the estimates in Table 3 were recomputed using an extended episode band of 15 months instead of 12 months to study effects of potential outliers. The first episode 1963:03 was removed as there were not enough months from the start of the dataset. The mean is found to be the most sensitive, indicating a lack of robustness. The median was not as sensitive, and the adjustment often brought it closer to the mode. The mode was left almost unaffected apart from the class-lag adjusted estimate in 2012:02, suggesting a shift by -9.2 months while the other estimates remained stable around -1.2 months. This alludes to the robustness of the mode as a turning point estimator but also emphasizes the caution needed when handling the adjusted estimates.

Third, the classes were changed to indicate different countries, rather than sectors. The weights for the class-lag and weight adjusted chronologies were taken as the relative GDP levels of the individual countries.[15] The GDP of the U.S. accounts for half of the total G7 output. Moreover, the influence of the U.S. is amplified thanks to the greater representation within the dataset even before the GDP weighting structure is introduced. Japan has a leading relationship with the reference cycle of approximately 1.5 months and France and Germany lag more than one month. The other countries are adjusted by less than 1 month. The UK doesn't have a significant adjustment, which is in contrast to Chauvet and Yu (2006), who found the recessions in the UK

[15] US = 0.50, CN = 0.05, JP = 0.12, UK = 0.075, BD = 0.11 FR = 0.076, IT = 0.069.

occur the earliest of all and last longer than most G7 countries. Upon deeper examination of the results, such trend does not seem to be observable.

The results are very similar to the ones in Table 2 but contain some valuable insights. The U.S. indicators have a very large impact on the final turning point due to the GDP weighting structure. In some cases, this leads to major discrepancies between the adjustment methods. Moreover, it seems that the sectors behave in a more consistent manner in the class-lag adjustment procedure, suggesting that the sectors across countries comove in a more synchronized way than the actual countries as a whole. This may be a consequence of an unbalanced representation of sectors in individual countries (e.g., Italy does not include personal expenditures) or an effect of globalization and interconnectedness of economies across borders.

On the other hand, some episodes behave more consistently in the country classification. For example, the 1997:10 episode estimates of mode are very close across all adjustment methods, compared to the unexpected jump in the class-lag adjusted mode visible in Table 2. This is again due to the large weight of the U.S. However, this episode occurred due to the East Asian financial crisis, while advanced economies mainly maintained their growth (Kose et al., 2020). As shown by Chauvet and Yu (2006), the Japanese economy moved together with the other G7 and OECD countries in the 1970s and 1980s. Japan experienced two severe and long recessions in the 1990s. Indeed, Japan records the most turning points, and their estimate matches the official dating committee. The amount of turning points triggered in other countries was not particularly high but they behaved in a similar fashion, influencing the final turning point estimate. Fushing et al. (2010) finds the corresponding recessions in Japan, France, and United Kingdom. The latter two lasted a shorter time and correspond closer to the dates found in the Adjusted chronology.

A noteworthy takeaway is the consistency of the unadjusted mode in a majority of episodes. Different weighting structures may be influenced by outliers but a combination of the country-based and sector-based usually point to the unadjusted mode as the most reliable measure. This is, however, a mere observation and a quantification of such a phenomenon could be an intriguing way to better estimate the turning points in international business cycles.

5.2 Boosting

5.2.1 Specifications

The boosting framework was utilized for two different reasons. First, is to determine, whether the OECD or the Adjusted chronology is a better reflection reality by forecasting the recession dates. Second, the boosting framework chooses the best predictors from a large set to best predict the binary target variable (1 = recession, 0 = no recession).[16] These importance measures can be extracted from the model to give an idea of which variable/sector is most valuable for recession dating.

To evaluate the performance of a machine learning model, the dataset must be separated into a training set and tests set. For each estimation, test sets of size 20%, 30%, 40%, and 50% of the total length of the series were considered. The model chooses the test dates at random to avoid training on “old” data to determine a “new” result. This is particularly important in economic research as each recession has different characteristics and drivers. Moreover, the model was tested for a different number of lags included for each predictor. Specifically, 3, 6, and 12 lags of each variable were incorporated to get a better understanding of the model’s behavior. This seemingly large number of characterizations were carried out to ensure stability.

Ng (2014) chooses to restrict the tree to a maximum dept of 1, transforming the tree to a stump. This method was tested and, although the results change slightly, it does not seem to have a systematic effect on the forecasting performance. The number of nodes within a decision tree was left at 6, which is the default setting of the XGBoost. The maximum number of boosted decision trees was set to 1000 to avoid overfitting.

It must be noted that no out-of-sample forecasting was carried out and all of the presented results originate from the in-sample fit. The Receiver Operating Characteristics (ROC) curve was created for each forecasting iteration, with the area under the curve (AUC) being the main evaluation metric of performance.

[16] The target variable was chosen to represent the current state of the economy rather than just turning points. The number of occurrences of turning points is very low and it would make the evaluation almost impossible.

5.2.2 Results

From the extensive tests carried out on the boosting algorithm it is clear that the results are not completely stable and are highly dependent on the amount of lags chosen, size of the test set, and the composition of the test set. One observation constant across the forecasting framework of boosting is that the way of classifying a recession has a direct impact on the forecasting performance. If peaks are included in the recession dummy variable, the model reports more false negatives as it probably recognizes the peak as a part of the expansion and the AUC scores are systematically worse. Therefore, the peak date is not classified as a recession in any of the presented results. The results of the general model and all sensitivity exercises are available in Table 5.

The new Adjusted chronology was compared to the OECD chronology. They were used as a target variable and the disaggregated dataset was trained to predict the chronologies. The comparison using only the disaggregated dataset as predictors is shown in the first panel of Table 5. With very few exceptions, the Adjusted chronology performed better in terms of the AUC score in all scenarios. The OECD chronology received AUC scores ranging from 0.87 to 0.91 (only 2 out of 12 surpassed 0.90), with smaller test sizes receiving usually greater scores. For the Adjusted chronology the scores are often similar but for specific settings perform better, they range from 0.87 to 0.94. For example, if 6 lags are included, which seemingly produces the best results, the AUC scores are on average 0.915 across the different test sizes. This suggests that the Adjusted chronology is a better reflection of the indicators of real economic activity.

5.2.3 Importance

The boosting framework presents the importances of factors used for predictions as it gradually learns to fit the model. Assessing the importances of individual series is not very practical due to large inconsistencies and the focus is instead set on classes of series. Though the importance of individual classes changes depending on the specifications of the model, there is a clear indication that some sectors and countries are more significant than others. As the classes contain a different number of series, the average importance is taken.

In the sector classification, the most important sector is consistently employment, with manufacturing and trade sales, and industrial production having slightly lower but similar scores. The other three sectors (personal income, housing starts, and personal expenditure) are generally less important and often report similar scores to each other. The importance of the sectors is reflected in the weights used in the weighted estimation of the Adjusted chronology in Table 3. Through a boosting framework, Berge (2015) finds industrial production and employment to be the best short-term predictors for the U.S. business cycle, which is a reassuring result. The general classification of housing starts and personal expenditure as leading indicators (Stock & Watson, 1989) does not seem to have an impact on their predictive power as they are usually assigned a low importance.

The country classification also offers interesting results, albeit less consistent. U.S. is the most important or very close to being the most important in a majority of the cases. This generally confirms the findings in previous literature. Aruoba et al. (2011) identify U.S. as having the most effect on the G7 business cycle. Brandt et al. (2021) show that U.S. factors are important drivers of the euro area asset prices, but not the other way around. It was also found to be the main driver of decreased volatility in the global economy (Dept., 2007).

Indicators in France are often found important, with a strong emphasis on employment. The UK statistics are also often an important factor for making predictions. The other countries also seem to be important predictors, but only for a specific test size or number of lags included. An illustration of the average importances is presented in Figure 3 and Figure 4. Note that 3 lags and a test set of size 50% was used as scores differ with a change of specifications, especially for the country classification.

5.2.4 Inclusion of Other Variables

Better prediction scores using the adjusted chronology are an expected result as the dataset, which created the chronology, should be better at predicting it. Analogously, one would expect the OECD composite leading indicators to be better at predicting the OECD chronology. Therefore, other variables should be included to evaluate the actual performance. Five additional indicators are included, which describe the movements in financial markets and the global business cycle (See Section 4.3). The results are presented in the second panel of Table 5. The forecasting performance is systematically better if the financial and global variables are included, and predictions of the Adjusted chronology remain better compared to the OECD chronology. The range of AUC scores for different specifications is 0.88 – 0.94 if the OECD chronology is used, and 0.90 – 0.95 in the case of the Adjusted chronology.

An interesting insight is the consistency of the AUC scores across different test sizes when only 3 lags are incorporated into the predictor set. On the other hand, sizeable inconsistencies are observable with 12 lags included, where one of the highest scores is recorded for a test size of 20% and the worst is recorded for the test size of 50%. This bad performance could be an instance of overfitting, although a change of iterations does not resolve the issue. Alternatively, the machine learning algorithm is learning to predict a value and is relying more on the lags than the actual values. Together with the decreasing test size the machine learning algorithm learns wrong patterns. The outcome does not change if only important variables are considered and if only variables more important than a certain value are considered, the performance diminishes. This trend is observable also if just the disaggregated dataset used as a predictor set.

If only financial and global variables are used as predictors, the AUC scores are lower, but the performance is very similar between the two considered chronologies. The decrease is expected as only VXO is available for the first eight years of the sample. The AUC scores are between 0.83 – 0.89 in the case of the OECD chronology and 0.83 – 0.90 for the Adjusted chronology. The trend of lower scores for a 12 lags and large test set model specification is not as apparent in this case because the smaller test sizes also report lower scores.

The financial variables are by far the most significant if average importances are used. Therefore, total importances are displayed in Figure 5 and Figure 6 as an illustration of scale. There are some differences depending on which chronology is used as the target variable. The OECD chronology is best predicted by the GCI, with GECON and EBP receiving smaller but still very significant weights. The GFC and VXO generally have a very low importance. Using the Adjusted chronology, the importances are more equally distributed among the indicators, with VXO receiving the lowest importance in most cases.

5.2.5 Cross-validation

So far, the composition of the test set was randomly sampled. To check for consistency, a cross-validation was carried out by splitting the train and test sets into k equally sized folds as illustrated in Figure 7. With each additional fold, the training set grows by including the test set of the previous fold. For $k = 5$, the test samples are split into five periods, approximately nine years long, ranging from 1975:03 to 2022:08. The AUC scores for all model specifications are reported in Table 5.

Only disaggregated dataset. Predictions of the adjusted chronology obtain a greater or equal AUC score in most cases. The model performs better for earlier time periods (the first three folds, 1975:03 – 2003:8), after which the performance diminishes. This trend is visible in all tested scenarios but is much more pervasive if more lags are included.

Disaggregated dataset with financial and global variables. Adjusted chronology reports better or very similar AUC scores in the first three folds (1975:03 – 2003:8) but significantly worse performance in the following 2 folds (2003:09 – 2022:08). If the beginning of the training set is moved to 1993 and is tested only on the remaining two folds (2003:09 – 2022:08), the performance increases dramatically for the Adjusted chronology and remains relatively stable with different number of lags included. The OECD experiences an improvement in the last fold but a slight decrease in performance in the period 2003:09 - 2013:02. Though it does not reach the levels of the OECD chronology, it suggests the economic climate and the threshold of what constitutes a recession have changed throughout the years. This phenomenon is not observable if only the disaggregated dataset is considered. The boosting model does not recognize most of the year 2015 as recessionary, thus producing more false negatives in the Adjusted chronology case as it extends the episode peak to 2014:12. Moreover, the algorithm produces many false positives in late 2018 and most of 2019 as the global economic activity was slowing down across a long period of time (See Section 5.1.4). By adjusting the peak to 2020:02, the diminished performance does not come as a surprise.

Only financial and global variables. The AUC scores are very similar in the first four folds (1975:03-2013:02). However, the performance in the last decade is consistently better when predicting the OECD chronology. If the beginning of the training set is moved to 1993 the performance of the Adjusted chronology increases significantly for both tested periods and becomes more stable, indicating a change of economic conditions. In the period 2003:09 - 2013:02 the Adjusted chronology performs better the one of OECD. On the other hand, the OECD performs better for the period 2013:03 – 2022:08 due to the same reason discussed previously.

6 Conclusion

This thesis studies the classification of the G7 business cycle into expansions and recessions. From the results in Sections 5.1 and 5.2 it is clearly visible that the use of a large, disaggregated dataset provides very useful insights to the process of dating turning points. The OECD chronology, which relies on highly aggregated composite leading indicators, is reviewed using the Stock and Watson (2014) algorithm. Relying on the distribution mode obtained from the simple random sampling procedure, an Adjusted chronology, more reflective of real macroeconomic activity, is provided. The algorithm suggests five significant recession episodes which align with the research of Kose et al. (2020) and Fushing et al. (2010), suggesting not all OECD reference dates are instances of actual global recessions. As the OECD has a wide range of available statistics at hand, some of which are available for a long historical period, the use of a date-then-average dating mechanism would introduce greater precision to the reference chronologies currently provided.

The XGBoost machine learning algorithm provided a very insightful assessment of the in-sample fit, while being fast and easy-to-use. The predictions of the Adjusted chronology generally reported better AUC scores than when the OECD chronology was used as a target variable. The model's scores of importance of individual sectors and countries usually reflected findings of previous literature, ranking United States as the most influential country. Employment, followed by industrial production, and manufacturing and trade sales were labeled as the most important sectors for dating turning points. Moreover, the algorithm also provided a brief assessment of financial and global indicators, with GECON by Baumeister et al. (2020) and Global Conditions Index by Cuba-Borda et al. (2018) receiving the greatest importance on average.

Inclusion of a larger number of variables and countries would make the estimated chronology more precise and potentially resolve some of the contested episodes as discussed in the Appendix. For example, the 37-country dataset of 392 monthly indicators used by Ferrara and Marsilli (2019) would make for an interesting analysis of global recessions.

7 Bibliography

Aguiar-Conraria, L., & Joana Soares, M. (2011). Business cycle synchronization and the Euro: A wavelet analysis. *Journal of Macroeconomics*, *33*(3), 477–489.

