Which Credit Gap Is Better at Predicting Financial Crises? A Comparison of Univariate Filters*

Mathias Drehmann and James Yetman
Bank for International Settlements

The credit gap, defined as the deviation of the credit-to-GDP ratio from a one-sided HP-filtered trend, is a useful indicator for predicting financial crises. Basel III therefore suggests that policymakers use it as part of their countercyclical capital buffer frameworks. Hamilton (2018), however, argues that you should never use an HP filter, as it results in spurious dynamics, has endpoint problems, and its typical implementation is at odds with its statistical foundations. Instead he proposes the use of linear projections. Some have also criticized the normalization by GDP, since gaps will be negatively correlated with output. We agree with these criticisms. Yet, in the absence of clear theoretical foundations, *all* proposed gaps are but indicators. It is therefore an empirical question which measure performs best as an early-warning indicator for crises. We run a horse race using expanding samples on quarterly data from 1970 to 2017 for 41 economies. We find that credit gaps based on linear projections in real time perform poorly when based on country-by-country estimation, and are subject to their own endpoint problem. But when we estimate as a panel, and impose the same coefficients on all economies, linear projections perform marginally better than the baseline credit-to-GDP gap, with somewhat larger improvements

*We thank an anonymous referee, Claudio Borio, Itamar Caspi, Stijn Claessens, Jim Hamilton, Daniel Leff, Moritz Schularick, and seminar participants at the BIS and the Reserve Bank of New Zealand for helpful comments. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Bank for International Settlements. Author contact: Drehmann: Centralbahnplatz 2, CH-4002 Basel, Switzerland, mathias.drehmann@bis.org; Yetman: Representative Office for Asia and the Pacific, 78th Floor, Two IFC, 8 Finance Street, Central, Hong Kong, james.yetman@bis.org.

concentrated in the post-2000 period and for emerging market economies. The practical relevance of the improvement is limited, though. Over a 10-year horizon, policymakers could expect one less wrong call on average.

JEL Codes: E44, G01.

1. Introduction

Excessive credit growth has long been recognized as integral to financial booms and busts (Minsky 1982; Kindleberger 2000). However, what constitutes growth being "excessive" remains undefined. Borio and Lowe (2002) propose a credit-to-GDP gap measured by the deviations of the credit-to-GDP ratio from a one-sided Hodrick-Prescott (HP) filter with a large smoothing parameter (400,000 for quarterly data). Borio and Drehmann (2009), Drehmann et al. (2010), and Drehmann, Borio, and Tsatsaronis (2012) revisit the gap in light of the crisis and do extensive comparisons of its early-warning indicator (EWI) properties for systemic banking crises with other variables. They identify the credit-to-GDP gap as the best single EWI across those that they examine. Their work underpins the choice of the Basel Committee for Banking Supervision (BCBS) to single out the credit-to-GDP gap as a useful guide for setting countercyclical capital buffers (BCBS 2010b).

But the credit-to-GDP gap is only one possible indicator of excessive credit growth.[1] Following the work of Jordà, Schularick, and Taylor (2011), for example, the academic literature has mainly relied on medium-term growth rates in credit-to-GDP. In addition, the HP-based gap has been challenged on conceptual grounds. We address two such challenges here.

Most importantly, many have criticized the use of the HP filter to derive the gap. It has long been known that the HP filter has serious problems. These are succinctly summarized by Hamilton (2018). In particular, the HP filter results in spurious dynamics that

[1] Since Borio and Lowe (2002), Bank for International Settlements (BIS) authors have always been careful in emphasizing this point (for an overview see Drehmann and Tsatsaronis 2014). This is also one reason why there is no mechanical link between the credit gap and the countercyclical capital buffer under the Basel III rules (BCBS 2010b).

are not found in the underlying data, results in filtered data with properties that differ between the middle and ends of the sample, and its typical implementation is at odds with its statistical foundations.[2] Hamilton therefore concludes that you should never use the HP filter for any purpose, including for deriving credit-to-GDP gaps. He proposes the use of linear projections as an alternative to derive deviations from trends.

In addition, some authors have criticized the use of GDP to normalize the level of credit in the economy. For instance, Repullo and Saurina (2011) point out that the credit-to-GDP gap will tend to be negatively correlated with GDP, and its use could exacerbate the procyclicality of macroprudential policy. Similar problems were highlighted by the Basel Committee (BCBS 2010b). Real credit per capita has been proposed as an alternative measure to overcome this potential drawback.

From a conceptual perspective, we agree with these criticisms. But, in the absence of clear theoretical foundations, *any* proposed gap measure is nothing more than an indicator, including when derived with more sophisticated empirical methods.[3] What should matter to policymakers is the relative performance of different possible measures, which can be assessed empirically.

In this paper, we therefore run a horse race between different proxies for excessive credit. Given that excessive credit is unobservable, we assess performance based on how well different credit gaps predict systemic banking crises. Performance is judged by the "area under the curve" (AUC), a summary measure of its predictive power. And we focus on (quasi) real-time information, which is the relevant case for policymakers who can only use the information they have available at each point in time to predict a crisis.[4]

[2] Relatedly, Edge and Meisenzahl (2011) document a large difference between real-time and full-sample estimates of credit-to-GDP gaps due to the endpoint problem of the HP filter.

[3] For instance, Buncic and Melecky (2014) and Juselius and Drehmann (2020) derive credit gaps based on multivariate VARs.

