Is this Time Different? Financial Follies across Centuries*

Jérémy Fouliard Hélène Rey Vania Stavrakeva

July, 2022

Abstract

Carmen Reinhart and Ken Rogoff write that "no matter how different the latest financial frenzy or crisis always appears, there are usually remarkable similarities with past experience from other countries and from history". Can we really use the same models to predict the 20th century Great Depression and the 21st century Great Recession? We find that we can predict *out-of-sample* the 1929 Great Depression and all the 20th and 21st century systemic financial crises in a panel of countries. We also generally do not over predict crises during the Bretton Woods period.

Keywords: Financial crises, early warning indicators, financial history

*Jérémy Fouliard, Paris School of Economics, jfouliard@london.edu. Rey: London Business School, CEPR, and NBER, hrey@london.edu.Stavrakeva: London Business School and CEPR, vstavrakeva@london.edu This working paper should not be reported as representing the views of the French Macroprudential Authority. We thank for very valuable comments and suggestions Barry Eichengreen and our discussant Philip Lane as well as seminar participants at UC Berkeley and the ASSA meetings. All errors are our own.

1 Introduction

According to Eichengreen and Portes (1987), "much as the study of disease is one of the most effective ways to learn about human biology, the study of financial crises provides one of the most revealing perspectives on the functioning of monetary economies." Crises can be defined as strong disturbances to financial markets, accompanied by sharp falls in asset prices and widespread insolvencies among debtors and intermediaries, which disrupts the capacity to allocate capital within the economy and leads to a decrease in real activity. As shown in the classic Reinhart and Rogoff (2009) book "This Time is Different, Eight hundred years of Financial Follies", financial crises have occurred repeatedly in emerging markets and advanced economies alike, and they exhibit some remarkable similarities. Crises are often, but not always, credit booms gone bust as described by Minsky (1986) and Kindleberger (1978). As the Great Depression unfolds, Fisher (1933) states that "over-investment and over-speculation are often important; but they would have far less serious results were they not conducted with borrowed money. That is, over-indebtedness may lend importance to over-investment or to over-speculation." From a theoretical point of view, there are many different models in macroeconomics and in finance which have been developed to understand the causes and mechanics of crises. Some emphasise runs as in Diamond and Dybvig (1983). Many models in macro-finance focus on the bust phase of the crisis and on amplification mechanisms such as fire sales of assets. However, as argued by Sufi and Taylor (2021) who provide a very useful survey of the literature, crises do not occur randomly and as a result, it is important to understand better the booms that precede them. A few papers analyze theoretically the boom phase of the financial cycle and emphasise limited liability and asset overvaluations due to risk-shifting (Coimbra and Rey (2017)), search-for-yield in low interest rates environments (Martinez-Miera and Repullo (2017)), or deviations from rational expectations and financial constraints (Gennaioli et al. (2012)) as key mechanisms. From an empirical point of view, a number of variables have been used to predict financial crises -mostly in-sample. Following Kaminski and Reinhart (1999), the literature has very usefully described the behaviour of a number of key variables around crisis episodes. Lowe and Borio (2002) and Gourinchas and Obstfeld (2012) emphasise the role of credit growth. Mian and Sufi (2009) and Mian et al. (2017) underline the importance of household debt in making financial crises more severe, while Müller and Verner (2021) shows that excess credit extended to the non-tradable sector (in particular real estate) is more likely to be followed by a crisis. Schularick and Taylor (2012) and Greenwood et al. (Forthcoming) find some evidence of predictability in post war data using past credit and asset price growth. But are the similarities between crises large enough to make them predictable even when they are

many years apart? Can we take Reinhart and Rogoff (2009) literally and use information from the end of the 19th and the very beginning of the 20th century to predict the Great Depression and all subsequent systemic crises in a panel of countries? The answer provided by this paper is positive. We use the Jorda-Schularick-Taylor macro-history database, which contains standard macroeconomic and financial variables at the annual frequency between 1870 and 2017 (Jordà et al. (2017)). Based exclusively on the models estimated on the period 1870-1922, we can predict out-of-sample the 1929 Great Depression three years ahead and, updating the coefficient estimates as we move forward in time we can predict out-of-sample all the 20th and 21st century systemic crises three years ahead for the United States, France, Italy, Japan, Spain and Netherlands. The relevant set of models we uncover is wider than what can be found in the existing literature. We show in particular that credit growth and asset prices are outperformed by a broader set of variables as far as their forecasting ability is concerned. In particular, it is often very important to complement them with housing market variables and with real economy, exchange rate and current account variables.

The paper is structured as follows. In section 2 we introduce and discuss the online learning methodology, which forms the basis of our predictive strategy. In section 3, we present our data and a set of empirical models of financial crises. Results and their interpretation are in section 4. Section 5 concludes.

2 Sequential learning about crises over a hundred years

Most of the literature on crises uses standard econometric methods such as panel data analysis or event studies in order to identify early warning indicators of financial crises. Some recent attempts to introduce new forecasting methods imported from the machine learning literature can be found in Ward (2017) who uses classification trees and Bluwstein et al. (2020) who compare the forecasting performance of decision trees, random forests, extremely randomised trees, support vector machines, and artificial neural networks. Instead, we adopt the *framework of sequential prediction or online machine learning* (see Cesa-Bianchi and Lugosi (2006)), which aggregates optimally a large number of models. As such it is more general, since if any model has a superior forecasting power, it will get picked, but if some models perform better at different points in time, the methodology will increase or decrease their weights in the forecast accordingly. Furthermore, unlike in the classical statistical theory of sequential predictions, where the sequence of outcomes is assumed to be a realization of a stationary stochastic process, in our framework, crises are the product of some unknown and unspecified mechanism, which could be deterministic, stochastic, or even adversarially adaptative to our own behavior. This allows us to make no assumptions

on how the data are generated, which is a big advantage as there is no consensus on a theory of financial crises. Online machine learning is specifically geared at real-time prediction in situations where the true models driving outcomes are not known and can be different over time. We aim at predicting pre-crises (3 year before a crisis) as this is the relevant horizon for macro prudential policies aiming at fostering financial stability. Our approach can be described as "meta-statistic" since the aim is to make the best prediction by aggregating models (also called experts)' predictions. The forecaster's error is then the sum of two errors: an *estimation error* measured by the error of the best combination of experts, known *ex post*, representing the best prediction the forecaster can make using the available information and an *approximation error* measuring the difficulty to approach *ex ante* the best combination of experts. Though based on model averaging with time varying weights, *on-line learning* is more general than Bayesian Model Averaging[1]; importantly and as already mentioned, it does not make any assumption on the data generating processes; furthermore it allows for time-varying learning rates. To our knowledge, there are only two papers applying online machine learning to economics: Amat et al. (2018) for models of exchange rates and our own work on post 1980 crises in Europe on a rich set of quarterly macro-financial variables (Fouliard et al. (2020)). In contrast online learning has been used in a number of applications outside economics, for example to forecast electricity consumption (Devaine et al. (2013)), to track the performance of climate models (Monteleoni et al. (2011)), to model the network traffic demand (Dashevskiy and Luo (2011)), to forecast air quality (Mallet et al. (2009)) and to predict of outcomes of sports games Dani et al. (2012). An advantage of the methodology is that it also allows us to track which models perform well over time in a given country. This is an important characteristic which sets it apart from "data mining" or black box approaches. This is often enlightening to understand sources of instability and suggest economic mechanisms -though of course we cannot formally identify any *causal relationship* between variables having good forecasting power and the origins of the crisis.

2.1 Online Learning Methodology

Consider a bounded sequence of observations (the occurence or non-occurrence of pre-crisis periods) $y_1, y_2, ..., y_T$ in an outcome space $\mathcal{Y}$. The goal of the forecaster is to make the predictions $\hat{y_1}, \hat{y_2}, ..., \hat{y_T}$ in a decision space $\mathcal{D}$.

This framework has two main specificities. First, the observations $y_1, y_2, ...,$ are revealed in a sequential order. At each step $t = 1, 2, ..$, the forecaster makes a prediction $\hat{y_t}$ on the

[1] In some cases, even very simple ones (see Grunwald and van Ommen (2014)), Bayesian Model averaging does not converge due to heteroskedasticity.

basis of the previous $t-1$ observations before the tth observation is revealed. This is why this approach is said to be "online" since the forecaster sequentially receives information. The optimal forecasting model is adaptable over time which is very convenient when the predictive content is unstable over time. This lack of stability is indeed a stylized fact in the forecasting literature (Stock and Watson (2012) and Rossi (2011)). In contrast to the stochastic modelling approach, we do not assume that $y_1, y_2, ...$ are the product of a stationary stochastic process. The sequence $y_1, y_2, ...$ could be the result of the interactions of very complex non linear processes.

The forecaster predicts the sequence $y_1, y_2, ...$ using a set of "experts". Experts are predictive models. They can be statistical models, an opinion on y_t using private sources of information or a black box of unknown computational power (neural networks for example). Each expert $j = 1, ..., N \in \mathcal{E}$ makes the prediction $f_{j,t}$ based only on information available until date t-1. The methodology of *online learning* is therefore extremely flexible and general as any forecasting model can be used to contribute to the optimal forecast. But of course there is no magic, if all forecasting models are bad, the optimal forecast will also be bad. If we put "garbage in", we will get "garbage out".

To combine experts' advice, the forecaster chooses a sequential aggregation rule $\mathcal{S}$ which consists in picking a time-varying weight vector $(p_{1,t}, ..., p_{N,t}) \in \mathcal{P}$. The forecaster's outcome is the linear combination of experts' advice :

$$\hat{y_t} = \sum_{j=0}^{N} p_{j,t} f_{j,t}$$

After having computed $\hat{y_t}$ (based on information available until t-1), the forecaster and each expert incur a loss defined by a non-negative loss function : $\ell : \mathcal{D} \times \mathcal{Y}$.

How do we measure the sequential aggregation rule's performance ? If the sequence $y_1, y_2, ...$ were the realisation of a stationary stochastic process, it would be possible to estimate the performance of a prediction strategy by measuring the difference between predicted value and true outcome. But we do not have any idea about the generating process of the observations. However, one possibility is to compare the forecaster's strategy with the best expert advice. Let's define the difference between the forecaster's loss and the loss of a given expert, cumulated over time:

$$R_{j,T} = \sum_{t=1}^{T} (\ell(\hat{y_t}, y_t) - \ell(f_{j,t}, y_t)) = \hat{L_T} - L_{j,T}$$

where $\hat{L_T} = \sum_{t=1}^{T} \ell(\hat{y}_t, y_t)$ denotes the forecaster's cumulative loss and $L_{j,T} = \sum_{t=1}^{T} \ell(f_{j,t}, y_t)$ is the cumulative loss of the expert j.

The *regret* of a sequential aggregation rule $\mathcal{S}$ is given by :

$$R(\mathcal{S}) = \hat{L}_T(\mathcal{S}) - \inf_{q \in \mathcal{P}} L_T(q)$$

where $\inf_{q \in \mathcal{P}} L_T(q) = \inf_{q \in \mathcal{P}} \sum_{t=1}^{T} \ell(\sum_{j=0}^{N} q_{j,t} f_{j,t}, y_t)$ is the cumulative loss of the best convex combination of experts (known *ex post*).

This difference is called "regret" since it measures how much the forecaster regrets not having followed the advice of this particular combination of experts. The regret is a way of measuring the performance of a forecaster's strategy by comparing the forecaster's predictions (based on information at date t-1) with the best prediction which could have been done had she followed a certain combination of experts based on realised value at date t.

Knowing that $\hat{y}_t = \sum_{j=0}^{N} p_{j,t} f_{j,t}$, the regret can be written as :

$$R(\mathcal{S}) = \sum_{t=1}^{T} \ell(\sum_{j=1}^{N} p_{j,t} f_{j,t}, y_t) - \inf_{q \in \mathcal{P}} \sum_{t=1}^{T} \ell(\sum_{j=1}^{N} q_{j,t} f_{j,t}, y_t)$$

Minimizing the regret is for the forecaster a robustness requirement. When the regret is close to 0, it ensures that forecaster's strategy (determined at date t-1) is close to the best combination of experts, which is known at the end of the round (at date t). To get a robust aggregation rule, the forecaster wants, in addition of having the smallest bound possible for the regret, to obtain a "vanishing per-round regret" so that when T goes to infinity the superior limit of the regret taken over all possible observation and prediction sequences goes to zero:

$$\lim_{T \to \infty} \sup \left\{ \frac{R(\mathcal{S})}{T} \right\} \leq 0$$

In this case, the forecaster's cumulative loss will converge to the loss of the best linear combination of experts known *ex-post*. This approach can be described as "meta-statistic" since the aim is to find the best sequential linear combination of experts. Indeed, the following decomposition:

$$\hat{L}_T(\mathcal{S}) = \inf_{q \in \mathcal{P}} L_T(q) + R(\mathcal{S})$$

indicates that the forecaster's cumulative loss is the sum of an estimation error, given by the cumulative loss of the best linear combination of experts (known *ex post*), and by the regret which measures the difficulty to approach *ex ante* the best combination of experts[2]. In our case we also have to take into account the delayed informational feedback as we know if we are in a pre-crisis period only 3 years later (when the crisis starts) so we update our weights of the aggregation rule with a 3 year delay. We describe the algorithm used in more details in Appendix.

2.2 Loss function

The loss function can take different forms. The only constraint is that it should be convex and bounded for minimizing the regret. In our case, we are seeking to predict a binary outcome so there is no issue. We use a squared loss function $\ell(\hat{y}_t, y_t) = (\hat{y}_t - y_t)^2$ (but could also use an absolute loss function $\ell(\hat{y}_t, y_t) = |\hat{y}_t - y_t)|$). Which of them is more appropriate for a given problem is an empirical question though the squared loss function tends to have better out-of-sample performance.

2.3 Time Varying Aggregation Weights

We use the Exponentially weighted average aggregation rule. This convex aggregation rules combine experts' predictions with a time-varying vector $p_t = (p_{1,t}, ..., p_{N,t})$ in a simplex $\mathcal{P}$ of $\mathbb{R}^N$:

$$\forall j \in \{1, ..., N\}, p_{j,t} \geq 0 \text{ et } \sum_{k=1}^{N} p_{k,t} = 1$$

We use the exponentially weighted average (EWA) aggregation rule as it presents key advantages. First, the weights are computable in a simple incremental way. Second, the forecaster's predicted probability only depends on the past performance of the experts and not on his past prediction. The forecaster predicts at each time t :

$$\hat{y}_t = \frac{\sum_{j=1}^{N} e^{-\eta_t L_{j,t-1}} f_{j,t}}{\sum_{i=1}^{N} e^{-\eta_t L_{i,t-1}}}$$

where η_t is the learning rate, the speed at which weights are updated.

[2]The bound of the regret guarantees that forecasters performance will compete with the performance of the best convex combination of experts when T goes to ∞. The forecaster's strategy is often worse than the performance of the best convex combination of experts since the best convex combination is known *ex post*, but it is not a theoretical necessity. With time-varying weights, an excellent online strategy could be able to beat the best (fixed) convex combination of experts.

We use the gradient-based version of the EWA aggregation rule $\mathcal{E}_{\eta}^{\text{grad}}$ where weights are defined by :

$$p_{j,t} = \frac{\exp(-\eta_t \sum_{s=1}^{t-1} \tilde{L_{j,s}})}{\sum_{k=1}^{N} \exp(-\eta_t \sum_{s=1}^{t-1} \tilde{L_{k,s}})}$$

where $\tilde{L_{j,s}} = \nabla \ell(\sum_{k=1}^{N} p_{k,s} f_{k,s}, y_s) \cdot f_{j,s}$ and ∇ is the gradient operator.