Alquist, R., Bhattarai, S., & Coibion, O. (2020). Commodity-price comovement and global economic activity. *Journal of Monetary Economics*, *112*, 41–56.

Aruoba, S. B., Diebold, F. X., Kose, M. A., & Terrones, M. E. (2011). Globalization, the Business Cycle, and Macroeconomic Monitoring. *NBER International Seminar on Macroeconomics*, *7*(1), 245–286.

Astolfi, R., Gamba, M., Guidetti, E., & Pionnier, P.-A. (2016). *The use of short-term indicators and survey data for predicting turning points in economic activity : A performance analysis of the OECD system of CLIs during the Great Recession*.

Baumeister, C., & Guérin, P. (2020). *A Comparison of Monthly Global Indicators for Forecasting Growth*.

Baumeister, C., & Hamilton, J. D. (2019). *Structural Interpretation of Vector Autoregressions with Incomplete Identification: Revisiting the Role of Oil Supply and Demand Shocks **.

Baumeister, C., Korobilis, D., & Lee, T. K. (2020). *Energy Markets and Global Economic Conditions*.

Berge, T. (2015). Predicting Recessions with Leading Indicators: Model Averaging and Selection Over the Business Cycle. *Journal of Forecasting*, *34*, 455–471.

Bloom, N. (2007). The Impact of Uncertainty Shocks. *Econometrica*, *77*.

Bodman, P., & Crosby, M. (2005). Are business cycles independent in the G7? *International Economic Journal*, *19*, 483–499.

Brandt, L., Guilhem, A. Saint, Schröder, M., & Robays, I. Van. (2021). *What drives euro area financial market developments? The role of US spillovers and global risk* (2560; ECB Working Paper Series).

Bry, G., & Boschan, C. (1971). *Cyclical Analysis of Time Series: Selected Procedures and Computer Programs*. National Bureau of Economic Research, Inc.

Burns, A. F., & Mitchell, W. C. (1938). *Statistical Indicators of Cyclical Revivals*. National Bureau of Economic Research, Inc.

Burns, A. F., & Mitchell, W. C. (1946). *Measuring Business Cycles*. NBER.

Chauvet, M. (1998). An Econometric Characterization of Business Cycle Dynamics with Factor Structure and Regime Switching. *International Economic Review*, *39*(4), 969–996.

Chauvet, M., & Hamilton, J. D. (2006). *Dating Business Cycle Turning Points*.

Chauvet, M., & Piger, J. (2008). A Comparison of the Real-Time Performance of Business Cycle Dating Methods. *Journal of Business Economics and Statistics*, *26*, 42–49.

Chauvet, M., & Potter, S. (2005). Forecasting Recessions Using the Yield Curve. *Journal of Forecasting*.

Chauvet, M., & Yu, C. (2006). International Business Cycles: G7 and OECD Countries. *Economic Review*.

Chen, T., & Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System*.

Cuba-Borda, P., Mechanick, A., & Raffo, A. (2018). Monitoring the World Economy: A Global Conditions Index. *IFDP Notes*, *2018*(45).

de Bondt, G., & Vermeulen, P. (2021). Business cycle duration dependence and foreign recessions. *Scottish Journal of Political Economy*, *68*(1), 1–19.

Delgado, P. C., Congregado, E., Golpe, A. A., & Vides, J. C. (2022). The Yield Curve as a Recession Leading Indicator. An Application for Gradient Boosting and Random Forest. *International Journal of Interactive Multimedia and Artificial Intelligence*, *7*(3), 7–19.

Dept., I. M. Fund. R. (2007). *World Economic Outlook, October 2007: Globalization and Inequality*. International Monetary Fund.

Diebold, F., & Rudebusch, G. (1996). Measuring Business Cycles: A Modern Perspective. *The Review of Economics and Statistics*, *78*(1), 67–77.

Enea, S.-F., Palaşcă, S., & Ţigănaş, C. (2015). G7 Countries – Advocates of the Global Business Cycle. *Procedia Economics and Finance*, *20*, 193–200.

Favara, G., Simon Gilchrist, Kurt F. Lewis, & Egon Zakrajšek. (2016). Updating the Recession Risk and the Excess Bond Premium. In *FEDS Notes*. Washington: Board of Governors of the Federal Reserve System.

Ferrara, L., & Marsilli, C. (2019). Nowcasting global economic growth: A factor-augmented mixed-frequency approach. *The World Economy*, *42*(3), 846–875.

Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, *29*(5), 1189–1232. http://www.jstor.org/stable/2699986

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, *38*(4), 367–378.

Fushing, H., Chen, S.-C., Berge, T. J., Jordà, Ò., & Berge, T. (2010). *A Chronology of International Business Cycles Through Non-parametric Decoding* (11–13).

Gilchrist, S., & Zakrajšek, E. (2012). Credit Spreads and Business Cycle Fluctuations. *American Economic Review*, *102*(4), 1692–1720.

Giusto, A., & Piger, J. (2017). Identifying business cycle turning points in real time with vector quantization. *International Journal of Forecasting*, *33*(1), 174–184.

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

Hamilton, J. D. (2005). What's real about the business cycle? In *Federal Reserve Bank of St. Louis Review* (Vol. 87, Issue 4, pp. 435–452). Federal Reserve Bank of St.Louis.

Hamilton, J. D. (2011). Calling Recessions in Real Time. *International Journal of Forecasting*, *27*(4), 1006–1026.

Harding, D., & Pagan, A. (2006). Synchronization of cycles. *Journal of Econometrics*, *132*(1), 59–79.

Jain, Y., & Bhasin, V. (2022, September 30). *Predicting Recession: Forward looking recession forecast based on real time variables*. Global Banking and Finance Review.

Jorda, O., & Berge, T. (2011). The Classification of Economic Activity into Expansions and Recessions. *American Economic Journal: Macroeconomics*, *3*, 246–277.

Juglar, C. (1862). Des crises commerciales et de leur retour périodique en France: en Angleterre et aux États-Unis. *Guillaumin e Cie*.

Kilian, L. (2009). *Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market*.

Kitchin, J. (1923). Cycles and Trends in Economic Factors. *The Review of Economics and Statistics*, *5*(1), 10–16.

Kondratiev, N. D. (1925). The Major Economic Cycles. *Voprosy Konjunktury*, *1*(1), 28–71.

Kose, M. A., Sugawara, N., & Terrones, M. E. (2020). *Global Recessions* (10/2020).

Kumar, R. R., Stauvermann, P. J., & Vu, H. T. T. (2021). The Relationship between Yield Curve and Economic Activity: An Analysis of G7 Countries. *Journal of Risk and Financial Management*, *14*(2).

Leventis, D. (2018, November 11). *XGBoost Mathematics Explained*. Medium.Com.

Liu, J., Li, J., Sun, X., Xie, Y., Lei, J., & Hu, Q. (2014). An Embedded Co-AdaBoost based construction of software document relation coupled resource spaces for cyber-physical society. *Future Generation Computer Systems*, *32*(1), 198–210.

Miranda-Agrippino, S., & Rey, H. (2020). U.S. Monetary Policy and the Global Financial Cycle. *Review of Economic Studies*, *87*(6), 2754–2776.

Miranda-Agrippino, S., & Ricco, G. (2021). The Transmission of Monetary Policy Shocks. *American Economic Journal: Macroeconomics*, *13*(3), 74–107.

Mitchell, W. C. (1927). *Business Cycles: The Problem and Its Setting*. National Bureau of Economic Research, Inc.

Nalewaik, J. J. (2007). *Estimating Probabilities of Recession in Real Time Using GDP and GDI*.

Ng, S. (2014). Viewpoint: Boosting Recessions. *Canadian Journal of Economics*, *47*(1), 1–34.

Ojo, M. O., Aguiar-Conraria, L., & Soares, M. J. (2023). The performance of OECD's composite leading indicator. *International Journal of Finance and Economics*.

Otsu, T. (2017). Abridged Bry-Boschan Procedure. *Seijo Economic Research*, *215*.

Pacella, C. (2021). Dating the Euro Area Business Cycle: An Evaluation. In *SSRN Electronic Journal* (1332).

Racocha, K. (2017). *Impact of Drought on Commodity Market Forecasting*. Charles University.

Ravazzolo, F., & Vespignani, J. (2020). *World Steel Production: A New Monthly Indicator of Global Real Economic Activity*.

Schumpeter, J. A. (1939). *Business cycles* (Vol. 1). Mcgraw-hill New York.

Statista Research Department. (2023, April 18). *G7 countries' share of world's GDP 2000-2023*. Statista.Com.

Stock, J. H., & Watson, M. W. (2014). Estimating Turning Points Using Large Data Sets. *Journal of Econometrics*, *178*(Part 2), 368–381.

Stock, J., & Watson, M. (1993). A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems. *Econometrica*, *61*(4), 783–820.

Stock, J., & Watson, M. W. (1989). New Indexes of Coincident and Leading Economic Indicators. *NBER Macroeconomics*, *4*, 351–409.

Stock, J., & Watson, M. W. (1991). A Probability Model of the Coincident Economic Indicators. In G. Moore & K. Lahiri (Eds.), *The Leading Economic Indicators: New Approaches and Forecasting Records* (pp. 63–90). Cambridge University Press.

Vrontos, S. D., Galakis, J., & Vrontos, I. D. (2021). Modeling and predicting U.S. recessions using machine learning techniques. *International Journal of Forecasting*, *37*(2), 647–671.

Watson, M. (1994). Business cycle durations and postwar stabilization of the U.S. economy. *American Economic Review*, *84*(1), 24–46.

Yousuf, K., & Ng, S. (2020). Boosting High Dimensional Predictive Regressions with Time Varying Parameters. *Journal of Econometrics*.

8 Tables

The Disaggregated Dataset

	United States		Class	Datastream Mnemonic	Start Date	End Date
1	Industrial Production		IP	USIPMMATG	1965:01	2023:04
2	Industrial Production	Equipment	IP	USIPMEQPH	1965:01	2023:04
3	Industrial Production	Consumer Goods	IP	USIPMCOGG	1965:01	2023:04
4	Industrial Production	Energy	IP	USIPMENTG	1968:01	2023:04
5	Industrial Production	Manufacturing	MT	USIPMAN.G	1972:01	2023:02
6	Total Retail Trade		MT	USOSLI07E	1990:05	2023:04
7	Wholesale trade	Inventories	MT	USINSW..B	1992:01	2023:04
8	Wholesale trade	Sales	MT	USSWTOT.B	1992:01	2023:04
9	Housing Starts		HOU	USHOUSE.O	1977:01	2023:03
10	Earnings (NAICS)	Manufacturing (Hourly)	PI	USWAGMANA	1965:01	2023:04
11	Earnings (NAICS)	Trade, Transport & Utilities (Hourly)	PI	USWRIT..B	1965:01	2023:04
12	Earnings (NAICS)	Other services (Hourly)	PI	USWR81..B	1965:01	2023:04
13	Unemployment Rate		EMP	USUN%TOTQ	1965:01	2023:03
14	Employment (NAICS)	Manufacturing	EMP	USEMPMANO	1965:01	2023:03
15	Employment (NAICS)	Trade, Transport & Utilities	EMP	USEMIT..O	1965:01	2023:04
16	Employment (NAICS)	Other services	EMP	USEM81..O	1965:01	2023:04
17	Personal Consumption Expenditure	Goods	EXP	USDGDSRCB	1965:01	2023:04
18	Personal Consumption Expenditure	Services	EXP	USCONSRVB	1965:01	2023:04
19	Personal Consumption Expenditure	Nondurables	EXP	USCONNDRB	1965:01	2023:04
20	Personal Consumption Expenditure	Durables	EXP	USCONDURB	1965:01	2023:04

Canada

21	Industrial Production		IP	CNCIND..G	1965:01	2023:03
22	Manufacturing		MT	CNOPRI38B	1965:01	2023:03
23	Total Retail Trade		MT	CNOSLI07E	1990:05	2022:12
24	Wholesale trade	Inventories	MT	CNINVWT.B	1992:08	2023:04
25	Wholesale trade	Sales	MT	CNSWTOT.B	1992:08	2023:04
26	Housing Starts		HOU	CNHOUSE.O	1977:01	2023:03
27	Earnings	Goods-Producing Industry (Weekly)	PI	CN186864	1991:01	2023:02
28	Earnings	Service-Producing Industry (Weekly)	PI	CN186872	1991:01	2023:02
29	Unemployment Rate		EMP	CNUN%TOTQ	1966:01	2023:03
30	Employment	Goods-Producing Industry	EMP	CNEMPGPLO	1976:01	2023:04
31	Employment	Service-Producing Industry	EMP	CNEMPSPDO	1976:01	2023:04

Japan

32	Industrial Production		IP	JPIPCONGG	1978:01	2023:04

33	Industrial Production	Manufacturing	MT	JPIPTOT.G	1965:01	2023:04
34	Total Retail Trade		MT	JPOSLI07E	1990:01	2023:02
35	Wholesale Trade	Sales	MT	JPWHOLSLA	1969:12	2023:03
36	Housing Starts		HOU	JPHOUSE.O	1977:01	2023:03
37	Earnings	Manufacturing (Weekly)	PI	JPWAGMANE	1969:12	2023:03
38	Earnings	Wholesale & Retail Trade (Weekly)	PI	JPWAWRREE	2000:01	2023:03
39	Earnings	Construction (Weekly)	PI	JPWACSREE	1965:01	2023:03
40	Earnings	Medical, Health Care & Welfare (Weekly)	PI	JPWAMHWEE	2000:01	2023:03
41	Unemployment Rate		EMP	JPUN%TOTQ	1965:01	2023:03
42	Employment	Manufacturing	EMP	JPEMPMANP	1965:01	2023:03
43	Employment	Wholesale & Retail Trade	EMP	JPEMPWRRP	1998:01	2023:03
44	Employment	Construction	EMP	JPEMPCONP	1965:01	2023:03
45	Employment	Medical, Health Care & Welfare	EMP	JPEMPMHWP	1998:01	2023:03
46	Household Living Expenditure		EXP	JPHLEXPWA	1965:01	2023:04