[4] We shorten "quasi real time" to "real time" for the remainder of the paper. Our real-time estimates use only data up to time t to estimate gaps at time t, with the sample expanding with each observation. But the data we use are those available at the time of estimation, rather than those available at time t, and hence are not truly "real time."

To keep the analysis concise, we split it into two parts. First, we compare different possible formulations of the linear projection, given the lack of exploratory work elsewhere. We examine a wide range of different combinations of lags in the underlying regression, and also consider projections based on equations estimated both economy-by-economy and in a panel, where the coefficients on the lags are constrained to be the same for all economies.

In a second step, we select the best performing of these linear projections and compare it against two alternative means of deriving "gaps": the HP trend and 20-quarter changes in credit. For each of these measures, we consider two means of normalizing credit, either by using nominal GDP or by transforming it into real-credit-per-capita terms.

The key finding for the different ways to derive projection gaps is that it is crucial to estimate the underlying linear equation as a panel instead of running economy-by-economy regressions. The panel approach results in a material improvement in performance across many forecast horizons and subsample specifications. The analysis also points to a potential "endpoint" problem of the linear projection gap, especially for small samples. If we compare the forecast performance of the full sample versus the real-time gaps, the performance of the real-time gaps in the economy-by-economy specification is much weaker. The reason is that during a credit boom—for example, in the early 2000s in the United States—the estimated coefficients increase in real time so that the residuals that the projection gap is based on don't increase sharply, and are hence less likely to signal the impending crisis.

The panel helps to alleviate this endpoint problem. More generally, it points to the benefits of including international data in the assessment of credit gaps for individual economies: perhaps because of the relative rarity of financial crises, there are material gains in using the experiences of other economies to calibrate and assess early-warning indicators.

For practical purposes, it is also interesting to note that different lag structures in the linear projections have a limited effect on crisis prediction performance, provided the included lags are sufficiently long (generally 20 or more quarters).

When we compare the alternative approaches to generate gaps, two findings stand out. First, normalizing by GDP results in superior forecast performance over normalizing by the population. Second, while the estimated projection gap generally has the highest AUC, differences in performance relative to a gap derived by an HP filter or 20-quarter changes tend to be quantitatively small, albeit in many cases statistically significant. Larger differences are only found for the post-2000 period and for emerging market economies.

But despite the statistical results, differences between different methods to derive gaps (at least when normalized by GDP) are not meaningful from a practical perspective. Dealing with the inherent uncertainty in identifying credit booms is more important by an order of magnitude. Across the different specification, and independent of the gap method, around 30 percent of signals are incorrect. And the higher AUC of the projection gap relative to an HP-filtered gap results in issuing 2–3 percentage points less incorrect signals in normal times. If policymakers would mechanically follow these gaps this would imply that, over a 10-year period, they could expect that the indicators would give wrong signals for around three years, independent of the gap they chose. Over the same period, the 2–3 percentage points difference of fewer wrong calls in normal times for the projection GDP gap relative to the HP GDP gap amounts to making the right call in just one additional quarter.

Addressing the underlying uncertainty about predicting crises, rather than the choice between these indicators, is therefore the key challenge. One possible source of improvement is to take a broader range of indicators into account. We do not do so in this paper, since we wish to focus on the debate about different methods to derive credit gaps as one fundamental component in early-warning indicator models, in light of its importance in the Basel III framework.[5] Therefore, alternative methods, including those focused on multivariate measures, are beyond the scope of this paper.[6] But even in

[5]The Basel III framework recognizes that the credit-to-GDP gap can only be a starting point of discussions about countercyclical capital buffers, as authorities should consider all available information (BCBS 2010b).

[6]Multivariate measures have been shown to have the potential to improve forecast performance, starting with Borio and Lowe (2002). Band-pass filters have been used in both univariate (Aikman, Haldane, and Nelson 2015) and multivariate (Drehmann, Borio, and Tsatsaronis 2011) contexts. Galati et al. (2016) extract a financial cycle using a multivariate unobserved-components model on the credit-to-GDP ratio, total credit, and house prices for six economies, and find

these cases, indicators provide incorrect signals, requiring judgment in practice and the recognition that policymaking based on these indicators is fraught with uncertainty.

In the next section, we outline the two challenges to the HP credit gap measure that we examine. Section 3 contains our methodology for comparing the different measures in light of the objective, and section 4 introduces the data. Section 5 provides a detailed analysis of the performance of different linear-projection based gaps, and section 6 compares the best of these against alternative measures. Robustness exercises are discussed in section 7. In section 8, we consider the practical implications of the differences before we conclude.

2. Critiques of the Baseline Credit Gap

Our baseline credit gap was proposed by Borio and Lowe (2002). They suggested measuring the credit gap as deviations of the credit-to-GDP ratio from a one-sided Hodrick-Prescott (HP) filter with a large smoothing parameter (400,000 for quarterly data). This measure has been subject to a number of criticisms. Here we outline two prominent ones: namely that the normalization is problematic, and the HP filter has undesirable properties.