An important advantage of the gradient-based version of the EWA aggregation rule is that weights are easy to interpret. If expert j's advice $f_{j,s}$ points in the direction of the largest increase of the loss function, i.e. if the inner products $\nabla \ell(\sum_{k=1}^{N} p_{k,s} f_{k,s}, y_s) \cdot f_{j,s}$ has been large in the past, the weight assigned to expert j will be small. The learning rate is optimised empirically. The algorithm is described in more detail in Appendix.

3 Data on Macro-History and Models of Crises

We use the macro-history data set of Jordà et al. (2017) which covers the period 1870-2016 for seventeen countries. The data are annual. Given our application which requires our algorithm to learn at the beginning of the sample we can only use a limited set of variables as we cannot incorporate those with missing values. We are therefore left with variables describing real economic activity such as Real GDP per capita, GDP, Consumer prices, Investment-to-GDP ratio, government expenditure; monetary aggregates: Broad money, Narrow money; lending and borrowing variables: loans, mortgages, Debt-to-GDP; asset prices: short-term interest rate, long-term interest rate, stock prices, house prices; the external sector: exports, current account, imports, exchange rate. We also use the crisis variables of the data set and have therefore a total of eighteen variables.

To predict the 1929 crisis, our methodology requires to learn on at least two other crises before. We therefore focus on countries which satisfy two criteria : i) it has two crises before 1929 ii) there is enough data so that it is possible to compute each expert.

In our database, six countries satisfy this requirement: France, Germany, Japan, Netherlands, Spain and the United States. The batch samples, i.e. the samples on which the learning takes place are indicated in Table 1. They start in 1896 due to data availability and end as early as 1906 for the US or as late as 1922 for Italy depending on the timing of 19th century or early 20th century crises. From 1922 at the latest, we predict financial crises out-of-sample three years ahead for all the seven countries.

Country	Batch sample	Online sample
US	1880 - 1913	1914- 2017
France	1880-1909	1909-2017
Italy	1880-1909	1909-2017
Japan	1880-1909	1909-2017
Spain	1880-1917	1917-2017
Netherlands	1880-1909	1909-2017

Table 1: Batch and Online Samples. Samples are defined so that the batch sample contains two pre-crisis period and the online sample has enough observations according to data availability.

3.1 Models of crises (experts)

We use two kinds of experts: elastic-net logits and machine learning experts.

3.1.1 Elastic-net logits Experts

Elastic-net logits are a regularized regression method that combines linearly the penalties of the LASSO and the Ridge with certain weights[3]. They have been found to be particularly well-suited for out-of-sample forecasts. First introduced by Zou and Hastie (2005), the good performance of elastic-net penalty compared to other regularization methods has been confirmed in various applications (Mol et al. (2009); Mol et al. (2009); Destrero et al. (2009))[4]. We design the elastic-net logits by grouping variables by themes in order to ease the economic interpretation of our results[5]. Two logits describe the real economy (**lre**: Real GDP per capita (PPP), Real GDP per capita (index), Real consumption per capita (index), GDP (nominal, local currency), Consumer prices (index). and **lre2**: GDP (nominal, local currency), Consumer prices (index), Investment-to-GDP ratio, Exports (nominal, local currency); another one is about valuations (**lval** : Short-term interest rate (nominal, percent per year), Long-term interest rate (nominal, percent per year), Stock prices (nominal index), House prices (nominal index); another is about the foreign sector (**lfor** : Current account (nominal, local currency), Imports (nominal, local currency), Exports (nominal, local currency); a Logit credit (Lcr) : Total loans to non-financial private sector (nominal, local currency), Mortgage loans to

[3]For more details see Appendix

[4]This is mainly due to the fact that, because it uses a penalty that is part ℓ_1 and part ℓ_2, this procedure works almost as well as the LASSO when the LASSO does best; but it also improves on the LASSO when the LASSO is dominated by the Ridge regression. This is usually the case if there exists high correlations among predictors, as in our case when we consider a large set of macroeconomic indicators (Tibshirani (1996)). As a consequence, the elastic-net penalty outperforms LASSO while preserving the sparse property (Zou and Hastie (2005); Mol et al. (2009)).

[5]a comprehensive list of our experts is provided in appendix

non-financial private sector (nominal, local currency), Total loans to business (nominal, local currency),House prices (nominal index); a Logit housing (Lho) : Total loans to non-financial private sector (nominal, local currency), House prices (nominal index). These six logits are common to all the countries. We also introduce elastic net logit combinations which are specific for each country. The variables included are the ones which have the highest AUROCs (Area Under the Receiving Operator Curve) on the batch sample, following Schularick and Taylor (2012) and Coudert and Idier (2016). The ROC curve represents the ability of a binary classifier by plotting the true positive rate against the false positive rate for all thresholds. The closer to one an AUROC is, the more accurate the classification is. For each country we have four combination logits. Their variables, which differ across countries, are detailed in the Appendix.

3.1.2 Machine Learning Experts

For each country we also include one Random-forest expert (**Rf**) with country-specific variables. We also include a General Additive Model (**GAM**) whose country specific variables are selected according to the same AUROC criteria as above (see Appendix for details on the machine learning experts). Our models incorporate various horizons of changes for the variables so that inflexion points can be captured. We have a total of twelve experts for each country. There is no a priori optimal number of experts and ultimately the judge is the quality of the out-of- sample forecast. This also means that, like open source software, our framework can be improved upon by researchers wishing to add their own favourite forecasting model in our set of models. If this is a valuable model, it will be highly weighted and this can only improve the forecast and

4 A Tale of two Centuries

Predicting financial crises from the 1920s out-of-sample in seven countries as diverse as France, Germany, Japan, Netherlands, Spain and the United States on a long period of time may seem like a hopeless enterprise, an economist's folly. Models are initially estimated using a limited set of variables on a short period a long time ago (between 1874 at the earliest and 1922 at the latest depending on data availability). Each country has different institutions and macroeconomic policies and crises have struck at very different points in time. Japan for example had crises in 1871, 1890, 1907, 1920, 1927 and 1997 while the US had crises in 1873, 1893, 1907, 1929, 1984 and 2007. A striking feature of the timing of crises is that many of them occurred at the end of the 19th century and until the 1930s or after 1990 but none during the post world war II period until the 1980s. Eichengreen and Portes (1987)

have noted the similarities between the 1930s and the 1980s in their work on the anatomy of financial crises. It will be important therefore to be able to predict the post war tranquil period as well as the sudden resurgence of financial instability in the US and in Japan and in Europe in the 80s and the 90s.

There is a first "global wave" of systemic crises in the 1930s as most of the countries in our sample had a financial crisis around then (US 1929, Spain 1931, Japan 1927, Italy 1930, France 1930; only the Netherlands experience a crisis much later in 1939). According to Eichengreen and Portes (1987) the 1920s were marked by two changes which increased the international financial system's susceptibility to destabilizing shocks: the end of the classical gold standard and the associated turmoils in the foreign exchange market with a "gold shortage" and the misalignment of currencies in the 1920s -the pound sterling in particular was largely overvalued; a rapid institutional change in the banking system as the US replaced the UK as the main external creditor with dramatic shifts in the volume and direction of international lending and rapidly mounting indebtedness in the periphery, in particular Central Europe and Latin America. US investments to the Far East were also important. Much of the business going through relatively inexperienced American hands may have increased the market's tendency to excessive risk taking. Each of these developments had its immediate origins in the dislocations associated with World War I. Massive capital flows out of periphery economies took place as the US stock market boomed. After peaking in the summer of 1928, capital flows fell by 46 per cent within a year. Then commodity exports declined abruptly following the U.S. cyclical downturn commencing in the summer of 1929. Even after suspending convertibility of their currencies, many countries found it impossible to service their external debt. The Great Depression prevented governments to generate enough tax revenues to service debt and after 1930, many Latin American and European countries defaulted. But a more serious threat to the stability of banking systems came as bank deposits started to be liquidated by foreign investors in order to repatriate liquidity to their domestic economies. The run on the Austrian Credit Anstalt in May 1931 created more turmoil. The US banking system suffered considerably unlike the UK banking system which turned out to be resilient despite a run on Sterling in 1931. If 1929 is the base at 100, UK bank shares were at 92 in 1930 and 104 in 1933 while US bank shares were at 67 in 1930 and 15 in 1933 (from the League of Nations 1934). The US banks were particularly vulnerable because they were very exposed the security market and to real estate. The Federal Reserve could not prevent widespread bank failures, nor break the doom loop between the financial markets and the real economy, which ultimately led to the Great Depression. Banking regulation, the second world war and Bretton Woods transformed the financial system and redrew the international monetary order. The creation of the IMF, a more regulated banking system and

restrictions to international capital mobility created the conditions of a quiet period between 1945 and the 1980s at least as far financial crises are concerned.

After that tranquil post world war II period, crises started to appear again after the collapse of Bretton Woods, just like the collapse of the classical gold standard preceded the Great Depression. Increased capital mobility is likely to have played a major role. Crises resumed in the 1980s and 1990s but, unlike in 1929, their timing was idiosyncratic across countries. The Latin American debt crisis unfolded in the 1980s after the second oil shock and after the US real rate went through the roof, rendering the cost of servicing external debt in dollars really high. The US experienced the savings and loans crisis in 1985 while Japan experienced a systemic financial crisis in 1997 after the bursting of the real estate bubble at the beginning of the 1990s. Spain had a crisis in 1977 but not Portugal while Italy experienced a crisis in 1990 linked to an economic slowdown and a fragile banking system. Thus it will also be important for our algorithm to predict the idiosyncratic timing of the crisis in the different countries after the first "global wave" of financial crises around 1929 and the quiet Bretton Woods period.

The second "global wave of crises" occurred around 2008 (US 2007, France 2008, Italy 2008, Netherlands 2008, Spain 2008; only Japan is spared). It can be traced to the collapse of Lehman Brothers and the high degree of leverage in the banking system in the US and Europe due to lax regulation and cheap liquidity. Real estate bubbles in a number of countries (US and Spain in particular) made the financial system more fragile as well while highly indebted households amplified the economic downturns. That wave of crises, just like the 1930s, seem to have been caused by an unchecked credit boom among euphoric valuations in real estate markets and securities market more generally at a time of financial innovation and deregulation.

4.1 Countries experiences

We present the probability of systemic crises predicted out-of-sample from the beginning of the 1920s for our seven countries. We start with the United States as it has been the hegemon of the global financial system during a large part of the 20th and 21st centuries (see Gourinchas and Rey (2007a) and Gourinchas and Rey (2007b)) and it is at the centre of the Great Depression.

United States

Figure 1 shows our estimated probability of systemic financial pre-crisis *out-of-sample* in the US between 1914 and 2017, i.e. for *more than a century*. The purple areas are the year of

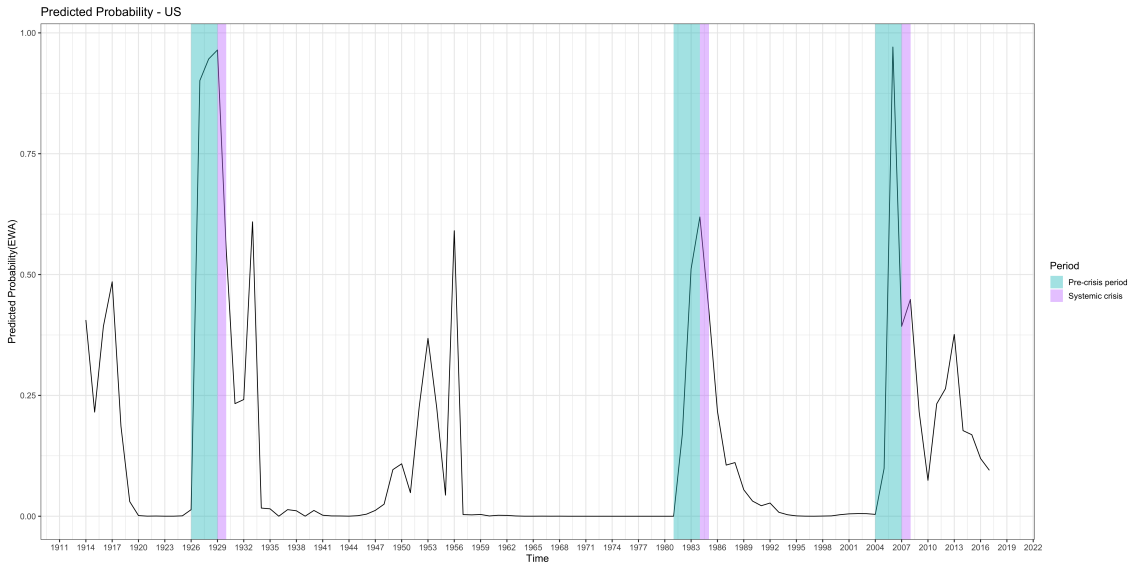


Figure 1: United States

systemic financial crises according to the macro history data set and the blue bars are the pre-crisis periods (three year before the crises). We are attempting to predict the blue bars: a perfect outcome would be to predict a probability close to 1 for the blue bars and close to 0 everywhere else (including during the actual crisis –purple bars). The results are very good: the probability spikes very steeply in 1926 to 1929; (there is a smaller spike as well in 1917). There are afterwards smaller alerts in 1933 and 1953 and more significant one in 1956. Remarkably, the probability of a pre-crisis is very low from 1956 till 1981 where its spikes again between 1981 and 1984 just before the Savings and Loans crisis. After coming down very significantly it then spikes massively between 2004 and 2007. There is then a smaller signal during the euro area crisis between 2011 and 2013. We show the RMSE and the AUROCs on the entire forecasting periods in Table 2. The aggregation strategy works well as our estimate is even better than the best convex combination of experts (known ex post). In theory the two converge asymptotically, but in sample, given that our EWA aggregation allows for time varying weights (while the best convex combination, although known ex post, has fixed weights, it is possible that EWA outperforms. It does better than the uniform aggregation as far as RMSE are concerned. The AUROC is close to 1 at 0.93 indicating that this methodology performs very well in terms of identifying crises and not having many false positives. By construction, our methodology outperforms all the papers in the literature (as if there is a better model, we can just throw it in our set and it would be picked by the algorithm). The excellent statistics show this is empirically the case. By plotting our estimator, we can also see that it provides a very reliable estimate of crises three

	PEWA	Best_convex	Uniform
AUROC	0.93	0.93	0.94
RMSE	0.22	0.24	0.24

Table 2: US

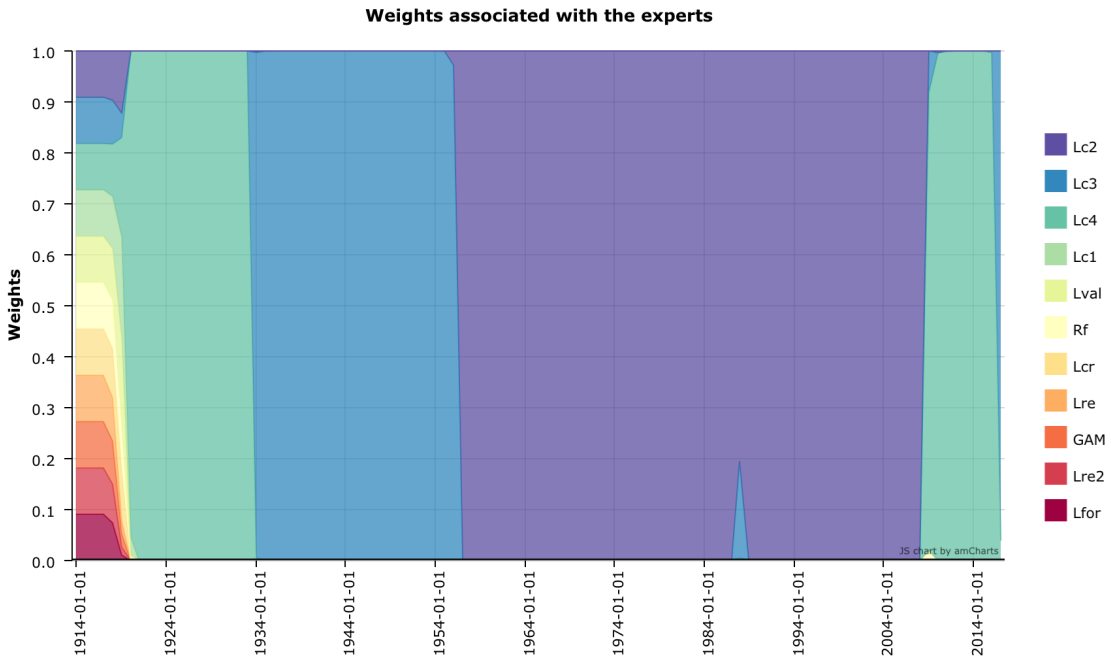


Figure 2: United States

year ahead, out-of-sample and that it is of value for policy makers[6]. In our case , probabilities increase sizeably at the right time and we predict 3 year ahead, not next days's crisis.

