Answering the Queen: Machine Learning and Financial Crises

JÉRÈMY FOULIARD¹ Michael Howell² & Hélène Rey^{3,4,5}

¹Paris School of Economics ²Cross Border Capital ³London Business School ⁴CEPR ⁵NBER

April 7, 2021

The views expressed here are those of the authors and do not represent those of the French Macroprudential Authority.

Forecasting the financial crisis of 2008

Visiting the LSE and being shown how terrible the situation was and had been, the Queen asked: "Why did nobody notice it?"



Motivation

▶ The Queen got a terse answer.

Motivation

▶ The Queen got a terse answer.

After several months, the Economic Section of the British Academy wrote a three-page missive to Her Majesty blaming the lack of foresight of the crisis on the "failure of the collective imagination of many bright people".

Motivation

▶ The Queen got a terse answer.

- After several months, the Economic Section of the British Academy wrote a three-page missive to Her Majesty blaming the lack of foresight of the crisis on the "failure of the collective imagination of many bright people".
- This paper aims at predicting systemic crises well in advance (12 quarters ahead) using cutting-edge machine learning tools.
 Macroprudential policies need advance information on risk build ups.

Short Literature Review

- Early Warning Indicators of financial crises : Kaminsky [1998], Kaminsky and Reinhart [1999], Borio and Lowe [2002], Gourinchas and Obstfeld (2012), Drehmann and Juselius [2013]; Demirgüç-Kunt and Detragiache [1998], Eichengreen and Rose [1997], Bongini et al. [2000], Frankel and Saravelos [2012], Schularick and Taylor [2012], Coudert and Idier [2017], Mian and Sufi [2018], Krishnamurthy and Muir [2018].
- Machine Learning : Davis and Karim [2008], Duttagupta and Cashin [2012], Ward [2014], Joy et al. [2017], Alessi and Detken [2018], Bluwstein et al. (2020).

Short Literature Review

- ▶ Theory of Financial Crises: Booms go bust (credit growth) [Kindleberger (1978), Shularick and Taylor (2012), Reinhart and Rogoff (2008), Coimbra and Rey (2018); Household debt, Mian and Sufi (2018); behavioural explanations, Bordalo Gennaioli Shleifer (2019) (bubbles in asset prices); excessive risk taking (leverage, moral hazard) Allen and Gale, Rajan; search for yield (Stein); real shock amplified by a capital constraint (macro finance); balance of payment (real exchange rate, capital flows); concentrated exposures of banking system (real estate, oil), ...
- Many different variables, non linear interactions, time varying effects, but commonalities known for a very long time.

We know a few things about crises

Irving Fisher 1933: ... I have, at present, a strong conviction that these two economic maladies, the debt disease and the price-level disease (or dollar disease), are, in the great booms and depressions, more important causes than all others put together. Some of the other and usually minor factors often derive some importance when combined with one or both of the two dominant factors. Thus 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. The same is true as to over-confidence. I fancy that over-confidence seldom does any great harm except when, as, and if, it beguiles its victims into debt.

We know a few things about crises

Charles Kindleberger 1978: ... By no means does every upswing in business excess lead inevitably to mania and panic. But the pattern occurs sufficiently frequently and with sufficient uniformity to merit renewed study. What happens, basically, is that some event changes the economic outlook. New opportunities for profits are seized, and overdone, in ways so closely resembling irrationality as to constitute a mania. Once the excessive character of the upswing is realized, the financial system experiences a sort of "distress," in the course of which the rush to reverse the expansion process may become so precipitous as to resemble panic. In the manic phase, people of wealth or credit switch out of money or borrow to buy real or illiquid financial assets. In panic, the reverse movement takes place, from real or financial assets to money, or repayment of debt, with a crash in the prices of commodities, houses, buildings, land, stocks, bonds -in short, in whatever has been the subject of the mania.