In order to turn the nominal level of credit into a magnitude that is comparable both across time and across countries, it must be normalized in some manner. In our baseline measure, the normalization is to divide nominal credit by nominal GDP. Repullo

---

that the resulting medium-term cycles vary in terms of length and amplitude across countries and over time. Rünstler and Vlekke (2018) conduct a similar exercise and identify medium-term cycles in credit volumes that are linked to GDP performance at longer frequencies than business cycles. Other recent examples include Schüler, Hiebert, and Peltonen (2015), who find that a financial cycle based on the common frequencies of credit and asset prices outperforms the credit-to-GDP gap in predicting systemic banking crises at horizons of one-to-three years; Aldasoro, Borio, and Drehmann (2018), who show that combining various indicators of excessive debt with property prices can help to improve financial crisis prediction; Alessi and Detken (2018), who use "random forest" machine learning methods based on a number of economic and financial indicators and find that this outperforms a logit model based on the same explanatory variables in terms of out-of-sample performance; and Lang et al. (2019), who develop a combined indicator that captures risks stemming from domestic credit, real estate markets, asset prices, and external imbalances and that outperforms univariate early-warning indicators.

and Saurina (2011) suggest that this could be problematic, since it would suggest reducing capital requirements when GDP growth is high and increasing them when GDP growth is low, hence exacerbating the procyclicality of regulations related to bank capital.[7] As discussed, this was already identified as a potential problem by the Basel Committee (2010b), which identified it as one of the reasons why policymakers' judgment is necessary when setting the countercyclical capital buffer. Jordà, Schularick, and Taylor (2017) and Richter, Schularick, and Wachtel (2017) use real credit per capita as their measure of normalized credit instead.

The other key component to measuring a credit gap is the definition of the gap—or, equivalently, defining the trend against which credit will be compared.

Following the original work by Borio and Lowe (2002), the long-term trend of the credit-to-GDP ratio is often calculated by means of a one-sided (i.e., real-time) HP filter. The filter is run in quasi real time, i.e., recursively, with an expanding sample each period. Thus, a trend calculated for, say, end-1998 only takes account of information up to 1998 even if this calculation is done in 2018. The HP filter also uses a much larger smoothing parameter—400,000 for quarterly data—than the one employed in the business cycle literature. This choice can be rationalized by the observation that credit cycles are on average about four times longer than standard business cycles and crises tend to occur once every 20–25 years (Drehmann et al. 2010).[8]

Hamilton (2018) points out some serious potential shortcomings with the HP filter in general—in particular, that

(i) it produces spurious dynamics that are not based on the underlying data-generating process;

---

[7] Also see the discussion in Jordà (2011).

[8] Hodrick and Prescott (1997) set λ equal to 1,600 for extracting business cycles in quarterly data. Ravn and Uhlig (2002) show that, for series of other frequencies (daily, annual, etc.), it is consistent to set λ equal to 1,600 multiplied by the fourth power of the observation frequency ratio, implying λ equal to 400,000 if credit cycles are four times longer than business cycles. Empirically, λ equal to 400,000 also delivers the credit-to-GDP gap with the best forecasting performance (Drehmann, Borio, and Tsatsaronis 2011).

(ii) the dynamics at the ends of the sample differ from those in the middle;[9] and

(iii) the standard implementation of the HP filter stands at stark odds from its statistical foundations.

To avoid these drawbacks, Hamilton suggests an alternative using a "linear projection" based on estimating the equation:

$$y_{t+h} = \beta_o + \sum_{j=1}^{J} \beta_j y_{t+j-1} + \nu_{t+h}. \tag{1}$$

The estimated residual from this equation, ν_{t+h}, is the projection gap that will be assessed as a predictor of financial crises. Richter, Schularick, and Wachtel (2017) implement this method, but with one alteration: they normalize the residuals by their standard deviation, (σ_v), to produce the projection gap.

Hamilton suggests that including four lags ($J = 4$) and a value of h corresponding to five years (i.e., $h = 20$ with quarterly data) for applications to debt (or credit) cycles may be appropriate. But given that the baseline HP-filter-based credit-to-GDP gap has already been carefully tested with different assumptions about the smoothing parameter (see discussion above), we first examine a range of possible formulations of the linear projection model, with varying numbers and lengths of lags, to see how sensitive the results are. We also compare the results when we estimate the underlying equation economy-by-economy versus in a panel with the β_j's constrained to be the same for all economies, while allowing separate fixed effects (β_0's) for each economy.

An alternative approach that we also examine is to detrend by computing growth rates. Taking the 20-quarter change in credit/GDP or real credit per capita provides a filter-free way of extracting a credit gap measure. This approach has been used, for

[9] The baseline credit gaps measure uses a one-sided filter, with observations added recursively. On the one hand, this means that we are never comparing an observation from the middle of the sample with one from the end, mitigating the second critique. On the other hand, given that the gaps are taken from samples of different sizes, their properties could still vary.

example, in Jordà, Schularick, and Taylor (2011) and Jordà et al. (2017).

In the following section, we outline the methodology to assess predictive performance that we use for the horse race between different measures of the credit gap to see how they compare.

3. Assessing Predictive Performance

As discussed in the introduction, all proposed gaps are intended to be indicators of excessive credit growth. In line with a long research tradition, we judge performance by how well the different measures predict systemic banking crises.

We follow the literature and use the area under the ROC curve (AUC) as a statistical measure to judge forecast performance.[10] It is a very intuitive measure. To fix ideas, assume a very simple economy that is in one of two states: $S = 0$, or $S = 1$. States are not directly observable, but a gap measure, G, carries imperfect information about the current state. In particular, the higher the value of G, the more likely it is that $S = 1$. In an ideal situation, there would be a threshold θ^i such that, if $G > \theta^i$, we would know that $S = 1$ (and $S = 0$ for $G \leq \theta^i$). But, if the signal is noisy, there is a tradeoff between the rate of true positives, $TPR\left[S\left(\theta^i\right)\right] = P(G > \theta^i|S = 1)$, and the rate of false positives, $FPR\left[S\left(\theta^i\right)\right] = P(G > \theta^i|S = 0)$.[11] For very low values of the threshold, for instance, the TPR will be close to one, but the same will also hold for FPR. We therefore look over all thresholds θ^i. And the mapping from FPR to TPR for all θ^i gives the ROC curve.