United Kingdom

47	Industrial Production		IP	UKIPTOT.G	1965:01	2023:04
49	Industrial Production	Consumer Goods	IP	UKK24P..G	1995:01	2023:04
48	Industrial Production	Manufacturing	MT	UKIPMAN.G	1965:01	2023:04
50	Total Retail Trade		MT	UKOSLI07E	1990:01	2023:03
51	Earnings	Public Sector (Weekly)	PI	UKKAC7..B	2000:01	2023:02
52	Earnings	Private Sector (Weekly)	PI	UKKAC4..B	2000:01	2023:02
53	Earnings	Whole Economy (Weekly)	PI	UKEARNTOT*	1965:01	1999:12
54	Earnings	Public Sector (Weekly)	PI	UKEARNPUB*	1990:01	1999:12
55	Earnings	Private (Weekly)	PI	UKEARNPRI*	1990:01	1999:12
56	Unemployment Rate		EMP	UKMLM006Q	1969:01	2021:08
57	Employment	Part-Time	EMP	UKYCCU..O	1992:04	2023:01
58	Employment	Full-Time	EMP	UKYCBK..	1992:04	2023:01
59	Employment	Total	EMP	UKLF2G..O	1971:02	2023:01
60	Personal Expenditure		EXP	UKCHBV..	1987:01	2023:04

Germany

61	Industrial Production		IP	BDI66..CE	1965:01	2023:02
62	Industrial Production	Manufacturing and Mining &quarying	MT	BDIPMMQ.G	1965:01	2023:04
63	Industrial Production	Construction	IP	BDIPBLD.G	1991:01	2023:04
64	Total Retail Trade		MT	BDOSLI07E	1990:01	2023:03
65	New Orders Recorded: Construction		HOU	BDHOUSE.G	1991:01	2023:03
66	Earnings	Manufacturing and Mining (Index)	PI	BDUSRB08G	1991:01	2023:03
67	Unemployment Rate		EMP	BDMLM006Q	1969:01	2023:04
68	Employment	Manufacturing and Mining	EMP	BDEMPPRDF	2005:01	2023:03
69	Employment	Wholesale & Retail Trade and MV Repair	EMP	BDEWRTMVF	2000:01	2023:03
70	Employment	Construction	EMP	BDSIBTOTP	1995:01	2023:03

France

71	Industrial Production		IP	FRI66..CE	1965:01	2023:02
72	Manufacturing		MT	FRIPMAN.G	1965:01	2023:04
73	Total Retail Trade		MT	FROSLI07E	1990:01	2023:03
74	Wholesale Trade	Turnover (Exc. MV, Motorcycles)	MT	FRESBRGDE	1999:01	2023:04
75	Housing Starts		HOU	FRHOUSE.P	2000:01	2023:03
76	Unemployment Rate		EMP	FRCUNP..Q	1975:02	2022:11
77	Employment	Manufacturing (Interpolated - FROBS080Q)	EMP	FREMPMFGO	1995:04	2023:04
78	Employment	Wholesale & Retail Trade and MV Repair (Interpolated - FR45.5.BQ)	EMP	FR577274O	1995:04	2023:04
79	Employment	Construction (Interpolated - FRSURFERQ)	EMP	FREMPCSTO	1995:04	2023:04
80	Household consumption	Goods	EXP	FRHCON..D	1980:01	2023:04
81	Household consumption	Engineered Products	EXP	FRHCONMGD	1980:01	2023:04
82	Household consumption	Energy	EXP	FRHCONE.D	1980:01	2023:04

Italy

83	Industrial Production		IP	ITI66..CE	1965:01	2023:02
84	Manufacturing		MT	ITOPRI38H	1965:01	2023:04
85	Total Retail Trade		MT	ITOSLI07E	1990:01	2023:02
86	Earnings	Manufacturing (Weekly)	PI	ITMLC007E	1965:01	2023:02
87	Earnings	Industry (Hourly)	PI	ITEPGEINE	2000:01	2023:02
88	Earnings	Services (Hourly)	PI	ITEPGESNE	2000:01	2023:02
89	Unemployment Rate		EMP	ITUN%TOTQ	2004:01	2023:03
90	Employment	Manufacturing (Interpolated - ITOBS080Q)	EMP	ITESOOYMO	1995:04	2023:04
91	Employment	Wholesale & Retail Trade (Interpolated - IT45.5.BQ)	EMP	ITES2TIVO	1995:04	2023:04

	Other Variables	Authors		
	Global Economic Conditions Indicator (GECON)	(Baumeister et al., 2020)	1973:02	2023:04
	Chicago Board of Options Exchange VXO index	(Bloom, 2007)	1965:01	2020:03
	Global Financial Cycle Indicator (GFC)	(Miranda-Agrippino & Rey, 2020)	1980:01	2019:04
	Excess bond premium (EBP)	(Favara et al., 2016)	1973:01	2023:04
	Global Conditions Index (GCI)	(Cuba-Borda et al., 2018)	1973:01	2017:12

*The three series reporting earnings in the UK are available at: https://webarchive.nationalarchives.gov.uk/ukgwa/20160105231310/http://www.ons.gov.uk/ons/guide-method/method-quality/specific/labour-market/articles-and-reports/index.html.

Table 1. The disaggregated dataset of 91 economic series. Each series is reported with a series number and the corresponding Datastream Mnemonic. Other variables are important indicators of economic and financial activity included in the boosting framework.

Reference Chronologies

Turning point	OECD	Baumeister et al. (2020)	Kose et al. (2020)	Fushing et al. (2010)
Peak	1966:03			
Trough	1967:07			
Peak	1969:03			
Trough	1971:03			
Peak	1973:05	1973:06	1974:01	1973:11
Trough	1975:05	1975:03	1975:01	1975:01
Peak	1979:08	1979:03	1981:10	1980:04
Trough	1982:11	1982:08	1982:10	1980:04
Peak	1985:09			
Trough	1987:02			
Peak	1990:04	1990:01	1990:10	
Trough	1993:08	1991:01	1991:01	
Peak	1997:10			
Trough	1999:03			
Peak	2000:06	2000:08		2000:04
Trough	2003:03	2003:04		2001:05
Peak	2008:01	2007:07	2008:07	2008:02
Trough	2009:05	2009:07	2009:01	2009:04
Peak	2012:02	2011:02		
Trough	2012:12	2012:08		
Peak	2015:03	2014:12		
Trough	2016:08	2016:11		
Peak	2019:07	2018:08		
Trough	2020:05	2020:05		
Peak	2022:01	2021:12		

Table 2. Reference chronologies. To create a chronology from Baumeister et al. (2020), each monthly value of the standardized GECON indicator of economic activity was added to the previous one and the local maximums (minimums) were taken as peaks (trough). Kose et al. (2020) provide a quarterly chronology and the first month of the quarter was taken as the turning point date.

OECD Dates		No Adjustments			Class-Lag Adjusted			Weighted Estimation		
		Mean	**Median**	**Mode**	**Mean**	**Median**	**Mode**	**Mean**	**Median**	**Mode**
1966:03	P	2.9 (1.6)	1.0 (1.8)	-0.2 (0.6)	2.4 (1.7)	0.1 (1.8)	-0.7 (0.8)	2.5 (2.3)	0.0 (0.7)	-0.3 (3.7)
1967:07	T	-0.3 (2.2)	-2.0 (1.7)	-1.5 (0.9)	-0.8 (2.1)	-2.5 (1.8)	-1.9 (1.1)	-0.9 (2.6)	-2.0 (1.0)	-1.4 (56.4)
1969:03	P	0.1 (1.6)	3.0 (1.6)	4.3 (0.4)	0.0 (1.6)	1.0 (1.6)	2.9 (1.2)	-0.1 (1.4)	3.0 (0.9)	4.2 (1.1)
1971:03	T	0.4 (1.8)	0.0 (4.2)	-4.4 (0.3)	0.2 (1.8)	-1.4 (2.3)	-3.5 (2.6)	-0.3 (2.0)	0.0 (3.8)	-4.5 (0.5)
1973:05	P	2.8 (1.4)	5.0 (1.4)	7.3 (0.8)	2.7 (1.4)	5.0 (1.2)	6.6 (1.6)	2.8 (1.1)	6.0 (0.9)	6.8 (1.3)
1975:05	T	-0.6 (0.8)	0.0 (0.8)	0.6 (0.5)	-0.8 (0.7)	-0.8 (0.7)	-0.5 (0.5)	-0.6 (0.8)	0.0 (0.9)	0.6 (0.9)
1979:08	P	0.7 (1.1)	0.0 (1.3)	-1.3 (0.5)	0.6 (1.0)	-0.5 (1.5)	-2.2 (8.4)	0.5 (1.2)	-1.0 (1.0)	-1.0 (1.6)
1982:11	T	-0.1 (1.0)	0.0 (0.8)	0.1 (0.3)	-0.1 (0.9)	-0.5 (0.8)	-0.7 (0.3)	-0.2 (1.0)	0.0 (0.8)	0.1 (1.3)
1985:09	P	0.9 (1.5)	-1.0 (3.6)	-4.0 (0.7)	1.0 (1.5)	-1.5 (2.3)	-3.4 (0.6)	0.6 (1.8)	-1.0 (2.0)	-3.7 (0.7)
1987:02	T	-2.2 (1.5)	-5.0 (1.6)	-6.1 (0.3)	-2.3 (1.7)	-4.0 (1.7)	-6.1 (1.0)	-1.9 (1.5)	-2.0 (1.7)	-5.9 (9.4)
1990:04	P	0.9 (1.1)	1.0 (1.5)	-1.1 (1.1)	1.0 (1.1)	1.5 (1.7)	-1.8 (1.1)	0.7 (1.3)	1.0 (1.7)	-1.1 (2.2)
1993:08	T	-0.4 (1.2)	0.0 (1.5)	-0.5 (0.6)	-0.7 (1.2)	-1.2 (1.5)	-0.9 (0.6)	-2.0 (1.4)	-2.0 (1.9)	-6.4 (0.8)
1997:10	P	-0.6 (1.5)	0.0 (3.7)	4.7 (0.5)	-0.6 (1.4)	-0.8 (3.1)	-8.5 (0.8)	-0.9 (1.4)	0.0 (3.2)	4.7 (1.0)
1999:03	T	-2.0 (1.3)	-4.0 (1.2)	-5.1 (1.6)	-2.3 (1.2)	-3.5 (1.3)	-4.3 (2.2)	-1.4 (1.3)	-3.0 (1.0)	-3.9 (0.6)
2000:06	P	2.7 (0.8)	5.0 (0.7)	6.0 (0.3)	2.6 (0.7)	4.5 (0.7)	5.5 (0.2)	2.3 (0.8)	4.0 (1.1)	5.8 (0.4)
2003:03	T	2.1 (1.1)	3.0 (1.1)	4.5 (0.9)	2.0 (1.0)	4.0 (0.9)	4.3 (0.5)	2.1 (1.1)	3.0 (1.3)	4.6 (1.0)
2008:01	P	0.8 (0.8)	0.0 (0.9)	-1.0 (0.5)	0.7 (0.8)	0.4 (0.8)	0.1 (1.5)	0.8 (0.8)	1.0 (1.3)	2.2 (0.7)
2009:05	T	0.8 (0.6)	1.0 (0.7)	-0.4 (1.2)	0.8 (0.6)	0.8 (0.7)	0.2 (1.4)	1.3 (0.5)	1.0 (1.1)	-0.2 (2.1)
2012:02	P	-1.2 (1.2)	-1.0 (1.4)	-1.2 (0.6)	-1.3 (1.1)	-1.2 (1.5)	-1.4 (0.5)	-1.8 (1.1)	-1.0 (1.5)	-1.3 (11.0)
2012:12	T	0.1 (1.0)	0.0 (1.2)	-0.6 (0.8)	-0.2 (1.0)	-0.8 (1.3)	-1.3 (0.6)	0.7 (1.3)	0.0 (1.2)	0.1 (1.4)
2015:03	P	-1.3 (1.3)	-3.0 (1.1)	-3.3 (0.3)	-1.5 (1.3)	-3.2 (1.1)	-3.4 (0.3)	-1.7 (1.4)	-3.0 (1.0)	-3.4 (0.9)
2016:08	T	-0.4 (1.0)	-1.0 (1.2)	-1.9 (0.7)	-0.6 (1.0)	-1.8 (1.0)	-2.3 (0.4)	0.4 (1.0)	0.0 (2.0)	6.2 (1.7)
2019:07	P	-0.6 (1.0)	-1.0 (1.4)	6.7 (0.4)	-0.7 (1.0)	-0.8 (1.3)	6.4 (0.4)	-0.8 (0.9)	-2.0 (1.6)	6.6 (1.5)
2020:05	T	-0.1 (0.7)	-1.0 (0.4)	-0.8 (0.1)	-0.1 (0.7)	-1.2 (0.5)	-0.7 (0.1)	0.0 (0.7)	-1.0 (0.6)	-0.8 (1.3)
2022:01	P	-0.6 (1.1)	0.0 (1.3)	1.4 (0.6)	-0.6 (1.1)	1.0 (1.2)	1.6 (0.9)	-0.2 (1.4)	2.0 (1.4)	1.9 (0.6)
Mean		0.2	0	0.09	0.06	-0.3	-0.65	0.08	0.12	0.39
MAE		1.02	1.52	2.76	1.07	1.76	2.85	1.1	1.56	3.1

Table 3. Chronologies and standard errors (in parenthesis) computed using 91 disaggregated series (sector classification) as a lead (positive value) or lag (negative value) of the OECD turning point. The mean and mean absolute error (MAE) in the final two rows summarize the discrepancies of the chronology for the column series, relative to the OECD chronology.

