

# Identifying Unsustainable Credit Gap

Nam Nguyen

July 28, 2022

## Abstract

To overcome model uncertainty in using credit gap as an early warning indicator (EWI) of systemic financial crises in a binary outcome setting, we propose using model averaging of different credit gap measurements to achieve a better averaged model fit and out-of-sample prediction. The methodology we use is logistic binary regression in a panel setup consisting of 40 countries. In this paper, we also propose a novel, superior criteria to judge the performance of an EWI than the one currently popularly used in the literature. Additionally, the empirical results showed that our Bayesian averaged model could synthesize a single credit gap that outperforms any other popularly studied credit gap measurements in terms of an early warning indicator.

## 1 Introduction

Credit-to-GDP gaps have been identified as the top candidate as an early warning indicator (EWI) for future financial crises. As the importance of the financial market to the real economy has increased, failing to predict a future systemic crisis and missing the correct timing to implement macroprudential instruments such as Countercyclical Capital Buffer (CCyB) can lead to significant economic losses. As a result, there is a need to improve the performance of our early warning indicator model, specifically when using the credit gap in univariate EWI models.

In the EWI literature, The Bank of International Settlement (BIS) Basel credit gap using a real-time (one-sided) version of the Hodrick-Prescott filter (HP) has been established as a reference point for policymakers. The BIS Basel gap is easy to implement and communicate; traditionally, it has performed well as an early indicator. However, recent literature has found evidence that optimizing parameters of various decomposition filters can improve those filters' performance beyond the predictive power of the BIS Basel gap as an EWI. The details are discussed in Drehmann & Yetman (2021) and Beltran, Jahan-Parvar, & Paine (2021).

The development of embedding optimization of decomposition filter into the process of building an EWI framework has significantly improved the performance of the credit gap filters. However, this also created another issue which is model uncertainty. Policymakers now face at least five decomposition methods, each with different feature parameters to consider optimizing for, such as smoothing parameters, robustness, and rolling window. Furthermore, changes in the sample window would change the performance of those filters with optimized parameters. Selecting a robust and best-performing decomposition filter and optimizing its feature parameters can be a paradox of choice for policymakers.

There is a need for a robust methodology to create a univariate EWI that adapts well to changes in dynamics of the underlying financial market but at the same time uses only the credit gap as an explanatory variable. This paper attempts to solve the model uncertainty issue by using Bayesian model selection and average methods. We first create a set of possible decomposition filters with rich features from the recent literature findings. Out of all possible initial credit gaps created, We then implement a series of model selection and discarding using an additional novel metric psAUC (partial standardized area under the curve of receiver operating characteristic), discussed in subsection 4.4.3. Subsequently, from a finalized subset of variables that passed the performance filters, we perform model averaging of all possible combinations of the variables. Raftery (1995) proposed the Bayesian model average method that averages its estimated parameters of interest across all possible combinations of variables. The model that uses the average estimate has proven to be more robust and performs better in out-of-sample forecasts than any other individual model in

the selected subset of model combinations.

Using data from 1970:Q4 to 2017:Q4 across 43 countries, we created 90 credit gaps with varying features found in the literature. We implemented the model selection and average methods described in the paragraph above. We then found evidence that the new EWI performance metric partially standardized AUC (psAUC), which constraints analysis to the region where Type II error  $< 1/3$ , to be a robust single value criterion. We also successfully incorporated the new metric into our model selection and averaging method. Lastly, to ease policy implications, we propose to create a single credit gap measurement from weighted-averaging other popularly studied credit gap measurements. The single weighted average credit gap is tested as more robust in resampling bootstrapping estimates. It performs better in out-of-sample forecasts k-fold cross-validation than any other individual model in the filtered subset of model combinations.

The paper is structured as follows. In the next section, we discuss relevant branches of literature. In Section 3, we describe the data used in this paper. Section 4 outlines the methodology I use for model selection, model average, and a weighted combination of variables of interest, as well as a novel metric for testing the performance of the selected variable and averaged model. Section 5 describes the empirical results of the selected variables in the horse race. Section 6 analyzes time series graphs of individual countries. Section 7 summarizes the main conclusions.

## 2 Literature Review

In this section, we discuss the branches of literature that our study is relevant and contributes to.

C. Borio & Lowe (2002) first documented credit gaps' property as a useful early warning indicator (EWI) for banking crises. Earlier literature on excessive credit leading to crises includes Minsky (1977), Manias (1978), and Kaminsky & Reinhart (1999).

C. E. V. Borio & Drehmann (2009) assessed different smoothing parameters for the one-sided HP filter and later determined  $\lambda = 400,000$  to be the ideal value for measuring credit gap based on its long smooth trend property and performance as an early warning indicator. Banking Supervision

(BCBS) (2010) later established this particular HP filter as the Basel credit gap (we call it BIS Basel gap in this paper), and it became the reference for all EWI models using credit gap filters. Drehmann & Tsatsaronis (2014) defended the Basel gap’s performance as an EWI against criticism it received, such as its failure to perform well in emerging market economies.

Drehmann & Juselius (2014) set out a list of criteria for an EWI. It needs to have the “right timing” of being able to detect future financial crises from 20-6 quarters ahead of a future systemic crisis; a “stable signal” that does not decrease as the periods approach the actual crises date; “robustness” that fits well with different data set; and “interpretability” which is ideal for a univariate model such as the credit gap to determine an optimized threshold of classifying excessive credit. Lastly, EWI  $S_i$  outperforms EWI  $S_j$  for horizon  $h$  if  $AUC(S_{i,h}) > AUC(S_{j,h})$ . This last criterion, however, receives criticism regarding the lower left region of the ROC curve that represents low predictive power of a model, which the AUC metric also measures.<sup>1</sup>

C. Borio (2014) established that credit cycles have a much lower frequency than business cycles. It also pointed out that credit cycle peaks are closely associated with the onset of financial crises. Because of that property, they can detect crises in real-time within macroprudential policies implementation window.

Hamilton (2018) argues against using HP filter for macroeconomics series as it introduces spurious dynamics. Drehmann & Yetman (2018) defended the HP filter again but will later admit other filters with additional features, such as a Hamilton filter estimated in a panel setting, outperformed the Basel gap in Drehmann & Yetman (2021).

Additionally, features for the decomposition methods include rolling sample windows of 15 and 20 years as proposed by Galán (2019) to help improve the performance of the decomposition filter. Beltran et al. (2021) embedded the optimization of features parameters into the construction of an EWI model and constraint the search range for the loss function to the region where True Positive rates  $\geq 2/3$  or Type II error rates  $< 2/3$  following suggestions in C. E. V. Borio & Drehmann (2009), the paper also introduced using Structural Time Series model (STM) gap, Moving average (MA)

---

<sup>1</sup>the details are discussed in subsection 4.4.3

gap as possible candidates for the EWI model.

Galán (2019) also proposed adding other features to the Basel gap that individual countries implement that are out of the scope of this paper's implementation. Such adjustments include adjusting for GDP decrease, combining one and two-sided filters for robustness, adjusting by projection as in Gerdrup, Kvinlog, & Schaanning (2013), and adjusting for currency fluctuations.

However, there is no consensus on which decomposition with which features would be the best performing and most robust EWI. In this paper, we aim to construct a methodology to find the best combination of decomposition and features to be used as an EWI and propose an average model that performs as well as the best combination, if not better. The methodology we use for the model selection, and average steps are discussed in the literature branch below.

Babecký et al. (2014) used a model selection method called Markov Chain Monte Carlo Model Comparison ( $MC^3$ ) developed by Madigan, York, & Allard (1995). This method allows for a feasible search for best-fitted combinations out of a large set of variables with similar features, which would not be correctly estimated using other model selection methods such as RIDGE, LASSO, and Elastic Net. Furnival & Wilson (2000) proposed a "leap and bound" algorithm to find the best-fitted subsets without examining all possible subsets of variables.

The Bayesian model averaging method is formally described in Raftery (1995). It aims to overcome model uncertainty by averaging the performance of a set of models. The averaged model performs better than the individual models on out-of-sample forecasts. Recently, Holopainen & Sarlin (2017) implemented ensemble methods to combine various machine learning model predictions and found significant improvement in the model performance.

The model average method could be applied to a multivariate setting, but that will cost us the ability to synthesize a combined weighted credit gap. Following the motivation from Chapter 3, we also attempt to create a combined weighted credit gap in this chapter that performs well in a horse race with other credit gap measurements.

Multivariate early warning system (EWS) is also a significant branch of early warning literature and is worth mentioning. The variables selected in Babecký et al. (2014) that perform well in EWS

are then used in the following papers. Drehmann & Juselius (2014) established that with the credit gap, the Debt-to-Service ratio is another EWI that performs well within 1-4 quarters of the future systemic financial crisis. Aldasoro, Borio, & Drehmann (2018) is an extension of Drehmann & Juselius (2014). Furthermore, Alessi & Detken (2018) used a Random Forest model to achieve a higher EWI performance metric than traditional generalized linear regression. Even though multivariate models can achieve higher model fit and prediction accuracy than univariate models, they also require richer data availability and limit the scope of the model implementation. Certain Emerging market economies with insufficient data would benefit more from a well-performing univariate model.

Lastly, we touch on the literature on partial estimating the Area Under the Curve of Receiver Operating Characteristic (psAUC). Detken et al. (2014) first introduced the use of partial standardized AUC in the EWI setting. The partial standardized AUC is popularly studied in the medical statistics literature, e.g., McClish (1989), but is relatively new in the EWI literature. We will use this novel (psAUC) metric, inspired by both Detken et al. (2014) and the policy function restriction in Beltran et al. (2021), to improve our model selection and averaging process in terms of satisfying policy implication requirements (e.g.  $TPR \geq 2/3$ ). The method for partial estimation of the ROC curve is documented in Robin et al. (2011).

### 3 Data Description

Our sample periods include quarterly data from 1970:Q4 to 2017:Q4 across 43 countries.<sup>2</sup> The sample periods were chosen based on the availability of systemic crisis data and total credit to GDP ratio data. The primary source of the data comes from the Bank of International Settlement (BIS). The credit data is measured as total credit to the private non-financial sector as a percentage of

---

<sup>2</sup>The list of countries with data available are: Argentina, Austria, Australia, Belgium, Brazil, Canada, Switzerland, Chile, China, Colombia, Czech Republic, Germany, Denmark, Spain, Finland, France, United Kingdom, Greece, Hong Kong SAR, Hungary, Indonesia, Ireland, Israel, India, Italy, Japan, Korea, Luxembourg, Mexico, Malaysia, Netherlands, Norway, New Zealand, Poland, Portugal, Russia, Saudi Arabia, Sweden, Singapore, Thailand, Turkey, United States, South Africa

GDP.<sup>3</sup>

Regarding data for dating systemic crises, there is no single consensus source of database. We follow the literature and use the crisis dates reported in the European Systemic Risk Board crisis data set (Lo Duca et al. 2017) and Laeven and Valencia (2018), which include systemic crisis data from after WWII. However, the first systemic crisis did not happen until the early 1970s.

There are three types of crises: banking, currency, and debt crises, all of which could lead to a systemic crisis. A crisis can be labeled as systemic when certain conditions are met, such as significant signs of financial distress in the banking system, bank runs, and significant policy intervention measurements are imposed. The authors also consulted with subject matter experts in each country to determine the dates of details of the systemic crises.

While the credit data is available until 2021:Q3, our systemic crisis data stopped at 2017:Q4. Therefore, we will limit our analysis till 2017:Q4. We will also omit periods of countries with shorter measurements of credit.

## **4 Empirical Methodology**

### **4.1 Overview**

The total credit series is nonstationary. In order to extract a useful stationary credit signal to use as an early warning indicator for future financial crises, we first decompose the series into a nonstationary trend and a stationary cycle using various popularly studied filters in subsection 4.2.

The idea behind using the credit gap as an early warning indicator is that a prolonged period of excessive credit growth could be linked to future financial crises. Therefore, we can use logistic regression to identify pre-crisis periods with unsustainable high credit gap values over a certain threshold.

In subsection 4.3, we discuss strategies for identifying the dependent variable (pre-crisis periods) and the logistic regression empirical method. In subsection 4.4 we then discuss the metrics to

---

<sup>3</sup>Source of the data could be accessed here: [https://www.bis.org/statistics/totcredit/credpriv\\_doc.pdf](https://www.bis.org/statistics/totcredit/credpriv_doc.pdf)

determine the performance of a credit gap as an early warning indicator EWI model such as BIC, AIC values, Area Under Curve (AUC) of receiver operating characteristic, how we determine a policy loss function, optimize credit gap threshold and estimate the minimized Type I and Type II error rates. We will then propose a novel metric - partial standardized AUC (psAUC) that, when used in conjunction with other metrics, will provide more desirable properties in terms of policy implication and overcome some of the criticism that the traditional AUC metric receives.

This paper aims to overcome model uncertainty when using credit gap filters as an early warning indicator for future financial crises. Since each decomposition filter of the total credit series provides a stationary series that performs qualitatively differently from other filters as an EWI, we propose selecting the best gap candidates and performing model averaging to achieve model stability and improved out-of-sample prediction as proposed in (Raftery, 1995).

In subsection 4.5, we will use the metrics discussed in subsection 4.3 to select credit gap filters that perform well individually and in combination with other credit gaps. Then in subsection 4.6, we discuss the theory behind the Bayesian model average method, our model customization to fit our need, and the new metric (psAUC).

Lastly, in subsection 4.7, for ease of policy implication, we propose the creation of a crisis-weighted credit gap that would inherit the information and features from the averaged model of 30 other credit gaps, and would perform as well as the averaged model prediction does.

## 4.2 Credit gap decompositions

$$100 * \frac{Credit}{GDP} = y_{it} = \tau_{itj} + c_{itj} \quad (1)$$

We started our variable selection process by creating 90 candidate one-sided credit gap measurements based on the literature.<sup>4</sup> The nonstationary Credit to GDP ratio series is decomposed

---

<sup>4</sup>List of filters created is can be viewed in Appendix A.1



into a nonstationary trend  $\tau_{yit}$  and a stationary cycle credit gap  $c_{itj}$  components. Once a country has more than 15 years of credit measurement available, we start storing its one-sided credit gap values onward. Our decomposition filter methods include one-sided Hodrick-Prescott, Hamilton (panel and non-panel setting), Moving Average, Structural Time-series model, Beveridge-Nelson, linear, quadratic, and polynomial decompositions. All are decomposed with full sample, rolling 15 years and 20 years rolling window feature when possible.

Apart from the filters that we already introduced in Chapter 3 - “Measuring credit gap” (Equations 3.1 - 3.5), the two additional filters we will use in this chapter are Moving Average, Structural Time-Series model filters that Beltran et al. (2021) introduced.<sup>5</sup>

### 4.3 Early Warning Indicator - GLM Logistic regression

We begin testing the fitness of the EWI model using credit gap measurements by implementing binary logistic regression. We identify our dependent variable as the periods of 5-12 quarters ahead of a systemic financial crisis. The underlying classification model assumes that the credit to GDP ratio gap will increase sharply as a response to an excessive credit increase due to expansionary monetary policy or a sharp decline in GDP, which is the ratio’s denominator. When the credit gap is above a certain threshold for a prolonged period, it signals that the economy is at risk of a future financial crisis. The periods with excessive credit before financial crises will be labeled pre-crisis periods.

It is also worth pointing out that there is no consensus on setting the periods to label as pre-crisis. The literature labeled pre-crisis periods with various ranges from 1-16 quarters before a crisis. We decided to use a labeling strategy with ideal policy implication properties. It is recommended that macroprudential policy takes more than four quarters to be effective because of policy implementation lag. As a result, a signal telling that the economy will experience a financial crisis in 1-4 quarters has little policy implication. Therefore, we start labeling the pre-crisis periods in the 5th quarter before a systemic crisis. We stop labeling in the 12th quarter before a crisis

---

<sup>5</sup>Refer to their equation representations here in Appendix A.2

because implementing counter-cyclical policy too early will be ineffective in lowering the risk of a future crisis and would only exacerbate economic conditions by unnecessary credit tightening.