Online learning: NOT big data but model AGGREGATION

This framework is very suitable for crisis prediction in real time:

- ▶ Multivariate : Which variables cause a financial crisis?
- ► **Time-varying weights** : Causes of financial crises may be different over time.
- **Statistically robust** : overfitting is a problem in the literature.
- ▶ Not "black-box" : assess the role each model plays to predict the pre-crisis.
- ► Theoretically grounded : asymptotic properties of our aggregation rules ensure convergence.
- More general than Bayesian Model Averaging
- ▶ This framework has been used to predict French electricity load (EDF); the tracking of climate models; the network traffic demand.

Sequential predictions

Online learning is performed in a sequence of consecutive rounds where at time instance t the forecaster:

- 1. Receives a question.
- 2. Uses expert advice $\{f_{j,t} \in \mathcal{D} : j \in \mathcal{E}\}$
- 3. Predicts $\hat{y}_t \in \mathcal{Y}$
- 4. Receives true answer $y_t \in \mathcal{Y}$
- 5. Suffers a loss $\ell(\hat{y}_t, y_t)$.

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

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

The forecaster and each expert incur a cumulative loss defined by :

$$L_T(\mathcal{S}) = \sum_{t=1}^T \ell(\sum_{j=0}^N p_{j,t} f_{j,t}) = \sum_{t=1}^T (\hat{y}_t - y_t)^2$$

- ▶ How can we measure the performance of a sequential aggregation rule ?
- We do not have any ideas about the generating process of the observations.
- ▶ Forecaster's performance is relative. We define the regret :

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

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 expert j.

We **minimize the regret** with respect to the best combination of experts:

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

We only select aggregation rules with a "vanishing per-round regret" (regret goes to zero asymptotically).

The Regret can be bounded (bound depends on T, on the learning rate and on log(number of experts)).

This approach is a **meta-statistic approach**: the aim is to find the best sequential combination of experts (who can be any economic models or judgement).

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

Forecaster's cumulative loss is the sum of :

- An approximation error : given by the cumulative loss of the best combination of experts.
- ► An estimation error : given by the regret. It measures the difficulty to approach the best combination of experts.

The rule of the game is to minimize the regret and to find the best experts.

Sequential prediction with expert advice and delayed feedback

- ▶ We adapt the approach to incorporate delayed feedback.
- ▶ We take into account the fact that we only know whether we are in a pre-crisis period after 12 quarters.
- Experts at t are estimated on the batch sample using information available at t-1.
- ► Aggregation rules at t use only t-12 information.

Aggregation rules

We used four aggregation rules :

- 1. The Exponentially weighted average aggregation rule (EWA) [Littlestone and Warmuth, 1994 ; Vovk,1990].
- 2. The Fixed Share aggregation rule (FS) [Devaine et al. (2013)]
- 3. The Online Gradient Descent aggregation rule (OGD) [Zinkevich, 2003].
- 4. The Ridge Regression aggregation rule (Ridge) [Azoury and Warmuth, 2001, Vovk, 2001].
- $\Rightarrow Rules$

Exponentially weighted average aggregation rule (EWA)

Convex aggregation rules combining experts' predictions with a time-varying vector $p_t = (p_{1,t}, ..., p_{N,t})$ in a simplex \mathcal{P} of \mathbb{R}^N :

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

▶ We use the gradient-based version of the EWA aggregation rule.

- ▶ The weights are computable in a simple incremental way.
- ▶ Easy to interpret.

Gradient-based version of the EWA

▶ 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 $L_{j,s} = \nabla \ell (\sum_{k=1}^{N} p_{k,s} f_{k,s}, y_s) \cdot f_{j,s}$ and where η_t is the learning rate.

• If j's advice $f_{j,s}$ points in the direction of the largest increase of the loss function (large inner products $\nabla \ell(\sum_{k=1}^{N} p_{k,s} f_{k,s}, y_s) \cdot f_{j,s}$ in the past) the weight assigned to expert j will be small.