OECD Peak			New Peak	OECD Trough			New Trough
1966:03				1967:07	-2	⇨	1967:05
1969:03	+4	⇨	1969:07	1971:03	-4	⇨	1970:11
1973:05	+7	⇨	**1973:12**	**1975:05**			
1979:08	-1	⇨	**1979:07**	**1982:11**			
1985:09	-4	⇨	1985:05	1987:02	-6	⇨	1986:08
1990:04				1993:08			
1997:10	+5	⇨	1998:03	1999:03	-5	⇨	1998:10
2000:06	+6	⇨	**2000:12**	2003:03	+5	⇨	2003:08
2008:01	-1	⇨	**2007:12**	**2009:05**			
2012:02	-1	⇨	2012:01	2012:12			
2015:03	-3	⇨	2014:12	2016:08	-2	⇨	2016:06
2019:07	+7	⇨	2020:02	2020:05	-1	⇨	**2020:04**
2022:01	+1	⇨	2022:02				
Mean	**+1,8**				**-2.1**		

Table 4. The Adjusted chronology computed using 91 disaggregated series (sector classification). The unadjusted mode estimates are rounded to the closest integer. Dates with more than half of the disaggregated series reporting turning points are shown in bold font. In the bottom row, the mean of the rounded adjustments is presented.

Boosting Results (AUC Scores)

	Only Disaggregated Dataset		All Variables		Only Financial and Global Variables	
	3 Lags					
Test Size	OECD	Adjusted	OECD	Adjusted	OECD	Adjusted
20%	0.89	0.89	0.91	**0.92**	0.83	**0.90**
30%	0.89	**0.90**	0.92	0.92	**0.86**	0.83
40%	0.88	0.88	0.92	0.92	**0.87**	0.84
50%	0.88	0.88	0.91	0.91	**0.87**	0.86
Mean	0.885	**0.888**	0.915	**0.918**	0.858	0.858
	6 Lags					
	OECD	Adjusted	OECD	Adjusted	OECD	Adjusted
20%	0.91	**0.94**	0.93	**0.95**	0.89	**0.90**
30%	0.87	**0.92**	0.90	**0.92**	**0.85**	0.83
40%	0.88	**0.91**	0.90	0.90	0.84	0.84
50%	0.88	**0.89**	0.91	0.91	0.85	**0.86**
Mean	0.885	**0.915**	0.910	**0.920**	0.858	0.858
	12 Lags					
	OECD	Adjusted	OECD	Adjusted	OECD	Adjusted
20%	0.90	**0.92**	0.94	0.94	0.85	0.85
30%	0.87	0.87	0.90	0.90	0.84	**0.86**
40%	0.87	**0.88**	0.89	**0.91**	0.83	**0.86**
50%	0.87	0.87	0.88	**0.90**	0.83	**0.85**
Mean	0.878	**0.885**	0.903	**0.913**	0.838	**0.855**

Time Series Split

Folds	Only Disaggregated Dataset		All Variables		Only Financial and Global Variables	
	3 Lags					
	OECD	Adjusted	OECD	Adjusted	OECD	Adjusted
1975:03 - 1984:08	**0.91**	0.90	**0.92**	0.90	**0.72**	0.71
1984:09 - 1994:02	0.82	**0.85**	**0.80**	0.77	**0.72**	0.69
1994:03 - 2003:08	0.74	**0.86**	0.81	**0.89**	0.92	**0.95**
2003:09 - 2013:02	**0.81**	0.80	**0.90**	0.79	0.82	**0.83**
2013:03 - 2022:08	**0.79**	0.70	**0.90**	0.75	**0.83**	0.81
Mean	0.814	**0.822**	**0.866**	0.820	**0.802**	0.798
	6 Lags					
	OECD	Adjusted	OECD	Adjusted	OECD	Adjusted
1975:03 - 1984:08	0.86	**0.90**	0.84	**0.90**	0.58	**0.66**
1984:09 - 1994:02	0.72	**0.77**	0.73	**0.77**	**0.72**	0.71
1994:03 - 2003:08	0.70	**0.91**	0.79	**0.95**	**0.92**	0.91
2003:09 - 2013:02	**0.74**	0.71	**0.84**	0.77	0.81	0.81
2013:03 - 2022:08	0.66	0.66	**0.87**	0.68	**0.83**	0.73
Mean	0.736	**0.790**	0.814	0.814	**0.772**	0.764

12 Lags						
	OECD	Adjusted	OECD	Adjusted	OECD	Adjusted
1975:08 - 1984:12	0.77	**0.84**	0.79	**0.84**	0.64	**0.69**
1985:01 - 1994:05	0.69	**0.75**	0.73	**0.75**	**0.79**	0.75
1994:06 - 2003:10	0.75	**0.89**	0.83	**0.94**	**0.86**	0.82
2003:11 - 2013:03	**0.67**	0.66	**0.82**	0.75	0.86	**0.87**
2013:08 - 2022:08	**0.69**	0.64	**0.81**	0.67	**0.82**	0.64
Mean	0.714	**0.756**	**0.796**	0.790	**0.794**	0.754

Start of training set: 1993:11

3 Lags						
	OECD	Adjusted	OECD	Adjusted	OECD	Adjusted
2003:11 - 2013:03	0.54	**0.81**	0.79	**0.87**	0.79	**0.89**
2013:04 - 2022:08	**0.65**	0.49	**0.84**	0.72	**0.88**	0.79
Mean	0.595	**0.650**	**0.815**	0.795	0.835	**0.840**

6 Lags						
	OECD	Adjusted	OECD	Adjusted	OECD	Adjusted
2003:11 - 2013:03	0.53	**0.76**	0.70	**0.85**	0.74	**0.88**
2013:04 - 2022:08	**0.65**	0.49	**0.87**	0.76	**0.87**	0.83
Mean	0.590	**0.625**	0.785	**0.805**	0.805	**0.855**

12 Lags						
	OECD	Adjusted	OECD	Adjusted	OECD	Adjusted
2004:05 - 2013:06	0.51	**0.70**	0.77	**0.81**	0.80	**0.84**
2013:07 - 2022:08	**0.66**	0.44	**0.86**	0.73	**0.88**	0.78
Mean	**0.585**	0.570	**0.815**	0.770	**0.840**	0.810

Table 5. Comparison of OECD and Adjusted chronology using AUC scores obtained from the boosting model. The model includes 3, 6, and 12 lags of each predictor. The results are split into three columns: Only disaggregated dataset, all available variables, and only financial and global variables. First section reports the AUC scores obtained with a randomly sampled test set. Second section reports the cross-validation results using the *TimeSeriesSplit* function in Python with 5 folds. The third section trims the dataset to start in 1993:11 and shows AUC scores for the last 2 folds. Means for each subsection of results are presented under the results.

9 Charts

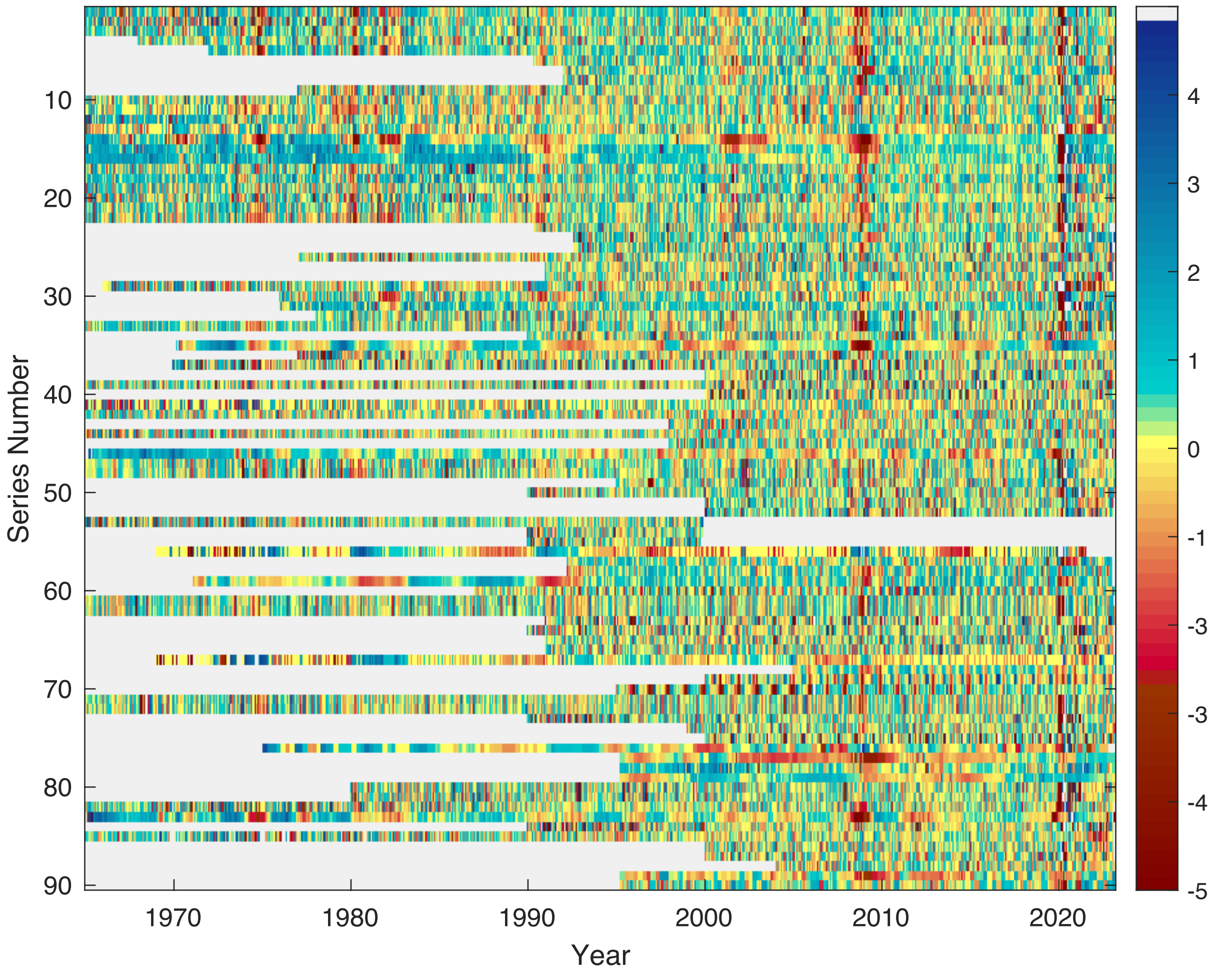


Figure 1. Heat map of monthly growth rates divided by the series standard deviation for the 91 series in the monthly dataset. The vertical axis is the series number as given in Table 1; the horizontal axis is the monthly time scale, 1965:1-2023:4. Negative monthly growth appears as red, positive monthly growth rates appear as blue. The gray sections represent missing data.

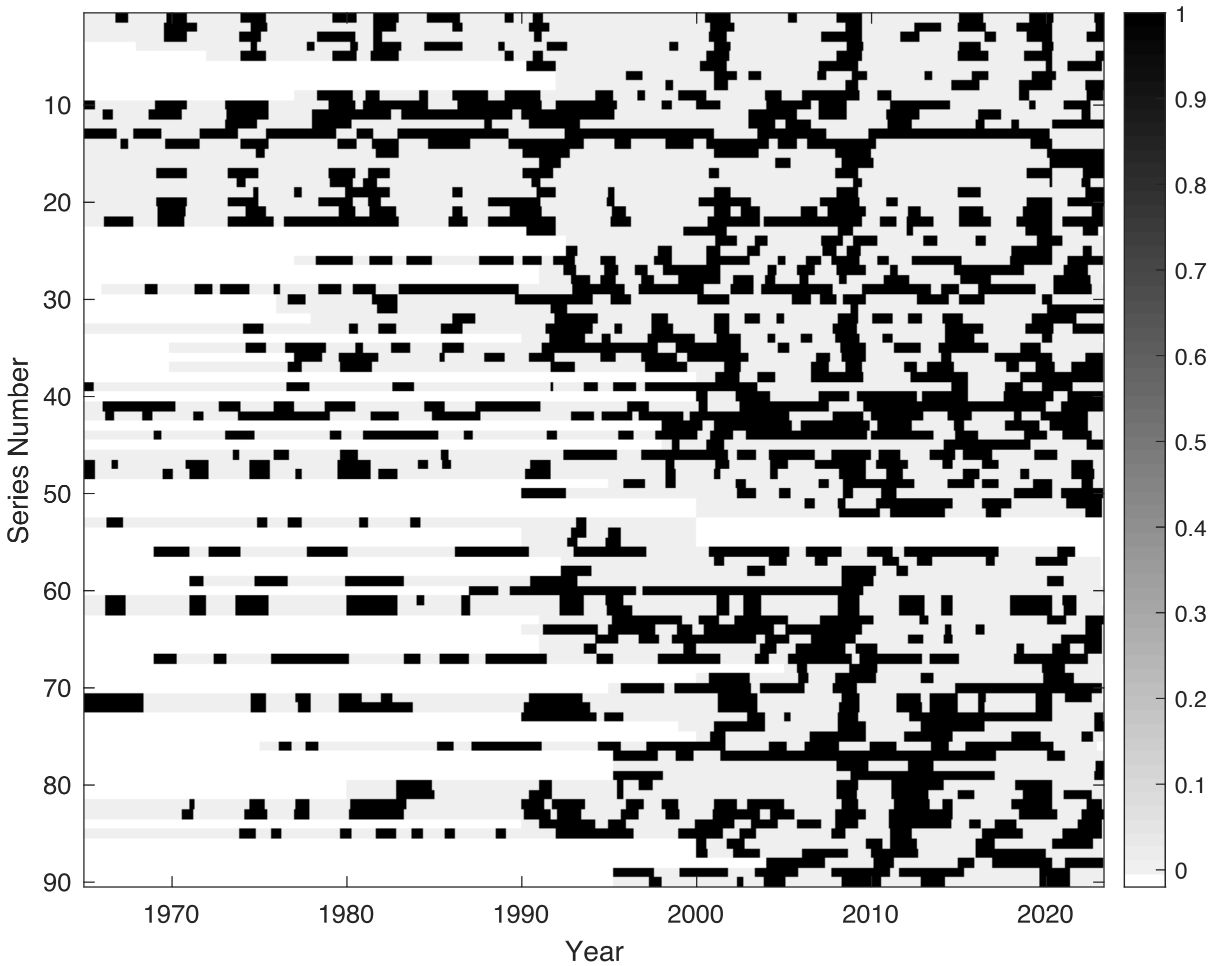


Figure 2. Bry-Boschan recessions computed using the monthly disaggregated dataset. Black denotes Bry-Boschan recessions (from a peak to a trough) and gray denotes Bry-Boschan expansions. White represents missing values.