The area under this curve, the AUC, can interpreted as the likelihood that the distribution of G when $S = 1$ is stochastically larger

[10] ROC stands for receiver operating characteristic. The somewhat awkward name goes back to its original use of trying to differentiate noise from signals of radars during World War II. Since then it has been used in many other sciences (e.g., Swets and Picket 1982). Over the last 10 years it has become increasingly popular in the context of crises or recession predictions, following in particular the work of Oscar Jordà (e.g., Berge and Jordà 2011, or Jordà, Schularick, and Taylor 2011).

[11] The FPR and the complement of the TPR correspond to the familiar type II and type I errors.

than when $S = 0$. It is a convenient and interpretable summary measure of the signaling quality. A completely uninformative indicator has an AUC of 0.5. Correspondingly, the AUC for the perfect indicator equals 1. The AUC of an informative indicator falls in between and is statistically different from 0.5. For two competing indicators, G_1 and G_2, it is also easy to test whether $AUC(G_1)$ is equal to $AUC(G_2)$ by using a Wald test.

We estimate the AUC nonparametrically with Stata. Standard errors are bootstrapped using 1,000 replications. We cluster at the country level. The Wald test for equality of AUCs also uses the joint bootstrap estimated variance-covariance matrix. As such we account for the very high correlation between different gap measures, often in the range of 0.9.[12]

For practical policy proposes, in addition to statistical power to predict crises, the right timing and stability of signals are important (Drehmann and Juselius 2014). EWIs need to signal a crisis early enough so that policy actions can be implemented in time to be effective. Yet, EWIs should not signal crises too early, as there are costs to macroprudential policies, and early adoption could undermine the support for necessary policy measures (e.g., Caruana 2010). EWIs should also be stable, as policymakers tend to base their decisions on trends rather than reacting to changes immediately (e.g., Bernanke 2004). A gradual implementation of policy measures may also allow policymakers to influence market expectations more efficiently, and to deal with uncertainties in the transmission mechanism (Committee on the Global Financial System 2012).

To assess the appropriate timing of a gap measure G_i, we follow Drehmann and Juselius (2014) and compute $AUC(G_{ij})$ for all horizons j within a three-year window before a crisis, i.e., j runs from -12 to -1 quarters.[13] When we compute $AUC(G_{ij})$, we ignore signals in all other quarters than j in the window. For example, at horizon -6, the rate of correctly predicted crises is solely determined by signals issued six quarters before crises. False alarms, on

[12]The high correlations are unsurprising given that all gaps are based on credit either normalized by GDP or population.

[13]By looking at each horizon separately, we wish to draw attention to the temporal stability of the EWIs, which is important for policymaking, rather than to the average time pattern.

the other hand, are based on all signals issued outside the three-year window before crises occur. We also do not consider signals issued during a crisis, as binary EWIs become biased if the crises periods are included in the analysis (Bussiere and Fratzscher 2006).

4. Data

Our data cover 41 economies.[14] We use quarterly data with samples from as early as 1970 (depending on data availability) to derive the trend. The sample ends in the third quarter of 2017.

In our baseline specification to test forecast performance we only include gaps for an economy once we have 15 years of quarterly data, leading to an earliest date of 1985:Q1 in the horse race.[15] This is necessary to ensure adequate data for the calculation of trends with the HP filter or regression coefficients with the linear projections.[16] This starting point also approximately coincides with when many countries liberalized their financial systems, which in turn affected the dynamics of financial cycles and their relation with financial crises (Borio 2014). For a small number of economies, we further delay their inclusion in the panel until an end of a crisis: there is little practical point in beginning to test for crises when an economy is already in one.

Our measure of credit is as published in the BIS database of total credit to the private nonfinancial sector (see Dembiermont, Drehmann, and Muksakunratana 2013), capturing total borrowing from all domestic and foreign sources. Our nominal GDP series used

[14] The sample includes Argentina, Australia, Austria, Belgium, Brazil, Canada, Chile, China, Colombia, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hong Kong SAR, Hungary, India, Indonesia, Ireland, Italy, Japan, Korea, Malaysia, Mexico, the Netherlands, Norway, the Philippines, Poland, Portugal, Russia, Saudi Arabia, Singapore, Spain, South Africa, Sweden, Switzerland, Thailand, Turkey, the United Kingdom, and the United States.

[15] For 20 economies in our sample, data are available from 1970:Q1, so they are included from the start in 1985. By 2000:Q1, 28 economies are included in the sample.

[16] As a robustness check we also used a run-in period of only 10 years with qualitatively similar results. However, we prefer the 15-year specification, as 10 years of data for country-specific projections are rather limited. On this basis, one could even argue for a longer sample such as 20 years to estimate stable trends. By doing so, the crises of the late 80s would, however, drop out of the horse race, severely limiting the number of observed crises episodes.

to generate credit-to-GDP are drawn from national sources. To generate the capita gaps, we use CPI from national sources and population numbers from the International Monetary Fund and the World Bank.