Equation (2) represents the binary logistic regression:

$$pre.crisis_{it} \sim c_{itj} \quad (2)$$

In which  $i$  is the country indicator, and  $j$  is the credit gap filter type. The binary dependent variable  $pre.crisis_{it}$  takes values of 1 when the economy is between 5-12 quarters before a systemic financial crisis labeling the periods as “pre-crisis” positive. We discard measurements between 1-4 quarters before a crisis, during, and post-crisis management periods identified in Lo Duca et al. (2017) and Laeven and Valencia (2018). The measurements in these periods are discarded to avoid biased estimation as credit gaps behave erratically during a crisis.

The indicator is set to 0 at other periods. Additionally, pre-crisis periods of imported crises identified in the dataset are set to 0 since the model aims to capture pre-crisis periods created by a country’s excess credit growth, not from exogenous shocks. However, for labeling conformity, we still discard measurements of periods during and post-crisis of these imported crises, as countries experiencing imported crises could implement reactionary monetary policies.

## 4.4 EWI performance metrics

### 4.4.1 Likelihood values and BIC

The logistic regression model parameters are estimated by maximizing the cross-entropy log-likelihood values. The relationship between the BIC (Bayesian Information Criterion) and likelihood value is expressed in Equation (7). The model selection criterion here is to choose the filter method with a lower BIC value associated with a higher likelihood value and fitness of the general linearized model or a higher degree of freedom by using fewer explanatory variables.

#### 4.4.2 Area Under the Curve (AUC) of receiver operating characteristic (ROC)

A receiver operating characteristic (ROC) curve in the EWI literature setting represents True Positive Rate (TPR) and False Positive Rate (FPR) Cartesian coordinates of different credit gap thresholds, which are used as indicators for identifying pre-crisis periods.<sup>6</sup> The credit gap thresholds are determined by the logistic regression model predicted response values which range from (0,1) and map increasingly with the credit gap magnitude values.

A logistic model is strictly preferred over another comparable model if all of the coordinates of its ROC curve are closer to the top-left region of the graph, which minimizes FPR and maximizes TPR. However, that is not always the case when we have such a clear-cut difference. Models can have overlapping ROC curves. To simplify the process of model selection in such cases, the area under the curve (AUC) of ROC metric is used.

$$AUC = \int_0^1 TPR d(FPR)$$

Each logistic regression with a different gap measurement yields an Area Under Curve (AUC) of receiver operating characteristic value. There is an underlying assumption that the higher the AUC value is, the better the overall performance of a credit gap is as an EWI since, on average, its ROC curve has coordinates with higher TPR and lower FPR.

True Positive Rate represents a model's predictive power (P) to identify pre-crisis periods correctly (positive case).  $TPR = TP/(TP + FN)$ . False Positive Rate represents the Type I error rate that a model misidentifies a calm, normal period (negative case) as a pre-crisis period (positive case).  $FPR = FP/(FP + TN)$ . Lastly, we discuss the False Negative Rate or the Type II error rate at which a model misidentifies a pre-crisis period (positive case) as a normal calm period (negative case).  $FNR = 1 - TPR = FN/(TP + FN)$ .<sup>7</sup>

Central planners working with classification problems have dual objectives to minimize Type I

---

<sup>6</sup>Refer to Figure 1 for illustration

<sup>7</sup>Our notation of Type I and Type II error follows Beltran (2021) which deviated from previous literature.

and Type II errors. To simplify this problem, a policy loss function with the input of the two error types are used:

$$L_{\theta, \rho} = \alpha TypeI(\theta) + (1 - \alpha) TypeII(\theta) \quad (3)$$

With  $\alpha = 0.5$  by default if policymakers are indifferent about the two types of errors. Here, the model selection criteria would be to select minimized policy loss function with the smallest combination of Type I and Type II error rates.

A binary logistic regression model with a substantially higher AUC value will have smaller sets of Type I and Type II metrics values. Therefore, its optimized credit gap threshold value that minimizes the policy loss function would represent superior Type I and Type II error rates.

However, because the AUC value is an aggregate value representing information of the whole ROC curve, the area on its lower left corner, where the predictive power of the threshold (TPR) is low, and type II error rates are high, is not relevant for policy implications discussion. Before the Great Financial Crisis of 2008, central planners tended to weigh Type I errors more than Type II errors since the cost of misspecifying a financial crisis was not apparent, and a Type I error meant costing the economy unnecessary credit constraints by implementing reactionary policies. However, since 2009, there has been a shift in weighing Type II errors more as the cost of misspecifying a systemic financial crisis became significantly heavier than a Type I error.

#### 4.4.3 Partial standardised AUC

To overcome the issue of unnecessary information included in the full AUC. An approach to estimate only a partial relevant area under the ROC was proposed in McClish (1989) and later implemented in the EWI setting by Detken et al. (2014). We will also discuss the standardization of the partial AUC metric (psAUC) in Equation (5).

Detken (2014) on partial standardized AUC:

“Instead of considering only the full AUROC (e.g. Drehmann and Juselius, 2014), this paper also presents a partial standardized AUROC (psAUROC) that cuts off the area associated with a preference parameter of  $\theta < 0.5$ .”...

“While the psAUROC has been used extensively in the area of medical statistics to assess the performance of a classifier only in specific regions of the ROC curve (e.g., McClish, 1989 and Jiang et al., 1996), it is a new approach in the literature evaluating EWMs”...

“The results reported in this paper show that the psAUROC can reveal useful additional information as long as the partial area does not become too restricted.”

Detken et al. (2014) proposed cutting off the area representing the preference response parameter in a multivariate regression setting. However, this application does not directly translate to a univariate credit gap model. More relevantly, (Beltran et al., 2021) constrained the policy loss function to regions where  $TPR \geq 2/3$  or Type II error rate  $< 1/3$ . They then estimated the policy loss function value at different points on the partial ROC curve by assigning different policy preference parameter values  $\alpha$ .

To merge these two ideas above and to solve the uncertainty issue of having to estimate policy loss function at different policy preference parameter values to determine model performance, we propose to restrict the consideration of the ROC curve to  $TPR \geq 2/3$ , then estimate the partial standardize psAUC of the restricted ROC curve region instead. Note in Equation (4). We estimated the partial area under to curve that represents TPR and TNR (instead of TPR and FNR) and took integral along the TPR axis.

$$pAUC = \int_{\frac{2}{3}}^1 TNR d(TPR) = \int_{\frac{2}{3}}^1 specificity d(sensitivity) \quad (4)$$

Because we are limiting our analysis only to the portion of the ROC curve that satisfies  $TPR \geq 2/3$ , the policy loss function (3) used in previous literature tends to have a corner solution of

optimized thresholds where  $TPR = 2/3$ . We propose using an alternative policy loss function that prioritizes points on the ROC closest to the top left regions where both Type I and II errors are the lowest:

$$L_{\theta, \rho} = TypeI(\theta)^2 + TypeII(\theta)^2 = (1 - sensitivity)^2 + (1 - specificity)^2$$

The last step is to standardize the partial AUC value:

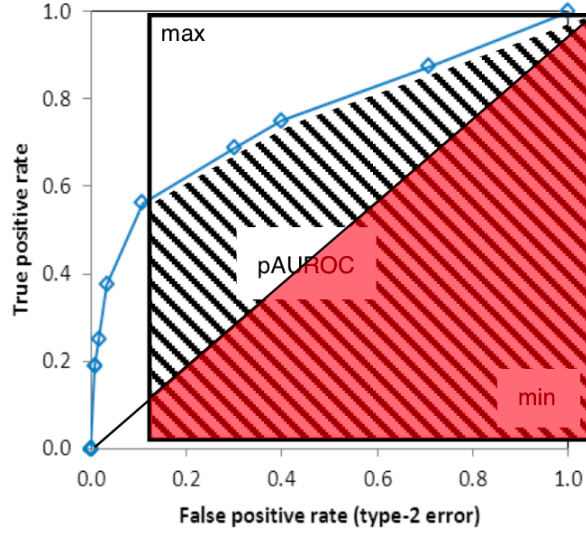


Figure 1: Standardize partial AUC

$$psAUC = \frac{1}{2} \left[ 1 + \frac{pAUC - min}{max - min} \right] \quad (5)$$

The standardization step helps with the comparison of EWI models' performance. Traditional AUC values range from 0 to 1, with 0.5 being the value for a null hypothesis model or an informationless guess. When only estimated partially, the partial AUC value will not retain the same range making it harder to interpret their meaning. Standardizing the partial AUC values restores the ranges of possible values to 0 and 1.

In section 5, we will discuss the relevance and performance of the psAUC as an EWI metric and the other metrics introduced in this subsection.

## **4.5 Variable Selection**

Our overall methodology goal is to implement model averaging of top candidate variables. In order to achieve that, we need to select top EWI candidates using the EWI performance criteria discussed in the previous section. We also search for models with combinations of variables that, when combined using positive weights, would have a better model fit than a univariate model would. We aim to select 29 credit gap measurements based on these two criteria.

Firstly, we compared performances of individual credit gaps using partial area under the curve (psAUC) values. Table 1 ranked decomposition filters by psAUC.

Table 1: Top 30 credit gap measurements ranked by psAUC

Cycles	BIC	AIC	AUC	psAUC	c.Threshold	Type.I	Type.II	Policy.Loss.Function
null	0.0000	0.0000	0.5000	<b>0.5000</b>		1.0000	0.0000	1.0000
c.bn6.r20	-108.0679	-114.4506	0.7048	<b>0.6379</b>	0.6581	0.3962	0.3019	0.2481
c.hamilton28.panel	-149.8518	-156.2346	0.7107	<b>0.6359</b>	9.7674	0.3912	0.3066	0.2470
c.hamilton13.panelr20	-150.2442	-156.6269	0.7036	<b>0.6333</b>	5.9895	0.4261	0.2547	0.2464
c.hamilton24.panel	-134.4093	-140.7920	0.6991	<b>0.6322</b>	7.1794	0.4383	0.2689	0.2644
c.hamilton20.panelr20	-151.5617	-157.9445	0.7048	<b>0.6313</b>	7.9350	0.4321	0.3066	0.2807
c.ma	-120.8108	-127.1936	0.6922	<b>0.6313</b>	5.7813	0.3989	0.3160	0.2590
c.hamilton20.panelr15	-135.3713	-141.7540	0.6985	<b>0.6312</b>	7.5244	0.4616	0.2689	0.2854
c.hamilton13.panelr15	-126.2968	-132.6796	0.6924	<b>0.6311</b>	6.5289	0.4297	0.2830	0.2647
c.hamilton28.panelr20	-164.6015	-170.9842	0.7158	<b>0.6302</b>	10.8558	0.3948	0.2925	0.2414
c.hamilton24.panelr20	-155.8638	-162.2466	0.7096	<b>0.6301</b>	9.1672	0.4251	0.2830	0.2608
c.hamilton24.panelr15	-143.2235	-149.6062	0.7033	<b>0.6299</b>	10.4963	0.3984	0.3160	0.2586
c.hamilton20.panel	-126.8625	-133.2452	0.6907	<b>0.6288</b>	5.6212	0.4686	0.2830	0.2997
c.hamilton28.panelr15	-154.4533	-160.8361	0.7091	<b>0.6270</b>	11.5510	0.3854	0.2972	0.2369
c.hamilton13.panel	-133.9347	-140.3175	0.6922	<b>0.6250</b>	4.9769	0.4285	0.2877	0.2664
c.bn2.r20	-109.3128	-115.6955	0.6963	<b>0.6218</b>	0.2776	0.4080	0.3255	0.2724
c.linear	-135.4069	-141.7896	0.6879	<b>0.6204</b>	3.9989	0.4616	0.2925	0.2986
c.bn2	-135.9914	-142.3741	0.6842	<b>0.6165</b>	0.1864	0.4530	0.3113	0.3021
c.bn6	-132.7915	-139.1742	0.6835	<b>0.6113</b>	0.4710	0.4371	0.2830	0.2712
c.bn6.r15	-54.9953	-61.3781	0.6756	<b>0.6070</b>	0.5680	0.4179	0.3255	0.2806
c.bn2.r15	-83.9469	-90.3297	0.6749	<b>0.6047</b>	0.1349	0.4761	0.3302	0.3357
c.poly4.r20	3.5738	-2.8090	0.5772	<b>0.6011</b>	0.1651	0.4980	0.3302	0.3570
BIS Basel gap	-121.5910	-127.9738	0.6733	<b>0.5960</b>	3.0578	0.4441	0.3255	0.3032
c.bn4	-169.1186	-175.5014	0.6892	<b>0.5943</b>	1.2840	0.3837	0.3255	0.2532
c.bn4.r15	-89.6147	-95.9975	0.6669	<b>0.5929</b>	0.4435	0.4792	0.2925	0.3152
c.bn5.r20	-99.7674	-106.1501	0.6744	<b>0.5928</b>	0.5016	0.4234	0.3302	0.2883
c.stm.r15	-79.5531	-85.9358	0.6575	<b>0.5924</b>	2.0027	0.4778	0.3160	0.3281
c.hp125k	-92.2897	-98.6725	0.6562	<b>0.5924</b>	2.5216	0.4547	0.3302	0.3158
c.hp221k	-106.8842	-113.2670	0.6656	<b>0.5921</b>	2.6641	0.4561	0.3160	0.3079
c.linear.r15	-66.3472	-72.7300	0.6474	<b>0.5916</b>	2.6344	0.4626	0.3302	0.3230
c.hp400k.r15	-67.1228	-73.5055	0.6472	<b>0.5912</b>	2.6223	0.4592	0.3255	0.3168

Secondly, we test for performances of multivariate models with combinations of different credit gaps.

$$Model_k : pre.crisis_{ti} \sim \sum_j \beta_j * credit.gap_{tij}$$

We implement our gaps combination model selection using Markov Chain Monte Carlo Model Comparison ( $MC^3$ ) method developed by Madigan et al. (1995). The method assigns a posterior probability for different credit gaps being selected in the most likely models/combinations with



the lowest BIC values. With 90 variables, there are  $2^{90} = 10^{26}$  possible subsets of combinations to choose from. Babecký et al. (2014) used this  $MC^3$  method to identify potential variables in multivariate EWI models. For more efficient search implementation, each variable is given increasing weights depending on their univariate EWI performance. We then estimated the posterior probability of each variable included in the most likely models using 4,000,000 MCMC iterations.

Table 2: Top 25 credit gap measurements ranked by MC3 probability

Variable	Pr(B!=0)	Variable	Pr(B!=0)
Intercept	1	c.bn2	0.3721
c.hamilton28_panel	0.999968	c.bn3_r10	0.2478
c.bn7_r15	0.93097125	c.hp400k	0.2309
c.poly3	0.915792	c.hamilton24_r20	0.2175
c.hamilton13_panel	0.733247	c.hamilton13_r20	0.1958
c.hamilton13_r10	0.70601225	c.bn2_r10	0.1916
c.hamilton28	0.65141375	c.stm	0.1894
c.hamilton13	0.6438865	c.hamilton28_r20	0.1799
c.bn3_r15	0.514728	c.hp_r20	0.1794
c.hp3k	0.4888055	c.hp	0.1761
c.hp3k_r20	0.4887615	c.bn8_r20	0.1561
c.poly6	0.39826475	c.linear_r20	0.1526
		c.hp221k	0.1440

*Note:*

Pr(B!=0) is the posterior probability of the variable being selected

Using the two results table above, We selected our candidate gaps not only by their psAUC values but also by their other metrics and Type of decomposition methods and features to ensure the robustness of the model averaging method.