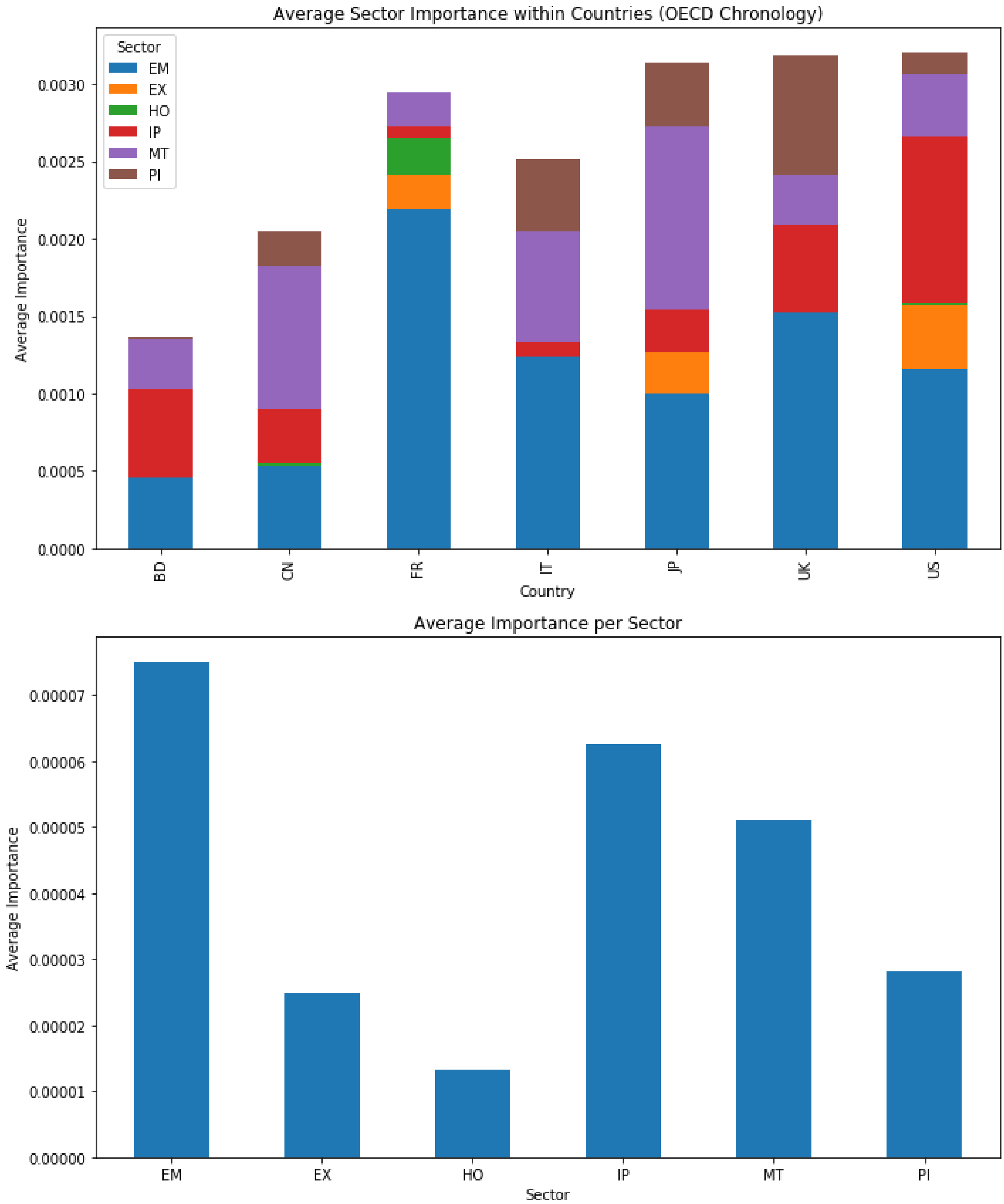


Figure 3. Average importances determined by the boosted prediction of the OECD chronology.

Specifications: 3 lags included, size of test set is 50%

Note: Employment (EM), Personal expenditure (EX), Housing starts (HO), Industrial Production (IP), Manufacturing and trade sales (MT), Personal income (PI).

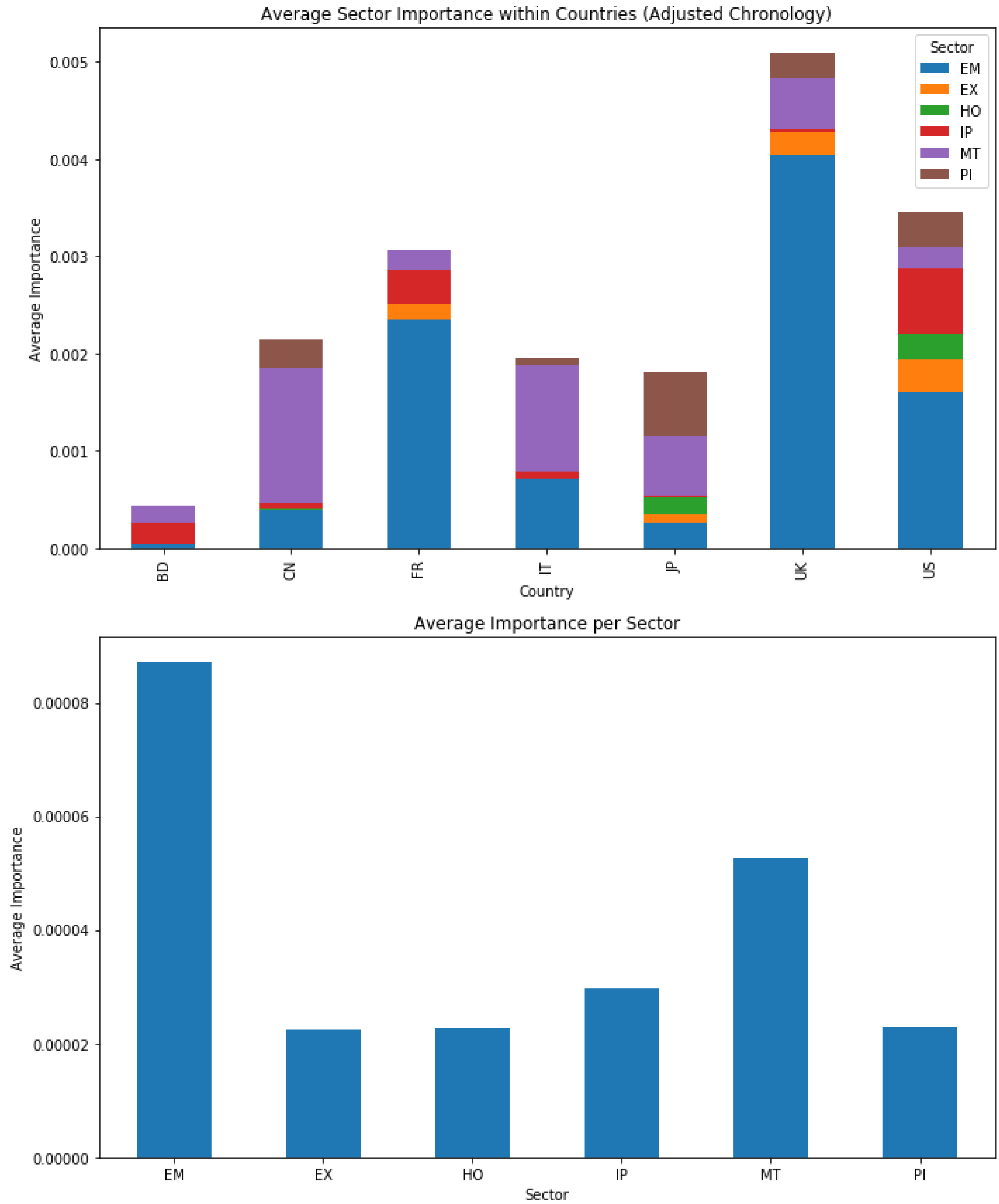
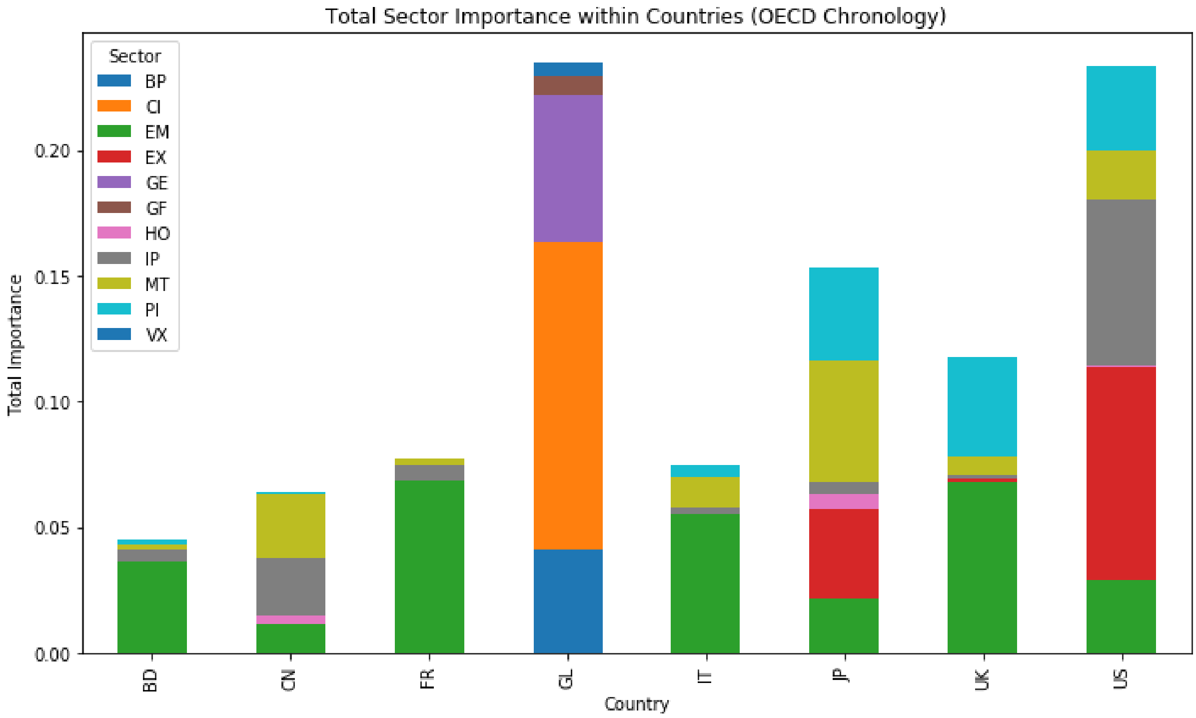


Figure 4. Average importances determined by the boosted prediction of the OECD chronology.

Specifications: 3 lags included, size of test set is 50%

Note: Employment (EM), Personal expenditure (EX), Housing starts (HO), Industrial Production (IP), Manufacturing and trade sales (MT), Personal income (PI).



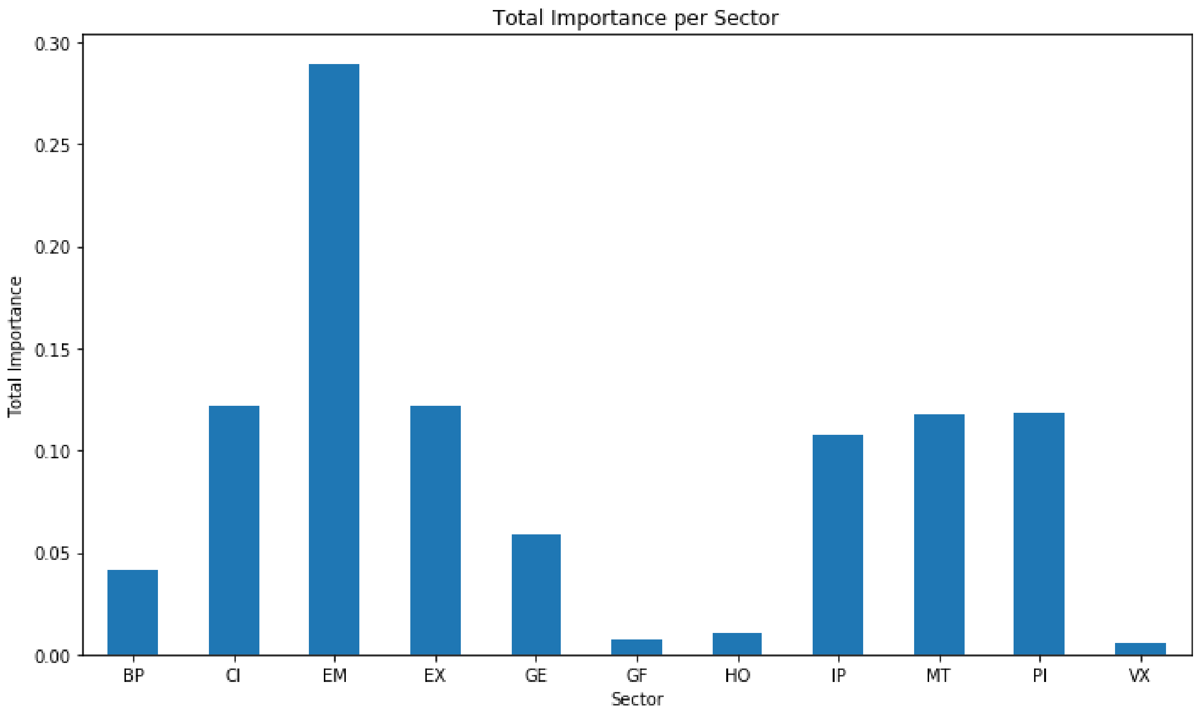
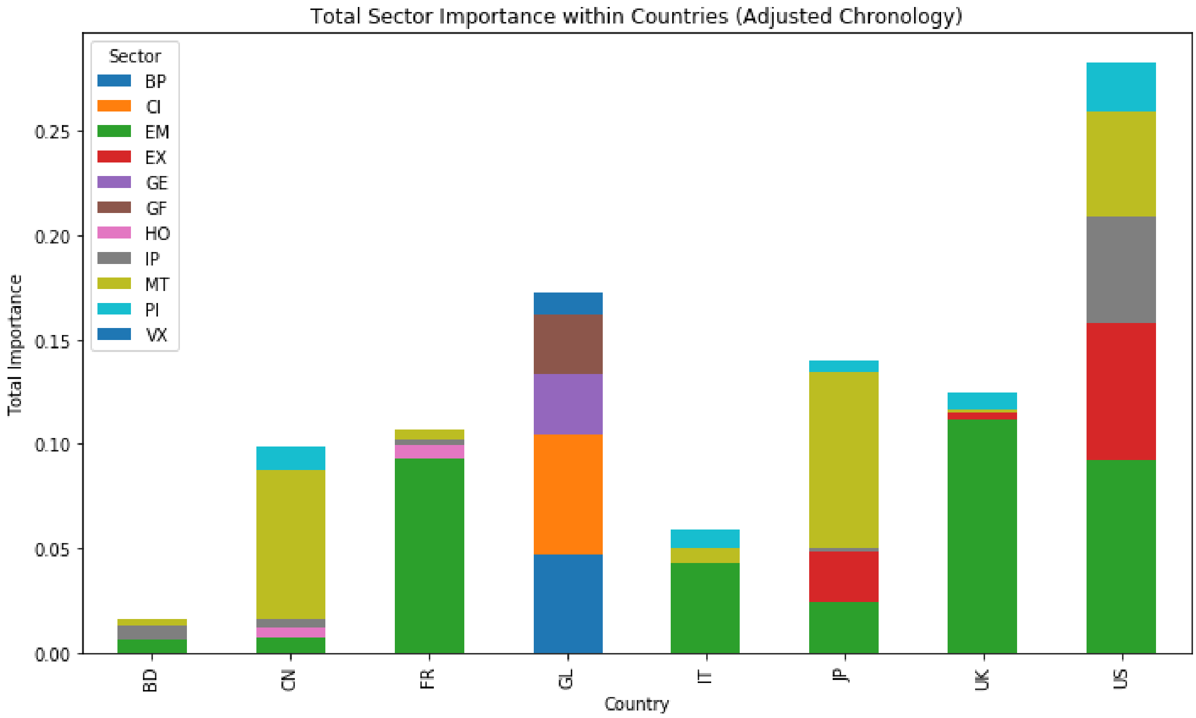