In total we have 27 crises in our sample. For crisis dating, we rely on the new European Systemic Risk Board crisis data set (Lo Duca et al. 2017) for European countries and on Drehmann et al. (2010) for the rest.[17] As discussed, we drop post-crisis periods as identified in Lo Duca et al. (2017) and Laeven and Valencia (2012) for European and non-European economies, respectively.

5. A Horse Race between Linear Projection Gaps

Our first exercise is to compare different linear projection gaps in order to get a sense of which performs best. We started with a broad set of options, with h ranging from 4 to 36 quarters, and one to eight lags included in the equation. In all cases, we considered two normalizations of credit, namely by GDP (i.e., the credit-to-GDP ratio, with both credit and GDP measured in nominal terms) and per capita (that is, nominal credit divided by the product of the level of the CPI and the population). These different normalizations are indicated by "GDP" and "capita," respectively.

When using real credit per capita, we face a scaling issue. The reason is that real credit per capita is measured in units of local currency, normalized by the CPI and population. National currencies have, however, very different units, as indicated by simple dollar exchange rates ranging from below one to multiples of thousands. While the growth gap method is invariant to scaling, this is not the case for the HP gap or the projection gap. To overcome the scaling problem for the per capita normalizations, we take natural logs of normalized credit. The gap measure may then be interpreted as the percentage difference between the level and the underlying trend.

We also perform the estimation both economy-by-economy and as a panel with economy fixed effects but with other coefficients constrained to be identical across all economies. We do this recursively,

[17] We exclude crises related to transitioning economies or that were imported from abroad based on Lo Duca et al. (2017). In addition, we classify the crisis in 2008 in Switzerland as imported.

adding one quarterly observation at a time to an expanding sample. With each recursion we take the final residual as a measure of the credit gap in that period. This approach is consistent with the idea that we require a measure that is useful in real time; in the same way, our HP-filter results will be based on a one-sided filter.

As well as adding observations with each recursion, we also add economies as data become available. For comparability between panel and economy-specific estimation, we only include an economy in the panel once we have 15 years of data for reasons discussed above.

Consistent with the intuition of Hamilton (2018), linear projections based on low values of h do not perform well.[18] In addition, performance generally drops off with additional lags. We hence report a range of results for $h \in \{20, 24, 28, 32, 36\}$, each with one, two, and four lags. Combined with two different normalizations and both economy-by-economy and panel estimation, we are comparing the AUCs of 60 different formulations of linear projections for each of 12 different horizons.

The key takeaways are summarized in figure 1.[19] For each panel, the solid line in figure 1 represents the AUC at different horizons, up to 12 quarters. Symmetric dotted lines indicate 95 percent confidence bands around the point estimates. In addition, the dot-dash black lines indicate the results for our, ultimately, preferred specification for the projections gaps, based on the panel estimates, normalized by GDP and with lags 28 and 29.

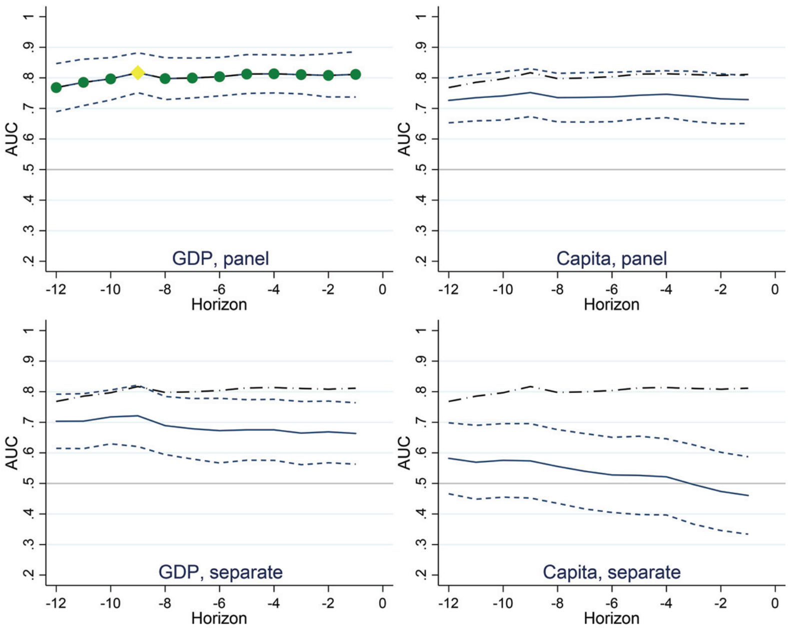
To highlight difference across the specifications, we add yellow diamonds and green dots (see online version of paper for figures in color). They are defined as follows:

- Yellow diamonds: Highest AUC across all of the 60 specifications at that given forecast horizon;
- Green dots: AUC is not statistically different from the highest AUC at a 95 percent confidence level, based on bootstrapped critical values using 1,000 replications.

[18] Full results are available on request.

[19] The figures and tables of underlying data for all 60 different formulations are shown in the online appendix (figure OA1 and table OA2), available at http://www.ijcb.org. Also available are all figures in color.

Figure 1. AUCs for Different Measures of the Linear Projection Gap Based on Lags 28–29



Notes: AUCs for different forecast horizons based on lags 28–29. A dot-dash line indicates the results for panel estimation with GDP normalization on lags 28–29, for ease of comparison. For the full set of graphs for h∈{20, 24, 28, 32, 36}, each with one, two, and four lags, please see the online appendix figure OA1. Horizon: quarters before crises. Solid line: point estimates; dashed lines: 95 percent confidence intervals. Yellow diamond: highest AUC across the 60 specifications at that given forecast horizon. Green dot: AUC is not statistically different from the highest AUC at this horizon at 95 percent confidence level, based on boot-strapped critical values using 1,000 replications.

