Designing gaps as early warning indicators: Accuracy, predictive power, and robustness

Beltran, Jahan-Parvar, Paine

Federal Reserve Board of Governors

February 27, 2019

EWIs for financial crises

Early warning indicators (EWIs) try to predict financial crises well before they materialize, to allow enough time for policymakers to react

Two big questions when designing EWIs:

- Which vulnerabilities matter for predicting crises? (Aldasoro et al. 2018, Lee et al. 2018)
- How do we define when such vulnerabilities are excessive?

We focus on the second question, and use credit-to-GDP gap as an example.

How do we create a good EWI using gaps?

- Take into account accuracy, predictive power, and *robustness*
- Consider different detrending methods
- Allow smoothing parameter in detrending methods to vary
- Optimize EWIs separately for AEs and EMEs

Why use credit-to-GDP gap?

Credit gap is the difference between the level of credit-to-GDP and its (estimated) long-run trend

- Financial crises as "credit booms gone wrong" (Schularick and Taylor, 2012; Minsky, 1977; Kindleberger, 1978)
- Credit gaps are good predictors of financial crises (Lowe and Borio, 2002; Borio and Drehmann, 2009; Drehmann and Juselius, 2014)
- Used to anchor countercyclical capital buffers for banks (Basel Committee on Banking Supervision, 2010)
- Also used to monitor financial stability

Preview of findings

- Allowing smoothing parameter to vary significantly improves accuracy
- Some detrending techniques for estimating gaps are more robust than others
- Credit gaps are poor predictors of crises in EMEs

Defining 'excessive'

Gap is deemed excessive when it breaches a certain critical threshold (θ)
Signal extraction approach for choosing threshold

- Kaminsky and Reinhart (1999); Edison (2003), choose θ that maximizes accuracy of the signal
- Borio and Drehmann (2009) maximize accuracy subject to predicting $2/3$ of crises

Estimating the gap: Detrending technique matters

- Borio and Drehmann (2009) recommend HP filter with $\lambda = 400,000$
- Edge and Meisenzahl (2011) show end-of-sample estimates from HP filter are unstable
- Hamilton (Forthcoming) also criticizes HP filter, proposes linear projection method ('Hamilton Filter')
- Gonzalez et al. (2017) use Bayesian structural times series model (STM) which are "less noisy"
- *We consider all three approaches plus a simple growth rate*

In each of these techniques, a parameter governs the smoothness of the estimated trend

Ad-hoc choices for smoothing parameter

- Parameter chosen based on prior that credit cycle is four times longer than business cycles (Hamilton, Forthcoming; Borio and Drehmann, 2009)
- Or fixed to an arbitrary value to ensure a smooth trend (Gonzalez et al, 2017)
- *We vary the smoothing parameter to improve the tradeoff between accuracy, predictive power, and robustness of the signal*

Data

- Total credit to the nonfinancial private sector (households and nonfinancial corporations) scaled by nominal GDP
- Quarterly, with earliest series starting in 1952
- 10 year "burn-in" period for detrending
- 39 crises across 33 countries (from Drehmann et al. 2011)

We investigate four detrending techniques (using real-time and full sample approaches):

- Hodrick-Prescott (HP) Filter, (Hodrick and Prescott, 1997)
- Hamilton Filter (Hamilton, Forthcoming)
- Bayesian structural time series model (STM) (Petris, 2010; Harrison 1997)
- Simple growth rate

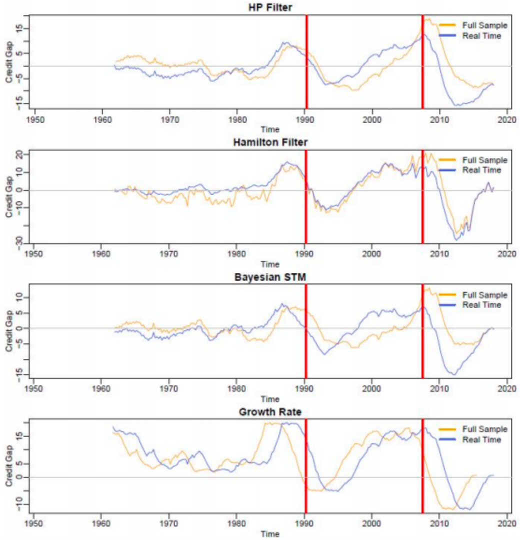

Figure 1: US credit cycle by trend-cycle decomposition method using literature recommended policy values. HP filter $\lambda = 400,00$. Hamilton filter $h = 20$. Bayesian STM $V = 600$. Growth rate $window = 20$.