Figure 5. Average importances determined by the boosted prediction of the OECD chronology including financial variables.

Specifications: 3 lags included, size of test set is 50%

Note: Employment (EM), Personal expenditure (EX), Housing starts (HO), Industrial Production (IP), Manufacturing and trade sales (MT), Personal income (PI), Excess Bond Premium (BP), Global Conditions Index (CI), GECON (GE), Global Financial Cycle Factor (GF), Volatility Index (VX).



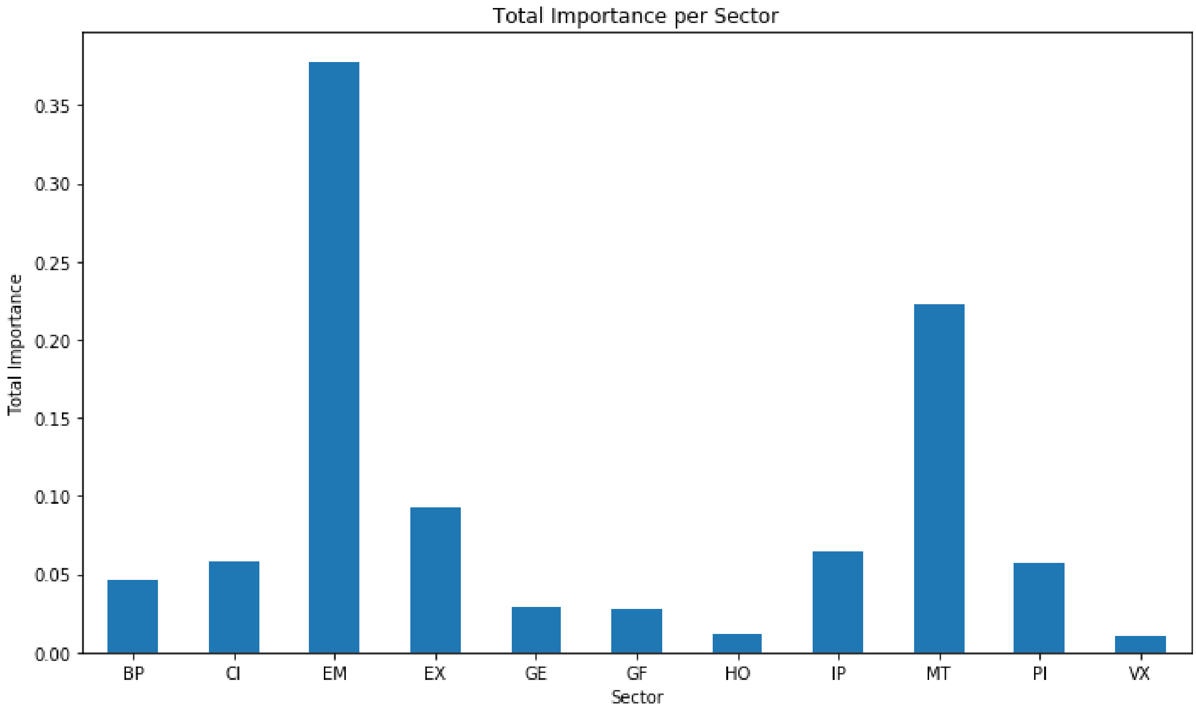


Figure 6. Average importances determined by the boosted prediction of the Adjusted chronology including financial variables.

Specifications: 3 lags included, size of test set is 50%

Note: Employment (EM), Personal expenditure (EX), Housing starts (HO), Industrial Production (IP), Manufacturing and trade sales (MT), Personal income (PI), Excess Bond Premium (BP), Global Conditions Index (CI), GECON (GE), Global Financial Cycle Factor (GF), Volatility Index (VX).

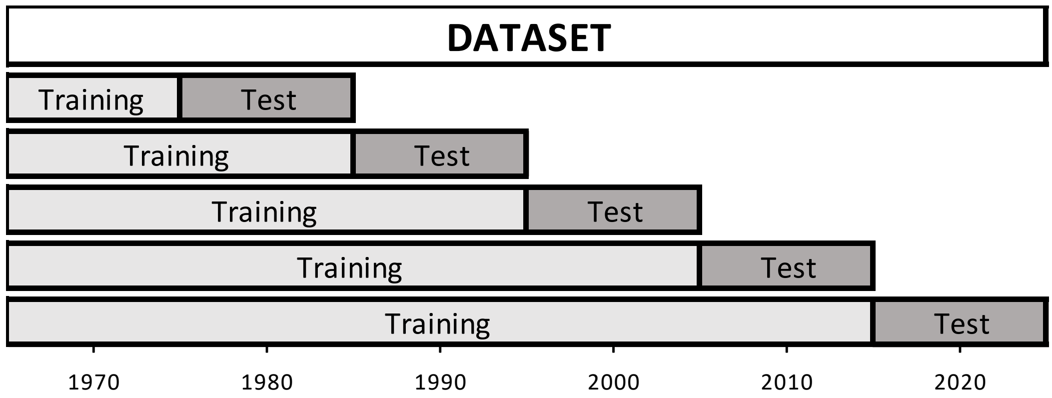


Figure 7. Illustration of the cross-validation scheme of the Boosting model, using the Python function *TimeSeriesSplit*.

10 Appendix

10.1 Baumeister Chronology

Table A.1 presents chronologies obtained using the 91 disaggregated series with the GECON indicator of Baumeister et al. (2020) as a reference chronology. Compared to Table 3, the adjustments made here are generally larger. This is expectable as peaks and troughs were chosen arbitrarily from the standardized GECON. The method of obtaining turning points was very simple, each value was added to the previous one and the local maximums (minimums) were taken as peaks (trough).

The estimates of mean, median and mode generally agree with each other when using different adjustment procedures. In two episodes, specifically the 2016:11 trough (same behavior as in the OECD chronology) and the 2018:08 peak, there is some disagreement when calculating the mode. In the unadjusted and class-lag adjusted chronologies the mode calls for a negative shift by several months in both episodes. On the contrary, the weighted estimation suggests a significant shift forwards, specifically, +9.3 months in the case of the 2018:08 peak. In the country classification the latter episode also behaves very irregularly. This is due to the distribution of the turning points in the two-year sample, being bimodal around 2016:11 and fairly flat around 2018:08. Considering this, a small change of weights can strongly impact the mode estimation in either direction. As the GECON is used for forecasting economic downturns it is expected to lead the actual recession dates. This is indeed the case as the peaks were on average adjusted by +3.4 months and the troughs by +0.7 months, which indicates it signals the starts of economic recessions sooner than the ends.

An important feature of these chronologies is the fact, then they coincide very closely with the ones from Table 3. This is a reassuring sign about the correct workings of the algorithm. A slight change in the reference date, and subsequently the sample from which the moving average is calculated, does not result in a significant change in the suggested chronology.

Baumeister Dates		No Adjustments			Class-Lag Adjusted			Weighted Estimation		
		Mean	**Median**	**Mode**	**Mean**	**Median**	**Mode**	**Mean**	**Median**	**Mode**
1973:05	P	3.8	5.0	6.3	3.5	4.4	5.6	3.0	5.0	5.5
		(1.3)	(1.3)	(0.8)	(1.2)	(1.2)	(0.5)	(1.0)	(1.0)	(1.1)
1975:03	T	0.1	2.0	2.6	-0.3	1.1	1.7	-0.6	1.0	2.6
		(0.8)	(0.8)	(0.5)	(0.8)	(0.8)	(1.1)	(0.8)	(1.0)	(0.6)
1979:03	P	2.1	4.0	3.7	2.2	2.8	2.8	1.5	3.0	3.8
		(1.3)	(1.2)	(0.5)	(1.2)	(1.5)	(4.7)	(1.3)	(1.2)	(3.0)
1982:08	T	0.8	3.0	3.1	0.9	2.2	2.5	0.9	3.0	3.1
		(1.1)	(0.8)	(0.3)	(1.0)	(0.8)	(0.4)	(1.0)	(0.7)	(0.7)
1990:01	P	1.8	2.0	1.9	2.0	2.3	2.0	1.9	2.0	1.9
		(1.3)	(1.4)	(1.1)	(1.2)	(1.5)	(2.1)	(1.4)	(1.6)	(0.8)
1991:01	T	2.6	2.0	1.1	2.8	3.0	1.6	2.2	2.0	1.3
		(1.2)	(1.1)	(0.5)	(1.3)	(1.3)	(1.2)	(1.4)	(1.0)	(0.9)
2000:08	P	1.5	3.0	4.0	1.2	2.7	3.4	1.0	2.0	3.8
		(0.8)	(0.8)	(0.3)	(0.7)	(0.7)	(0.2)	(0.8)	(1.1)	(2.2)
2003:04	T	1.5	3.0	3.6	1.4	2.7	3.0	1.5	2.0	3.4
		(1.0)	(1.0)	(0.9)	(1.0)	(0.9)	(1.4)	(1.1)	(1.2)	(0.9)
2007:07	P	2.0	4.0	4.9	1.8	4.2	5.8	1.4	4.0	4.5
		(0.9)	(0.9)	(0.5)	(0.9)	(0.9)	(0.4)	(1.0)	(1.2)	(1.7)
2009:07	T	-0.5	-1.0	-2.4	-0.5	-1.5	-2.0	-0.2	-1.0	-2.1
		(0.6)	(0.7)	(1.2)	(0.6)	(0.7)	(1.0)	(0.6)	(1.0)	(0.9)
2011:02	P	1.1	2.0	10.3	1.2	1.2	2.3	1.1	2.0	10.2
		(1.2)	(1.8)	(0.4)	(1.2)	(1.6)	(3.3)	(1.3)	(1.8)	(0.5)
2012:08	T	2.3	3.0	3.3	1.9	2.2	2.5	2.4	4.0	3.7
		(1.1)	(1.1)	(0.8)	(1.1)	(1.3)	(7.3)	(1.2)	(1.1)	(1.8)
2014:12	P	-2.7	-1.0	-0.3	-2.9	-1.9	-0.8	-3.1	-2.0	-0.4
		(1.2)	(1.2)	(0.3)	(1.2)	(1.3)	(0.3)	(1.2)	(1.5)	(13.0)
2016:11	T	-3.1	-4.0	-4.9	-3.5	-4.9	-5.6	-2.1	-3.0	3.2
		(0.9)	(1.1)	(0.7)	(0.9)	(1.1)	(1.3)	(0.9)	(1.9)	(0.7)
2018:08	P	0.6	0.0	-2.6	0.5	0.4	-3.6	0.9	1.0	9.3
		(1.0)	(1.6)	(2.9)	(0.9)	(1.6)	(0.4)	(1.0)	(2.1)	(0.3)
2020:05	T	-0.1	-1.0	-0.8	-0.2	-1.5	-1.1	0.0	-1.0	-0.8
		(0.7)	(0.4)	(0.1)	(0.7)	(0.5)	(0.1)	(0.7)	(0.6)	(0.9)
2021:12	P	-0.2	1.0	2.2	-0.2	0.9	1.8	0.4	3.0	2.9
		(1.1)	(1.2)	(0.6)	(1.1)	(1.2)	(0.3)	(1.4)	(1.4)	(0.6)
Mean		0.2	0	0.09	0.06	-0.3	-0.65	0.08	0.12	0.39
MAE		1.02	1.52	2.76	1.07	1.76	2.85	1.1	1.56	3.1

Table A.1. Chronologies and standard errors (in parenthesis) computed using 91 disaggregated series (sector classification) as a lead (positive value) or lag (negative value) of the Baumeister turning point. The mean and mean absolute error (MAE) in the final two rows summarize the discrepancies of the chronology for the column series, relative to the Baumeister chronology.