The 29 selected decomposition filters are reported in Figure 2. 29 is also the number of recommended variables to use to start the Model Averaging method in the following subsection.

## 4.6 Model averaging

The Bayesian Model Average method is formalized in Raftery (1995) to account for model uncertainty and achieve better prediction results in out-of-sample forecasts.

Equation (6) represents a Model k's posterior probability or weight in the averaging process:

$$P(M_k|D) = \frac{P(D|M_k)P(M_k)}{\sum_{l=1}^K P(D|M_l)P(M_l)} \approx \frac{\exp(-\frac{1}{2}BIC_k)}{\sum_{l=1}^K \exp(-\frac{1}{2}BIC_l)} \quad (6)$$

Where  $P(M_k)$  is model prior probability and can be ignored if all models are assumed to equal prior weights.  $P(D|M_k)$  is marginal likelihood. And  $P(D|M_k) \propto \exp(-\frac{1}{2}BIC_k)$

In which:

$$BIC_k = 2\log(Bayesfactor_{sk}) = \chi_{sk}^2 - df_k \log(n) \quad (7)$$

The subscript s indicates the saturated model.  $\chi_{sk}^2$  is the deviance of model K from the saturated model.  $\chi_{sk}^2 = 2(\log(M_s) - \log(M_k))$ . And  $\log(M_k)$  is the log-likelihood of model Mk given data D. A model with better fitness will have a higher log-likelihood value, hence smaller deviance and smaller BIC value.

#### 4.6.1 Alternate BIC measurement and three-pass filter.

We propose using psAUC in addition to the log-likelihood in the measurement of deviance. Hence, an alternative BIC value can be estimated at:

$$BIC_{alt,k} = 2\log(Bayesfactor_{alt,sk}) \quad (8)$$

$$= 2(1000 * (psAUC_s - psAUC_k)) - df_k \log(n) \quad (9)$$

We scaled the psAUC value by 1000 since  $0 < psAUC < 1$ . Also, by design, the saturated model has  $psAUC_s = 1$ .

#### 4.6.2 Final search steps for best models combinations

All of the following search steps described below are done in quasi-real time. As additional quarterly data are added, we will perform the search steps again to find the best possible combinations of models that fit the updated data.

From the set of 29 candidate filter gaps selected in the previous subsection, we add an intercept parameter to make a list of 30 possible variables from which to choose a subset of combinations. The number of possible subsets of combinations is still significantly high at  $2^{30} = 10^6$  possible combinations subset, which will cost us computation time and resources. We still need to limit the number of possible subsets further to consider. Furnival & Wilson (2000) proposed an algorithm for computing the residual sums of squares for all possible regressions with minimum computing power requirement. With a “leap and bound” technique, the paper found it possible to find the best subsets without examining all possible subsets. This feature reduced the number of operations required to find the best subsets by several orders of magnitude.

Using the list of reduced subsets using “leap and bound” algorithm, we further filter the number of possible gaps combinations using a method called Occam’s razor window described in Raftery (1995). The Bayesian Model Average (BMA) R package has an inbuilt function to filter out less likely models using Occam’s razor window method. The package also allowed for a multiple-pass filter through a custom function.

In this variable selection step using Occam’s Razor, we will first select the models with BIC values within an Occam’s Razor window of the best-fitted model with the lowest BIC value. Secondly, we repeat the same step above but regarding the alternative BIC value derived from psAUC as in Equation (8).

Lastly, we filter out models with negative weight combinations for model averaging stability. In order to reach feasible results in these three-pass filter steps, we relax the Occam’s razor (OC) value from the default value of 20 to 2000, corresponding to a change in the actual search window ( $2 * \ln(OC)$ ) that searches for models with six times less likelihood value to 15 times.

#### 4.6.3 Posterior distribution of coefficients of interest:

With all possible candidate models of credit gap combinations selected, we can start implementing the model averaging steps. First, we estimate each model  $k$  separately, then store their maximum likelihood estimated parameters  $\beta_j(k)$  and their standard deviations. We use the notation  $\beta_j(k)$  as the coefficient of credit gap  $j$  ( $c_j$ ) in a logistic regression model  $k$  against pre-crisis indicator.

We are interested in the averaged estimate of the parameter  $\beta_j(K)$  across all models  $K$ s, this is the estimate of the parameter of interest  $\beta_j$  in an averaged model. When considering a particular  $\beta_1$ , the parameter has this probability distribution in a Bayesian averaged model setting.

$$p(\beta_1|D, \beta_1 \neq 0) = \sum_{A_1} p(\beta_1|D, M_k) p'(M_k|D) \quad (10)$$

Where  $p'(M_k|D) = p(M_k|D)/pr[\beta_1 \neq 0|D]$ . And  $pr[\beta_1 \neq 0|D] = \sum_{A_1} p(M_k|D)$ .  $p(M_k|D)$  is the posterior probability discussed in Equation (6).

While  $pr[\beta_1 \neq 0|D]$  is the probability that  $\beta_1$  is in the averaged model; and  $A_1 = \{M_k : k = 1, \dots, K; \beta_1 \neq 0\}$  is the set of models that includes  $\beta_1$ .

#### 4.6.4 estimation of coefficients of interest:

From Equation (10), an expected value of the parameter of interest  $\beta_1$  could be estimated at:

$$\hat{\beta}_1 = E[\beta_1|D, \beta_1 \neq 0] = \sum_{A_1} \hat{\beta}_1(k) p'(M_k|D) \quad (11)$$

$$SD^2[\beta_1|D, \beta_1 \neq 0] = [\sum_{A_1} [se_1^2(k) + \hat{\beta}_1(k)] p'(M_k|D) - E[\beta_1|D, \beta_1 \neq 0]^2] \quad (12)$$

Where  $\hat{\beta}_1(k)$  and  $se_1^2(k)$  are respectively the MLE and standard error of  $\beta_1$  under the model  $M_k$ .

So far, we have established averaging estimated parameters using the Bayesian Model Average framework. The response values predicted by the generalized linear model regression can perform

better than any individual model in out-of-sample prediction. However, this approach only solves half of the model uncertainty problem. Since we use a multivariate model, deducting an optimized credit gap threshold is still impossible through minimizing the policy loss function. We propose a solution to this issue in the following subsection.

## 4.7 Weighted credit gap creation

Our motivation to create an averaged weighted gap is to find a solution to the issue of not being able to determine an optimized credit threshold discussed above. Since all credit gap variables measure one set of information: how much in absolute value the total credit series deviated from its long-run trend, we can strategically assign weights to them to create a weighted credit gap.

Our objective is to assign weights to the credit gaps to combine them and create a single credit gap that performs as well as the multivariate Bayesian averaged model in the previous subsection 4.6. After creating the weighted cycle, we can perform an optimized credit gap threshold analysis, assign a policy loss function value, and compare its EWI performance with other credit gap filters.

Multivariate GLM binary logistic predicted response values:

$$\widehat{pre.crisis}_{ti} = \widehat{probability}_{ti} = \frac{1}{1 + \exp(-(a + \sum_j \hat{\beta}_j c_{tij}))} \quad (13)$$

From the previous subsection 4.6, in a Bayesian averaged model:  $\hat{\beta}_j = E[\beta_j | D, \beta_j \neq 0] = \sum_{A_j} \hat{\beta}_j(k) p'(M_k | D)$ .

Using the information in Equation (13), we propose creating a single weighted credit gap  $\hat{c}_{ti}$  that satisfies:

$$\frac{1}{1 + \exp(-(a + \hat{\beta} \hat{c}_{ti}))} = \frac{1}{1 + \exp(-(a + \sum_j \hat{\beta}_j c_{tij}))}$$

Or

$$\hat{\beta} \hat{c}_{ti} = \sum_j \hat{\beta}_j c_{tij} \quad (14)$$

This condition allows us to create a weighted gap  $\hat{c}_{ti}$  that will have similar predicted response values and averaged characteristics of the averaged decomposition filters selected. Additionally, to further simplify construction of the weighted gap, we then propose  $\hat{\beta} = \sum_j \hat{\beta}_j$ .

Therefore,

$$\hat{c}_{ti} = \frac{\sum_j (\hat{\beta}_j c_{tij})}{\hat{\beta}} = \frac{\sum_j (\hat{\beta}_j c_{tij})}{\sum_j \hat{\beta}_j} = \sum_j w_j c_{tij} \quad (15)$$

The weight of each candidate credit gap  $j$  is  $w_j = \frac{\hat{\beta}_j}{\sum_j \hat{\beta}_j}$ , which is the fraction of the estimate of coefficient  $\beta_j$  in the averaged model over the sum of the parameters of all selected credit gaps.

#### 4.7.1 One-sided crisis-weighted averaged credit gap

We save the weights  $w_j$  at every incremental period  $t$  of available data to create a one-sided weight vector  $w_{tj}$ . In the next section, Figure 2 illustrates time series of changes in the weights of credit gaps. We estimate the first 15 years of data available for stable model implementation as a full sample regression and assign the weights as constant during that period.

To create one-sided crisis weighted averaged credit gap for each country  $i$  ( $\hat{c}_{ti}$ ), we compute:

$$\hat{c}_{ti,one-sided} = \sum_j w_{tj} * c_{tij} \quad (16)$$

At this point, we have established a framework to create a one-sided crisis weighted averaged credit gap. In the next section, Empirical Results, we will compare the performance as an EWI of this one-sided crisis weighted credit gap (one-sided weighted gap) with other top candidate credit

gaps.

## 5 Empirical Results

### 5.1 One sided weights stacked time series graph

In Figure 2, we see that for the first 15-18 years of the data sample, the parsimony (c.poly3) filter's weight dominated all the other filters. However, after about 20 years, the (c.poly3) filter lost all of its weight. This could be explained by the lack of initial data, leading to poorer performance of other credit gap measurements.

However, as more data were updated, the dynamics of the weights changed significantly. During the early 1990s, Hodrick-Prescott filters' weights overshadowed others, then Beveridge-Nelson, and the Structural Time-Series model did. As the Great Financial Crisis happened, the Hamilton filter in a panel setting gained weight. And lastly, at the end of the crisis data period 2017:Q4, we have 3 Beveridge-Nelson filters (c.bn2\_r20, c.bn2\_r15 and c.bn3\_r15) and 1 Hamilton filter (c.hamilton28\_panelr20) sharing nearly equal weights. To apply our findings in our limited sample, we fixed the weights for 2017:Q4 at constant and extrapolated the weights to 2021:Q3, the end of this paper's credit data availability.

### 5.2 Model Fitness Results

The standard deviation for each estimate is deducted from a 95% confidence interval of the bootstrapping results using 2000 stratified bootstrap replicates.<sup>8</sup>

---

<sup>8</sup>We estimate partial ROC and confidence interval using the R package "pROC" by Robin et al. (2011) whose methodology is based on Carpenter & Bithell (2000)

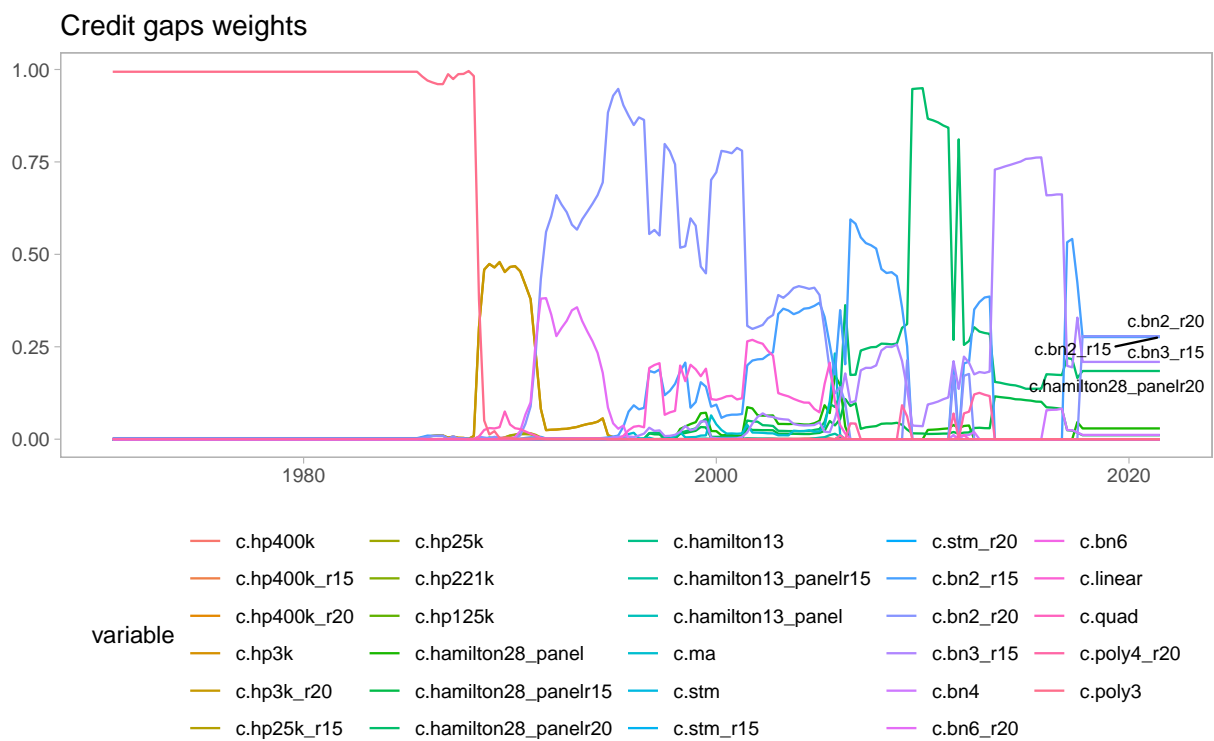
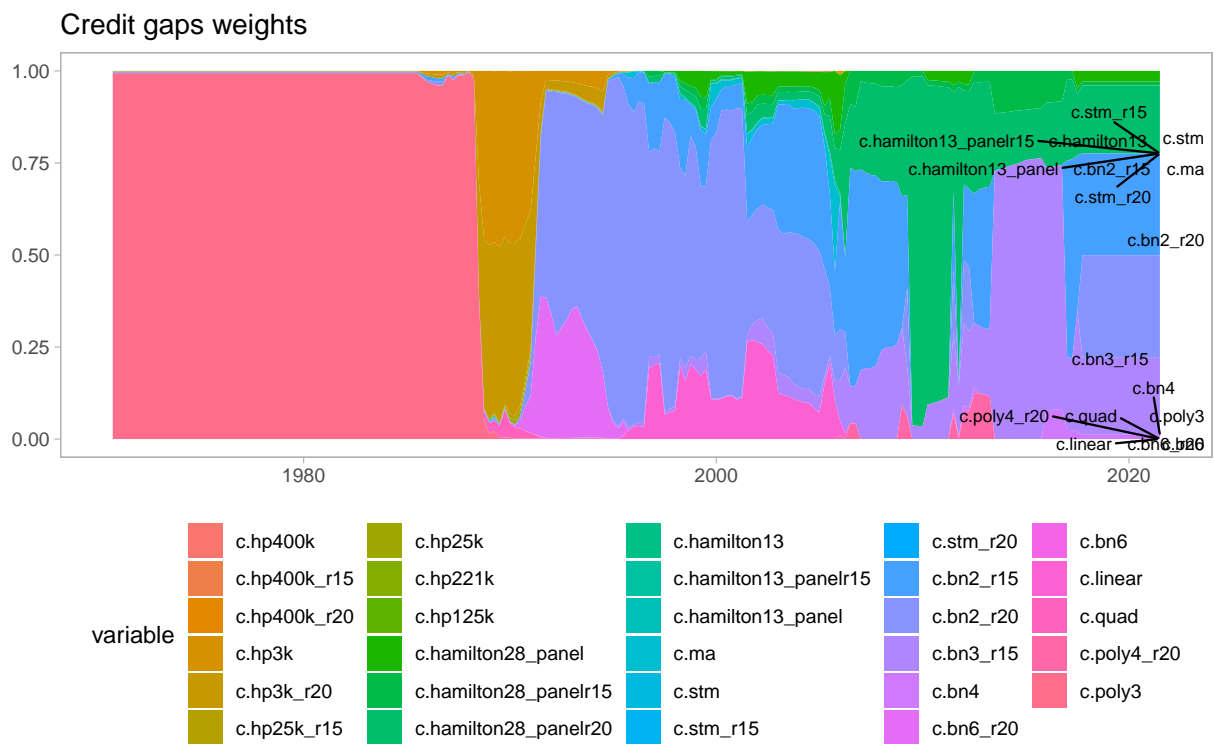