Figure 2 presents the weights carried by each model in the forecast. The higher the weights the more informative the model is to predict crisis. There is no causal implications but knowing which set of variables carries more weight may be informative to understand where imbalances in the financial sector lie and do further investigations. In the case of the US, **lc4**[7], **lc3**[8], **lc2**[9] seem the most important models. **lc4** appears both in the 1920s and in

[6]reporting an AUROC relatively close to 1 but not plotting the probability of crisis, as is often done in the literature is no guarantee that the forecast is of value for policy makers as it may give very weak signal (very small increase in probability at the "right time")

[7]**USlc4**'s variables are Long-term interest rate (nominal, percent per year), Mortgage loans to non-financial private sector (nominal, local currency), Real GDP per capita, Consumer prices.

[8]**US lc3**'s variables are Public debt-to-GDP ratio, Total loans to non-financial private sector (nominal, local currency), Stock prices (nominal index),Mortgage loans to non-financial private sector (nominal, local currency).

[9]**US lc2**'s variables are Broad money (nominal, local currency), Stock prices (nominal index), Government expenditure (nominal, local currency), Total loans to non-financial private sector (nominal, local currency).

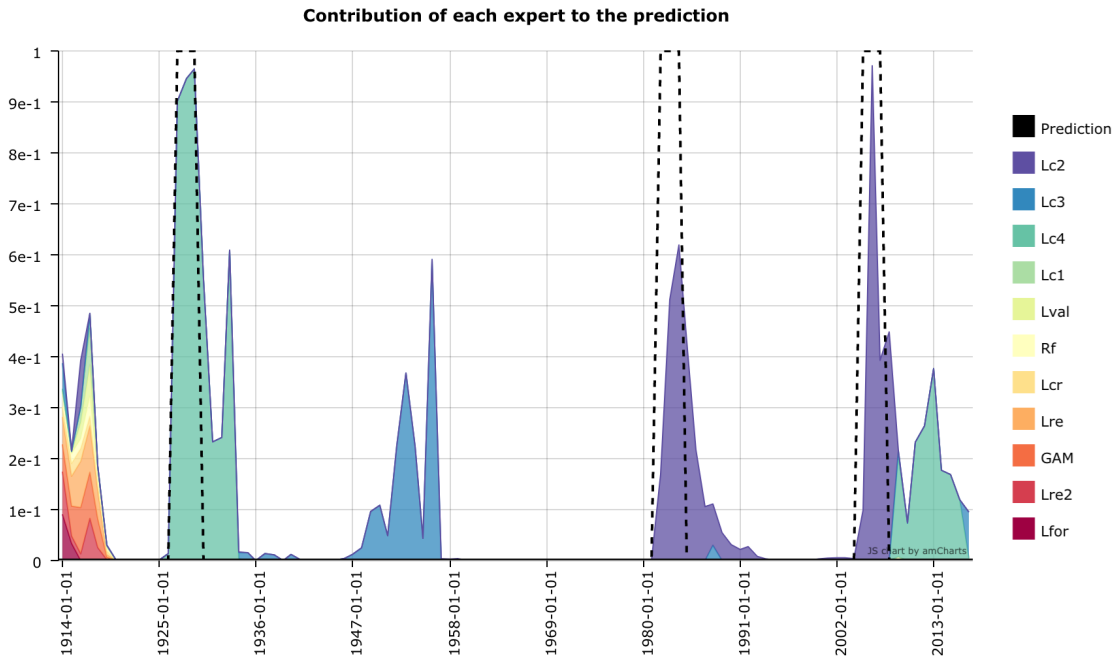


Figure 3: United States

the 2000s while **lc2** is dominant between the end of the 1950s and the 2000s. **lc3** is picked after the Great Depression and until 1956. It is dominant during the period of the second world war. In Figure 3, where the pre-crisis periods are shown in dashed lines, we show which models give the signal of the crisis three years ahead. For the 1929 crisis the dominant model is **lc4** underlining the importance of interest rates, credit (mortgages), inflation and the real economy; while for the 1984 and the 2007 crisis it is **lc2** which gives the signal, emphasizing the combined importance of valuations (stock markets) and credit (loans and broad money). During the euro area crisis, the best model reverts to **lc4**. For the 1950s the domination of **lc3** shows that the debt to GDP ratio may have played a role in the false alarm in 1956. These subsets of variables fit reasonably well with the ex post narrative of the 1929 crisis (credit boom and cyclical downturn), of the 1984 crisis (savings and loans debacle) and of the 2007 crisis (exuberant valuations and central role of leverage and credit).

In Figures 4 and 5 we show that the average loss and the dynamic loss suffered by our aggregation rule EWA is lower than the loss of any of our experts. It is also lower than the loss of a uniform aggregation rule. This shows that the our algorithm is able to improve on the forecasting ability of any of our experts, by varying weights across time periods. This also shows that the early warning literature emphasising specific subsets of variables (credit to GDP, asset prices or their combination) fails to incorporate a wealth of relevant information. As a result, our economic narratives of crises may be too simplistic. We now explore



Figure 4: United States

	PEWA	Best_convex	Uniform
AUROC	0.93	0.94	0.94
RMSE	0.22	0.20	0.24

Table 3: France

the other countries of the dataset. Among European countries France and the Netherlands have only one post world war II crisis (2008) according to the macro history data set.

France

For France, the crisis linked to the Great Depression, occurred in 1930; the subsequent crisis was in 2008.

Figure 6 shows a very good ability of the on-line learning framework to predict the 1930 and 2008 crisis. The probability of crisis does not go down quickly after 2008 and it is still elevated in 2012 and 2013. This can be linked to the euro area crisis. There is a small spike in 1984 (false alarm) which may have been linked to the second oil shock and the large external deficit following major changes in economic policies; this also coincides with the US savings and loans crisis. We have a large spike in 1939, at the beginning of the second world war. Remarkably, the probability of pre-crisis is very close to zero between 1945 and 2006, which coincides with the relatively quiet period of Bretton Woods and flexible exchange rate regime post Bretton Woods for advanced economies.

The AUROC and the RMSE are also excellent for France as shown in Table 3. In particular, the RMSE is small though there is still margin for improvement in terms of the aggregation algorithm as the best convex combination RMSE is still a bit lower. Just like for the US , the AUROC performance and the RMSE outperform the literature. For France and the subsequent countries, graphs of the average losses are relegated into the appendix.

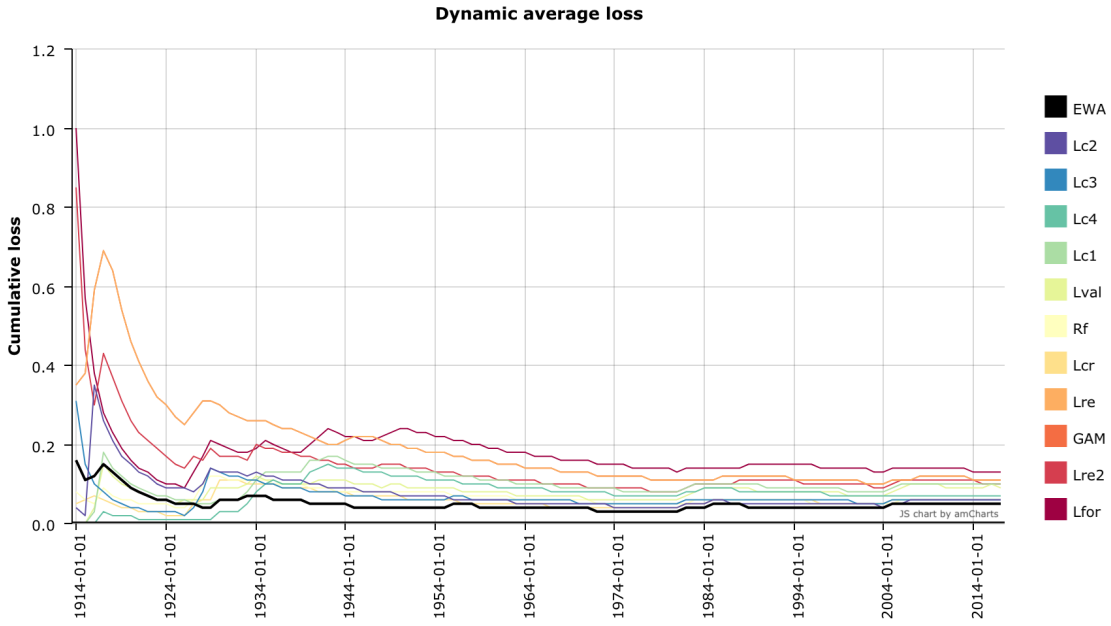


Figure 5: United States

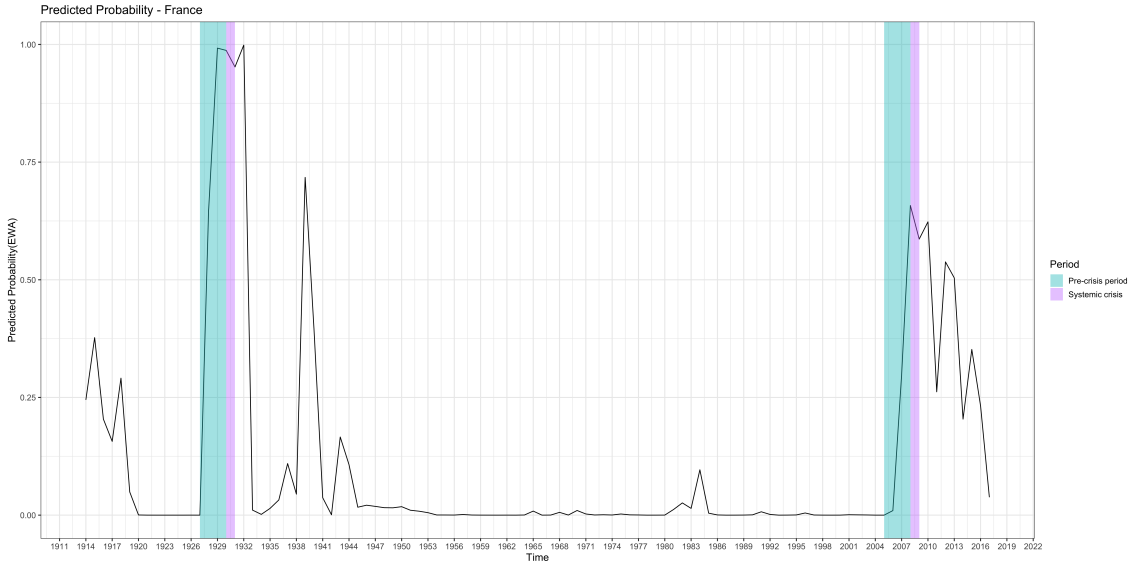


Figure 6: France

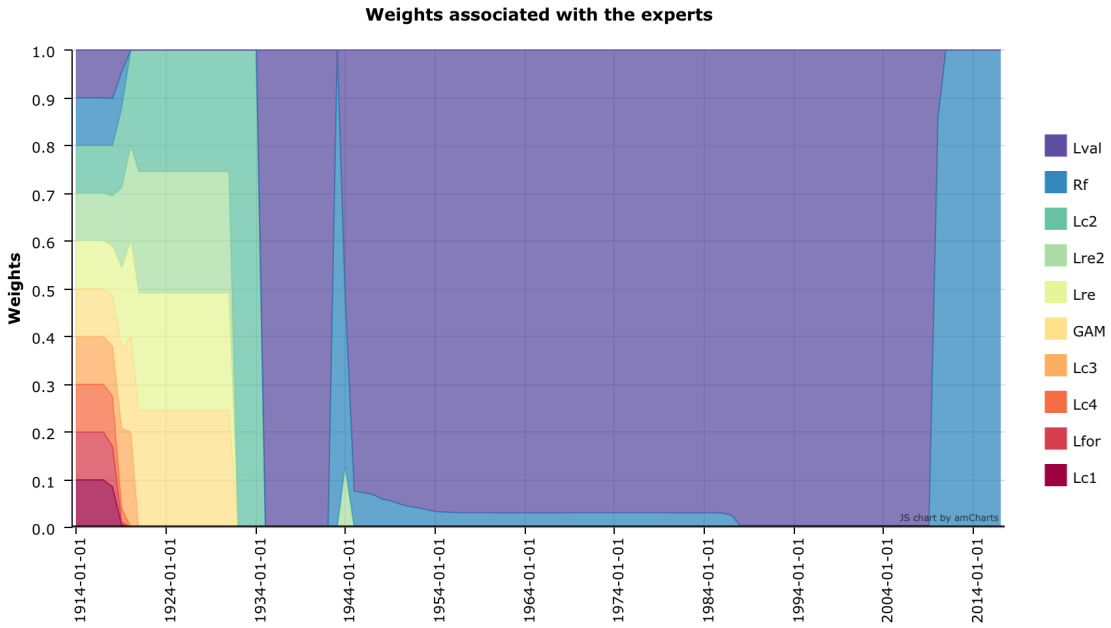


Figure 7: France

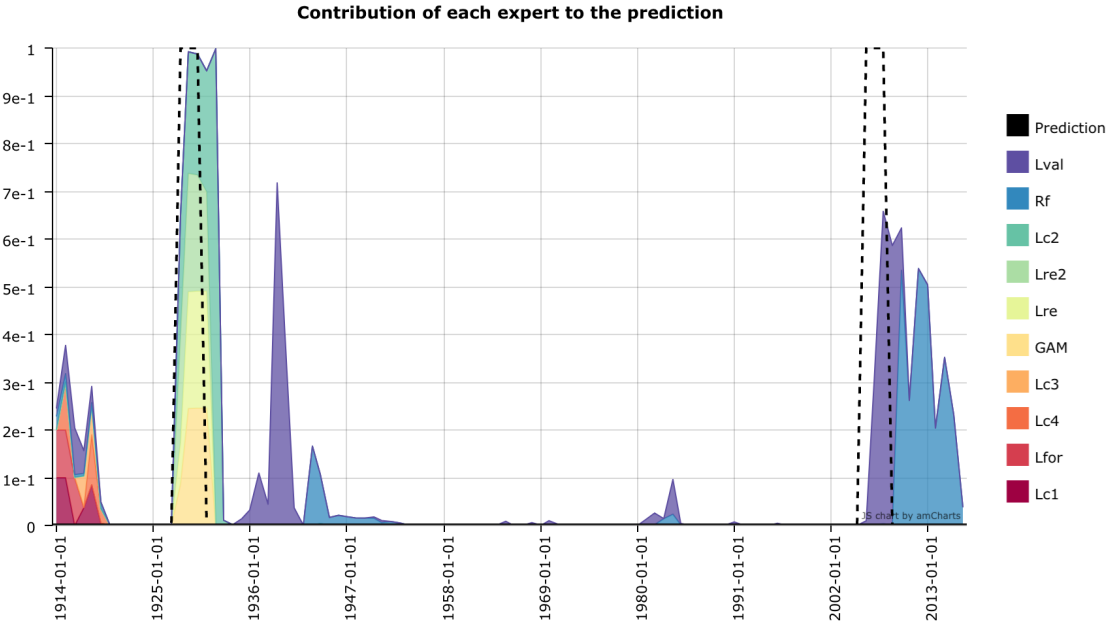


Figure 8: France

	PEWA	Best_convex	Uniform
AUROC	0.85	0.90	0.81
RMSE	0.27	0.24	0.29

Table 4: Netherlands

Figure 7 presents the most important models for the forecast and Figure 8 shows their contribution in terms of giving the signal of a crisis three year-ahead. For 1930, the real economy logits **lre2**[10]; **lre**[11], the **GAM** model [12] and **lc2** [13] are dominant. **lc2** becomes increasingly important. The most important warning in 1930 comes therefore from real economic activity, investment inflation and the external sector as experts containing external variables are systematically picked (exports, imports, exchange rate, current account). The stock market, government expenditure and house prices are also present to a lesser degree. For 2008, **lval**[14], gives the signal. This makes the 2008 crisis less linked to the real and the external sector variables and more linked to valuation variables (short and long interest rates, stock prices, house prices) compared to the 1930s. In both crises, the real estate sector has played a role.