10.2 Kose Chronology

The chronology of Kose et al. (2020) is different from the rest as it is built on a quarterly frequency. The first month of the quarter was taken as the reference turning point, which results to a slightly imprecise reference chronology. However, this should not be a problem for the algorithm or the following inference as the 12-month moving average covers a much broader time period and is able to identify the actual turning point. Instances of perfect match between the reference and the suggested turning point are very scarce as a consequence of the adjustment from quarters to months. The chronologies built on the Kose reference cycle are reported in Table A.2.

Another feature of the reference chronology created by Kose is the fact, that the average duration of a recession is less than one year. Therefore, there are several instances where a mismatch of reference date with the OECD chronology happens and the moving averages around the date are created using a very different, if not a completely separate, sample.

Kose et al. (2020) distinguish global recessions and downturns. They classify the less important recession episodes as global downturns where the global economy registered a very low growth rate, but the implications were not as severe and broad-based to be recognized as a global recession. In the final reference cycle, the focus is set on significant global recessions where many of the disaggregated series report turning points. This leads to more accurate estimates. And indeed, the dates agree across the different adjustment schemes in most cases.

One case, in particular, displays large inconsistencies in the mode using no adjustments. It states the 1974:01 episode to have occurred in 1974:07. By analyzing the individual series and their turning points it is clear that most series tracking industrial production experienced a downturn with a 5-to-6-month lag. On the other hand, a majority of turning points in the manufacturing and employment class were close to, or slightly before, the reference date. This resulted in a more conservative estimate, matching the one found using the OECD chronology.

Kose Dates		No Adjustments			Class-Lag Adjusted			Weighted Estimation		
		Mean	**Median**	**Mode**	**Mean**	**Median**	**Mode**	**Mean**	**Median**	**Mode**
1974:01	P	0.5 (1.1)	0.0 (1.4)	6.2 (0.6)	0.2 (1.1)	-0.4 (1.5)	-1.3 (0.7)	0.7 (1.2)	0.0 (1.1)	-1.4 (0.4)
1975:01	T	2.1 (0.8)	4.0 (0.8)	4.6 (0.5)	1.8 (0.8)	3.4 (0.8)	4.1 (0.4)	1.4 (0.8)	3.0 (0.8)	4.6 (1.7)
1981:10	P	-1.1 (1.2)	-3.0 (1.0)	-3.2 (2.4)	-1.1 (1.2)	-3.6 (0.9)	-4.2 (0.7)	-1.8 (1.4)	-3.0 (0.6)	-3.1 (1.7)
1982:10	T	0.1 (1.0)	1.0 (0.8)	1.0 (0.3)	0.1 (0.9)	0.6 (0.8)	0.8 (0.6)	0.1 (0.9)	1.0 (0.6)	1.1 (10.4)
1990:10	P	-1.1 (1.1)	-2.0 (1.8)	-7.1 (1.1)	-1.2 (1.1)	-2.1 (1.6)	-7.0 (1.3)	-0.7 (1.1)	-2.0 (1.6)	-7.1 (3.8)
1991:01	T	2.6 (1.2)	2.0 (1.1)	1.1 (0.5)	2.8 (1.2)	2.9 (1.2)	1.7 (0.6)	2.2 (1.4)	2.0 (0.8)	1.3 (1.5)
2008:07	P	-2.5 (0.8)	-3.0 (1.0)	-7.0 (0.5)	-2.7 (0.7)	-3.6 (1.0)	-6.2 (2.2)	-2.7 (0.8)	-3.0 (0.9)	-4.1 (1.0)
2009:01	T	3.0 (0.6)	3.0 (0.7)	3.6 (1.2)	3.0 (0.6)	3.5 (0.6)	3.8 (0.9)	3.1 (0.6)	3.0 (0.8)	3.9 (1.1)
Mean		0.45	0.25	-0.11	0.37	0.11	-1.04	0.3	0.12	-0.59
MAE		1.62	2.25	4.24	1.61	2.5	3.63	1.59	2.12	3.32

Table A.2. Chronologies and standard errors (in parenthesis) computed using 91 disaggregated series (sector classification) as a lead (positive value) or lag (negative value) of the Kose turning point. The mean and mean absolute error (MAE) in the final two rows summarize the discrepancies of the chronology for the column series, relative to the Kose chronology.

10.3 Fushing Chronology

Table A.3 present chronologies obtained using the global reference chronology by Fushing et al. (2010). They exhibit similar patterns to the results found through the Kose chronology, but adjustments tend to be even greater in this case as can be seen with the high mean and mean squared error reported at the bottom of Table A.3. These larger shifts serve as a reassuring metric as they mostly move the turning point to the same date found with the OECD chronology.

The measures of central tendency generally agree across all adjustment procedures, with the exception of the 1973:11 peak, which is analogous to the phenomenon explained in the previous section. The reference chronology dates the peak and trough of the 1980:4 episode in the same month, meaning that their model surpassed the recession threshold for only a brief moment in time. The mode from the disaggregated dataset dates of the peak nine months earlier and the trough 2 months later, suggesting a 11-month log recession.

Fushing Dates		No Adjustments			Class-Lag Adjusted			Weighted Estimation		
		Mean	**Median**	**Mode**	**Mean**	**Median**	**Mode**	**Mean**	**Median**	**Mode**
1973:11	P	2.5 (1.1)	2.0 (1.4)	8.2 (0.6)	2.1 (1.1)	1.5 (1.3)	0.8 (0.4)	2.7 (1.2)	2.0 (1.0)	0.6 (1.3)
1975:01	T	2.1 (0.8)	4.0 (0.8)	4.6 (0.5)	1.6 (0.8)	2.6 (0.9)	3.9 (0.3)	1.4 (0.8)	3.0 (0.8)	4.6 (0.8)
1980:04	P	-2.3 (1.3)	-4.0 (2.0)	-9.2 (0.3)	-2.5 (1.4)	-4.4 (1.5)	-9.5 (0.5)	-2.7 (1.3)	-4.0 (1.5)	-9.1 (1.2)
1980:04	T	0.5 (1.0)	2.0 (0.8)	2.3 (0.3)	0.4 (0.9)	1.5 (0.7)	2.0 (0.3)	0.5 (0.8)	2.0 (0.5)	2.1 (1.7)
2000:04	P	4.6 (0.8)	7.0 (0.7)	8.0 (0.4)	4.2 (0.8)	6.2 (0.7)	7.3 (0.3)	3.8 (0.9)	6.0 (0.8)	7.5 (0.4)
2001:05	T	2.9 (0.9)	5.0 (0.9)	6.1 (0.4)	2.6 (0.8)	4.2 (0.9)	5.2 (0.3)	2.0 (0.8)	4.0 (1.0)	6.1 (2.2)
2008:02	P	0.9 (0.7)	0.0 (0.9)	-2.0 (0.5)	0.8 (0.8)	-0.8 (0.9)	-2.2 (0.3)	0.6 (0.8)	0.0 (0.9)	1.0 (1.1)
2009:04	T	1.3 (0.6)	2.0 (0.7)	0.6 (1.2)	1.2 (0.6)	0.6 (0.7)	-0.1 (0.6)	1.7 (0.6)	2.0 (0.8)	0.8 (1.1)
Mean		1.55	2.25	2.32	1.31	1.43	0.91	1.25	1.88	1.71
MAE		2.14	3.25	5.12	1.94	2.73	3.88	1.92	2.88	3.99

Table A.3. Chronologies and standard errors (in parenthesis) computed using 91 disaggregated series (sector classification) as a lead (positive value) or lag (negative value) of the Fushing turning point. The mean and mean absolute error (MAE) in the final two rows summarize the discrepancies of the chronology for the column series, relative to the Fushing chronology.

10.4 Contested Episodes

By extending the moving average, from which the distribution of turning points is created, the different chronologies can be compared even in episodes where the turning point dates are more than two years apart. The methodology was created for a 12-month moving average. It was extended to 24 months to capture the trends in the data, but it should serve as an illustration of the problem rather than a robust inference measure. Moreover, the Bry-Boschan procedure identifies cycles of minimal length 15 months. This leads to a possible shift of the reported turning point of an individual series if it experienced two short cycles. The initial value was chosen to fit all potential turning points into the sample. The gray columns represent actual turning points in disaggregated series and the orange line is a kernel density estimate of the turning points. The blue column indicates the adjusted turning point obtained using the OECD chronology. The red column indicates the adjusted turning point obtained from an alternative chronology.

1973:05 Peak. U.S., Canada, Japan, and Germany all report the turning point occurred sometime during the autumn of 1973. The other countries report the mean turning points several months into 1974. The distribution and estimates are shown in Figure A.1. As Kose et al. (2020) reiterates, the Arab oil embargo initiated in October 1973 played a major role in this economic downturn. Although the embargo ended in March 1974, the supply shock and associated sharp rise in oil prices lead to a substantial increase in inflation and a significant weakening of growth in the G7 countries apart from Germany and Japan. It is not clear if this is the case using the disaggregated dataset as there are not many series available for the two mentioned countries at the time. In terms of the sectors, personal expenditure leads all of the other sectors on average and industrial production reports many instances of turning points occurring well into 1974.

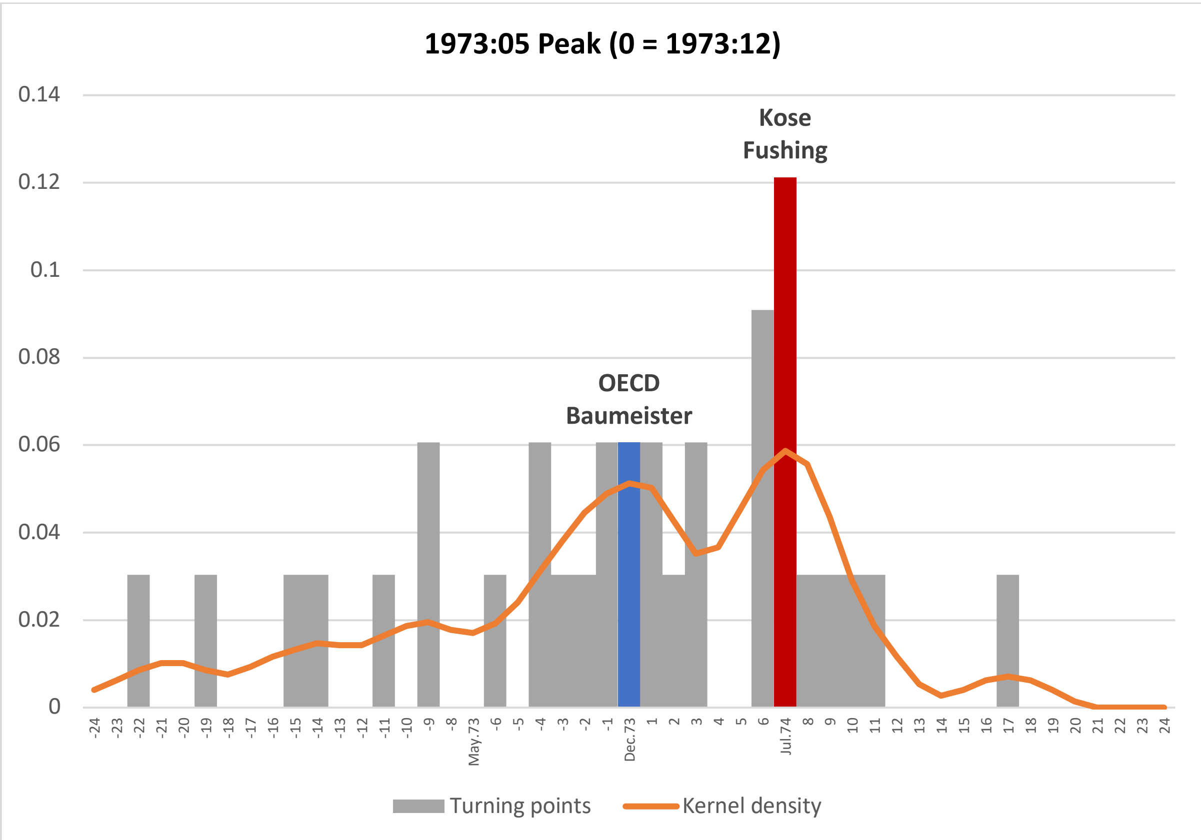


Figure A.1. Histogram and kernel estimate of unadjusted turning points of disaggregated series. The title represents the episodes in terms of the OECD date and the reference date for the dating algorithm is in parentheses.

1979:08 Peak. The CEPR dates the 1980s recession to have started in 1980:Q1 and ended in 1982:Q3, whereas the NBER identifies two short-lived recessions within that period. Specifically, the start of the second recession is 1981:07, the same as identified with the use of Kose chronology visible in Figure A.2. Again, the downturn in 1979 was driven by an oil shock, which led to monetary policy tightening in most of the G7 countries. This caused to declines of economic activity and increase in unemployment (Kose et al., 2020). Chauvet and Yu (2006) calculated recession probabilities for the G7 countries and found that U.S., Canada, and Italy had two recessions during this period while the other countries the other countries experienced a single, longer lasting recession.