Signal extraction approach

- EWI is triggered if real-time credit-to-GDP gap $>$ threshold
- 3-year forecasting horizon (before crisis)
- We ignore signals in 2-years after crisis

	Crisis within 3 years	**No crisis within 3 years**
Signal issued	a	b (type-2 error)
No Signal issued	c (type-1 error)	d

Table 1: EWI signaling

The perfect signal would result in a *and* d *only.*

Objective Function

We assume the policymaker wants to maximize the score (S) with respect to the threshold level (θ), and the smoothing parameter (ρ)

$$S_{\theta,\rho} = \begin{cases} (wR + (1-w)A) & |P \geq \frac{2}{3} \\ 0 & \text{otherwise} \end{cases}$$

where R, A, and P are the measures of robustness (ex-post), accuracy, and predictive power.

We consider scenarios where $w = 0$, $w = 0.2$, $w = 0.5$

Predictive Power

Predictive power (P) is the ratio of true positives to the total number of crisis episodes.

$$P = \frac{\#\ \textit{true positives}}{\#\ \textit{crises}}$$

In terms of the matrix, predictive power is equal to $a/(a+c)$.

	Crisis within 3 years	**No crisis within 3 years**
Signal issued	a	b (type-2 error)
No Signal issued	c (type-1 error)	d

Robustness

- Robustness is the difference between the real-time and full-sample gap estimates across countries and over time.
- Measured ex-post
- In a fully robust detrending technique the real-time and the full sample gap estimates are the same.

Given a panel of C countries, the robustness measure, R, is

$$R = 1 - \frac{\sum_{c=1}^{C} \sum_{t=1}^{N} |gap_{c,t}^{full} - gap_{c,t}^{rt}|}{2 * \sum_{c=1}^{C} \sum_{t=1}^{N} |gap_{c,t}^{full}|} \tag{1}$$

where at time t, $gap_{c,t}^{full}$ and $gap_{c,t}^{rt}$ represent the full-sample and real-time gaps for country c, respectively.

Accuracy

The accuracy, A, of the signal is determined by the noise-to-signal ratio (NSR), defined as

$$NSR = \frac{T_2}{1 - T_1} = \frac{\frac{\#\ \textit{false positives}}{\#\ \textit{non-crisis episodes}}}{\frac{\#\ \textit{true positives}}{\#\ \textit{crisis episodes}}} \tag{2}$$

$$\text{A} = 1 - NSR, \text{ where} \tag{3}$$

Assumption

The signal $EWI_{c,t}$ is turned 'on' (takes a value of 1) when the gap breaches the threshold θ, and it stays on for two years.

$$EWI_{c,t} = \begin{cases} 1 & \text{if for any h} \in [0, 7],\ gap_{c,t-h} \geq \theta \text{ \& no crisis in country c} \\ 0 & \text{otherwise} \end{cases}$$

An Illustrative Example

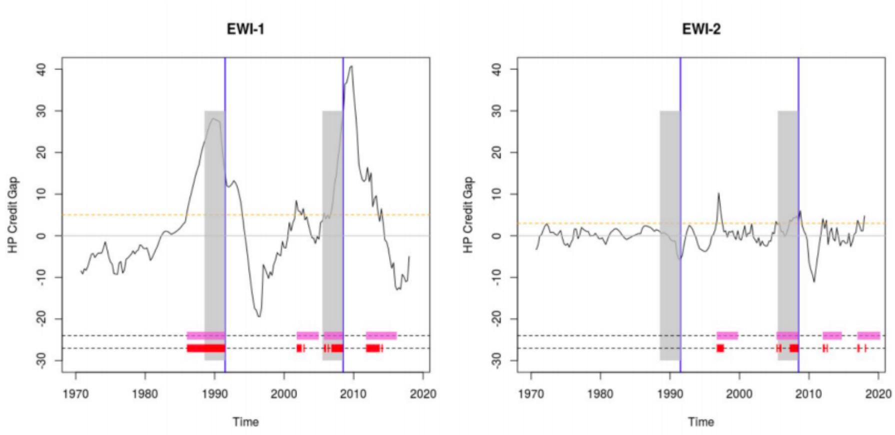

Figure 2: Sweden credit gap. Left panel calculated with $\lambda = 400,00$ and threshold of 5. Right panel calculated with $\lambda = 100$ and threshold of 3.