The euro area crisis is mostly predicted by the **Rf** which uses a number of different variables including real GDP per capita, house prices, government revenues and stock prices.

The Netherlands

The Netherlands experienced crises in 1921, 1939 and 2008. Our methodology gives an increase of crisis probability in 1914 (first world war) before 1921 but the spike is relatively small and there are some false positive around 1929-1933. But it picks up really well the 1939 crisis as well as the 2008 crisis. The probability remains somewhat elevated during the euro area crisis in 2011-2012.

The diagnostics (AUROC and RMSE) are very good as shown in Table 4. In that case, like for France, we seem to have some margin of improvement to approximate the best convex combination of experts ex ante as the RMSE of the EWA is at 0.28 above its theoretical asymptotic value of 0.23. We do better than a uniform aggregator.

[10] **lre2**: GDP (nominal, local currency), Consumer prices (index), Investment-to-GDP ratio, Exports (nominal, local currency).

[11] **lre**: Real GDP per capita (PPP), Real GDP per capita (index), Real consumption per capita (index), GDP (nominal, local currency), Consumer prices (index)

[12] **FranceGAM**: Real GDP per capita, GDP, Exchange Rate.

[13] **lc2**: House prices (nominal index, 1990=100), Real GDP per capita (PPP), GDP (nominal, local currency), Stock prices (nominal index), Consumer prices (index, 1990=100))

[14] **Lval**: Short-term interest rate, Long-term interest rate, Stock prices, House prices.

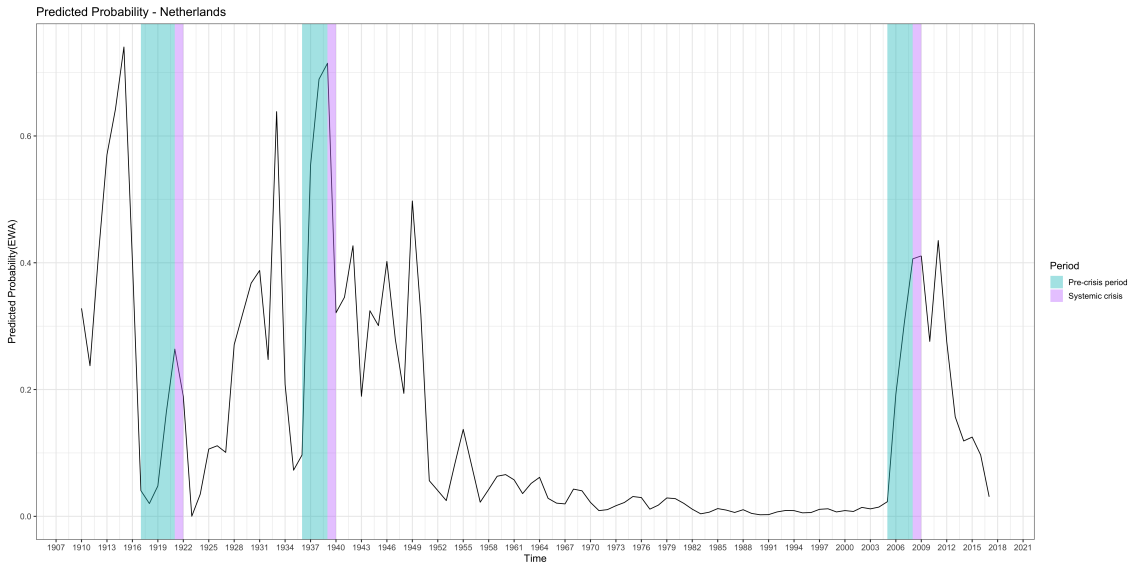


Figure 9: Netherlands

Figure 10 shows the weights of the different models while Figure 23 indicates which models have given the signal.

For 1921,1939 and 2008, the model **Lval**[15] is systematically flashing red. The 1921 crisis was in part due to the collapse of the German economy, a close trading partner. In 1939 and 2008 the random forest **Rf**. also contributes through the variables house prices, CPI, long term rate, government expenditures.

1939 is jointly predicted by **Lval**, **Rf** and **Lm**[16]. These variables reflect the economic shock of the Great Depression which affected the Netherlands a bit later than most countries.

The false positives around 1929 are mostly due to **GAM** (Real GDP per capita, House prices, Government expenditure). The false positives in the 1940s are due to **Lval** and **Lm**.

Spain

Spain's crisis pattern is different from France and Netherlands as it had two crises after the second world war: besides 2008, it also experienced a crisis in 1977. So, in our forecasting sample, Spain had crises in 1924, 1931, 1977 and 2008. The AUROCs and RMSE are shown in Table 5. As in the other cases, they show a high signal to noise ratio. In this particular case , uniform aggregation does about as well as our EWA rule.

The 1931 crisis is linked to the Great Depression. According to our algorithm, it is mostly

[15] **Lval**: Short-term interest rate, Long-term interest rate, Stock prices, House prices.
[16] **Lm**: Narrow Money, Broad Money, Short-term interest rate, Long-term interest rate.

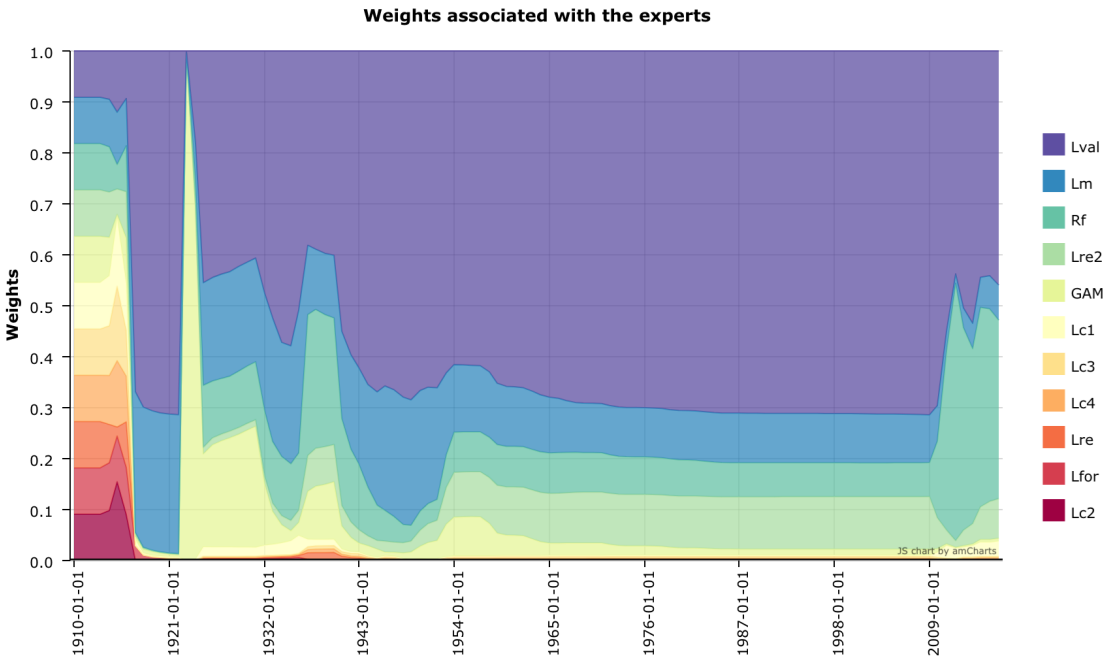


Figure 10: Netherlands

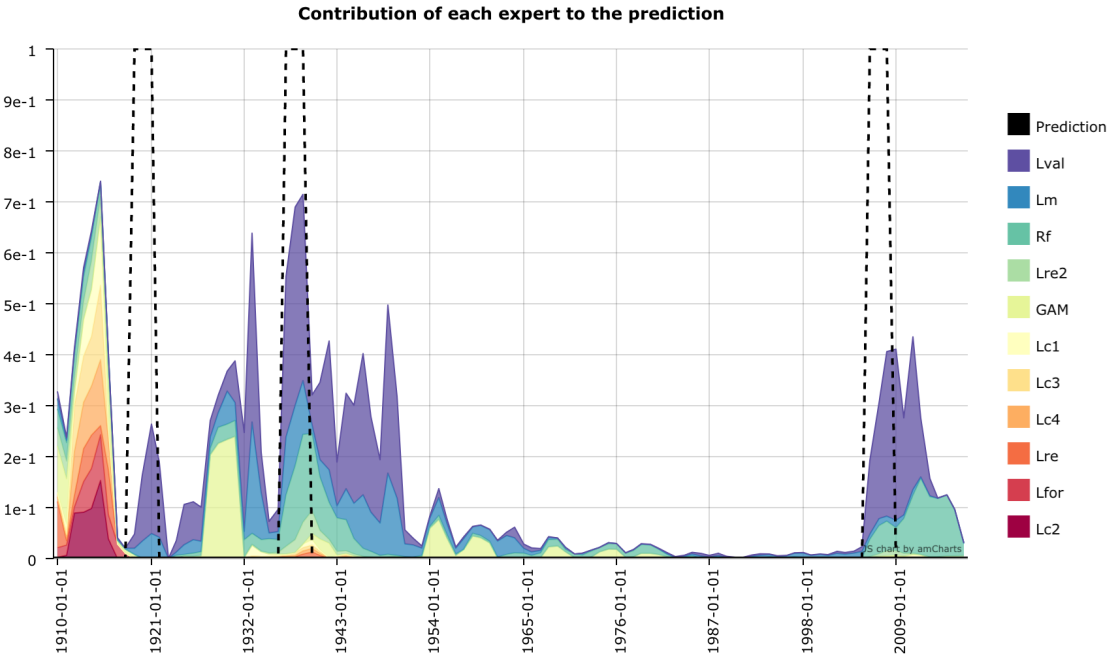


Figure 11: Netherlands

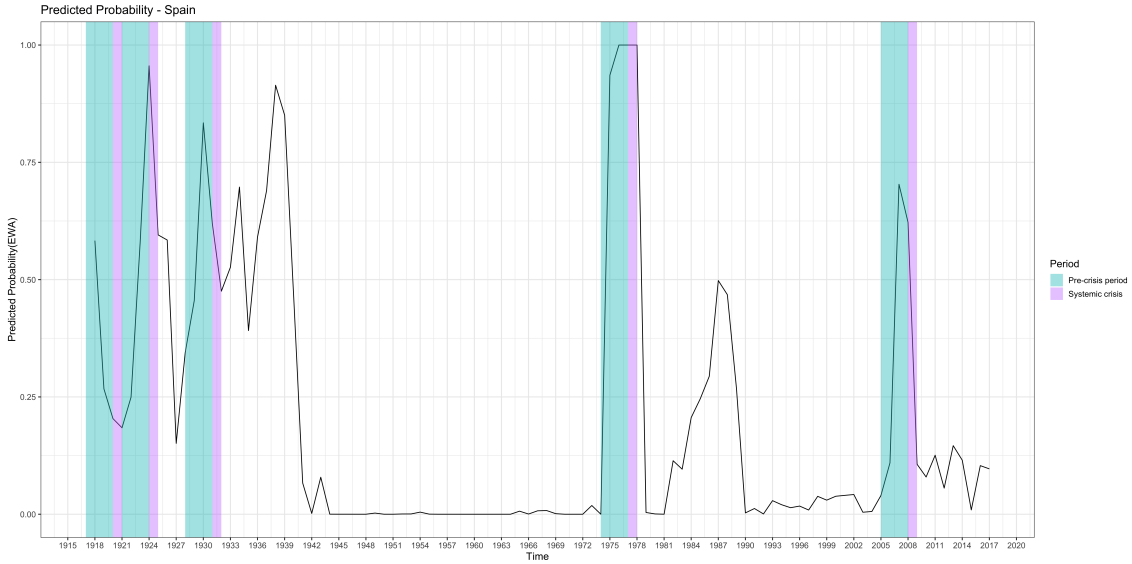


Figure 12: Spain

	PEWA	Best_convex	Uniform
AUROC	0.90	0.93	0.90
RMSE	0.30	0.30	0.30

Table 5: Spain

predicted by the expert **Lre**[17]. The 1977 crisis was caused by a deterioration of the current account due to the first oil shocks, a drop in investment and an exchange rate crisis with a big devaluation of the peseta. Interestingly, the expert picked by our algorithm ex ante is **Lc4**[18] which fits perfectly the ex post narrative of the crisis. The 2008 crisis is also signalled by the **Lc4** model.

Italy

Italy, like Spain had 4 crises in our forecasting sample out of which two are post world war 2. Italy had systemic crises in 1930, 1935, 1990 and 2008.

Just like for Spain we are able to forecast the first post war crisis even though it occurs at a very different times from Spain (1990 versus 1977). The diagnoses are again excellent with our aggregation rule having the same level of RMSE as the best convex combination (see 6), although with a somewhat lower AUROC. Though we get some false positives during the

[17] **Lre**:.Real GDP per capita (PPP), Real GDP per capita (index), Real consumption per capita (index), GDP (nominal, local currency), Consumer prices (index)

[18] **Lc4**: Real GDP per capita (index, 2005=100), Imports (nominal, local currency), USD exchange rate (local currency/USD), nvestment-to-GDP ratio..