Figure 2: One sided weights graph



Table 3: Credit gaps performance as EWIs - Advanced Economies

Cycle	BIC	AIC	AUC	psAUC	c.Threshold	Type I	Type II	Policy Loss Function
null	0.0000	0.0000	0.5000	<b>0.5000</b>	NA	1.0000	0.0000	1.0000
<b>1.sided weighted.cycle</b>	<b>-128.8749</b>	<b>-134.9430</b>	<b>0.7545</b> (0.0177)	<b>0.6889</b> (0.0195)	<b>3.5486</b>	<b>0.3177</b> (0.0088)	<b>0.2841</b> (0.0348)	<b>0.1817</b> (0.0254)
c.hamilton13.panelr15	-116.9260	-122.9941	0.7282 (0.0174)	<b>0.6846</b> (0.0178)	8.8000	0.3469 (0.0087)	0.3011 (0.0348)	0.2110 (0.0270)
c.hamilton28.panelr15	-141.5772	-147.6453	0.7491 (0.0184)	<b>0.6825</b> (0.0198)	11.5510	0.3834 (0.0088)	0.2216 (0.0319)	0.1961 (0.0212)
c.hamilton28.panelr20	-147.4322	-153.5003	0.7514 (0.0185)	<b>0.6811</b> (0.0195)	13.1072	0.3489 (0.0086)	0.2898 (0.0348)	0.2057 (0.0262)
c.bn6.r20	-90.1022	-96.1703	0.7338 (0.0183)	<b>0.6751</b> (0.0171)	0.6581	0.4126 (0.0094)	0.2273 (0.0319)	0.2219 (0.0222)
c.hamilton28.panel	-129.1730	-135.2411	0.7382 (0.0187)	<b>0.6740</b> (0.0189)	9.7674	0.3977 (0.0090)	0.2386 (0.0319)	0.2151 (0.0224)
c.hamilton13.panel	-122.0054	-128.0735	0.7208 (0.0182)	<b>0.6684</b> (0.0174)	7.2247	0.3509 (0.0090)	0.3239 (0.0362)	0.2280 (0.0296)
c.ma	-97.4868	-103.5549	0.7133 (0.0181)	<b>0.6677</b> (0.0174)	6.0375	0.4073 (0.0087)	0.2727 (0.0333)	0.2403 (0.0255)
c.bn6	-104.1404	-110.2085	0.7042 (0.0198)	<b>0.6426</b> (0.0184)	0.8256	0.4040 (0.0089)	0.2955 (0.0348)	0.2505 (0.0277)
c.hp125k	-92.8331	-98.9012	0.6977 (0.0195)	<b>0.6397</b> (0.0185)	3.6972	0.3788 (0.0089)	0.3239 (0.0348)	0.2484 (0.0293)
<b>BIS Basel gap</b>	<b>-113.9826</b>	<b>-120.0507</b>	<b>0.7124</b> (0.0200)	<b>0.6390</b> (0.0197)	<b>3.9706</b>	<b>0.3847</b> (0.0085)	<b>0.3011</b> (0.0348)	<b>0.2387</b> (0.0275)
c.hp221k	-104.1190	-110.1871	0.7070 (0.0198)	<b>0.6388</b> (0.0197)	3.8869	0.3755 (0.0086)	0.3011 (0.0362)	0.2317 (0.0285)
c.stm.r15	-86.6063	-92.6744	0.6989 (0.0195)	<b>0.6382</b> (0.0204)	4.0784	0.3486 (0.0090)	0.3295 (0.0348)	0.2301 (0.0292)
c.stm	-89.9073	-95.9754	0.6929 (0.0197)	<b>0.6357</b> (0.0193)	3.3071	0.3924 (0.0091)	0.3295 (0.0348)	0.2626 (0.0300)
c.hp400k.r15	-73.5902	-79.6582	0.6869 (0.0194)	<b>0.6357</b> (0.0201)	4.2355	0.3496 (0.0085)	0.3125 (0.0348)	0.2199 (0.0277)
c.linear	-111.9570	-118.0251	0.7069 (0.0206)	<b>0.6356</b> (0.0177)	3.9964	0.4620 (0.0092)	0.2443 (0.0319)	0.2732 (0.0241)
c.hp400k.r20	-93.8576	-99.9257	0.6991 (0.0197)	<b>0.6340</b> (0.0197)	4.3393	0.3622 (0.0090)	0.3295 (0.0362)	0.2398 (0.0306)
c.bn2.r20	-84.0800	-90.1481	0.6955 (0.0201)	<b>0.6297</b> (0.0168)	0.2773	0.4216 (0.0090)	0.3182 (0.0348)	0.2790 (0.0301)
c.stm.r20	-87.6992	-93.7673	0.6879 (0.0200)	<b>0.6282</b> (0.0178)	3.1055	0.4073 (0.0089)	0.3239 (0.0348)	0.2708 (0.0298)

The standard deviation for each estimates are deducted from 95% confidence interval of the bootstrapping results using 2000 stratified bootstrap replicates.

Table 4: Credit gaps performance as EWIs - Emerging Market Economies

Cycle	BIC	AIC	AUC	psAUC	c.Threshold	Type I	Type II	Policy Loss Function
null	0.0000	0.0000	0.5000	<b>0.5000</b>	NA	1.0000	0.0000	1.0000
c.bn3.r15	-46.2774	-51.3507	0.7365 (0.0471)	<b>0.6308</b> <b>(0.0451)</b>	0.6244	0.3059 (0.0134)	0.3333 (0.0779)	0.2047 (0.0623)
c.poly3	5.3862	0.3129	0.5737 (0.0358)	<b>0.6046</b> <b>(0.0264)</b>	1.8089	0.5280 (0.0152)	0.3056 (0.0709)	0.3721 (0.0593)
c.bn2.r15	-13.2062	-18.2795	0.6879 (0.0505)	<b>0.5879</b> <b>(0.0508)</b>	0.2952	0.3566 (0.0140)	0.3333 (0.0779)	0.2383 (0.0642)
c.poly4.r20	7.0732	1.9999	0.5040 (0.0339)	<b>0.5816</b> <b>(0.0160)</b>	-0.9609	0.5962 (0.0147)	0.3333 (0.0779)	0.4665 (0.0717)
<b>1.sided weighted.cycle</b>	<b>6.2094</b>	<b>1.1361</b>	<b>0.5325</b> <b>(0.0402)</b>	<b>0.5811</b> <b>(0.0183)</b>	<b>-1.0639</b>	<b>0.6827</b> <b>(0.0136)</b>	<b>0.1111</b> <b>(0.0496)</b>	<b>0.4784</b> <b>(0.0310)</b>
c.linear	-9.3676	-14.4409	0.5787 (0.0471)	<b>0.5774</b> <b>(0.0208)</b>	-0.9783	0.6294 (0.0143)	0.2222 (0.0709)	0.4455 (0.0495)
c.bn2.r20	-16.6411	-21.7144	0.6760 (0.0519)	<b>0.5751</b> <b>(0.0462)</b>	0.1470	0.4510 (0.0154)	0.3056 (0.0709)	0.2968 (0.0572)
c.hamilton13	6.5749	1.5016	0.5468 (0.0386)	<b>0.5710</b> <b>(0.0229)</b>	3.6354	0.6206 (0.0143)	0.3056 (0.0709)	0.4785 (0.0610)
c.ma	-11.6401	-16.7133	0.5572 (0.0546)	<b>0.5457</b> <b>(0.0164)</b>	-0.2250	0.7220 (0.0134)	0.1667 (0.0567)	0.5491 (0.0382)
c.hamilton28.panel	-7.8687	-12.9420	0.5392 (0.0522)	<b>0.5384</b> <b>(0.0170)</b>	-1.7750	0.6958 (0.0132)	0.2778 (0.0709)	0.5613 (0.0577)
c.quad	6.1997	1.1264	0.4654 (0.0474)	<b>0.5334</b> <b>(0.0140)</b>	-6.4882	0.7456 (0.0127)	0.1944 (0.0638)	0.5938 (0.0455)
c.hamilton13.panelr15	-1.6660	-6.7393	0.5087 (0.0550)	<b>0.5274</b> <b>(0.0130)</b>	-1.1064	0.7002 (0.0136)	0.3333 (0.0779)	0.6014 (0.0732)
c.hp25k.r15	3.6420	-1.4313	0.5018 (0.0558)	<b>0.5265</b> <b>(0.0101)</b>	-3.5672	0.7850 (0.0125)	0.1111 (0.0496)	0.6285 (0.0320)
c.hp25k	3.9466	-1.1267	0.4975 (0.0560)	<b>0.5247</b> <b>(0.0099)</b>	-3.7339	0.7893 (0.0118)	0.1111 (0.0496)	0.6354 (0.0310)
c.hp3k	3.3678	-1.7054	0.5276 (0.0556)	<b>0.5235</b> <b>(0.0146)</b>	-1.1119	0.7019 (0.0136)	0.3333 (0.0850)	0.6038 (0.0758)
c.hp3k.r20	3.3703	-1.7030	0.5276 (0.0556)	<b>0.5235</b> <b>(0.0149)</b>	-1.1125	0.7028 (0.0132)	0.3333 (0.0779)	0.6050 (0.0726)
c.hamilton13.panel	-1.6294	-6.7027	0.5166 (0.0554)	<b>0.5222</b> <b>(0.0134)</b>	-2.9398	0.7500 (0.0132)	0.2778 (0.0709)	0.6397 (0.0591)
<b>BIS Basel gap</b>	<b>-0.7015</b>	<b>-5.7748</b>	<b>0.4928</b> <b>(0.0544)</b>	<b>0.5217</b> <b>(0.0100)</b>	<b>-5.3969</b>	<b>0.7920</b> <b>(0.0118)</b>	<b>0.1389</b> <b>(0.0638)</b>	<b>0.6465</b> <b>(0.0382)</b>
c.hamilton28.panelr20	-4.9986	-10.0719	0.5123 (0.0544)	<b>0.5213</b> <b>(0.0147)</b>	-1.5578	0.6932 (0.0136)	0.3333 (0.0710)	0.5916 (0.0663)

The standard deviation for each estimates are deducted from 95% confidence interval of the bootstrapping results using 2000 stratified bootstrap replicates.

Table 5: Credit gaps performance as EWIs comparison - Full sample

Cycle	BIC	AIC	AUC	psAUC	c.Threshold	Type I	Type II	Policy Loss Function
null	0.0000	0.0000	0.5000	<b>0.5000</b>	NA	1.0000	0.0000	1.0000
<b>1.sided weighted.cycle</b>	<b>-127.5308</b>	<b>-133.9135</b>	<b>0.7182</b> <b>(0.0174)</b>	<b>0.6454</b> <b>(0.0165)</b>	<b>2.8892</b>	<b>0.3532</b> <b>(0.0076)</b>	<b>0.3255</b> <b>(0.0313)</b>	<b>0.2307</b> <b>(0.0257)</b>
c.bn6.r20	-108.0679	-114.4506	0.7048 (0.0184)	<b>0.6379</b> <b>(0.0176)</b>	0.6581	0.3962 (0.0077)	0.3019 (0.0313)	0.2481 (0.0250)
c.hamilton28.panel	-149.8518	-156.2346	0.7107 (0.0185)	<b>0.6359</b> <b>(0.0169)</b>	9.7674	0.3912 (0.0075)	0.3066 (0.0313)	0.2470 (0.0251)
c.ma	-120.8108	-127.1936	0.6922 (0.0181)	<b>0.6313</b> <b>(0.0170)</b>	5.7813	0.3989 (0.0077)	0.3160 (0.0313)	0.2590 (0.0259)
c.hamilton13.panelr15	-126.2968	-132.6796	0.6924 (0.0180)	<b>0.6311</b> <b>(0.0175)</b>	6.5289	0.4297 (0.0079)	0.2830 (0.0313)	0.2647 (0.0245)
c.hamilton28.panelr20	-164.6015	-170.9842	0.7158 (0.0189)	<b>0.6302</b> <b>(0.0181)</b>	10.8558	0.3948 (0.0074)	0.2925 (0.0313)	0.2414 (0.0241)
c.hamilton28.panelr15	-154.4533	-160.8361	0.7091 (0.0190)	<b>0.6270</b> <b>(0.0184)</b>	11.5510	0.3854 (0.0077)	0.2972 (0.0325)	0.2369 (0.0254)
c.hamilton13.panel	-133.9347	-140.3175	0.6922 (0.0184)	<b>0.6250</b> <b>(0.0169)</b>	4.9769	0.4285 (0.0075)	0.2877 (0.0301)	0.2664 (0.0239)
c.bn2.r20	-109.3128	-115.6955	0.6963 (0.0188)	<b>0.6218</b> <b>(0.0170)</b>	0.2776	0.4080 (0.0079)	0.3255 (0.0313)	0.2724 (0.0268)
c.linear	-135.4069	-141.7896	0.6879 (0.0191)	<b>0.6204</b> <b>(0.0145)</b>	3.9989	0.4616 (0.0079)	0.2925 (0.0325)	0.2986 (0.0264)
c.bn6	-132.7915	-139.1742	0.6835 (0.0194)	<b>0.6113</b> <b>(0.0188)</b>	0.4710	0.4371 (0.0075)	0.2830 (0.0313)	0.2712 (0.0242)
c.bn2.r15	-83.9469	-90.3297	0.6749 (0.0192)	<b>0.6047</b> <b>(0.0149)</b>	0.1349	0.4761 (0.0074)	0.3302 (0.0313)	0.3357 (0.0277)
c.poly4.r20	3.5738	-2.8090	0.5772 (0.0154)	<b>0.6011</b> <b>(0.0144)</b>	0.1651	0.4980 (0.0077)	0.3302 (0.0313)	0.3570 (0.0284)
<b>BIS Basel gap</b>	<b>-121.5910</b>	<b>-127.9738</b>	<b>0.6733</b> <b>(0.0197)</b>	<b>0.5960</b> <b>(0.0164)</b>	<b>3.0578</b>	<b>0.4441</b> <b>(0.0077)</b>	<b>0.3255</b> <b>(0.0325)</b>	<b>0.3032</b> <b>(0.0281)</b>
c.bn4	-169.1186	-175.5014	0.6892 (0.0204)	<b>0.5943</b> <b>(0.0206)</b>	1.2840	0.3837 (0.0075)	0.3255 (0.0325)	0.2532 (0.0268)
c.stm.r15	-79.5531	-85.9358	0.6575 (0.0193)	<b>0.5924</b> <b>(0.0167)</b>	2.0027	0.4778 (0.0075)	0.3160 (0.0325)	0.3281 (0.0279)
c.hp125k	-92.2897	-98.6725	0.6562 (0.0194)	<b>0.5924</b> <b>(0.0164)</b>	2.5216	0.4547 (0.0077)	0.3302 (0.0337)	0.3158 (0.0293)
c.hp221k	-106.8842	-113.2670	0.6656 (0.0196)	<b>0.5921</b> <b>(0.0163)</b>	2.6641	0.4561 (0.0077)	0.3160 (0.0325)	0.3079 (0.0274)
c.hp400k.r15	-67.1228	-73.5055	0.6472 (0.0190)	<b>0.5912</b> <b>(0.0164)</b>	2.6223	0.4592 (0.0075)	0.3255 (0.0313)	0.3168 (0.0272)

The standard deviation for each estimates are deducted from 95% confidence interval of the bootstrapping results using 2000 stratified bootstrap replicates.