Comparing the results, two results are evident from figure 1:

(i) Normalizing by GDP statistically dominates normalizing by population.

(ii) The panel estimation dominates estimation by each economy separately.

We therefore focus on cases where credit is normalized by GDP, and the linear model is estimated as a panel.

Reading across all different specifications in the online appendix (figure OA1), it is also clear that:

(iii) The lag length and choice of h make little difference for the predictive performance of the different projection gaps, at least when h is between five and nine years.

This is not surprising, as the different gaps share very similar cyclical properties, with an average cycle length of 16 years.[20] Given that we judge performance by the AUC, we ultimately chose the specification with the highest AUC on average. As such we focus on the model with $h = 28$ and two lags as our preferred linear projection model for the remainder of the paper.[21] However, as a robustness check, we will also assess the original specification for the projection gap suggested by Hamilton (2018) later.

To further uncover the sensitivity of the linear projection gap's performance to modeling assumptions, figure 2 reports the AUCs for the projection gaps based on real-time information versus over the full sample; estimated economy-by-economy (labeled "separate") versus as a panel; and normalized by GDP versus population. All panels in figure 2 are based on our preferred projection specification.

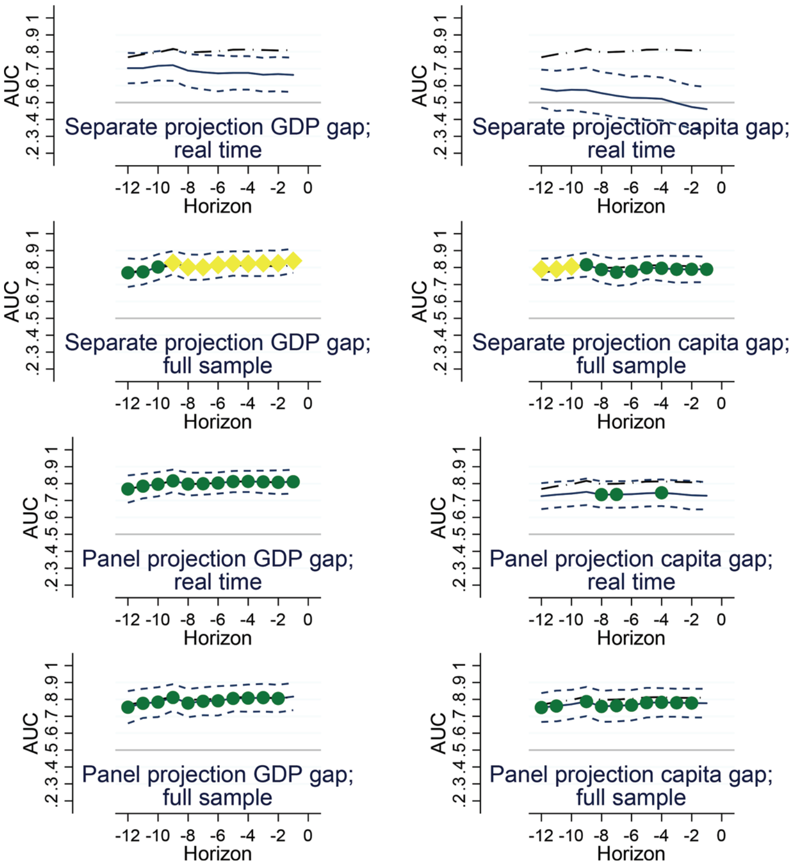
In line with the results above, it is clear that normalizing by GDP generally generates higher AUCs than normalizing by the population, especially when estimating in real time. The improvement based on full sample averages 0.06, whereas for real time it is a larger 0.11.

The figure also highlights that estimating using a panel instead of on each country individually makes little difference when applied to the full sample (fourth row of the figure versus the second row;

[20]Cyclical properties, based on a turning-point analysis, are presented in table OA1 in the online appendix for all different gaps discussed in the paper. The average cycle length of 16 years is similar to that of the baseline HP-filtered credit-to-GDP gap. It also in line with the financial cycle literature (e.g., Claessens, Kose, and Terrones 2012; Drehmann, Borio, and Tsatsaronis 2012; and Aikman, Haldane, and Nelson 2015).

[21]The average AUC for including only one lag of 28 quarters is marginally higher at the fourth decimal (the average AUC for using lags 28 and 29 or only lag 28 are 0.803 to three decimal places). We prefer two lags, mindful of the original justification of Hamilton (2018) for proposing four lags for the linear projection: d lags should in principle work with any process up to order $I(d)$.

Figure 2. AUCs for Different Measures of the Linear Projection Gap for $h = 28$ with Two Lags



Notes: AUCs for different forecast horizons. A dot-dash line indicates the results for real-time panel estimation with GDP normalization, for ease of comparison. Horizon: quarters before crises. Solid line: point estimates; dashed lines: 95 percent confidence intervals. Linear projections based on economy-by-economy real-time estimates (top row), economy-by-economy full-sample estimates (second row), panel real-time estimates (third row), and panel full-sample estimates (bottom row). Left column is credit normalized by GDP, and right column is based on real credit per capita. Yellow diamond: highest AUC across the eight specifications at that given forecast horizon. Green dot: AUC is not statistically different from the highest AUC at this horizon at 95 percent confidence level, based on bootstrapped critical values using 1,000 replications.

the average AUC declines by 0.02) but dramatically improves performance when applied to expanding samples (first versus third rows; the average improvement in AUC is 0.16).