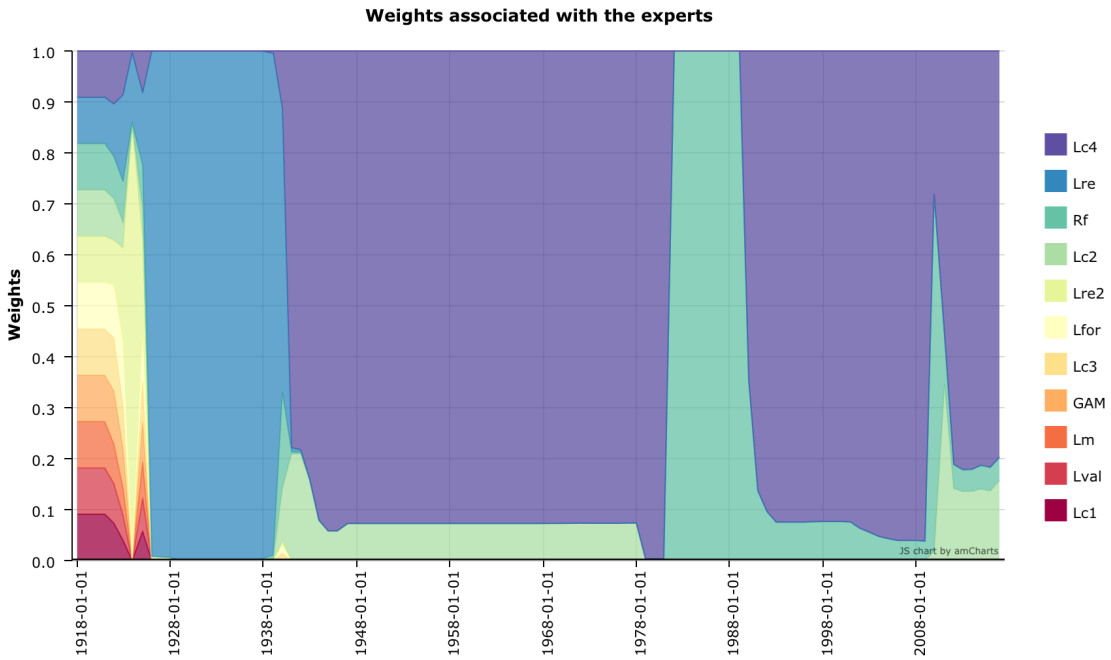


Figure 13: Spain

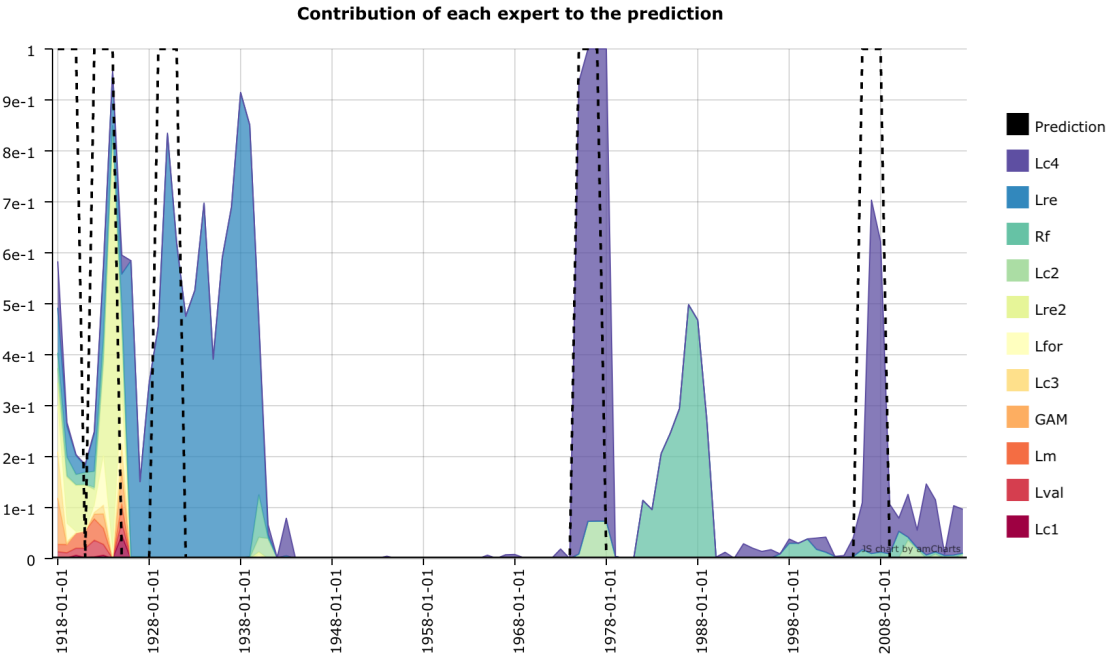


Figure 14: Spain

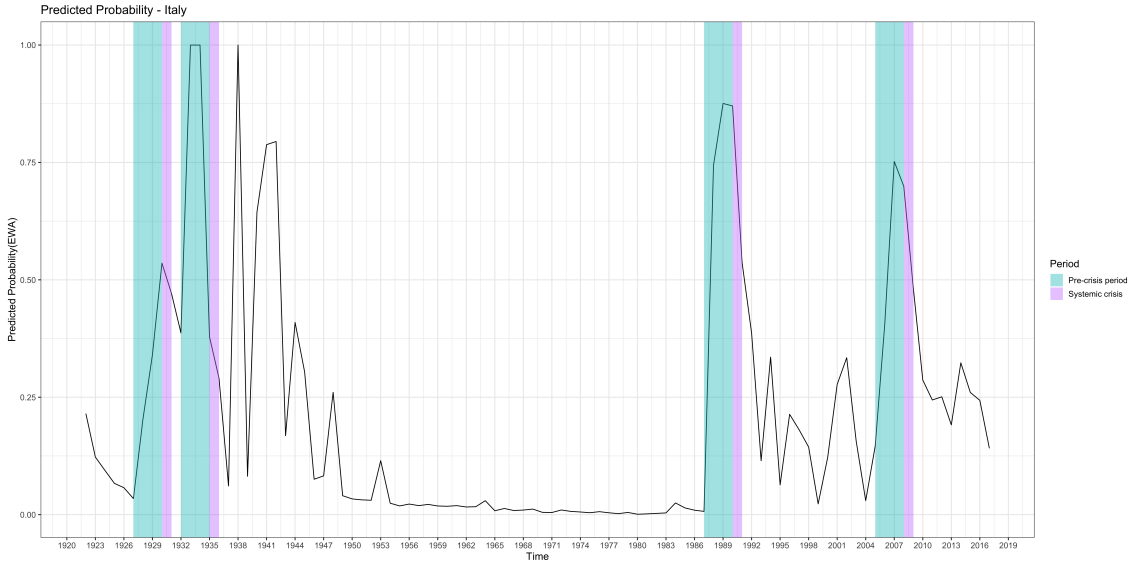


Figure 15: Italy

	PEWA	Best_convex	Uniform
AUROC	0.93	0.96	0.91
RMSE	0.28	0.28	0.29

Table 6: Italy

euro area crisis time (which is not counted as a systemic crisis) in the macro history data set and in the 1930s.

The model which predict 1930 (crisis linked to the Great Depression) and 1935 is mostly **Lc3**[19]. A false alarm in 1938 corresponds to **lfor**[20]

For the 1990 crisis, the models which contains the most information are again **Lc3** and also **Lc1**[21]. The 1990 crisis is Italy was due to the distress of many small banks especially in the South in a context of slow economic growth. For the 2008 crisis it is **Lc1** mostly but also **Lc3** textbfRf who give the signals and keep the probability of a crisis relatively elevated during the euro area crisis.

[19]**ItalyLc3**: Stock prices (nominal index), Real GDP per capita (PPP), USD exchange rate (local currency/USD). Narrow money (nominal, local currency).

[20]Current account (nominal, local currency), Imports (nominal, local currency), Exports (nominal, local currency).

[21]**ItalyLc1**: GDP (nominal, local currency), Total loans to non-financial private sector (nominal, local currency), Narrow money (nominal, local currency), USD exchange rate (local currency/USD)



Figure 16: Italy

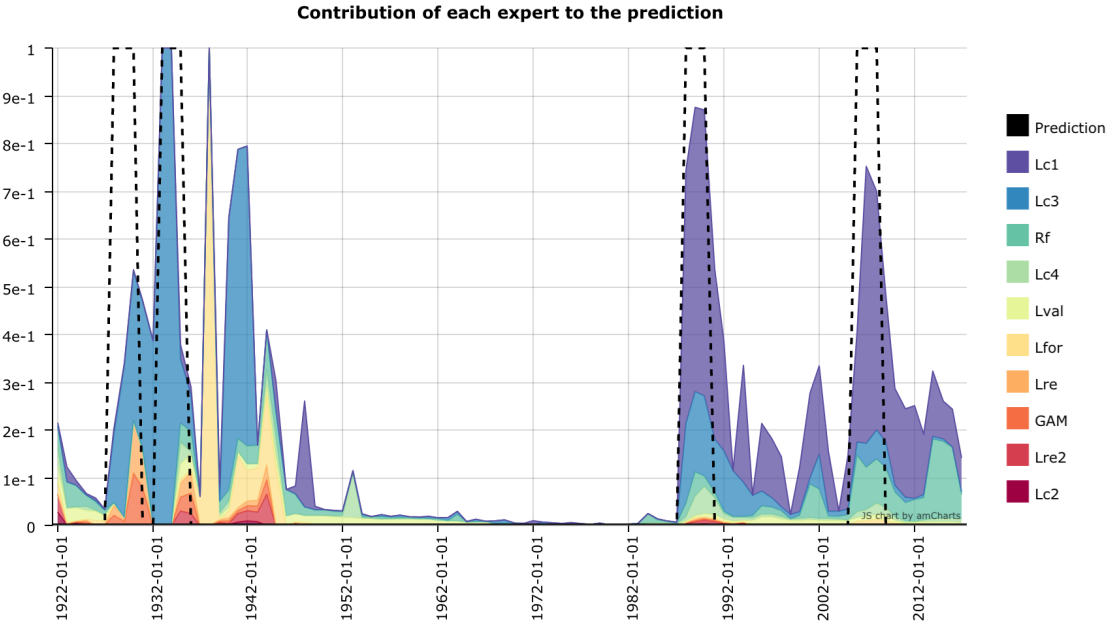


Figure 17: Italy

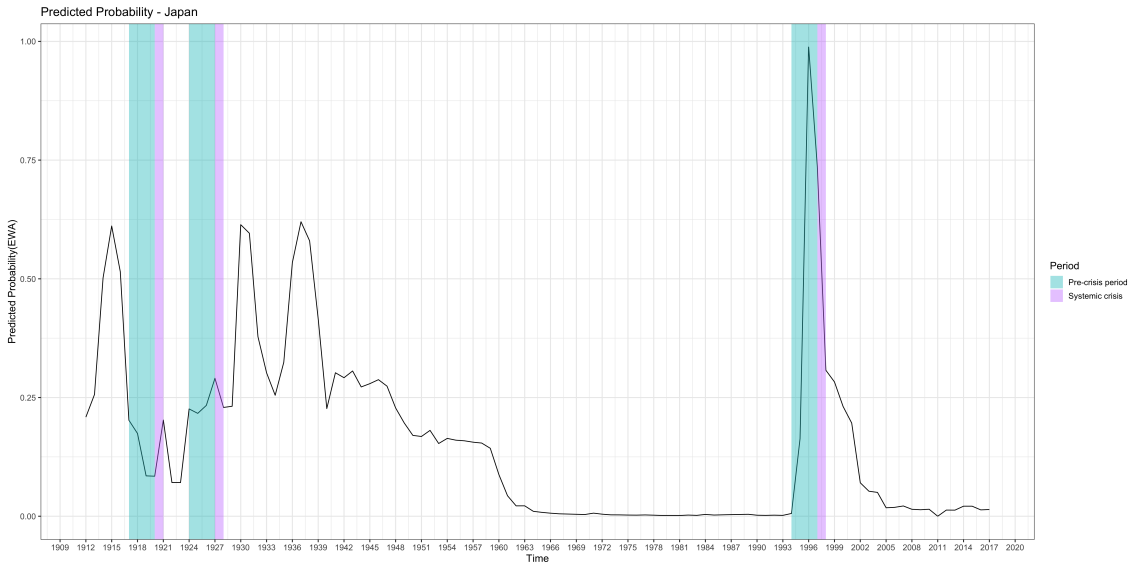


Figure 18: Japan

	PEWA		Best_convex	Y
AUROC	0.74	0.86	0.79	
RMSE	0.30	0.20	0.29	

Table 7: Japan

Japan

The final country we study is Japan, which is a very interesting case as it is well known that the burst of the land price bubble in Japan around 1990 led to to a very prolonged period of low growth and has been seen by many as very comparable to 2008. But Japan was very idiosyncratic in the timing of its crises. On our forecasting sample, the Japanese crises are in 1921, 1928 and 1998. Our methodology works relatively well towards the end of the sample but misses crises at the beginning. Our diagnosis are less strong than for the other countries as documented in Table 7.

The models tend to predict numerous crises in 1915, 1930 and 1939 so the performance at the beginning of our sample is not perfect. But after predicting low crisis probability post war, the probability of crisis picks up significantly in 1995. The most important model is **Lc3**[22] This is consistent with a crisis due to a real estate bubble bursting and affecting the financial sector and he real economy.

[22] **Lc3**: Long-term interest rate (nominal, percent per year), Government expenditure (nominal, local currency), House prices (nominal index, 1990=100), Real GDP per capita (index, 2005=100)..