The EWA Aggregation rule

Theorem 1 [Cesa-Bianchi and Lugosi, 2006]

We assume that

- Functions $L(\cdot, y)$ are differentiable.
- ▶ The losses $L_{j,t}$ are bounded.

Therefore, for all learning rate $\eta_t > 0$,

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

For the gradient-based EWA aggregation rule, the forecaster chooses the parameter η_t with the best past performance :

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

To find the value of η_t which minimizes the cumulative loss, at each time instance, we minimize the cumulative loss on a grid.

Data of systemic crisis episodes: off-the-shelf

- ▶ The ECB provides an official database of systemic crisis episodes [Lo Duca et al., 2017q] and additional smaller crises. Judgement of national authorities is involved.
- Our sample starts in 1985q1 (depending on data availability) and ends in 2018q1.
- Our sample includes 7 countries: France, Germany, Italy, Spain, Sweden, UK, US.
- We focus on France, the UK, Germany and Italy. We predict systemic pre-crises.

Data of crisis episodes: off-the shelf

The ECB data uses a characteristic function $C_{n,t}$:

$$C_{n,t} = \begin{cases} 1 & \text{If there is a systemic crisis in country } n \text{ at time } t \\ 0 & \text{Otherwise} \end{cases}$$

Let's define the pre-crisis indicator $I_{n,t}$:

$$I_{n,t} = \begin{cases} 1 & \text{if } \exists h \in H = [0, 12] \text{ such that } C_{n,t+h} = 1\\ 0 & \text{otherwise} \end{cases}$$

Variables I

Our database contains commonly used Early Warning Indicators with transformations (1-y, 2-y, 3-y change and gap-to-trend).

- Macroeconomic indicators : GDP, GDP per person employed, GDP per capita, GDP per hour worked, Unemployment rate, Consumer Price Index, General Government Debt, Golden rule (gap of real long term interest rate to real GDP), Political Uncertainty Index, Oil price index, Consumption, Investment, Multifactor Productivity.
- Credit and Debt indicators : Total credit (to households, to private non-financial sector, to non-financial firms), Debt Service Ratios (household, non-financial corporations, private non-financial sector), Household Debt, General Government Debt.
- **Banking sector indicators**: Banking credit to private sector, Bank assets, Bank equity.

Variables II

- ▶ Interest rates and monetary indicators : 3-month rate, 10-year rate, slope of the yield curve (10y-3m), monetary aggregate M3.
- ▶ Real estate indicators : Loans for House purchase, Residential real estate prices, Price-to-income ratio, Price-to-rent ratio, rent price index, house price forecasts.
- Market indicators: Share prices, Financial Conditions Index, Risk Appetite Index, oil price, Equity holdings, Financial assets, VXO, Global Factor in Asset Prices.
- ▶ External condition indicators: Cross-border flows, Real effective exchange rate, Dollar effective exchange rate, Current account, Shipping indicator; export growth, import growth, terms of trade, growth of Foreign Exchange Reserves, External Debt.
- Liquidity Indicators: Total Liquidity, Domestic Liquidity, Policy Liquidity.

Ecumenical Choice of Experts I

We take some standard models:

- Expert P1. Dynamic Probit Model: variables selected with a country-specific AUROC on the batch sample panel.
- **Expert P2**. Panel logit fixed effect: variables selected with a country-specific PCA Analysis on the batch sample panel.
- **Expert P3**: Panel logit fixed effect. Exact specification from literature.
- **Expert BMA**: Bayesian Model Averaging. Variables selected with a country-specific AUROC on the batch sample panel.