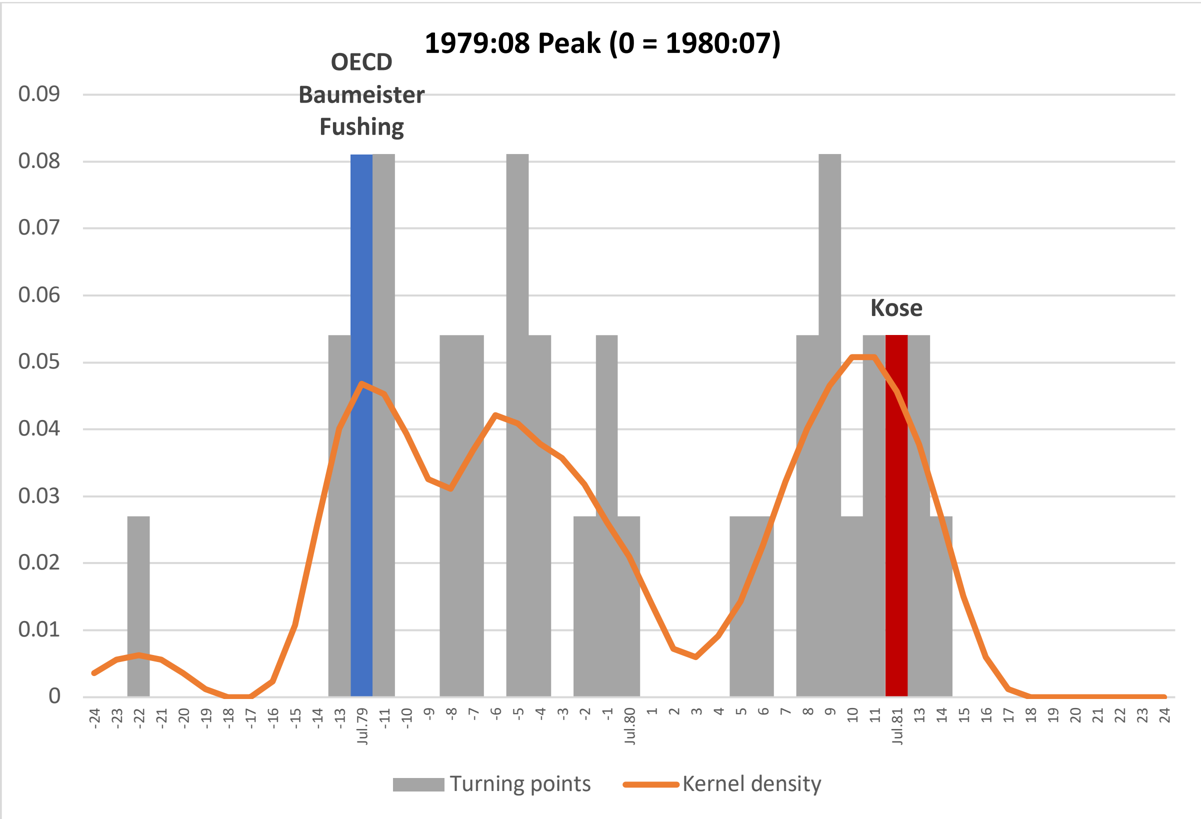


Figure A.2. Histogram and kernel estimate of unadjusted turning points of disaggregated series. The title represents the episodes in terms of the OECD date and the reference date for the dating algorithm is in parentheses.

1993:08 Trough. NBER dates the trough to be 1991:03, whereas CERP identifies 1993:Q3 as the trough. This is observable in the dates presented in Figure A.3. There is a large cluster of turning points in 1991, which would suggest a trough, however, this was mainly driven by the early recovery of the U.S. while other countries remained in a recession for a longer period.

The CEPR dates the start of the recession as 1992:Q1, meaning the U.S. and the European have experienced downturns in different years. Germany, France, and Italy follow the chronology of the CEPR and reflect the problems with the exchange rate mechanism of the European Monetary System in 1992.

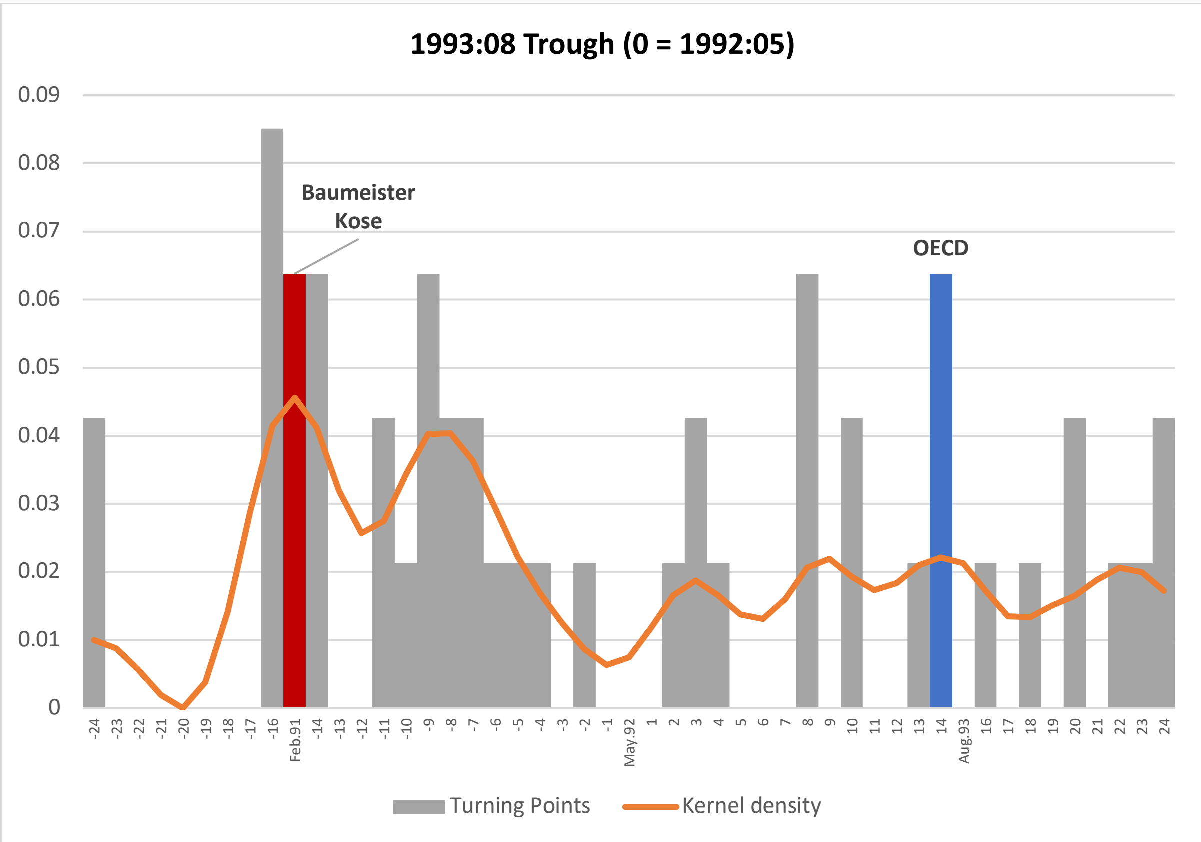


Figure A.3. Histogram and kernel estimate of unadjusted turning points of disaggregated series. The title represents the episodes in terms of the OECD date and the reference date for the dating algorithm is in parentheses.

2003:03 Through. As discussed in Section 5.1.4, the earlier date is analogous to the NBER announcement and relies on the U.S. chronology. However, many more countries have been affected. The Japanese Committee for Business Cycle Indicators of the Economic and Social Research Institute (ESRI) dates the trough to be 2002:1, indicating a shift to prior years compared to the OECD. The OECD dates the U.S. trough to be 2003:02 indicating the chronology does not align with the official cycle. The OECD trough for Europe is dated to 2003:06.

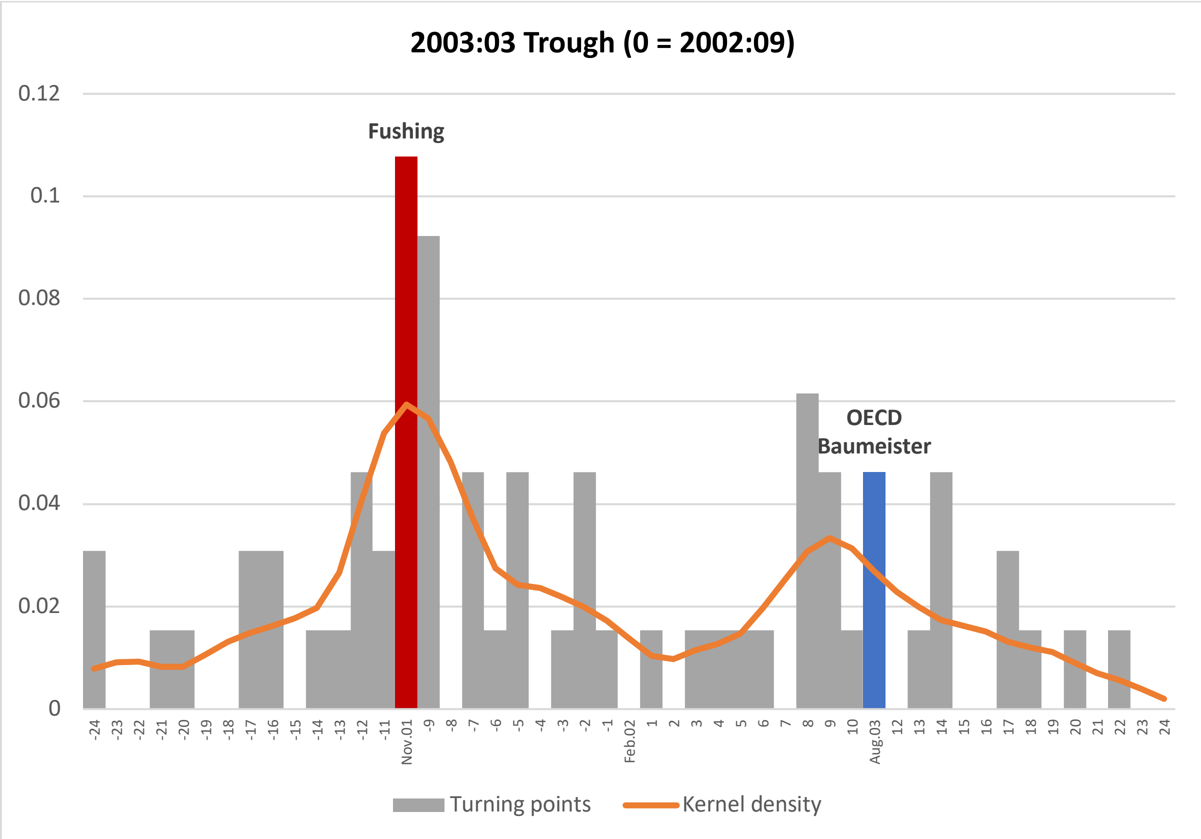


Figure A.4. Histogram and kernel estimate of unadjusted turning points of disaggregated series. The title represents the episodes in terms of the OECD date and the reference date for the dating algorithm is in parentheses.

2019:07 Peak. The OECD dates the European peak in 2018:01, the U.S. peak 2020:01, the Japan peak in 2019:05, and the UK and Canada peak in 2019:08. The NBER peak is dated in 2020:02, the CEPR peak in 2019:Q4, and the Japanese ESRI peak in 2018:Q4. There are many different inputs, but the official dating committees of Europe and U.S. more towards the end of year 2019 and beginning of 2020.

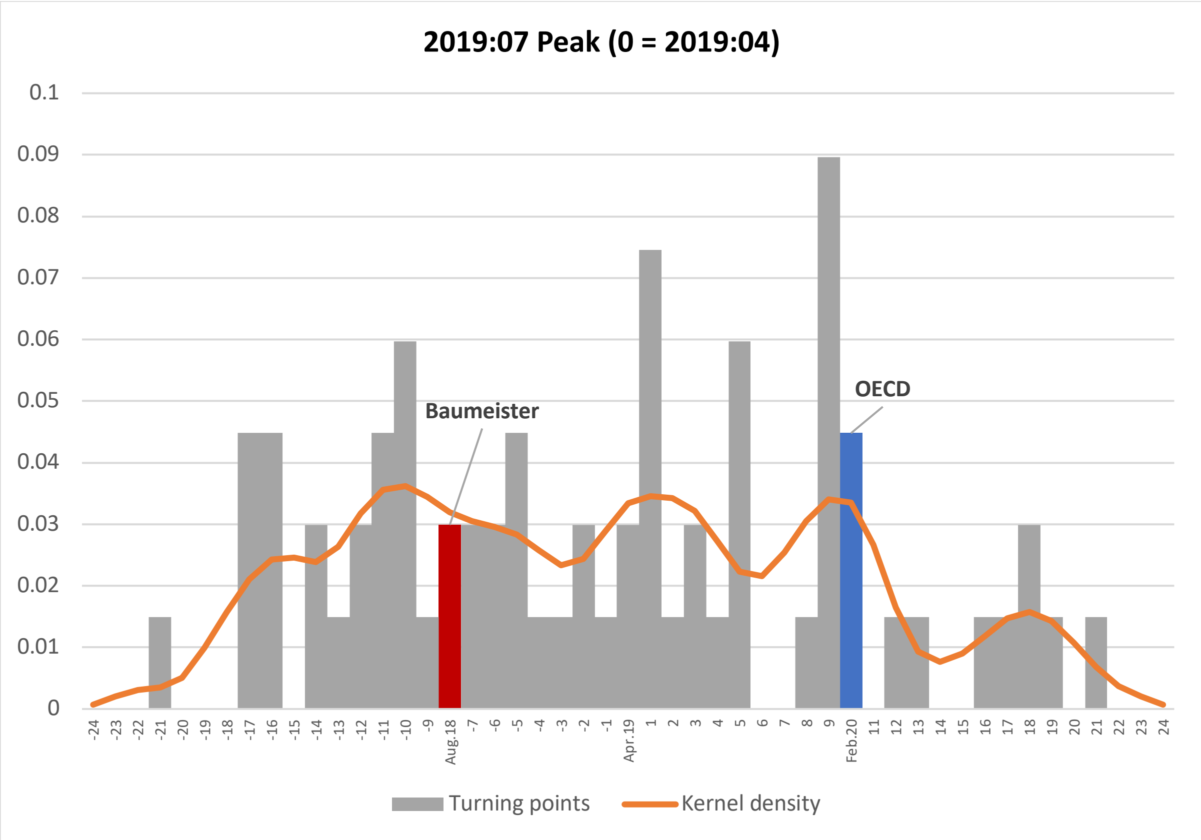


Figure A.5. Histogram and kernel estimate of unadjusted turning points of disaggregated series. The title represents the episodes in terms of the OECD date and the reference date for the dating algorithm is in parentheses.