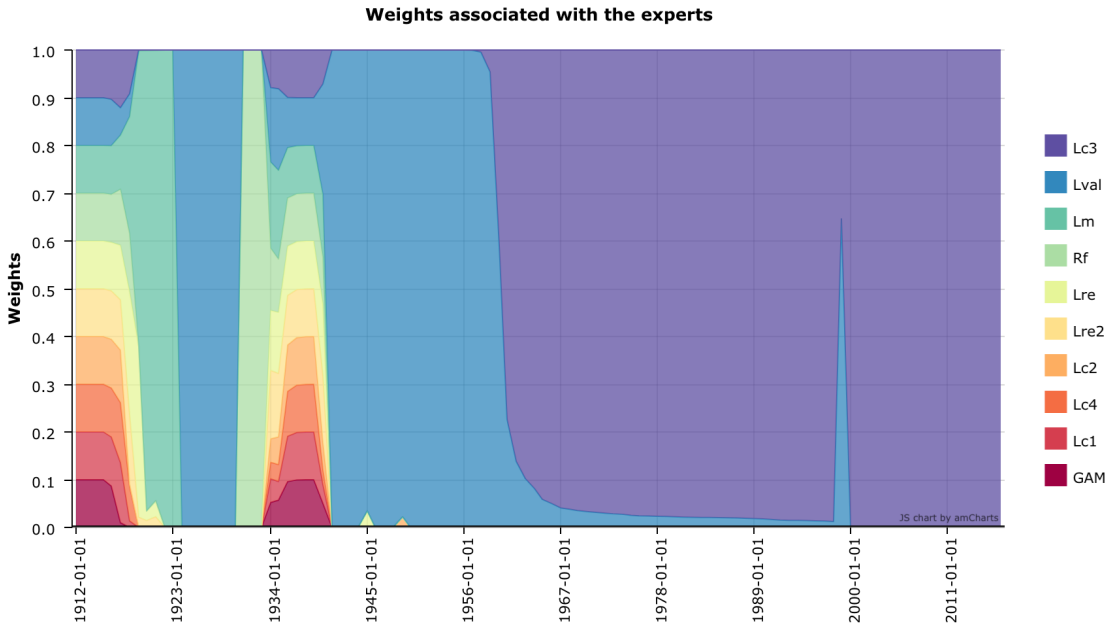


Figure 19: Japan

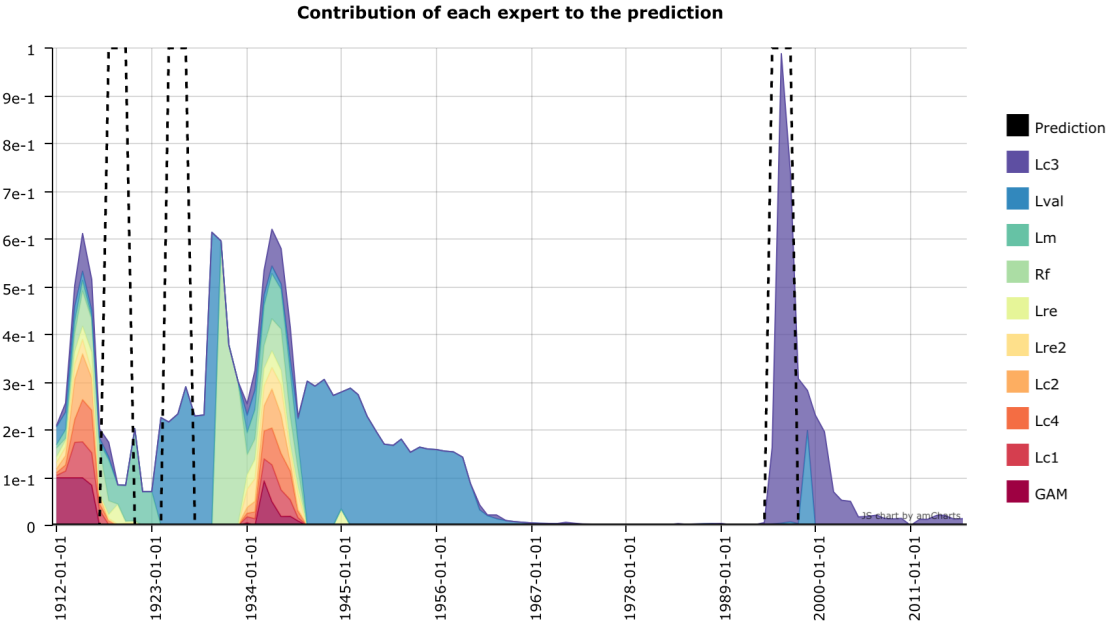


Figure 20: Japan

5 Conclusions

We find that, surprisingly, optimally aggregating limited sets of models based on conventional historical macroeconomic variables enables us to predict out-of-sample systemic crises three years ahead from the Great Depression to the Great Recession. AUROCs and RMSEs vastly outperform the literature. Moreover, these results hold for countries as diverse as the US, Japan and European countries including France and Italy. They hold even though those economies have faced idiosyncratic timings for some systemic crises (Japan 1997, Spain 1977, US 1984 for example). Our estimated probabilities provide very clear signals for policy makers and we do not over predict many crises. It is remarkable for example that our crisis probabilities are usually extremely low during the Bretton Woods period, as in the data. Hence Reinhart and Rogoff are right: there is enough information in the 1896-1922 crises to predict the Great Depression and the North Atlantic financial crisis of 2008 out-of-sample in all countries, as well as more country specific crises with the same information. There are few false positives but often they correspond to turbulent times: world wars or the euro area crisis (which is not counted as a systemic crisis). Interestingly we have no false positive during the entire period post war up to the 1980s, which indeed was a very tranquil period as far as financial crises are confirmed. What is also true is that the variables and models that are key to give an early warning are not necessarily the same in the 1930s and in the 1990s-2000s. But we have enough information to pin them down. The variables selected ex ante by our algorithm tend to fit the ex post narratives of the crises quite well. The set of variables we uncover is, despite obvious data limitations linked to the historical nature of our sample, wider than what can be found in the existing literature. We show in particular that credit growth and asset prices are outperformed by a broader set of variables as far as their forecasting ability is concerned. In particular, it is often very important to complement them with housing market variables and with real economy, exchange rate and current account variables. Our methodology is oecumenical as it can incorporate any model or human judgement. It can be applied in many different context, including for continuous variables. It is also open source as we invite anyone with a good candidate expert to throw it in our mix of models. If it is a good predictor, it will be selected and will further improve our forecasts.

References

Amat, Christophe, Tomasz Michalski, and Gilles Stoltz (2018) "Fundamentals and exchange rate forecastability with simple machine learning methods," *Journal of International Money and Finance*, Vol. 88, pp. 1–24.

Bluwstein, Kristina, Marcus Buckmann, Andreas Joseph, Miao Kang, Sujit Kapadia, and Özgür Simsek (2020) "Credit growth, the yield curve and financial crisis prediction: evidence from a machine learning approach," Bank of England Working Paper 848, National Bureau of Economic Research.

Cesa-Bianchi, Nicolo and Gabor Lugosi (2006) *Prediction, Learning and Games*: Cambridge University Press.

Coimbra, Nuno and Hélène Rey (2017) "Financial Cycles with Heterogeneous Intermediaries," NBER Working Papers 23245, National Bureau of Economic Research, Inc.

Coudert, Virginie and Julien Idier (2016) "An Early Warning System for Macro-prudential Policy in France."

Dani, Varsha, Omid Madani, David M Pennock, Sumit Sanghai, and Brian Galebach (2012) "An empirical comparison of algorithms for aggregating expert predictions," *arXiv preprint arXiv:1206.6814*.

Dashevskiy, Mikhail and Zhiyuan Luo (2011) "Time series prediction with performance guarantee," *IET communications*, Vol. 5, No. 8, pp. 1044–1051.

Destrero, Augusto, Christine Mol, Francesca Odone, and Alessandro Verri (2009) "A Regularized Framework for Feature Selection in Face Detection and Authentication," *International Journal of Computer Vision*, Vol. 83, pp. 164–177, 06.

Devaine, Marie, Pierre Gaillard, Yannig Goude, and Gilles Stoltz (2013) "Forecasting electricity consumption by aggregating specialized experts," *Machine Learning*, Vol. 90, pp. 231–260, 01.

Diamond, Douglas W and Philip H Dybvig (1983) "Bank Runs, Deposit Insurance, and Liquidity," *Journal of Political Economy*, Vol. 91, No. 3, pp. 401–19, June.

Eichengreen, Barry and Richard Portes (1987) "The anatomy of financial crises."

Fisher, Irving (1933) "The Debt-Deflation Theory of Great Depressions," *Econometrica*, Vol. 1, No. 4, pp. 337–357.

Fouliard, Jérémy, Michael Howell, and Hélène Rey (2020) "Answering the Queen: Machine Learning and Financial Crises," nber working paper.

Friedman, Jerome, Trevor Hastie, and Rob Tibshirani (2010) "Regularized Paths for Generalized Linear Models Via Coordinate Descent," *Journal of Statistical Software*, Vol. 33, 02.

Gennaioli, Nicola, Andrei Shleifer, and Robert Vishny (2012) "Neglected risks, financial innovation, and financial fragility," *Journal of Financial Economics*, Vol. 104, No. 3, pp. 452 – 468.

Gourinchas, Pierre-Olivier and Maurice Obstfeld (2012) "Stories of the Twentieth Century for the Twenty-First," *American Economic Journal: Macroeconomics*, Vol. 4, No. 1, pp. 226–65.

Gourinchas, Pierre-Olivier and Helene Rey (2007a) "From World Banker to World Venture Capitalist: US External Adjustment and the Exorbitant Privilege," in Richard Clarida ed. *G-7 Current Account Imbalances: Sustainability and Adjustment*, pp. 11–55, Chicago,: University of Chicago Press.

Gourinchas, Pierre-Olivier and Hélène Rey (2007b) "International Financial Adjustment," *Journal of Political Economy*, pp. 665–703, August.

Greenwood, Robin, Samuel G. Hanson, Andrei Shleifer, and Jakob Ahm Sørensen (Forthcoming) "Predictable Financial Crises," *Journal of Finance*.

Grunwald, Peter and Thijs van Ommen (2014) "Inconsistency of Bayesian Inference for Misspecified Linear Models, and a Proposal for Repairing It," *Bayesian Analysis*, Vol. 12, 12.

Hastie, T.J. and R.J. Tibshirani (1986) *Generalized additive models*, Vol. 1, pp.1-335.

Jordà, Òscar, Moritz Schularick, and Alan M Taylor (2017) "Macrofinancial history and the new business cycle facts," *NBER macroeconomics annual*, Vol. 31, No. 1, pp. 213–263.

Kaminski, Graciela and Carmen Reinhart (1999) "The Twin Crises: The Causes of Banking and Balance of Payments Problems," *American Economic Review*, Vol. 89, No. 3, pp. 473–500, June.

Kindleberger, Charles (1978) *Manias, Panic and Crashes*.

Lowe, Philip and Claudio Borio (2002) "Asset prices, financial and monetary stability: exploring the nexus," BIS Working Papers 114, Bank for International Settlements.

Mallet, Vivien, Gilles Stoltz, and Boris Mauricette (2009) "Ozone ensemble forecast with machine learning algorithms," *Journal of Geophysical Research: Atmospheres*, Vol. 114, No. D5.

Martinez-Miera, David and Rafael Repullo (2017) "Search for yield," *Econometrica*, Vol. 85, No. 2, pp. 351–378.

Mian, Atif and Amir Sufi (2009) "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis," *The Quarterly Journal of Economics*, Vol. 124, No. 4, pp. 1449–1496, 11.

Mian, Atif, Amir Sufi, and Emil Verner (2017) "Household Debt and Business Cycles Worldwide*," *The Quarterly Journal of Economics*, Vol. 132, No. 4, pp. 1755–1817, 05.

Minsky, Hyman (1986) "Global Consequences of Financial Deregulation,"Technical Report 96, Washington University Working Paper.

Mol, Christine, Ernesto De Vito, and Lorenzo Rosasco (2009) "Elastic-Net Regularization in Learning Theory," *Journal of Complexity*, Vol. 25, pp. 201–230, 04.

Mol, Christine, Sofia Mosci, Magali Traskine, and Alessandro Verri (2009) "A Regularized Method for Selecting Nested Groups of Relevant Genes from Microarray Data," *Journal of computational biology*, Vol. 16, pp. 677–90, 06.

Monteleoni, Claire, Gavin A Schmidt, Shailesh Saroha, and Eva Asplund (2011) "Tracking climate models," *Statistical Analysis and Data Mining: The ASA Data Science Journal*, Vol. 4, No. 4, pp. 372–392.

Müller, Karsten and Emil Verner (2021) "Credit Allocation and Macroeconomic Fluctuations," *Available at SSRN 3781981*.

Reinhart, Carmen M. and Kenneth S. Rogoff (2009) *This Time Is Different: Eight Centuries of Financial Folly*, Princeton, New Jersey: Princeton University Press.

Rossi, Barbara (2011) "Advances in Forecasting Under Instability," *Handbook of Economic Forecasting*, Vol. 2, 09.

Schularick, Moritz and Alan M. Taylor (2012) "Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crises, 1870-2008," *American Economic Review*, Vol. 102, No. 2, pp. 1029–61.

Stock, James and Mark Watson (2012) "Disentangling the Channels of the 2007â€"09 Recession [with Comments and Discussion]," *Brookings Papers on Economic Activity*, pp. 81–156, 01.

Stoltz, Gilles (2010) "AgrÃ©gation sÃ©quentielle de prÃ©dicteurs : mÃ©thodologie gÃ©nÃ©rale et applications Ã la prÃ©vision de la qualitÃ© de l'air et Ã celle de la consommation Ã©lectrique," *Journal de la SociÃ©tÃ© Francaise de Statistique*, Vol. 151, 01.

Sufi, Amir and Alan M Taylor (2021) "Financial crises: A survey," Working Paper 29155, National Bureau of Economic Research.

Tibshirani, Robert (1996) "Regression Shrinkage and Selection Via the Lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 58, pp. 267–288, 01.

Ward, Felix (2017) "Spotting the danger zone: Forecasting financial crises with classification tree ensembles and many predictors," *Journal of Applied Econometrics*, Vol. 32, pp. 359–378.

Zou, Hui and T. Hastie (2005) "Regularization and Variable Selection via the Elastic Nets," *Journal of the Royal Statistical Society: Series B*, Vol. 67, pp. 301–320, 01.

Appendix

6 Sequential learning

Algorithm 1. ***Prediction with expert advice and delayed feedback***

1. *The expert advice* $\left\{f_{j,t} \in \mathcal{D} : j \in \mathcal{E}\right\}$ *is revealed to the forecaster.*
2. *The forecaster makes the prediction* $\hat{y}_t \in \mathcal{D}$.
3. *The t-3 observation* y_t *is revealed.*
4. *The forecaster and each expert respectively incur losses* $\ell(\hat{y_{t-3}}, y_{t-3})$ *and* $\ell(f_{j,t-3}, y_{t-3})$.

Algorithm 2. ***Gradient-based EWA***

1. *Parameter : Choose the learning rate* $\eta_t > 0$.
2. *Initialization :* p_1 *is the first uniform weight,* $p_{j,1} = \frac{1}{N} \forall j \in \{1, ..N\}$.
3. *For time instances* $t = 2, 3, ..., T$ *the weights vector* p_t *is defined by :*

$$p_{j,t} = \frac{\exp(-\eta_t \sum_{s=1}^{t-1} \tilde{L_{j,s}})}{\sum_{k=1}^{N} \exp(-\eta_t \sum_{s=1}^{t-1} \tilde{L_{k,s}})}$$

where $\tilde{L_{j,s}} = \nabla \ell(\sum_{k=1}^{N} p_{k,s} f_{k,s}, y_s) \cdot f_{j,s}$

The strategy $\mathcal{E}_\eta^{grad}$ competes with the best convex combination of experts. The following theorem is stated in Stoltz (2010):

Theorem 1. *If* $\mathcal{D} = [0, 1]$ *is convex,* $\mathcal{L}(\cdot, y)$ *are differentiable on* $\mathcal{D}$ and $\tilde{\mathcal{L}_{j,t}}$ *are in* $[0, 1]$*, for all* $\eta_t > 0$:

$$\sup\{R_T(\mathcal{E}_\eta^{grad})\} \leq \frac{\ln(N)}{\eta_t} + \eta_t \frac{T}{2} \tag{1}$$

The bound of the regret depends on three parameters, two exogeneous (N and T) and one endogenous (η_t). An interesting property of the theorem is that the bound does not depend linearly on the number of experts, but on $\ln(N)$. A large number of experts will not drastically increase the difference between the forecaster's cumulative loss and the cumulative loss of the best combination of experts. The last parameter of the bound η_t is the learning rate. For the gradient-based EWA aggregation rule, the forecaster chooses the parameter η_t with the best past performance :

$$\eta_t \in \arg\min_{\eta > 0} \hat{L}_{t-1}(\mathcal{E}_\eta)$$

7 Experts

Elastic-net Logits

The logits with elastic net penalty are constructed following Friedman et al. (2010) as:

$$\min_{\beta_0,\beta}\left\{\frac{1}{N}\sum_{i=1}^{N}F(\beta_0+x_i\beta)-\lambda P_\alpha(\beta)\right\} \text{ and } P_\alpha(\beta)=\sum_{j=1}^{p}[\tfrac{1}{2}(1-\alpha)\beta_j^2+\alpha\,|\,\beta_j\,|\,]$$

Since there is a risk of correlation, we pick $\alpha = 0.7$. All the Logits include each variable in level as well as the 1-year change, the 2-year change and the 3-year change.

Machine Learning Experts

Random Forest
A random forest (**RF**) consists in three steps :

- Build a number of decisions trees on bootstrapped training samples.
- Each time a split in a tree is considered, a random sample of m predictors is chosen as split candidate.
- Aggregate the prediction of each tree.

General Additive Model
Generalized additive models (**GAM**) provide a general framework for extending a standard linear model by allowing non-linear functions of each of the variables, while maintaining additivity. We consider here a General Additive Model such as :

$$y_t = \beta_0 + f_1(x_{1,t}) + f_2(x_{2,t}) + f_2(x_{12t})$$

The model is fitted with smoothing splines [see Hastie and Tibshirani (1986)].

1 US

Missing series : house price, total loans to households, total loans to business.

Common experts :

- Logit real economy (Lre) : Real GDP per capita (PPP), Real GDP per capita (index), Real consumption per capita (index), GDP (nominal, local currency), Consumer prices (index).
- Logit real economy 2 (Lre2) : GDP (nominal, local currency), Consumer prices (index), Investment-to-GDP ratio, Exports (nominal, local currency).