We add machine learning experts:

- **Expert GAM**: General Additive Model
- **Expert RF**: Random Forest
- **Expert SVM**: Support Vector Machine

Ecumenical Choice of Experts II

We add Logits with elastic-net penalty:

- **Expert Lre**. Logit real economy: GDP; GDP per person; GDP per hours work; unemployment rate; import, export, public debt.
- Expert Lre2. Logit real economy 2: consumer prices; unemployment rate; GDP per person, GDP per hours work; GDP per capita; public debt; consumption; investment.
- ▶ Expert Lval . Logit valuation: Share Price Index; Real Estate Price; Global Factor in Asset Prices; Short-term interest rate; Long-term interest rate; Dollar effective exchange rate
- Expert Lfor. Logit foreign: Cross Border Flows; Real Effective Exchange Rate; Dollar Effective Exchange Rate; Current Account; Terms of Trade.
- Expert Lba. Logit bank: Risk Appetite; Share price Index; Equity holdings; Total Liquidity Index.
- Expert Lcr. Logit credit: Total credit to non-financial sector; Banking Credit to non-financial sector; Total Credit to Households; Total Credit to non-financial corporations
- Expert Lbis. Logit BIS: Logit credit + DSR Households; DSR Non Financial corporations; DSR Total.

Ecumenical Choice of Experts III

- Expert Lm. Logit monetary: M3; Short-term interest rate; Long-term interest rate; Consumer Prices; Slope of the Yield Curve
- Expert Lho. Logit housing: Price-to-rent; Price to income; Rent Price Index; Real Estate Price
- Expert Lfgo. Logit Foreign Global: Logit Foreign + Global Factor in Asset Prices
- **Expert Lfgho**. Logit Foreign Global + Housing
- **Expert Lhore**. Logit housing + real economy
- **Expert Lbfo**. Logit bank + foreign
- **Expert Lrisk**. Logit Risk: VXO, Risk Appetite; Equity Holdings.
- Expert Lc1 to Expert Lc5. They are obtained by using the variables with the highest AUROC for a given country on the batch sample.

Les crises systémiques en France



Forecasting the pre-crisis period out-of-sample



Figure: Predicted probability - EWA

Forecasting the pre-crisis period out-of-sample



Figure: Predicted probability - FS



Weights associated with the experts

Figure: Weights - EWA



Contribution of each expert to prediction

Figure: Contribution to Forecasts - EWA



Weights associated with the experts

Figure: Weights- FS



Contribution of each expert to prediction

Figure: Contribution to Forecasts - FS

France: pre-crisis period out-of-sample



Figure: Predicted probability - EWA- 2 year

France: contributions to crisis prediction

- EWA: Expert Logit combination 4 (Lc4) : Real estate price, GDP, Total Credit to Households, Rent Price Index, Loans, Banking Credit to private non-financial sector, Price-to-income, Investment, Share price index, Equity Holdings
- **EWA -2 year**: Logit Bank / Foreign : Share price index, Equity Holdings, Risk Appetite, Total Liquidity Index, Crossborder flows, Real effective exchange rate, dollar effective exchange rate, current account, Terms of Trade.
- ▶ **FS**: Expert Logit combination 4 (Lc4) : Real estate price, GDP, Total Credit to Households, Rent Price Index, Loans, Banking Credit to private non-financial sector, Price-to-income, Investment, Share price index, Equity Holdings.

AUROC

- ▶ 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 AUROC is the area under the ROC curve :



Figure: ROC curve

France: Root Mean Square Errors and AUROCs

Online Aggregation Rule	RMSE	AUROC
EWA	0.26	0 975
FS	$0.20 \\ 0.31$	0.92
OGD	0.33	0.85
Ridge	0.52	0.70
Best fixed convex combination	0.28	0.97
Uniform	0.36	0.79

Table: RMSE and AUROC of different aggregation rules. France.

Systemic crises in the United Kingdom



Figure: Contribution to forecasts - EWA

UK: contributions to crisis prediction

Crisis date different from France. No residual event. Crisis in 1990s linked to real estate. Two experts are doing most of the work:

- ► GAM: GDP (2y), Long-term interest rate, Price-to-rent.
- ▶ Logit risk (Lrisk) VXO, Risk Appetite, Equity Holding.
- ▶ It is the **GAM** expert that gives the signal before the 2008 crisis.