### 5.2.1 Performance Metrics

Along with the results from Table 3 - 5, we will discuss psAUC as a performance measurement criteria of a credit gap series as an EWI.

Firstly, the psAUC metric overall gives a consistent signal about the quality of the credit gap decomposition filters. A filter with a higher psAUC value, in general, will have a higher quality set of possible policy loss function values to optimize. The standard deviation values of the weighted cycle metrics are also comparable to the other gap measurements.

In some scenarios, psAUC can be a deciding factor if policymakers are indifferent about the AUC value of a filter. In Table 5, the Moving Average (c.ma) and Hamilton 13th lag using panel setting (c.hamilton13.panel) filters have the same AUC values. While the BIC and AIC values, the overall linearized binary logistic regression fitness criteria, would favor (c.hamilton13.panel) filter. The psAUC indicates otherwise. If we only focus on the region where FPR or the optimized threshold's predictive power is at least  $2/3$ , the minimized policy loss function would favor c.ma.

However, there are also instances where psAUC failed to be a single criterion for determining the performance of a filter. In Table 5, the 4th-degree polynomial filter with a rolling sample of 20 years (c.poly4.r20) have a higher psAUC value than the BIS Basel gap. However, all of (c.poly4.r20) filter's other metrics have less preferred values than the BIS Basel gap. As a result, we will deem (c.poly4.r20) less preferred than the BIS Basel gap.

In conclusion, even though psAUC is an overall good measurement for a specific case of binary regression model, where Type II errors are costlier than Type I errors, we focus our optimized threshold on the region where Type II errors are less than  $1/3$ . We still have to use it in conjunction with other criteria such as BIC, AUC, and loss function to determine the performance of a filter. Using it alone will not give biased results of the model performance.

### 5.2.2 Weighted gap and other gaps performance comparison

In this part, we will discuss the performance of the weighted credit gap series as an EWI and compare it with other selected series.

From Table 3 and Table 5 regarding Advanced Economies and full sample results, the weighted credit cycle outperformed other cycles in AUC, psAUC, and Policy Loss function values. Specifically, while maintaining approximately the same optimized threshold as the BIS Basel gap, the weighted credit cycle can achieve lower Type I and II error rates than the BIS Basel gap in all samples.

The findings in Advanced Economies and full sample results also align with the literature findings. Hamilton, moving average filters with optimized parameters perform well, which agrees with the findings in Beltran et al. (2021). Drehmann & Yetman (2021) proposed using Hamilton filter in a panel setting with fixed sloping coefficients across countries, and Galán (2019) found rolling sample with 15 and 20 years window helps with improvement of the early indicator performance.

Across the three tables, optimizing and changing parameters on the one-sided HP filter does not improve its performance over the BIS Basel gap. Contrary to the finding in Beltran et al. (2021) and Galán (2019). The author did not use a full sample analysis but by individual countries. The structural time series filter proposed in Beltran et al. (2021) also did not perform better than the BIS Basel gap.

Unsurprisingly, parsimony filters such as linear and polynomial does not perform well in AE and full sample. However, they can outperform other well-studied gaps in EMEs. This confirmed our finding in the previous section 5.1 that when data are limited, and the market economies are not well established, parsimony filters outperformed other complex filters.

We want to reiterate the importance of avoiding the pitfall of using a single criterion for selecting models, even for the psAUC metric. In Table 4. The (c.poly3) filter has a higher psAUC value than its next filter on the table (c.bn2.r15). However, because (c.poly3)'s other metrics have lower quality than its counterpart, the policy loss function, Type I and II errors rates are higher for (c.poly3).

Last but not least, the emerging market economies (EMEs) have proven challenging to predict crises using the credit gap as EWI because of their more limited sample sizes and less established financial systems. Both Beltran et al. (2021) and Drehmann & Yetman (2021) found that credit gap-based EWIs do not perform well in emerging market economies. In this paper, we found evidence

that adding Beveridge-Nelson decomposition filters to the model selection process improved our model performance for the EMEs, specifically the (c.bn3.r15) Beveridge-Nelson with smoothing parameter 3 and 15 years rolling sample filter. Later graphical analysis in this paper on EMEs will include (c.bn3.r15) filter as another reference. Another Beveridge-Nelson filter with specific features such as (c.bn6.r20) also performs well in both AE and full samples.

### **5.3 Out of sample forecast**

We followed Alessi & Detken (2018) in implementing k-fold leave-one-out cross-validation to perform an out-of-sample forecast analysis. To assess the out-of-sample predictive power of the credit gaps, we perform a 3-fold cross-validation, in which the data is randomly split into three partitions. For each partition, the model is estimated using the other two groups. The resulting estimated parameter is then used to predict the dependent variable in the unused group. We then compute psAUCs and other metrics values and obtain the average metrics and their standard deviations from 10 repeats.

Table 6: Credit gaps performance as EWIs - Out of sample prediction - AE

Cycle	BIC	AUC	psAUC	c.Threshold	Type I	Type II	Policy Loss Function
c.null	0.0000 (1.4201)	0.4729 (0.0099)	<b>0.4906</b> <b>(0.0032)</b>	0.0000	1.0000	0.0000	1.0000
c.hamilton13.panelr15	-92.1757 (3.1133)	0.7247 (0.0050)	<b>0.6801</b> <b>(0.0053)</b>	8.0229 (1.3409)	0.3694 (0.0218)	0.2909 (0.0245)	0.2220 (0.0085)
<b>1.sided weighted.cycle</b>	<b>-83.3823</b> <b>(5.0485)</b>	<b>0.7469</b> <b>(0.0084)</b>	<b>0.6776</b> <b>(0.0127)</b>	<b>3.7972</b> <b>(0.8036)</b>	<b>0.3261</b> <b>(0.0200)</b>	<b>0.2915</b> <b>(0.0253)</b>	<b>0.1923</b> <b>(0.0158)</b>
c.hamilton28.panelr15	-102.2043 (3.1634)	0.7442 (0.0036)	<b>0.6755</b> <b>(0.0059)</b>	12.1456 (1.7511)	0.3595 (0.0202)	0.2744 (0.0300)	0.2058 (0.0070)
c.hamilton28.panelr20	-106.5651 (3.8537)	0.7469 (0.0040)	<b>0.6748</b> <b>(0.0061)</b>	12.2954 (1.8840)	0.3625 (0.0151)	0.2722 (0.0213)	0.2061 (0.0037)
c.hamilton28.panel	-89.7510 (2.9903)	0.7354 (0.0024)	<b>0.6705</b> <b>(0.0036)</b>	10.0439 (1.9885)	0.3930 (0.0157)	0.2585 (0.0226)	0.2220 (0.0041)
c.hamilton13.panel	-101.4619 (1.6108)	0.7184 (0.0022)	<b>0.6665</b> <b>(0.0022)</b>	6.7644 (1.0174)	0.3714 (0.0260)	0.2983 (0.0355)	0.2287 (0.0039)
c.ma	-69.1570 (3.2513)	0.7091 (0.0036)	<b>0.6630</b> <b>(0.0051)</b>	6.0750 (0.7204)	0.3994 (0.0244)	0.2818 (0.0405)	0.2410 (0.0065)
c.bn6.r20	-72.3756 (2.7401)	0.7062 (0.0122)	<b>0.6353</b> <b>(0.0158)</b>	0.6105 (1.4701)	0.4187 (0.0333)	0.3028 (0.0349)	0.2691 (0.0324)
c.hp125k	-71.8112 (5.1241)	0.6930 (0.0039)	<b>0.6351</b> <b>(0.0030)</b>	3.5090 (1.0440)	0.4030 (0.0250)	0.3057 (0.0217)	0.2568 (0.0098)
<b>BIS Basel gap</b>	<b>-87.6905</b> <b>(2.4274)</b>	<b>0.7089</b> <b>(0.0029)</b>	<b>0.6350</b> <b>(0.0049)</b>	<b>4.4951</b> <b>(0.9743)</b>	<b>0.3818</b> <b>(0.0255)</b>	<b>0.3131</b> <b>(0.0283)</b>	<b>0.2451</b> <b>(0.0059)</b>
c.stm.r15	-65.4667 (2.8230)	0.6950 (0.0018)	<b>0.6349</b> <b>(0.0017)</b>	3.7851 (0.7083)	0.3723 (0.0170)	0.3216 (0.0108)	0.2424 (0.0102)
c.hp400k.r15	-58.7622 (3.6325)	0.6830 (0.0028)	<b>0.6330</b> <b>(0.0037)</b>	4.2265 (0.9549)	0.3756 (0.0276)	0.3051 (0.0191)	0.2352 (0.0108)
c.linear	-87.1078 (2.9343)	0.7045 (0.0024)	<b>0.6328</b> <b>(0.0035)</b>	4.8208 (0.8031)	0.4265 (0.0104)	0.3040 (0.0184)	0.2747 (0.0091)
c.hp221k	-79.5029 (4.4132)	0.7001 (0.0051)	<b>0.6311</b> <b>(0.0070)</b>	3.6423 (1.1691)	0.3779 (0.0149)	0.3165 (0.0154)	0.2434 (0.0116)
c.stm	-71.3061 (2.9283)	0.6881 (0.0028)	<b>0.6298</b> <b>(0.0044)</b>	2.9453 (0.8258)	0.4040 (0.0143)	0.3153 (0.0140)	0.2631 (0.0071)
c.hp400k.r20	-74.2327 (2.0353)	0.6940 (0.0050)	<b>0.6286</b> <b>(0.0067)</b>	3.7081 (1.2202)	0.3932 (0.0258)	0.3165 (0.0085)	0.2554 (0.0208)
c.bn6	-76.6398 (2.4656)	0.6914 (0.0100)	<b>0.6243</b> <b>(0.0141)</b>	1.1094 (0.6884)	0.4392 (0.0417)	0.2960 (0.0324)	0.2830 (0.0268)
c.stm.r20	-70.6443 (2.8287)	0.6818 (0.0049)	<b>0.6219</b> <b>(0.0051)</b>	2.9958 (1.0397)	0.4207 (0.0069)	0.3227 (0.0128)	0.2813 (0.0047)
c.bn4	-94.6995 (1.7677)	0.6964 (0.0072)	<b>0.6174</b> <b>(0.0103)</b>	1.8741 (0.8777)	0.3953 (0.0306)	0.3125 (0.0149)	0.2549 (0.0286)

Note:

3-fold cross-validation results. Standard errors are reported in parentheses.

Table 7: Credit gaps performance as EWIs - Out of sample prediction - EME

Cycle	BIC	AUC	psAUC	c.Threshold	Type I	Type II	Policy Loss Function
c.null	0.0000 (2.1053)	0.4445 (0.0294)	<b>0.4828</b> <b>(0.0081)</b>	0.0000	1.0000	0.0000	1.0000
c.bn3.r15	-42.6016 (2.2377)	0.7211 (0.0293)	<b>0.6224</b> <b>(0.0301)</b>	0.3484 (0.4116)	0.3928 (0.0764)	0.3111 (0.0255)	0.2569 (0.0709)
c.linear	-18.9590 (4.6826)	0.5665 (0.0110)	<b>0.5675</b> <b>(0.0123)</b>	0.7792 (2.2166)	0.6094 (0.0510)	0.2889 (0.0559)	0.4599 (0.0407)
c.bn2.r20	-7.2882 (2.2840)	0.6564 (0.0162)	<b>0.5579</b> <b>(0.0164)</b>	0.1180 (0.3002)	0.4755 (0.0576)	0.3167 (0.0268)	0.3300 (0.0459)
c.bn2.r15	-3.5368 (1.8649)	0.6418 (0.0392)	<b>0.5570</b> <b>(0.0247)</b>	0.0252 (0.4690)	0.4911 (0.0757)	0.3306 (0.0088)	0.3557 (0.0745)
c.poly3	1.7710 (0.9026)	0.5143 (0.0404)	<b>0.5499</b> <b>(0.0367)</b>	3.0514 (7.6000)	0.6313 (0.0705)	0.2917 (0.0458)	0.4900 (0.1062)
c.hamilton28.panel	-23.8796 (4.1713)	0.5309 (0.0074)	<b>0.5267</b> <b>(0.0097)</b>	-0.8265 (4.2248)	0.6839 (0.0330)	0.3028 (0.0357)	0.5615 (0.0352)
c.ma	-26.0114 (7.1586)	0.5466 (0.0068)	<b>0.5244</b> <b>(0.0168)</b>	1.2194 (1.5300)	0.6885 (0.0302)	0.3000 (0.0450)	0.5666 (0.0401)
<b>1.sided weighted.cycle</b>	<b>1.6191</b> <b>(0.8146)</b>	<b>0.4787</b> <b>(0.0301)</b>	<b>0.5219</b> <b>(0.0306)</b>	<b>1.5348</b> <b>(4.9140)</b>	<b>0.6621</b> <b>(0.0877)</b>	<b>0.3056</b> <b>(0.0393)</b>	<b>0.5400</b> <b>(0.1028)</b>
c.hamilton13	1.6095 (0.8780)	0.4816 (0.0404)	<b>0.5197</b> <b>(0.0347)</b>	1.7494 (3.5850)	0.6581 (0.0923)	0.3111 (0.0410)	0.5391 (0.1137)
c.quad	1.6542 (0.9705)	0.4498 (0.0291)	<b>0.5040</b> <b>(0.0267)</b>	-8.3825 (9.5714)	0.7454 (0.0623)	0.2806 (0.0357)	0.6389 (0.0948)
c.hamilton13.panel	-7.5078 (7.0689)	0.4985 (0.0153)	<b>0.5037</b> <b>(0.0107)</b>	-0.4802 (2.9623)	0.7158 (0.0277)	0.3139 (0.0187)	0.6119 (0.0342)
c.hamilton28.panelr20	-18.0060 (5.4148)	0.5006 (0.0075)	<b>0.5034</b> <b>(0.0176)</b>	-5.4259 (2.7504)	0.7463 (0.0456)	0.2833 (0.0410)	0.6407 (0.0542)
<b>BIS Basel gap</b>	<b>-6.5423</b> <b>(5.8202)</b>	<b>0.4791</b> <b>(0.0128)</b>	<b>0.5023</b> <b>(0.0138)</b>	<b>-4.3181</b> <b>(2.1378)</b>	<b>0.7816</b> <b>(0.0374)</b>	<b>0.2417</b> <b>(0.0764)</b>	<b>0.6759</b> <b>(0.0408)</b>
c.hamilton28.panelr15	-16.2657 (6.9303)	0.4841 (0.0190)	<b>0.5022</b> <b>(0.0214)</b>	-2.1619 (3.4804)	0.7701 (0.0471)	0.2583 (0.0508)	0.6641 (0.0625)
c.hp3k	1.0131 (0.9964)	0.5102 (0.0157)	<b>0.5017</b> <b>(0.0232)</b>	-0.2747 (1.0254)	0.7488 (0.0234)	0.2861 (0.0322)	0.6440 (0.0445)
c.hamilton13.panelr15	-11.6105 (6.0026)	0.4688 (0.0612)	<b>0.4976</b> <b>(0.0259)</b>	1.2963 (6.5172)	0.7628 (0.0461)	0.2861 (0.0541)	0.6682 (0.0660)
c.hp25k.r15	1.0960 (1.5266)	0.4604 (0.0283)	<b>0.4957</b> <b>(0.0135)</b>	-1.7970 (2.1943)	0.7691 (0.0419)	0.2917 (0.0419)	0.6797 (0.0437)
c.hp221k	-3.6880 (3.7753)	0.4692 (0.0089)	<b>0.4955</b> <b>(0.0139)</b>	-3.9673 (1.9937)	0.7913 (0.0415)	0.2306 (0.0859)	0.6874 (0.0354)
c.bn6.r20	-6.2230 (4.0505)	0.5007 (0.0238)	<b>0.4948</b> <b>(0.0125)</b>	-0.1197 (0.6508)	0.6855 (0.0541)	0.3028 (0.0306)	0.5650 (0.0699)

Note:

3-fold cross-validation results. Standard errors are reported in parentheses.