This points to a potential "endpoint" problem of the linear projection gap, especially for small samples. The reason is that during a credit boom—for example, in the early 2000s in the United States—the estimated coefficients increase in real time so that the residuals that the projection gap is based on don't increase sharply, and are hence less likely to signal the impending crisis (see figure OA2 in the online appendix).

Finally, using the full sample rather than expanding sample regressions generally improves AUCs, although this is not a practical option for policymakers seeking to construct an EWI. Comparing analogous panels between rows 1 and 2, and also rows 3 and 4, full-sample estimation dramatically improves the AUC when the normalization is by population (by an average of 0.15) or the estimation is economy-by-economy (by 0.19) or both (by 0.26). By contrast, the difference is trivially negative when panel estimation is applied to credit normalized by GDP (−0.01).

These results suggest caution in interpreting some implementations of the linear projection. For example, Richter, Schularick, and Wachtel (2017) use the linear projection-based gap in country-by-country estimation on the full sample based on credit normalized by population with $h = 20$ and four lags. In the context of our figure 2, their results are closest to the second panel on the right column. However, if the objective is to assess the usefulness of measures of the credit gap to policymakers, the real-time results are the relevant ones to focus on. These are given in the top-right panel. The point AUCs here are less than 0.5 at some horizons and never statistically significantly different from an uninformative indicator, indicating that this implementation of the linear projection has no statistical power for predicting crises in real time in our panel.

6. Widening the Field

Given our preferred linear projection model, we now compare it against alternatives. We consider six gaps, as summarized in table 1. As in the previous section, we focus on two different normalizations

Table 1. Different Credit Gap Measure Labels

		Gap Measure		
		Difference from One-Sided HP Trend	**Five-Year Growth**	**Residual from Linear Projection**
Normalization	**GDP**	HP GDP Gap (Baseline)	Growth GDP Gap	Projection GDP Gap
	Real Credit per Capita[a]	HP Capita Gap	Growth Capita Gap	Projection Capita Gap

Note: [a]To overcome the scaling problem of real credit per capita, ln(real credit per capita) is used.

of credit, namely by GDP and per capita. For each ratio, we apply three possible gap measures:

(i) the difference from a one-sided HP-filtered credit with a smoothing parameter of 400,000 (the HP gap);

(ii) 20-quarter (five year) growth rates (the growth gap); and

(iii) the residual from real-time linear projections with $h = 28$ and two lags (the projection gap).

As before, we include the gaps for a country for each of the measures once we have 15 years of underlying credit data for the country.[22]

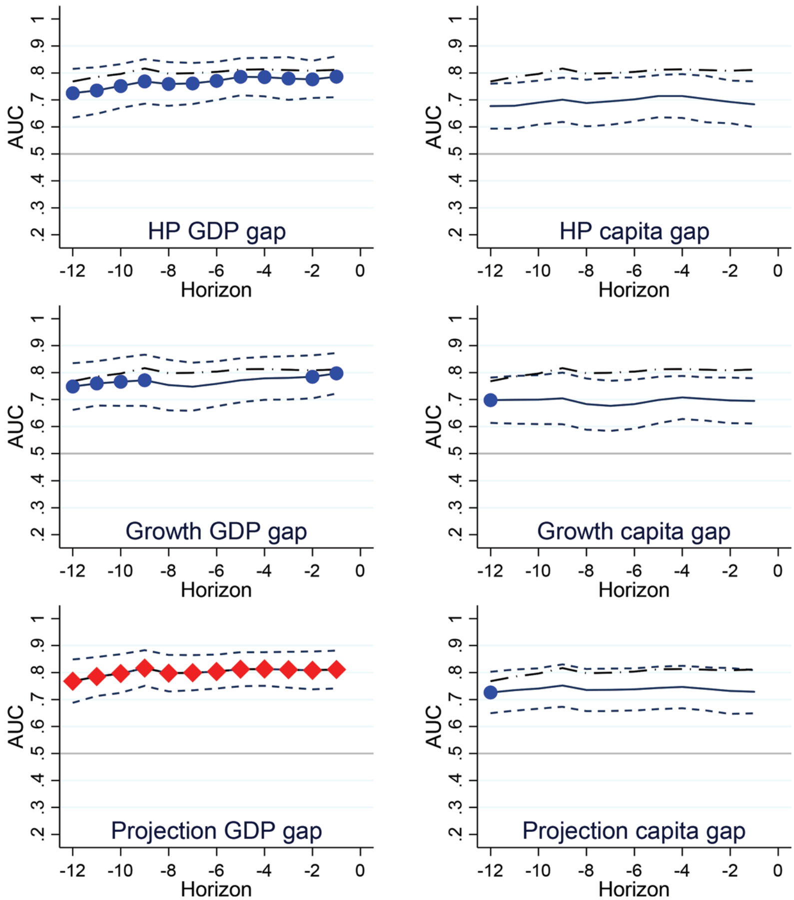
Figure 3 presents the main results (table OA3 in the online appendix shows the underlying statistics). Panels in the left-hand column are based on credit-to-GDP ratios, and the right-hand column on real credit per capita. The top row shows the HP gaps, the middle row the growth gaps, and the bottom row the projection gaps. For each panel, the solid line represents the AUC at different horizons, up to 12 quarters. Symmetric dotted lines indicate 95 percent confidence bands around the point estimates.