UK: Root Mean Square Errors and AUROCs

Online Aggregation Rule	RMSE	AUROC
EWA	0.29	0.92
Best convex combination	0.29	0.94
Uniform	0.43	0.6

Table: RMSE and AUROCs of different aggregation rules. UK

Systemische Krisen in Deutschland



Contribution of each expert to prediction

Figure: Contribution of experts - EWA

Germany: contributions to crisis prediction

Crisis dates different from France and UK. First systemic crisis later (2001Q1-2003Q4). One expert is doing the work:

Panel model P1: Long term interest rate, Price-to-rent, Real estate price, Banking credit to private non-financial sector, .

Germany: Root Mean Square Errors and AUROCs

Online Aggregation Rule	RMSE	AUROC
EWA Best convex combination	$0.21 \\ 0.21$	$0.84 \\ 0.84$
Uniform	0.41	0.78

Table: RMSE and AUROCs of different aggregation rules. Germany

Le crisi sistemiche in Italia



Figure: Contribution to prediction EWA

Italy: contributions to crisis prediction

Crisis dates different from France and UK. First systemic crisis beginning of the 1990s linked to banking distress. One expert is doing the work:

- Logit combination 2 Lc2: Consumption, Investment, Housing 1, Housing 2, Total Credit to Households, Global Capital Factor.
- Logit combination 4 Lc4: : Consumption, Investment, Housing 1,Housing 2, Total Credit to Households, Global Capital Factor, Dollar Effective Exchange Rate, Real Effective Exchange Rate, Terms of Trade.
- ▶ This is **Lc2** who predicts the pre-crisis.

Italy: Root Mean Square Errors and AUROCs

Online Aggregation Rule	RMSE	AUROC
EWA	0.34	0.77
Best convex combination	0.28	0.94
Uniform	0.42	0.70

Table: RMSE and AUROCs of different aggregation rules. Italy

France: pre-crisis period real time

Predicted Probability - France Swaternic crisie Pre-crisis period (Residual event) 2004อากา 2006อากา 2006อากา 2006อากา 2006อากา 2006อากา 2012อากา 2012อากา 2012อากา 2012อากา 2016อากา 2016อากา

Figure: Predicted probability - EWA

France: pre-crisis period real time



Figure: Predicted probability - experts

France contributions to crisis prediction

Fewer variables available. All credit and in general quantity variables missing. One expert gives the signal mostly.

 GAM: Short-term interest rate 2y; Risk Appetite 2y; Private Sector Liquidity 2y

UK: pre-crisis period real time



Figure: Predicted probability - EWA

UK: pre-crisis period real time



Contribution of each expert to prediction

Figure: Predicted probability - experts

UK contributions to crisis prediction

Fewer variables available. All credit and in general quantity variables missing. T

- ► **GAM** : Dollar effective exchange rate, Private Sector Liquidity Stock (2y), Domestic Liquidity Stock local (2Y).
- \blacktriangleright Support Vector Machine ${\bf SVM}$
- ▶ This is GAM who gives the signal.

Conclusions

- This approach gives strong out-of-sample forecasting results to predict financial crises. Gives different crisis flavours for different countries.
- ▶ We now work on the historical database as well (we predict the Great Depression out-of-sample). We can investigate which models are better for Great Depression versus Lehman Brothers.
- ▶ Will not help to predict "out-of-the blue" crisis. Cyber attack crisis?
- Open questions :
 - Using more microeconomic data from bank databases?
 - ► Causality?
 - How to test for the effect of macroprudential policies on crisis probabilities?