Dynamics

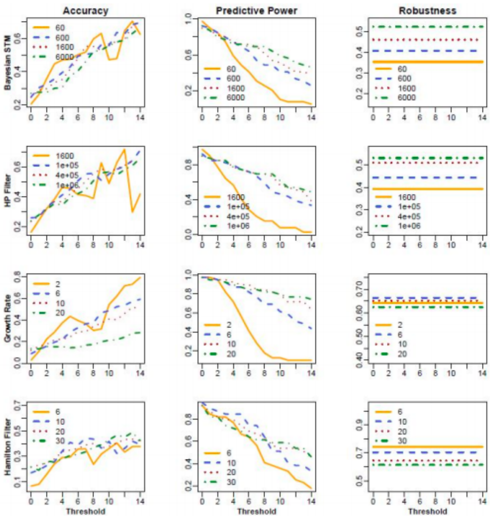

Figure 3: Policy tradeoff dynamics across panel of all countries.

Results: All Economies

Panel A: All Economies								
	w	Smoothing Parameter	Threshold	Score	Accuracy	Robustness	Predictive Power
HP Baseline	0	400,000	7	0.501	0.501	0.509	0.692
HP Filter	0	451,000/ 461,000	9.3	0.598	0.598	0.514	0.667
	0.2	451,000/ 461,000/ 471,000	9.3	0.581	0.598	0.514	0.667
	0.5	581,000/ 651,000/ 661,000	9.4	0.558	0.594	0.521	0.667
Hamilton Filter	0	10.0/ 13.0	7.4/ 7.6	0.466	0.466	0.704	0.718
	0.2	10	7.4	0.513	0.466	0.704	0.718
	0.5	10	7.4	0.585	0.466	0.704	0.718
Bayesian STM	0	1100	7.3	0.574	0.574	0.439	0.667
	0.2	1100/ 1500	7.3/ 7.6	0.547	0.574	0.439	0.667
	0.5	5400	8.4	0.536	0.553	0.520	0.692
Growth Rate	0	8	12.4	0.562	0.562	0.663	0.667
	0.2	8	12.4	0.582	0.562	0.663	0.667
	0.5	8	12.4	0.613	0.562	0.663	0.667

Results: Advanced Economies

Panel B: Advanced Economies

	w	Smoothing Parameter	Threshold	Score	Accuracy	Robustness	Predictive Power
HP Baseline	0	400,000	7	0.586	0.586	0.496	0.786
HP Filter	0	651,000/ 671,000/ 681,000	12.7/ 12.8	0.726	0.726	0.510	0.679
	0.2	671,000/ 681,000	12.8	0.683	0.726	0.510	0.679
	0.5	651,000 - 731,000	12.7/ 12.8	0.618	0.726	0.509	0.679
Hamilton Filter	0	31	14	0.640	0.640	0.600	0.679
	0.2	13	11.9	0.636	0.622	0.693	0.679
	0.5	13	11.9	0.657	0.622	0.693	0.679
Bayesian STM (by 100)	0	5000, 5200, 5300	11.7, 11.8	0.712	0.712	0.503	0.679
	0.2	5200, 5300	11.8	0.671	0.712	0.504	0.679
	0.5	5300	11.8	0.609	0.712	0.505	0.679
Growth Rate	0	10	13.8	0.654	0.654	0.683	0.679
	0.2	10	13.8	0.660	0.654	0.683	0.679
	0.5	8	11.5	0.671	0.650	0.693	0.679

Results: Emerging Markets

Panel C: Emerging Economies

	w	Smoothing Parameter	Threshold	Score	Accuracy	Robustness	Predictive Power
HP Baseline	0	400,000	7	0.000	0.124	0.552	0.455
HP Filter	0	11,000	4.3	0.400	0.400	0.419	0.727
	0.2	11,000	4.3	0.404	0.400	0.419	0.727
	0.5	1,041,000 - 1,091,000	0	0.412	0.241	0.583	0.909
Hamilton Filter	0	15	4.8	0.333	0.333	0.669	0.818
	0.2	15	4.8	0.400	0.333	0.669	0.818
	0.5	1	1.8	0.540	0.254	0.826	0.727
Bayesian STM	0	300	3.8	0.305	0.305	0.428	0.727
	0.2	300	3.8	0.330	0.305	0.428	0.727
	0.5	4700	0.5	0.395	0.232	0.559	0.909
Growth Rate	0	6	11.5	0.326	0.326	0.621	0.727
	0.2	6	11.5	0.385	0.326	0.621	0.727
	0.5	6	11.5	0.474	0.326	0.621	0.727

Summary of results

- HP filter with more smoothing and a higher threshold improves accuracy by 20% relative to baseline HP filter
- As the weight on robustness increases, growth rate improves on the HP filter.
- AE subsample: better performance overall, optimal to smooth more and use a higher threshold
- EME subsample: poor performance overall, generally optimal to smooth less, and use a lower threshold

Why do our EWIs perform poorly for EMEs?