Table 8: Credit gaps performance as EWIs - Out of sample prediction - Full sample

Cycle	BIC	AUC	psAUC	c.Threshold	Type I	Type II	Policy Loss Function
c.null	0.0000 (1.8172)	0.4767 (0.0128)	<b>0.4907</b> <b>(0.0049)</b>	0.0000	1.0000	0.0000	1.0000
<b>1.sided weighted.cycle</b>	<b>-80.9638</b> <b>(5.0654)</b>	<b>0.7134</b> <b>(0.0033)</b>	<b>0.6404</b> <b>(0.0037)</b>	<b>2.9939</b> <b>(0.6513)</b>	<b>0.3707</b> <b>(0.0282)</b>	<b>0.3208</b> <b>(0.0092)</b>	<b>0.2411</b> <b>(0.0164)</b>
c.hamilton28.panel	-121.3928 (1.7471)	0.7078 (0.0025)	<b>0.6318</b> <b>(0.0032)</b>	9.6472 (0.3081)	0.4043 (0.0195)	0.3071 (0.0155)	0.2583 (0.0113)
c.hamilton13.panelr15	-120.5975 (2.5998)	0.6903 (0.0016)	<b>0.6291</b> <b>(0.0019)</b>	6.6028 (0.5314)	0.4120 (0.0158)	0.3071 (0.0187)	0.2646 (0.0056)
c.ma	-98.5015 (7.0341)	0.6892 (0.0018)	<b>0.6276</b> <b>(0.0030)</b>	5.5675 (0.6929)	0.4045 (0.0146)	0.3113 (0.0201)	0.2611 (0.0059)
c.hamilton28.panelr20	-139.0052 (3.0642)	0.7134 (0.0018)	<b>0.6272</b> <b>(0.0027)</b>	10.5537 (1.1148)	0.3910 (0.0119)	0.3042 (0.0165)	0.2458 (0.0070)
c.hamilton13.panel	-123.8627 (4.9205)	0.6894 (0.0021)	<b>0.6220</b> <b>(0.0034)</b>	5.1997 (0.6877)	0.4206 (0.0194)	0.3132 (0.0199)	0.2757 (0.0092)
c.hamilton28.panelr15	-132.7462 (5.5529)	0.7052 (0.0041)	<b>0.6217</b> <b>(0.0053)</b>	11.6613 (0.5448)	0.3874 (0.0245)	0.3127 (0.0135)	0.2486 (0.0141)
c.linear	-106.5647 (4.8347)	0.6846 (0.0017)	<b>0.6158</b> <b>(0.0028)</b>	4.0160 (1.5341)	0.4621 (0.0110)	0.3113 (0.0156)	0.3108 (0.0080)
c.bn2.r20	-76.7863 (2.1205)	0.6908 (0.0043)	<b>0.6137</b> <b>(0.0066)</b>	0.2458 (0.2248)	0.4327 (0.0239)	0.3208 (0.0116)	0.2907 (0.0149)
c.bn6.r20	-86.4636 (2.8925)	0.6757 (0.0160)	<b>0.5987</b> <b>(0.0205)</b>	0.2144 (1.1466)	0.4713 (0.0471)	0.3231 (0.0114)	0.3286 (0.0445)
c.bn6	-101.2737 (4.0553)	0.6693 (0.0130)	<b>0.5935</b> <b>(0.0176)</b>	0.3956 (0.6125)	0.4611 (0.0472)	0.3250 (0.0061)	0.3203 (0.0459)
<b>BIS Basel gap</b>	<b>-110.9026</b> <b>(4.6247)</b>	<b>0.6707</b> <b>(0.0018)</b>	<b>0.5932</b> <b>(0.0025)</b>	<b>3.3012</b> <b>(0.7148)</b>	<b>0.4490</b> <b>(0.0080)</b>	<b>0.3250</b> <b>(0.0052)</b>	<b>0.3073</b> <b>(0.0059)</b>
c.hp221k	-100.8887 (2.0973)	0.6632 (0.0020)	<b>0.5908</b> <b>(0.0015)</b>	3.4199 (0.5762)	0.4530 (0.0103)	0.3292 (0.0020)	0.3137 (0.0098)
c.hp125k	-88.2934 (2.9343)	0.6540 (0.0012)	<b>0.5905</b> <b>(0.0019)</b>	2.4639 (0.7551)	0.4691 (0.0079)	0.3236 (0.0067)	0.3248 (0.0089)
c.bn2.r15	-47.5774 (2.8755)	0.6637 (0.0110)	<b>0.5901</b> <b>(0.0152)</b>	0.1127 (0.2168)	0.5010 (0.0270)	0.3212 (0.0090)	0.3550 (0.0294)
c.bn4	-143.6399 (1.8203)	0.6835 (0.0043)	<b>0.5889</b> <b>(0.0061)</b>	0.9797 (0.7268)	0.4146 (0.0233)	0.3236 (0.0116)	0.2772 (0.0186)
c.stm.r15	-66.3874 (2.6907)	0.6530 (0.0024)	<b>0.5879</b> <b>(0.0025)</b>	2.4833 (0.8012)	0.4742 (0.0311)	0.3179 (0.0190)	0.3272 (0.0208)
c.stm	-84.2480 (3.8697)	0.6500 (0.0021)	<b>0.5879</b> <b>(0.0026)</b>	2.3616 (0.5334)	0.4767 (0.0063)	0.3250 (0.0068)	0.3330 (0.0056)
c.hp400k.r15	-61.6848 (2.2021)	0.6439 (0.0014)	<b>0.5876</b> <b>(0.0034)</b>	2.6897 (0.8146)	0.4627 (0.0125)	0.3226 (0.0102)	0.3184 (0.0075)

Note:

3-fold cross-validation results. Standard errors are reported in parentheses.

The overall results of the out-of-sample prediction exercise show similar findings as in the previous subsection 5.2 that used bootstrapping method. The ranking of the filters does not change significantly. These results overall affirm the robustness of using credit gaps as EWI models.

On Table 6 regarding the AEs, the weighted cycle has a lower value than the (c.hamilton.panelr15) filter. However, the weighted cycle has a lower loss function value and lower overall Type I and Type II error rates.

Expectedly, for EMEs, on Table 7 the parsimony (c.poly3) 3th power polynomial filter dropped its psAUC in out-of-sample prediction context compared to its bootstrapped value on Table 4. This could be predicted by its poor metrics such as BIC and Loss Function on the model fitness results in the previous subsection.

Through the two subsections 5.2 and 5.3, the model proved to perform well in identifying at-risk periods before a crisis and normal periods of countries in the sample, especially regarding criteria for selecting robust EWI early warning indicators or weighted-averaging between the top candidates for improved model certainty and out-of-sample forecast. With this evidence, We will move our discussion to specific individual countries' analysis in the next section.

## **6 Analyzing time series graphs of individual countries**

This section will discuss the BIS Basel gap and weighted gap as EWIs for the US and UK. Based on the estimated optimized gap threshold values for the weighted gap and BIS Basel gap reported in Table 5 and 8 for full sample results, we will use an optimized credit gap threshold of 3.00 for ease of comparison. The black shaded area on Figures 3 and 4 ranges from 2018:Q1 to 2021:Q3, representing periods where we do not have sufficient crisis data, but credit data are available.

For the US, the weighted gap and BIS Basel gap performed similarly. They were able to predict the late 80s - early 90s financial crisis. They raised above the threshold starting early 2000s, which would result in misclassification of pre-crisis periods (Type I error) since our model wants to only

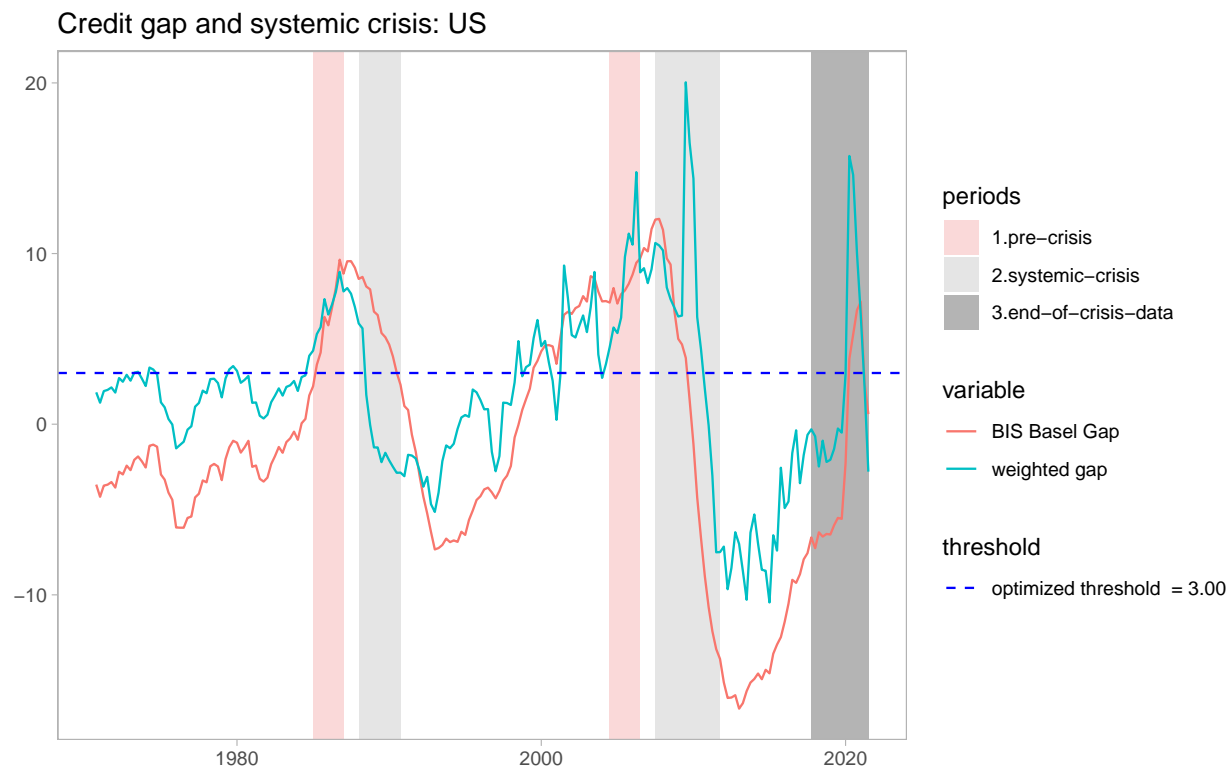


Figure 3: US time series graph

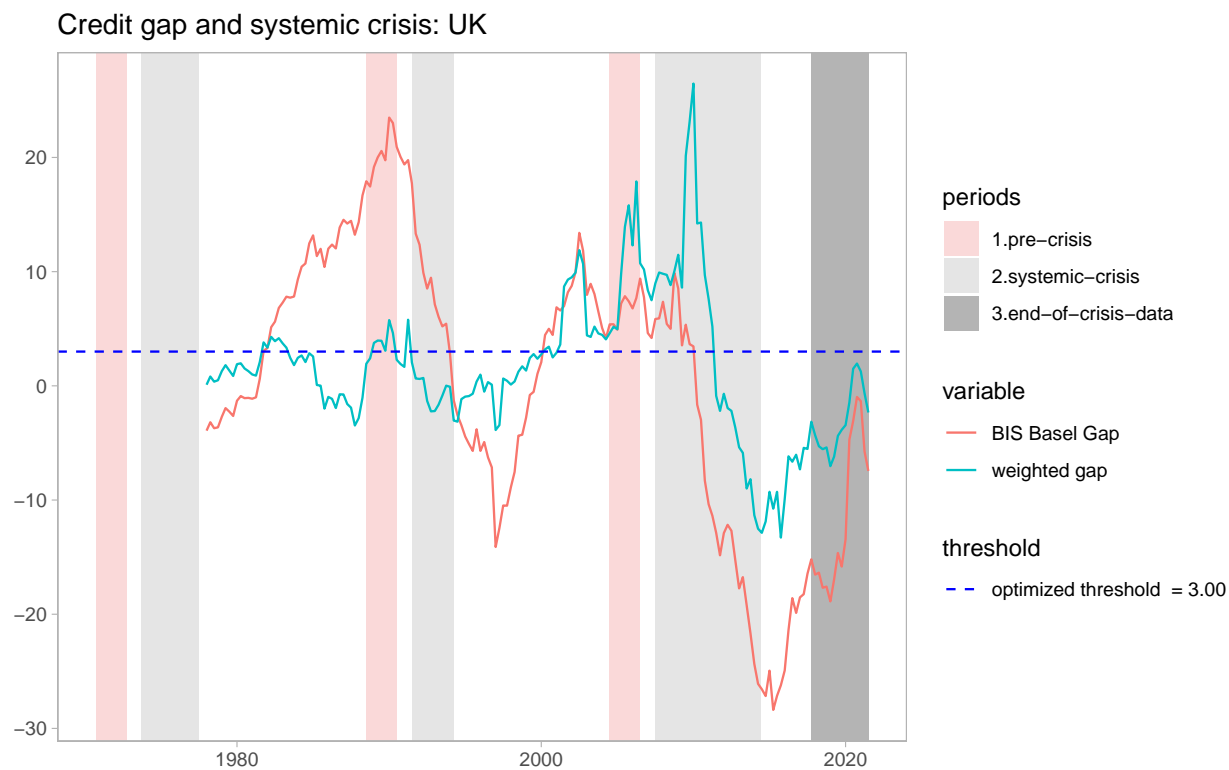


Figure 4: UK time series graph

identify periods of 5-12 months before a systemic financial crisis that would not happen until 2008. Nevertheless, the two gap series showed a build-up of an unsustainable credit gap for a prolonged period before the Great Financial Crisis (GFC) in the US. It is also worth pointing out that the weighted gap dipped below the threshold again before climbing again to identify the pre-crisis period of the GFC correctly.

The graph for the UK tells quite a similar story to the US. Both the credit gaps were able to detect the pre-crisis periods in the early 1990s and the GFC on 2008 for the UK. However, regarding the 90s crisis, the BIS Basel gap went above its threshold too early, for about five years. This could trigger unnecessary counter-cyclical buffer measurements from the central bank and cause the economy to slow its growth rate because of credit constraints. The weighted gap performed much better than the BIS Basel gap as it only sent out the signal during the pre-crisis period of the early 1990s for the UK.

From the graphical features of the weighted gap, we can compare the magnitude and length of the weighted gaps to that of the BIS Basel gap, which has long, smooth trends. Additionally, the weighted gap also retains the more volatile features and short spikes of sudden credit growth that the Beveridge-Nelson and other shorter-term cycle decomposition methods have. These averaged features help explain the averaged method's improved performance as an EWI over other traditional decomposition filters.

From here, to continue our analysis beyond 2017:Q4, we extrapolated our weights on one-sided credit gaps as constants to create a weighted credit gap outside of the crisis data range. This allows us to extrapolate our prediction until 2021:Q3. In response to COVID-19, most countries increased their total credit through multiple monetary policies, or their GDP growth slowed down, both of which created a significant upswing in the magnitude of the credit gap measurement above its optimized pre-crisis detection threshold. Per Drehmann & Yetman (2021), a good EWI signal would have the ideal characteristic of a stable signal meaning the credit gap need to be above the threshold and remain stable above the threshold during the pre-crisis period. We assume from our model finding that an economy with a credit gap above its optimized threshold for 8 consecutive quarters

has a 2/3rd chance of having a financial crisis after the next 4 quarters.