Aggregation rules

2 kinds of aggregation rule :

▶ Follow-The-Leader : the forecaster chooses the convex weight vector p_t to minimize the cumulative loss :

$$p_t = \arg\min_{p \in \mathcal{P}} \widehat{L_{t-1}}(\mathcal{S}) = \arg\min_{p \in \mathcal{P}} \sum_{i=1}^{t-1} \widehat{\ell_i}(\sum_{j=0}^N p_{j,t} f_{j,t})$$

▶ Follow-The-Regularized leader : the forecaster chooses the weight vector *p_t* to minimize the cumulative loss plus a regularization term :

$$p_t = \arg\min_{p \in \mathcal{P}} \widehat{L_{t-1}}(\mathcal{S}) + R(p_t) = \arg\min_{p \in \mathcal{P}} \sum_{i=1}^{t-1} \widehat{\ell_i}(\sum_{j=0}^N p_{j,t}f_{j,t}) + R(p_t)$$

In the last case, $R(p_t)$ can be :

▶ a linear/convex function : **Online Gradient Descent**.

• a square- ℓ_2 -norm regularization : **Ridge regression**. $\Leftarrow Contents$

Online Gradient Descent aggregation rule - General Framework

We define the \mathcal{O}_{η} OGD aggregation rule :

- ▶ Parameter : a learning rate η_t .
- Initialization : an arbitrary vector p_1 .

For each round t = 1, 2, ..., T, the vector p_{t+1} is selected according to :

$$p_{j,t+1} = P_j(p_{j,t} - \eta_t \partial \ell(\sum_{j=1}^N p_{j,t} f_{j,t}, y_t))$$

where $P_{j} = \arg \min_{p_{j}} d(p, y) = \arg \min_{p_{j}} || \sum_{j=1}^{N} p_{j,t} f_{j,t} - y_{t} ||$

The Online Gradient Descent aggregation rule Theorem

Theorem 2 [Zinkevich, 2003] If $\eta_t = t^{\frac{1}{2}}$, the regret is :

$$R(\mathcal{O}_{\eta}^{grad}) \le \frac{\sqrt{T}}{2} + 4(\sqrt{T} - \frac{1}{2})$$

The Online Gradient Descent aggregation rule

The bound of the regret is obtained using two facts :

$$\blacktriangleright \max |\hat{y}_t - y_t| = 1$$

$$\max \frac{\partial \ell(\hat{y}_t)}{\partial \hat{y}_t} = \max 2(\hat{y}_t - y_t) = 2 \in Contents$$

The ridge regression aggregation rule

In this framework, weights $u_t = (u_{1,t}, ..., u_{N,t}) \in \mathbb{R}$ are not choosen in a simplex \mathcal{P} anymore. The forecaster prediction is defined by :

$$\tilde{y_t} = \sum_{j=1}^N u_{j,t} f_{j,t}$$

The ridge regression aggregation rule

We define by :

▶ $f_t = (f_{1,t}, ..., f_{N,t})$ the vector of forecaster predictions.

▶ the euclidian norm of a vector $u \in \mathbb{R}$:

$$||u_2|| = \sqrt{\sum_{j=1}^{n} u_j^2}$$

The rest of the framework does not change.

The ridge regression aggregation rule

We choose u_t as follows :

$$u_t \in \arg\min_{v \in \mathbb{R}^N} \left\{ \lambda ||v||_2^2 + \sum_{s=1}^{t=1} (y_s - \sum_{j=1}^N v_j f_{j,s})^2 \right\}$$

An explicit solution is given by :

$$u_t = (\lambda I_N + M_{t-1})^{-1} \sum_{s=1}^{t-1} y_s f_s$$

Ridge Regression

Theorem 4 [Vovk, 2001] Since $\hat{y}_t \in [0, 1]$: $R(\mathcal{R}_{\eta}) \leq \inf_{v \in \mathbb{R}^N} \left\{ \lambda ||v_2^2|| \right\} + N \times \ln(1 + \frac{T}{\lambda N})$

Tuning

As for η , we calibrate λ with a grid such as :

$$\lambda \in \arg\min_{\eta>0}\widehat{L_{t-1}}(\mathcal{R}_{\lambda})$$

 $\Leftarrow Contents$