- Similar frequency of crises across both subsamples
- Average EME sample length (43 years) is adequate, and similar to that of AE subsample (53 years)
- Truncated AE sample produces similar results as the full-length AE sample

Differences between AE and EME subsample results likely reflect importance of other vulnerabilities for predicting EME crises (e.g. excessive government or current account deficits)

Conclusion

- Current BIS recommendation for calculating credit gaps are not optimal for EWIs
- If policymaker values robustness, simple growth rate produces best results overall
- EWIs perform much better for AE subsample, with more smoothing and a higher threshold (relative to full sample)
- EWIs perform poorly for EME subsample, more research is needed to understand why
- Our approach can be extended to other gap-based EWIs (including composite EWIs)

Appendix: HP Filter

The HP filter decomposes a time series y_t into a trend (g_t) and a cyclical component (c_t). Thus, we have $y_t = g_t + c_t$.
The two-sided (full sample) HP filter solves

$$\arg\min \sum_{t=1}^{T} (y_t - g_t)^2 + \lambda \sum_{t=2}^{T-1} (g_{t+1} - 2g_t + g_{t-1})^2. \tag{4}$$

The smoothness of g_t is determined by the λ parameter; the larger the choice of the λ, the smoother the trend.
The one-sided version of the filter (using data up to time $T' \leq T$) solves a similar equation

$$\arg\min \sum_{t=1}^{T'} (y_t - g_t)^2 + \lambda \sum_{t=2}^{T'} (g_t - 2g_{t-1} + g_{t-2})^2. \tag{5}$$

The Hamilton filter uses ordinary least squares regression on four lags of the observed quarterly values of a data series plus a constant to predict the series h-steps ahead.

$$y_{t+h} = \nu + \sum_{j=0}^{J} \beta_j y_{t-j} + u_{t+h},$$

where $J = 4$. The cyclical component is estimated using the residuals as follows,

$$\hat{u}_{t+h} = y_{t+h} - \hat{\nu} - \hat{\beta}_0 y_t - \hat{\beta}_1 y_{t-1} - \hat{\beta}_2 y_{t-2} - \hat{\beta}_3 y_{t-3}$$

Similar to λ in the HP filter case, higher values of h will result in a smoother trend. [1]

[1] With quarterly data, Hamilton recommends $h = 8$ for detrending variables that co-move with the business cycle, and $h = 20$ for detrending variables that co-move with the credit cycle.

Appendix: Bayesian STM

A state vector $\zeta_t = (u_t, \beta_t)'$ where u_t is a time-varying local level and β_t is the time-varying local growth rate. The dynamics are given by the equations

$$\begin{aligned} Y_t &= u_t + v_t \quad v_t \sim N(0, V) \\ u_t &= u_{t-1} + \beta_{t-1} + w_{1,t} \quad w_{1,t} \sim N(0, \sigma^2_{w1}) \\ \beta_t &= \beta_{t-1} + w_{2,t} \quad w_{2,t} \sim N(0, \sigma^2_{w2}) \end{aligned}$$

Error variance V is the smoothing parameter. It affects the size of the deviations credit-to-GDP from its trend. The system variances $\sigma^2_{w1} = 1$ and $\sigma^2_{w2} = 0.01$ where $W = diag(\sigma^2_{w1}, \sigma^2_{w2})$. [2]
The trend u_t is estimated using the Kalman filter. For a full sample (two-sided) trend computation, we use a Kalman smoothing algorithm on the Kalman filter output.

[2] The behavior of Y_t is dependent on the ratio of W/V. Thus fixing the numerator (W) and varying the denominator (V) decreases the dimensionality of the problem.

Appendix: Growth Rate

The 'smoothing parameter' in this case is the number of lags (the size of the differencing window). [3]

For a series c in real time, a q-quarter differencing would yield:

$$g_t = \frac{c_{t-q} - c_t}{c_{t-q}}. \tag{6}$$

The full sample would yield:

$$g_t = \frac{c_{t-q/2} - c_{t+q/2}}{c_{t-q/2}}. \tag{7}$$

[3] For a full sample version, we use a centered window, while in the real time version we use a purely backward looking window.