As for the US, even though its credit gap was above the threshold for a few quarters, the gap eventually dropped below the threshold. This does not give us enough evidence to suggest that the country is at risk of a systemic financial crisis in the short future. However, there was an increase in the risk of crisis as the credit gap went beyond the threshold substantially during COVID-19 period. For the UK, the credit gaps did not exceed its threshold. Therefore, there is no evidence to show that the UK is at risk of a future financial crisis.

We performed a graphical analysis of all the countries in the sample and identified countries with significant credit gaps above the optimized 3.00 credit gap threshold beyond the 2017:Q4 crisis data limit window. We selected nine countries with signs of being at pre-crisi risk and should have performed pre-emptive macroprudential measures. They are Canada, France, Hong Kong (SAR), Japan, South Korea, Saudi Arabia, Switzerland, Sweden, and Thailand.<sup>9</sup>

There are three outliers in those nine countries: Canada, Switzerland, and Hong Kong (SAR). Since the end of the GFC, the credit gaps in these three countries have been substantially above their optimized threshold. Additionally, those three countries never experienced systemic financial crises. Therefore, the EWI model performance for those three countries is not robust, and the evidence only comes from abroad. The other six countries: France, Switzerland, Japan, South Korea, Saudi Arabia, and Thailand, have evidence to show that pre-emptive macroprudential measures are to be performed to reduce the risk of having a systemic crisis.

## 7 Conclusion

Our study is based on the idea that excessive credit growth measured by the total credit to GDP deviations could be used as an early warning indicator for future systemic financial crises. We build a novel framework to select relevant credit gap decomposition filters, average their combinations, and create a crisis-weighted credit gap that inherits the combined features of the selected filters.

---

<sup>9</sup>graphs of those countries are included in Appendix A.3.1

We also proposed a novel performance metric for early warning indicator psAUC with a specific constraint that has appealing policy implication properties.

Our findings in this paper provide robust evidence to show the superior predictive performance of our proposed weighted credit gap over other univariate early warning indicator models that use individual credit gap filters. The novel proposed metric psAUC also showed that it has appealing properties for policy maker. It simplifies the steps to search for a decomposition filter with minimized loss function in the range of credit gap thresholds with Type II error rates smaller than  $1/3$ . Therefore, the weighted gap and novel metric can be regarded as useful additional reference points for policy implications.

## References

- Aldasoro, I., Borio, C. E. V., & Drehmann, M. (2018). *Early Warning Indicators of Banking Crises: Expanding the Family* (SSRN Scholarly Paper No. ID 3139160). Rochester, NY: Social Science Research Network.
- Alessi, L., & Detken, C. (2018). Identifying excessive credit growth and leverage. *Journal of Financial Stability*, 35, 215–225.
- Babecký, J., Havránek, T., Matějů, J., Rusnák, M., Šmídková, K., & Vašíček, B. (2014). Banking, debt, and currency crises in developed countries: Stylized facts and early warning indicators. *Journal of Financial Stability*, 15, 1–17.
- Banking Supervision (BCBS), B. C. for. (2010). Guidance for national authorities operating the countercyclical capital buffer. *Bank for International Settlements, December*.
- Beltran, D. O., Jahan-Parvar, M. R., & Paine, F. A. (2021). Optimizing Credit Gaps for Predicting Financial Crises: Modelling Choices and Tradeoffs. *International Finance Discussion Paper*, 2021, 1–40.
- Borio, C. (2014). The financial cycle and macroeconomics: What have we learnt? *Journal of Banking & Finance*, 45, 182–198.
- Borio, C. E. V., & Drehmann, M. (2009). *Assessing the Risk of Banking Crises – Revisited* (SSRN Scholarly Paper No. ID 1513316). Rochester, NY: Social Science Research Network.
- Borio, C., & Lowe, P. (2002). Assessing the risk of banking crises. *BIS Quarterly Review*, 7, 43–54.
- Campagnoli, P., Petrone, S., & Petris, G. (2009). *Dynamic Linear Models with R*. New York, NY: Springer New York.
- Carpenter, J., & Bithell, J. (2000). Bootstrap confidence intervals: When, which, what? A practical guide for medical statisticians. *Statistics in Medicine*, 19, 1141–1164.
- Detken, C., Weeken, O., Alessi, L., Bonfim, D., Boucinha, M., Castro, C., . . . Welz, P. (2014). Operationalising the Countercyclical Capital Buffer: Indicator Selection, Threshold Identification and Calibration Options. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3723336>
- Drehmann, M., & Juselius, M. (2014). Evaluating early warning indicators of banking crises:

- Satisfying policy requirements. *International Journal of Forecasting*, 30, 759–780.
- Drehmann, M., & Tsatsaronis, K. (2014). *The Credit-to-GDP Gap and Countercyclical Capital Buffers: Questions and Answers* (SSRN Scholarly Paper No. ID 2457108). Rochester, NY: Social Science Research Network.
- Drehmann, M., & Yetman, J. (2018). *Why you should use the Hodrick-Prescott filter - at least to generate credit gaps*. 24.
- Drehmann, M., & Yetman, J. (2021). Which Credit Gap Is Better at Predicting Financial Crises? A Comparison of Univariate Filters. *International Journal of Central Banking*, 31.
- Furnival, G. M., & Wilson, R. W. (2000). Regressions by leaps and bounds. *Technometrics*, 42, 69–79.
- Galán, J. (2019). Measuring Credit-To-GDP Gaps. The Hodrick-Prescott Filter Revisited. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3384613>
- Gerdrup, K., Kvinlog, A. B., & Schaanning, E. (2013). *Key indicators for a countercyclical capital buffer in Norway - Trends and Uncertainty*. 44.
- Hamilton, J. D. (2018). Why You Should Never Use the Hodrick-Prescott Filter. *The Review of Economics and Statistics*, 100, 831–843.
- Holopainen, M., & Sarlin, P. (2017). Toward robust early-warning models: A horse race, ensembles and model uncertainty. *Quantitative Finance*, 17, 1933–1963.
- Kaminsky, G. L., & Reinhart, C. M. (1999). The Twin Crises: The Causes of Banking and Balance-of-Payments Problems. *THE AMERICAN ECONOMIC REVIEW*, 89, 132.
- Madigan, D., York, J., & Allard, D. (1995). Bayesian Graphical Models for Discrete Data. *International Statistical Review / Revue Internationale de Statistique*, 63, 215.
- Manias, K. C. (1978). Panics, and crashes: A history of financial crises. *Kindleberger—Basic Books.—1978.—272*.
- McClish, D. K. (1989). Analyzing a Portion of the ROC Curve. *Medical Decision Making*, 9, 190–195.
- Minsky, H. P. (1977). The financial instability hypothesis: An interpretation of Keynes and an



alternative to “standard” theory. *Challenge*, 20, 20–27.

Raftery, A. E. (1995). Bayesian Model Selection in Social Research. *Sociological Methodology*, 25, 111.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12, 77.

## A Appendix

### A.1 List of complete decomposition filters used in the model selection process

All filters are in (quasi-real time) one-sided fashion. We store the value of the decomposed cycles for the current period permanently as data becomes available and will not change it when new data comes in the next period.

#### **Hodrick Prescott (HP) filters with different smoothing parameters $\lambda$ :**

- c.hp, c.hp3k, c.hp25k, c.hp125k, c.hp221k, c.hp400k

#### **Hamilton filters with different smoothing parameters $\theta$ (distance of past lags):**

- c.hamilton13, c.hamilton20, c.hamilton24, c.hamilton28

#### **Linear and polynomial filter models:**

- c.linear, c.quad, c.poly3, c.poly4, c.poly5, c.poly6

#### **Beveridge-Nelson decomposition filters with different smoothing parameters (number of lags):**

- c.bn2, c.bn3, c.bn4, c.bn5, c.bn6, c.bn7, c.bn8

#### **Structural time series model:**

- c.stm

#### **Rolling sample with 15 and 20 years window (of all previous filters in 1-5):**

- c.hp\_r15, c.hp3k\_r15, c.hp25k\_r15, c.hp125k\_r15, c.hp221k\_r15, c.hp400k\_r15, c.hamilton13\_r15, c.hamilton20\_r15, c.hamilton24\_r15, c.hamilton28\_r15, c.linear\_r15, c.quad\_r15, c.poly3\_r15,

c.poly4\_r15, c.poly5\_r15, c.poly6\_r15, c.bn2\_r15, c.bn3\_r15, c.bn4\_r15, c.bn5\_r15, c.bn6\_r15, c.bn7\_r15, c.bn8\_r15, c.stm\_r15 - c.hp\_r20, c.hp3k\_r20, c.hp25k\_r20, c.hp125k\_r20, c.hp221k\_r20, c.hp400k\_r20, c.hamilton13\_r20, c.hamilton20\_r20, c.hamilton24\_r20, c.hamilton28\_r20, c.linear\_r20, c.quad\_r20, c.poly3\_r20, c.poly4\_r20, c.poly5\_r20, c.poly6\_r20, c.bn2\_r20, c.bn3\_r20, c.bn4\_r20, c.bn5\_r20, c.bn6\_r20, c.bn7\_r20, c.bn8\_r20, c.stm\_r20

**Hamilton filters in panel setting (with rolling samples):**

- c.hamilton13\_panel, c.hamilton20\_panel, c.hamilton24\_panel, c.hamilton28\_panel, c.hamilton13\_panelr15, c.hamilton20\_panelr15, c.hamilton24\_panelr15, c.hamilton28\_panelr15, c.hamilton13\_panelr20, c.hamilton20\_panelr20, c.hamilton24\_panelr20, c.hamilton28\_panelr20

**Moving Average filter:**

- c.ma

All models are given equal prior weights.

## A.2 Statistical Methods for Trend-Cycle Decomposition

### Bayesian structural time series model (STM)

$$y_t = u_t + v_t, v_t \sim N(0, V)$$

$$u_t = u_{t-1} + \beta_{t-1} + w_{1,t}, w_{1,t} \sim N(0, \sigma_{w1}^2)$$

$$\beta_t = \beta_{t-1} + w_{2,t}, w_{2,t} \sim N(0, \sigma_{w2}^2)$$

The implementation of this filter is discussed in (Campagnoli, Petrone, & Petris, 2009). One feature of this filter that allows for smoother trend component is its inclusion of a time-varying local growth rate  $\beta_t$ . Beltran et al. (2021) estimated the optimized the value of the smoothing parameter at  $\sigma_{w1}^2 = 1$ ,  $\sigma_{w2}^2 = 0.01$  and  $V = 1100$ .

## Moving Average

$$MA_t = y_t - \frac{\sum_{i=t-q+1}^t y_i}{q}$$

With  $q$  being the smoothing parameter and the length of the moving average window. Beltran et al. (2021) estimated the optimized value for parameter  $q$  to be 16.