- Logit valuation (Lval) : Short-term interest rate (nominal, percent per year), Long-term interest rate (nominal, percent per year), Stock prices (nominal index), House prices (nominal index, 1990=100).
- Logit foreign (Lfor) : Current account (nominal, local currency), Imports (nominal, local currency), Exports (nominal, local currency).
- Logit credit (Lcr) : Total loans to non-financial private sector (nominal, local currency), Mortgage loans to non-financial private sector (nominal, local currency), Total loans to business (nominal, local currency),House prices (nominal index, 1990=100).
- Logit housing (Lho) : Total loans to non-financial private sector (nominal, local currency), House prices (nominal index, 1990=100).

Country-specific experts :

- Logit combination 1 (Lc1) : Long-term interest rate (nominal, percent per year), Real GDP per capita (index, 2005=100), Real consumption per capita (index, 2006=100), Consumer prices (index, 1990=100)
- Logit combination 2 (Lc2) : Broad money (nominal, local currency), Stock prices (nominal index), Government expenditure (nominal, local currency), Total loans to non-financial private sector (nominal, local currency).
- Logit combination 3 (Lc3) : Public debt-to-GDP ratio, Total loans to non-financial private sector (nominal, local currency), Stock prices (nominal index),Mortgage loans to non-financial private sector (nominal, local currency)
- Logit combination 4 (Lc4) : Long-term interest rate (nominal, percent per year), Mortgage loans to non-financial private sector (nominal, local currency), Real GDP per capita (index, 2005=100), Consumer prices (index, 1990=100).
- GAM : Imports (nominal, local currency), Long-term interest rate (nominal, 2y change), Public debt-to-GDP ratio (3y)
- Random Forest.

Variable selected (at least partly during the period) by elastic-net regressions experts :

- Logit combination 1 (Lc4) : Consumer prices (index, 1y,2y,3y), long-term interest rate (nominal, 1y,2y,3y), Real GDP per capita (index,1y,2y,3y), Mortgage loans to non-financial private sector (nominal, local currency),
- Logit combination 2 (Lc2) : Broad money (nominal, local currency), Stock prices (nominal index), Government expenditure (nominal, local currency), Total loans to non-financial private sector (nominal, local currency).
- Logit combination 3 (Lc3) : Public debt-to-GDP ratio, Total loans to non-financial private sector (nominal, local currency), Stock prices (nominal index),Mortgage loans to non-financial private sector (nominal, local currency)

- Logit combination 4 (Lc4) : Long-term interest rate (nominal, percent per year), Mortgage loans to non-financial private sector (nominal, local currency), Real GDP per capita (index, 2005=100), Consumer prices (index, 1990=100).

2 France

Missing series : total loans to non-financial private sector, total loans to households, total loans to business.

Common experts :

- Logit real economy (Lre) : Real GDP per capita (PPP), Real GDP per capita (index), Real consumption per capita (index), GDP (nominal, local currency), Consumer prices (index).
- Logit real economy 2 (Lre2) : GDP (nominal, local currency), Consumer prices (index), Investment-to-GDP ratio, Exports (nominal, local currency).
- Logit valuation (Lval) : Short-term interest rate (nominal, percent per year), Long-term interest rate (nominal, percent per year), Stock prices (nominal index), House prices (nominal index, 1990=100).
- Logit foreign (Lfor) : Current account (nominal, local currency), Imports (nominal, local currency), Exports (nominal, local currency).
- Logit credit (Lcr) : Total loans to non-financial private sector (nominal, local currency), Mortgage loans to non-financial private sector (nominal, local currency), Total loans to business (nominal, local currency),House prices (nominal index, 1990=100).
- Logit housing (Lho) : Total loans to non-financial private sector (nominal, local currency), House prices (nominal index, 1990=100).

Country-specific experts :

- Logit combination 1 (Lc4) : Real GDP per capita (PPP), GDP (nominal, local currency), USD exchange rate (local currency/USD), Exports (nominal, local currency).
- Logit combination 2 (Lc2) : House prices (nominal index, 1990=100), Real GDP per capita (PPP), GDP (nominal, local currency), Stock prices (nominal index), Consumer prices (index, 1990=100).
- Logit combination 3 (Lc3) : Government expenditure (nominal, local currency), Real GDP per capita (PPP), Consumer prices (index, 1990=100),Population.
- Logit combination 4 (Lc4) : Stock prices (nominal index),Government expenditure (nominal, local currency), House prices (nominal index, 1990=100), USD exchange rate (local currency/USD).

- GAM : Real GDP per capita (PPP), GDP (nominal, local currency), USD exchange rate (local currency/USD).
- Random Forest.

Variable selected (at least partly during the period) by elastic-net regressions experts :

- Logit combination 2 (Lc2) : Real GDP per capita (PPP,1y,2y,3y), GDP (nominal, 2y,3y), USD exchange rate (local currency/USD), Stock price (index, 1y,2y,3y), House prices (nominal index, 1990=100, 1y, 2y, 3y)
- Logit valuation(Lval) : House prices (nominal index, 1990=100, 1y, 2y, 3y), long-term interest rate (nominal, 1y,2y,3y), stock prices (index, 1y, 2y,3y)
- Random Forest : (at the end of the sample) GDP (nominal), House prices (nominal), Real GDP per capita (PPP), stock price (index)

3 Italy

Missing series : short-term interest rate, stock price, total loans to households, total loans to business, house price.

Common experts :

- Logit real economy (Lre) : Real GDP per capita (PPP), Real GDP per capita (index), Real consumption per capita (index), GDP (nominal, local currency), Consumer prices (index).
- Logit real economy 2 (Lre2) : GDP (nominal, local currency), Consumer prices (index), Investment-to-GDP ratio, Exports (nominal, local currency).
- Logit valuation (Lval) : Short-term interest rate (nominal, percent per year), Long-term interest rate (nominal, percent per year), Stock prices (nominal index), House prices (nominal index, 1990=100).
- Logit foreign (Lfor) : Current account (nominal, local currency), Imports (nominal, local currency), Exports (nominal, local currency).
- Logit credit (Lcr) : Total loans to non-financial private sector (nominal, local currency), Mortgage loans to non-financial private sector (nominal, local currency), Total loans to business (nominal, local currency),House prices (nominal index, 1990=100).
- Logit housing (Lho) : Total loans to non-financial private sector (nominal, local currency), House prices (nominal index, 1990=100).

Country-specific experts :

- Logit combination 1 (Lc1) : GDP (nominal, local currency), Total loans to non-financial private sector (nominal, local currency), Narrow money (nominal, local currency), USD exchange rate (local currency/USD).
- Logit combination 2 (Lc2) : Consumer prices (index, 1990=100), Exports (nominal, local currency), Total loans to non-financial private sector (nominal, local currency), Government revenues (nominal, local currency).
- Logit combination 3 (Lc3) : Stock prices (nominal index), Real GDP per capita (PPP), USD exchange rate (local currency/USD). Narrow money (nominal, local currency).
- Logit combination 4 (Lc4) : Population, Total loans to non-financial private sector (nominal, local currency), Real GDP per capita (index, 2005=100), Narrow money (nominal, local currency), Current Account.
- GAM : Population (3y change), Real GDP per capita (3y change), Total loans to non-financial private sector (1y change-.
- Random Forest.

Variable selected (at least partly during the period) by elastic-net regressions experts :

- Logit combination 1 (Lc1) : all the variables (partially).
- Logit combination 3 (Lc3) : Current Account (2y,3y), Real GDP per Capita (PP,1y,2y,3y), USD exchange rate (local currency/USD) , Narrow money (2y,3y).
- Random Forest (Rf) : [from 1980] mostly exports (3y), Population (3y), Total loans to non-financial private sector (2y and 3y).

4 Japan

Missing series : investment, mortgage loans to non-financial private sector, total loans to households, total loans to business, house price.

Common experts :

- Logit real economy (Lre) : Real GDP per capita (PPP), Real GDP per capita (index), Real consumption per capita (index), GDP (nominal, local currency), Consumer prices (index).
- Logit real economy 2 (Lre2) : GDP (nominal, local currency), Consumer prices (index), Investment-to-GDP ratio, Exports (nominal, local currency).
- Logit valuation (Lval) : Short-term interest rate (nominal, percent per year), Long-term interest rate (nominal, percent per year), Stock prices (nominal index), House prices (nominal index, 1990=100).

- Logit foreign (Lfor) : Current account (nominal, local currency), Imports (nominal, local currency), Exports (nominal, local currency).
- Logit credit (Lcr) : Total loans to non-financial private sector (nominal, local currency), Mortgage loans to non-financial private sector (nominal, local currency), Total loans to business (nominal, local currency),House prices (nominal index, 1990=100).
- Logit housing (Lho) : Total loans to non-financial private sector (nominal, local currency), House prices (nominal index, 1990=100).

Country-specific experts :

- Logit combination 1 (Lc1) : Broad money (nominal, local currency), USD exchange rate (local currency/USD), Government revenues (nominal, local currency), Real GDP per capita (PPP).
- Logit combination 2 (Lc2) : Broad money (nominal, local currency), Total loans to non-financial private sector (nominal, local currency), Population
- Logit combination 3 (Lc3) : Stock prices (nominal index), Real GDP per capita (PPP), USD exchange rate (local currency/USD). Narrow money (nominal, local currency), Real consumption per capita (index, 2006=100).
- Logit combination 4 (Lc4) : Narrow money (nominal, local currency), Total loans to non-financial private sector (nominal, local currency), Consumer prices (index, 1990=100), Government expenditure (nominal, local currency).
- Logit Monetary (Lm) : Narrow money (nominal, local currency), Broad money (nominal, local currency).
- GAM : Population (3y change), Real GDP per capita (3y change), Total loans to non-financial private sector (1y change-.
- Random Forest.

Variable selected (at least partly during the period) by elastic-net regressions experts :

- Logit combination 4 (Lc1) : all the variables (partially). Loans is the most used variable at the end of the sample
- Logit valuation (Lm) : only long-term interest rate.

5 Spain

Missing series : Total loans to non-financial private sector (nominal, local currency), Mortgage loans to non-financial private sector (nominal, local currency), Total loans to households (nominal, local currency), Total loans to business (nominal, local currency), House prices (nominal index, 1990=100)

Common experts :

- Logit real economy (Lre) : Real GDP per capita (PPP), Real GDP per capita (index), Real consumption per capita (index), GDP (nominal, local currency), Consumer prices (index).
- Logit real economy 2 (Lre2) : GDP (nominal, local currency), Consumer prices (index), Investment-to-GDP ratio, Exports (nominal, local currency).
- Logit valuation (Lval) : Short-term interest rate (nominal, percent per year), Long-term interest rate (nominal, percent per year), Stock prices (nominal index), House prices (nominal index, 1990=100).
- Logit foreign (Lfor) : Current account (nominal, local currency), Imports (nominal, local currency), Exports (nominal, local currency).
- Logit monetary (Lm) : Narrow money (nominal, local currency), Broad money (nominal, local currency), Short-term interest rate (nominal, percent per year), Long-term interest rate (nominal, percent per year).

Country-specific experts :

- Logit combination 1 (Lc1) : Investment-to-GDP ratio, USD exchange rate (local currency/USD), GDP (nominal, local currency), Population.
- Logit combination 2 (Lc2) : Exports (nominal, local currency), Stock prices (nominal index), Investment-to-GDP ratio, USD exchange rate (local currency/USD).
- Logit combination 3 (Lc3) : Real GDP per capita (PPP), Imports (nominal, local currency), USD exchange rate (local currency/USD), Consumer prices (index, 1990=100).
- Logit combination 4 (Lc4) : Real GDP per capita (index, 2005=100), Imports (nominal, local currency), USD exchange rate (local currency/USD), nvestment-to-GDP ratio.
- Logit Monetary (Lm) : Narrow money (nominal, local currency), Broad money (nominal, local currency).
- GAM : USD exchange rate (local currency/USD), Population, Imports(2y).
- Random Forest.

Variable selected (at least partly during the period) by elastic-net regressions experts :

- Logit combination 4 (Lc1) : all the variables (partially).
- Random Forest : the most used variable (all sample) : exchange rate. At the end of sample, consumer prices (3y) and investment-to-GDP ratio (3y).

6 Netherlands

Missing series : GDP, Investment-to-GDP ratio, Current Account, Total loans to households (nominal, local currency), Total loans to business (nominal, local currency).

Common experts :

- Logit real economy (Lre) : Real GDP per capita (PPP), Real GDP per capita (index), Real consumption per capita (index), GDP (nominal, local currency), Consumer prices (index).
- Logit real economy 2 (Lre2) : GDP (nominal, local currency), Consumer prices (index), Investment-to-GDP ratio, Exports (nominal, local currency).
- Logit valuation (Lval) : Short-term interest rate (nominal, percent per year), Long-term interest rate (nominal, percent per year), Stock prices (nominal index), House prices (nominal index, 1990=100).
- Logit foreign (Lfor) : Current account (nominal, local currency), Imports (nominal, local currency), Exports (nominal, local currency).
- Logit monetary (Lm) : Narrow money (nominal, local currency), Broad money (nominal, local currency), Short-term interest rate (nominal, percent per year), Long-term interest rate (nominal, percent per year).

Country-specific experts :

- Logit combination 1 (Lc1) : Government revenues (nominal, local currency), Long-term interest rate (nominal, percent per year), House prices (nominal index, 1990=100), USD exchange rate (local currency/USD).
- Logit combination 2 (Lc2) : Government expenditure (nominal, local currency), Short-term interest rate (nominal, percent per year), House prices (nominal index, 1990=100), Real GDP per capita (index, 2005=100).
- Logit combination 3 (Lc3) : Long-term interest rate (nominal, percent per year), Government expenditure (nominal, local currency), House prices (nominal index, 1990=100), Real GDP per capita (index, 2005=100).
- Logit combination 4 (Lc4) : Government expenditure (nominal, local currency), Real GDP per capita (index, 2005=100), House prices (nominal index, 1990=100), Consumer prices (index, 1990=100).
- GAM : Real GDP per capita (3y) + House prices (3y), Government expenditure (3y)
- Random Forest.