### A.3 Out of sample forecast graphs

#### A.3.1 Graphs of selected countries

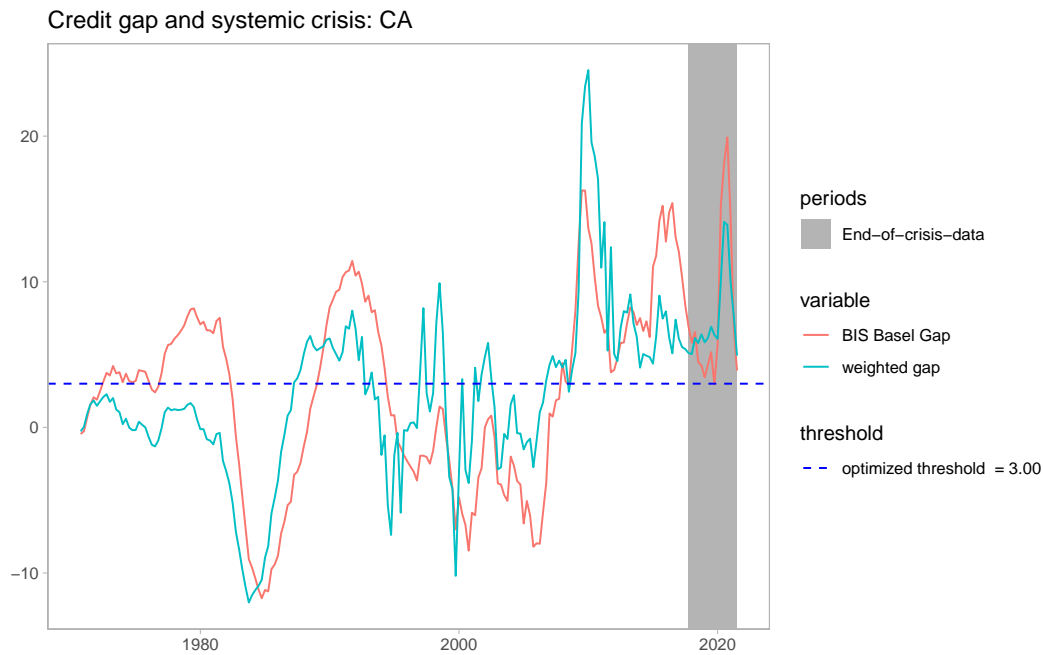


Figure 5: Canada time series graph

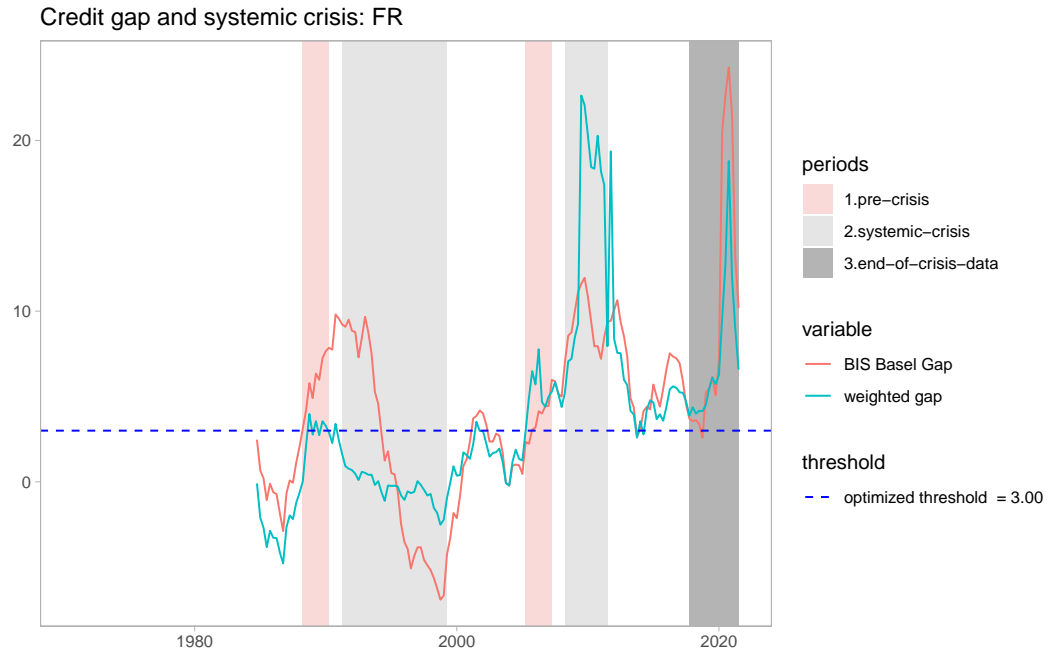


Figure 6: France time series graph

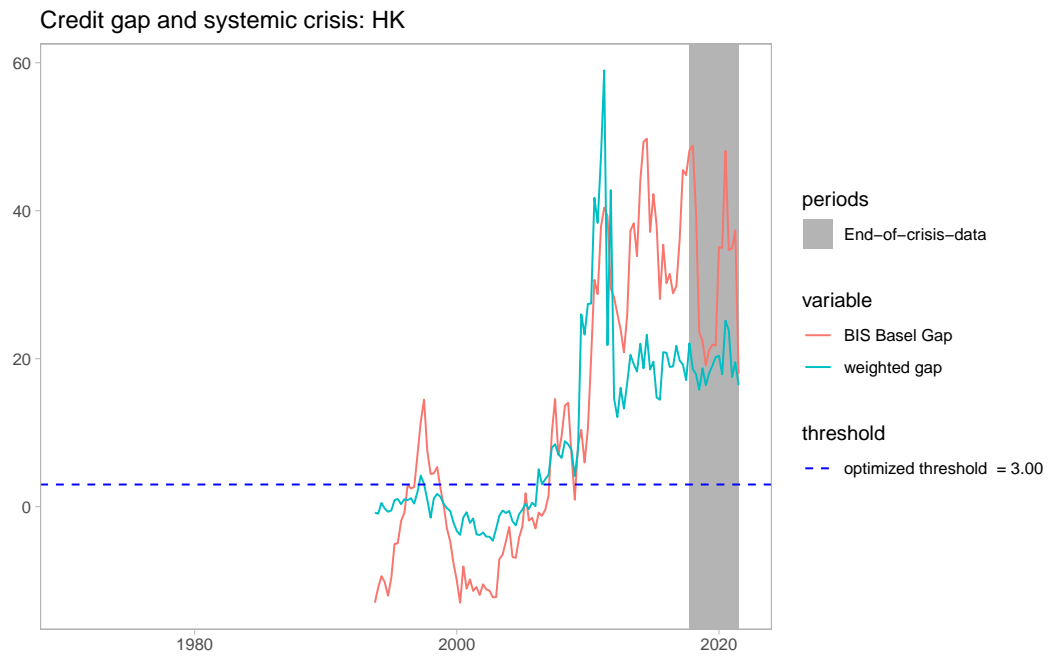


Figure 7: Hong Kong time series graph

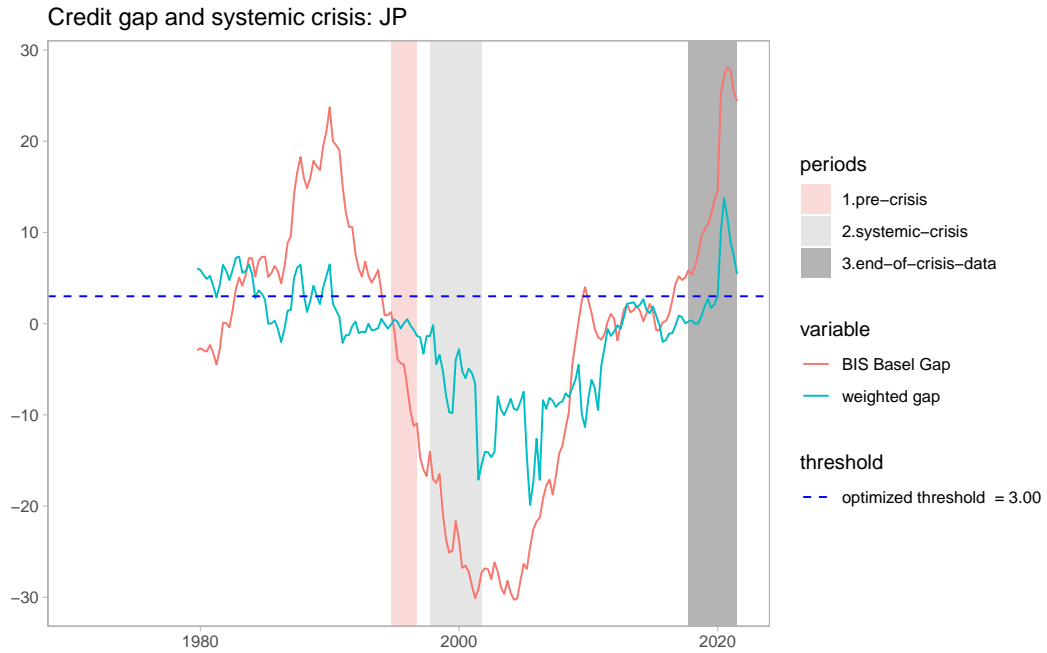


Figure 8: Japan time series graph

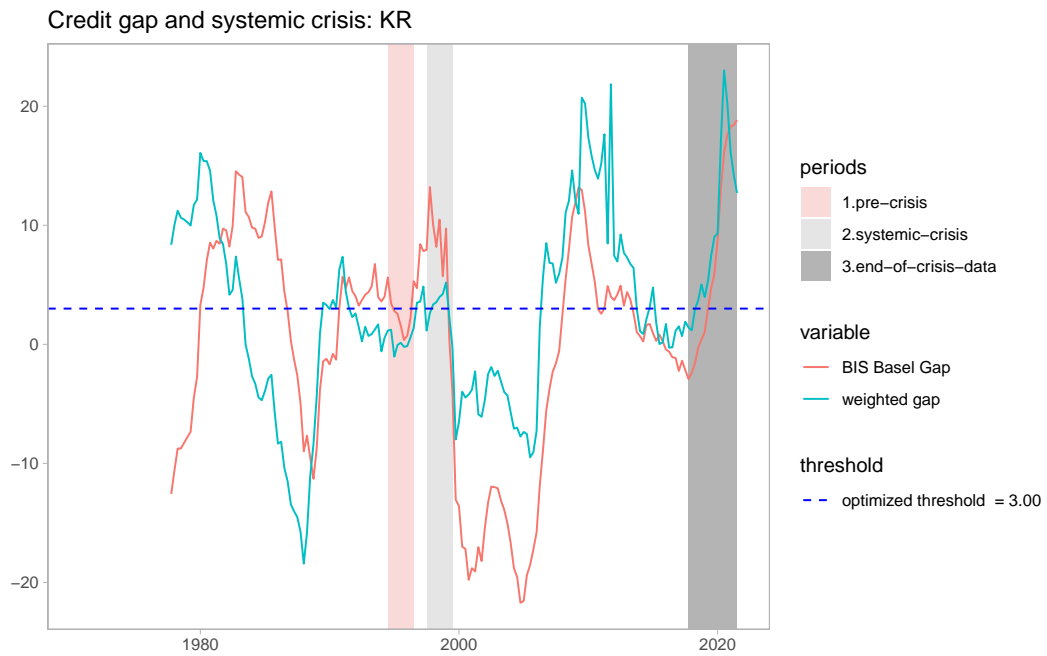


Figure 9: South Korea time series graph

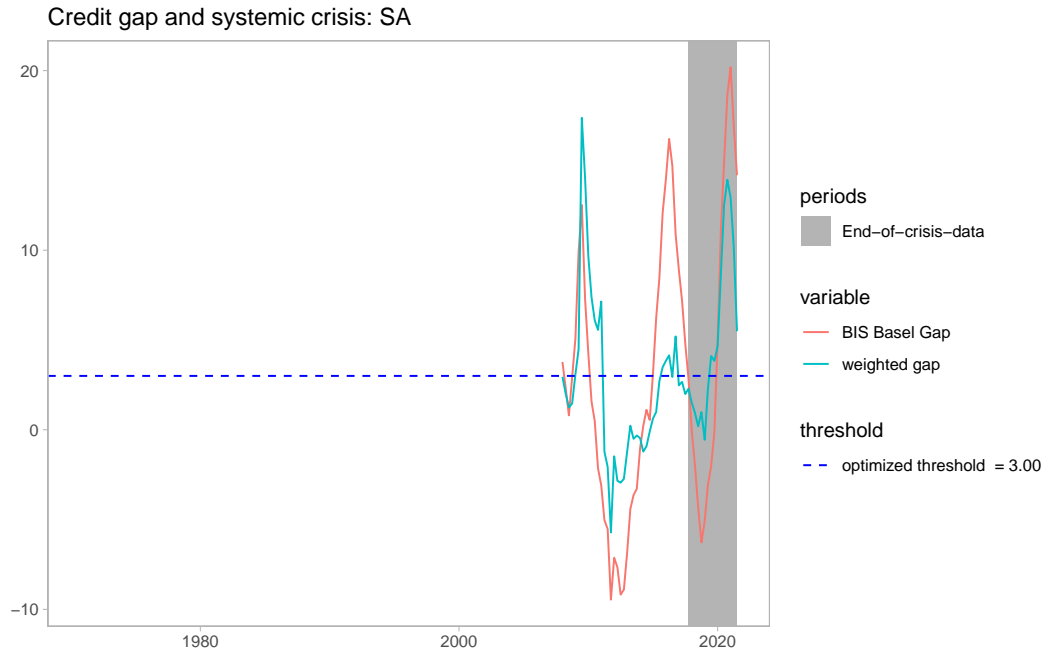


Figure 10: Saudi Arabia time series graph

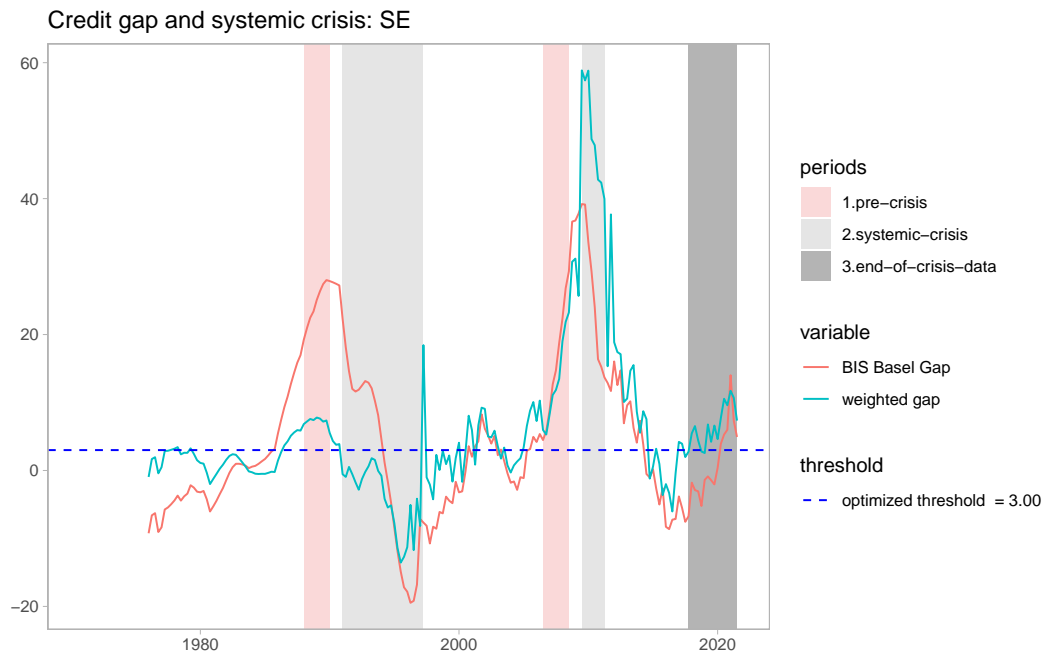


Figure 11: Sweden time series graph

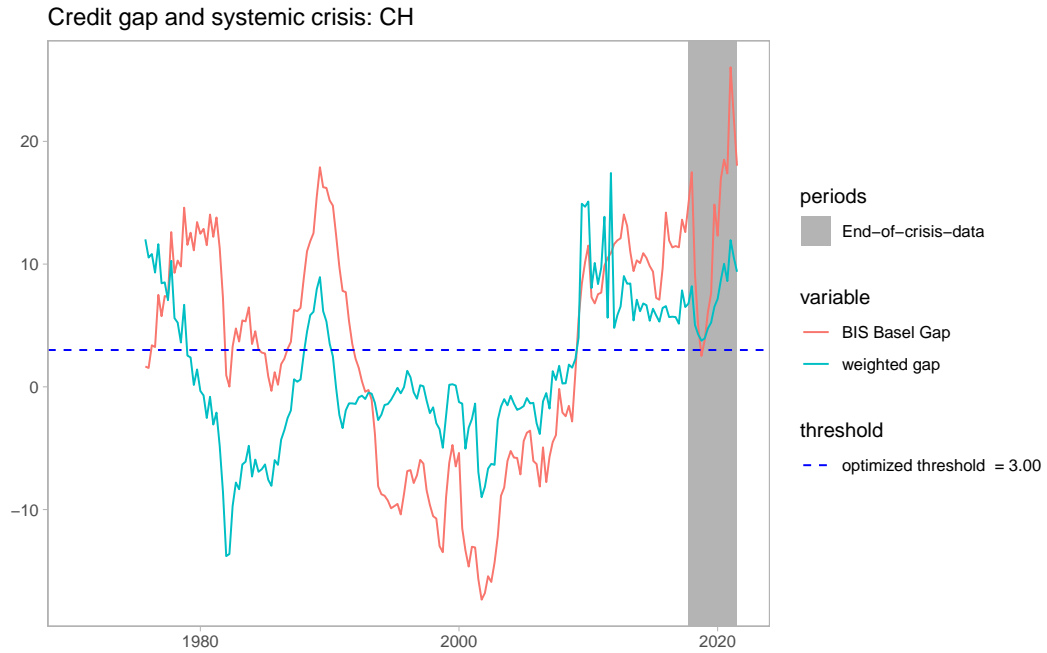


Figure 12: Switzerland time series graph

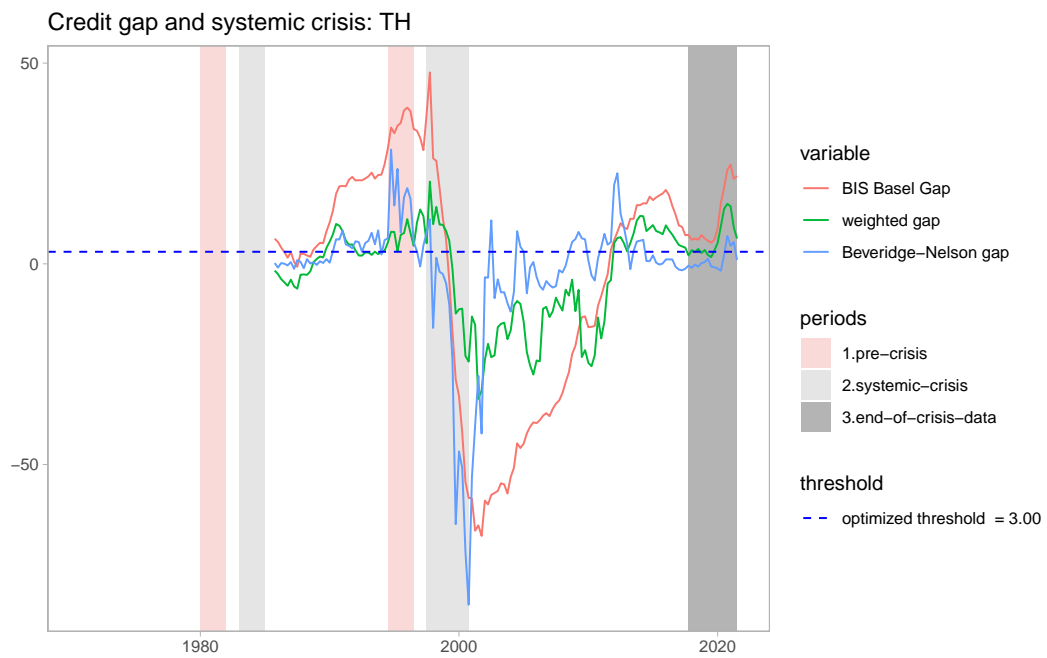


Figure 13: Thailand time series graph