Variable selected (at least partly during the period) by elastic-net regressions experts :

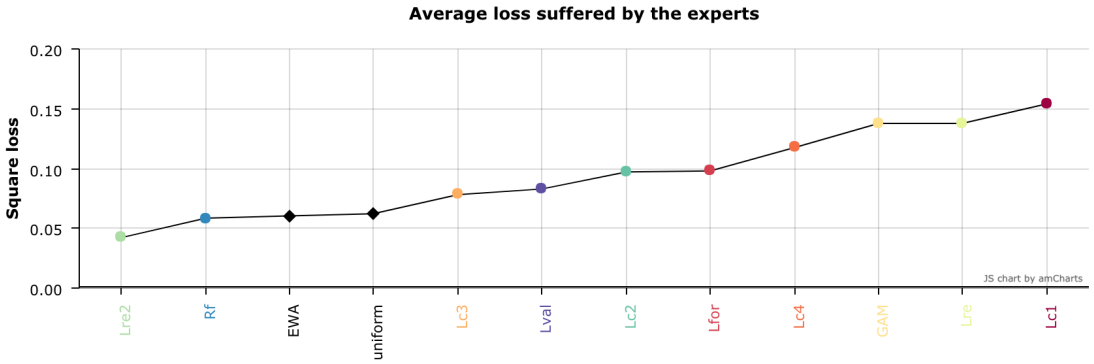


Figure 21: France

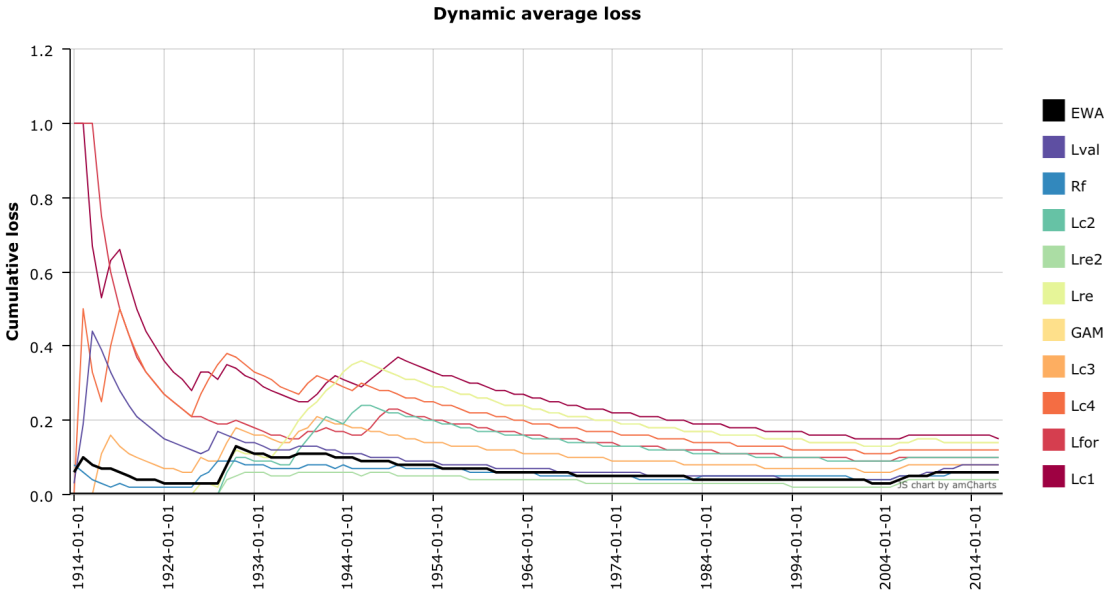


Figure 22: France

- Logit valuation (Lval): all the variables except stock prices. House prices and Long-term interest rate are very consistently used.
- Random Forest : the most used variable (all sample) : House price is the most used variable (from 1939), before a mix (consumer price, 3y very used)/

*More diagnostics

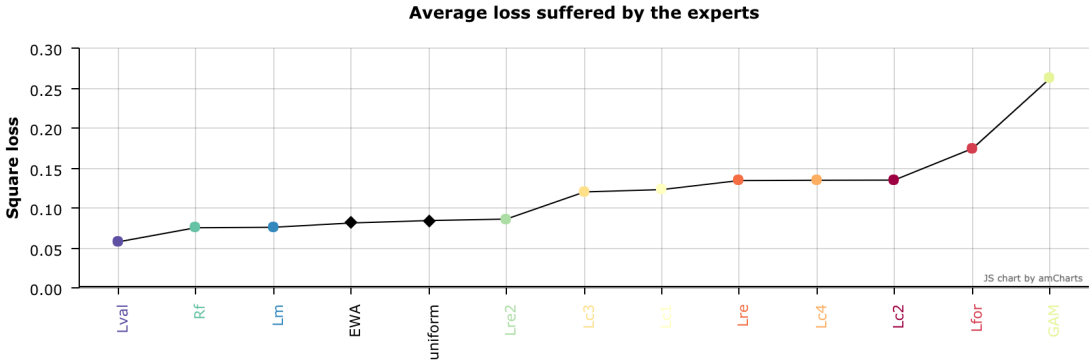


Figure 23: Netherlands

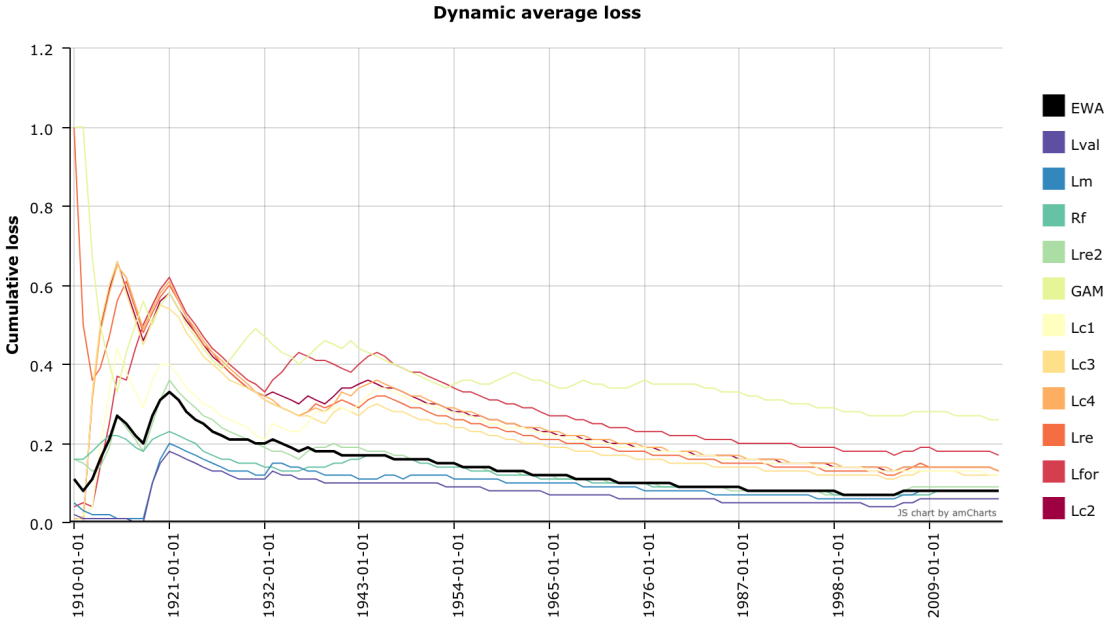


Figure 24: Netherlands



Figure 25: Spain

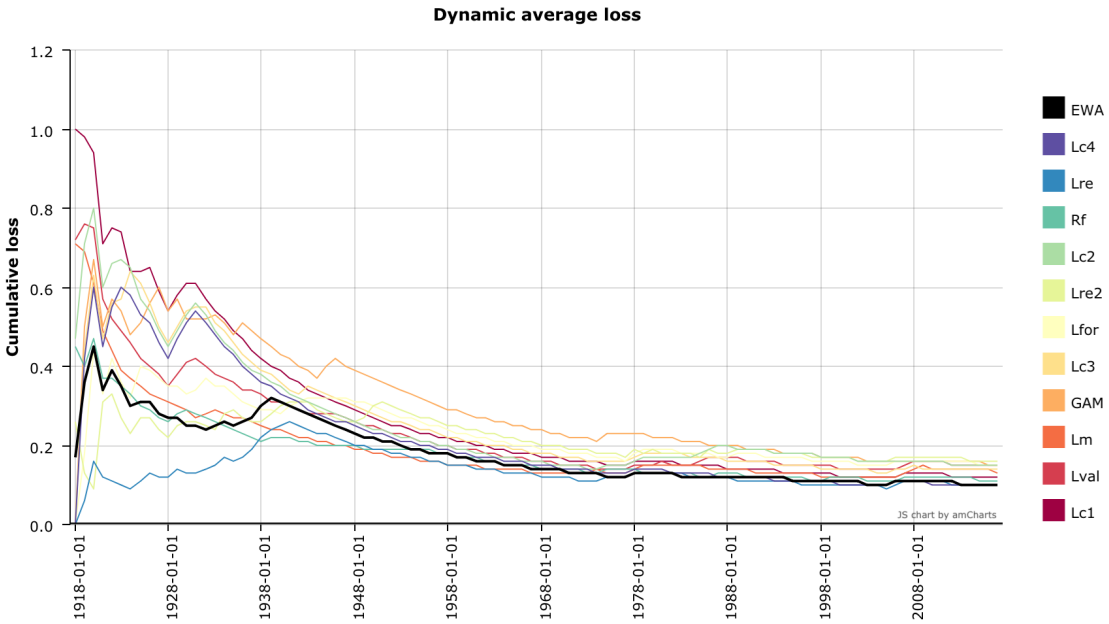


Figure 26: Spain

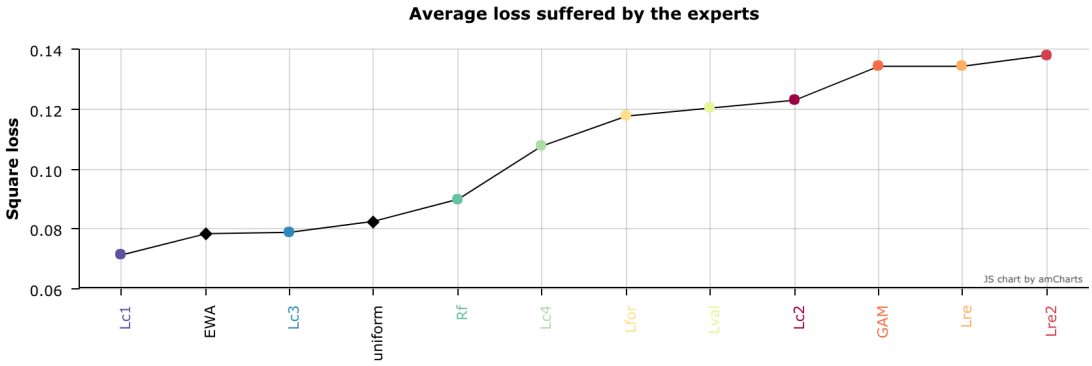


Figure 27: Italy

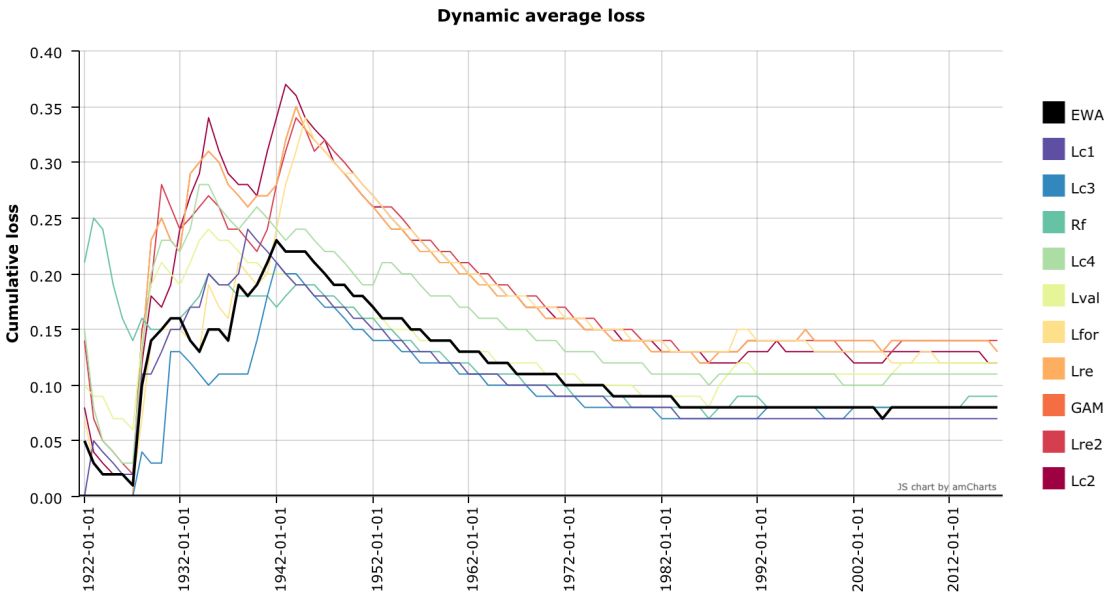


Figure 28: Italy

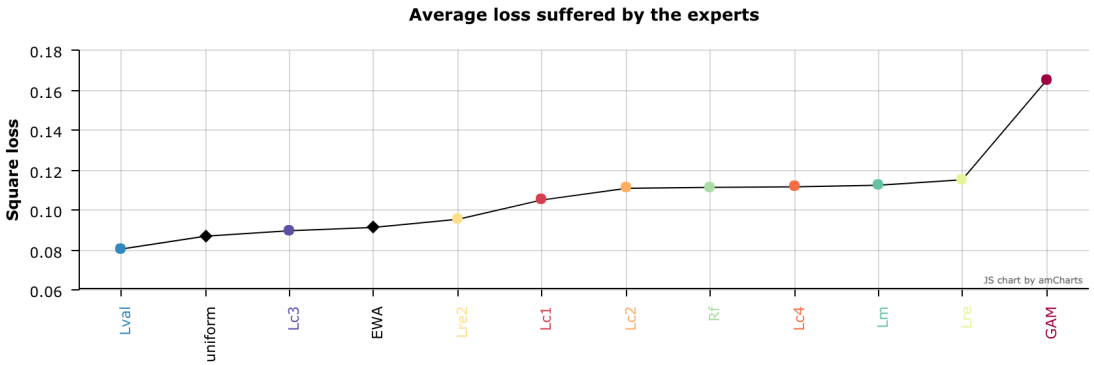


Figure 29: Japan

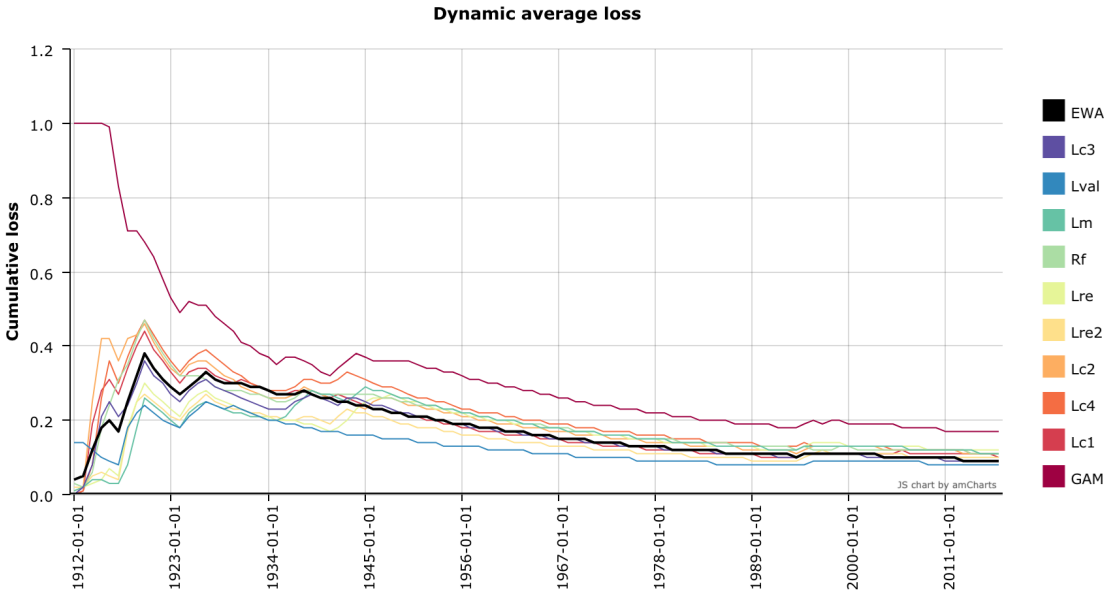


Figure 30: Japan

Variables selection for logit combinations

1. The variables are selected thanks to an AUROC procedure performed on the batch sample, following Schularick and Taylor [2012] and Coudert and Idier [2017].

2. The number of selected variables in the logit combinations depends on the variables' AUROC. Regularization is a way to decrease the risk of overfitting when the model complexity increases; adding to much variables decreases the forecasting ability [Hastie and Zou, 2005]. As in Coudert and Idier[2017], we restrain the sets of variables included in the "logit combinations" experts. In our case, 3 to 5 variables, usually having AUROCs superior to 0.7, are included corresponding to 12 to 20 variables since we always include 1y, 2y and 3y transformations. If several variables have a large AUROC, i.e. superior to 0.7 [Pepe, 2003; Schularick and Taylor, 2012] [23], more variables are included in the logit combinations.