The figure summarizes the key takeaways from the horse race: First, normalizing by GDP results in superior forecast performance over normalizing by the population. This holds across all methods to derive the gaps and for most forecast horizons.

Second, the panel projection GDP gap has the highest AUCs of all the different gap measures for all horizons, although the differences vis-à-vis the AUCs of the HP GDP gap are very small and never statistically significant. This is in stark contrast to the comparative performance of the same model when applied economy-by-economy, as we have shown in an earlier version of this paper (Drehmann and Yetman 2018). Then, the HP gap consistently outperforms the projection gap. The forecast performance of the growth GDP gap is also not much worse, albeit with some significant differences to the projection gap.

[22] Graphs of the underlying credit gap data, by economy, are available in the online appendix (figure OA3).

Figure 3. AUCs for Different Measures of the Credit Gap



Notes: AUCs for different forecast horizons. A dot-dash line indicates the results for the projection GDP graph, for ease of comparison. Horizon: quarters before crises. Solid line: point estimates; dashed lines: 95 percent confidence intervals. Red diamond: highest AUC across the six specifications at that given forecast horizon. Blue dot: AUC is not statistically different from the highest AUC at this horizon at 95 percent confidence level, based on bootstrapped critical values using 1,000 replications. See the online appendix for the underlying statistics (table OA3).

Third, while the differences are sometimes statistically significant, they are generally not large from a policy perspective. The average AUC differences across horizons between the best performer and the other gaps are less than 0.04 for both the HP GDP gap and the growth GDP gap and 0.06 for the projection capita gap.

7. Robustness Checks

As robustness checks, we consider splitting the sample in three different ways: by time, between advanced and emerging market economies, and between countries that experienced a (domestically driven) crisis during the GFC and those that did not. We also compare the result with those obtained using the original specification for the projection gap suggested by Hamilton (2018) for credit gap calculations, based on lags 20–23 (instead of 28–29). To preserve space, we only show the gaps normalized by GDP; full versions of the graphs and the underlying data are reported in the online appendix.

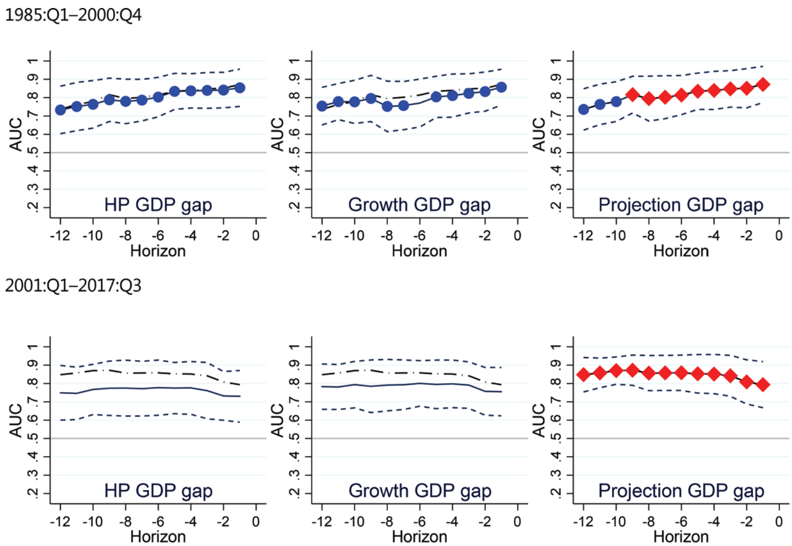
For the first exercise, we split the sample at the end of 2000. The results are reported in figure 4. They illustrate the key role that later periods play in the strong performance of the projection GDP gap. This measure no longer has the highest AUC at the longest horizons for the early sample split, although it is never statistically significantly different from the best performer. However, in the later subsample the projection GDP gap is the best-performing EWI at all horizons, and the difference is always statistically significant.

We next compare advanced and emerging market economies in figure 5. For the advanced economies there is little to choose between any of the measures statistically at most horizons. By contrast, for emerging market economies (EMEs) AUC performance is lower and more dispersed and confidence bands are much wider, suggesting that crisis prediction is inherently more difficult in EMEs.

Results also seem not to be driven by the global financial crisis (GFC): they are very similar for economies that had a domestically driven crisis during the GFC and those that did not (figure 6).

Finally, we compare the results with those based on projection gap parameters originally suggested by Hamilton (2018), using lags 20–23, to see how sensitive our results could be to the risk of overfitting of the projection equation. The results are displayed in figure 7. While there are differences between this specification and the one using lags 28–29, these are quantitatively small and the projection GDP gaps continues to perform best, by a small margin, at all horizons.

Figure 4. AUCs for Different Measures of the Credit Gap: Different Time Periods



Notes: AUCs for different forecast horizons. A dot-dash line indicates the results for the projection GDP gap for the respective time periods. Panels based on normalizing by population are excluded to save space, but are available in the online appendix (figures OA4.1 and OA4.2). Horizon: quarters before crises. Solid line: point estimates; dashed lines: 95 percent confidence intervals. Red diamond: highest AUC across the six specifications at that given forecast horizon for the respective time period. Blue dot: AUC is not statistically different from the highest AUC at this horizon at 95 percent confidence level, based on bootstrapped critical values using 1,000 replications (for the respective time period). See the online appendix for the underlying statistics (tables OA4.1 and OA4.2).

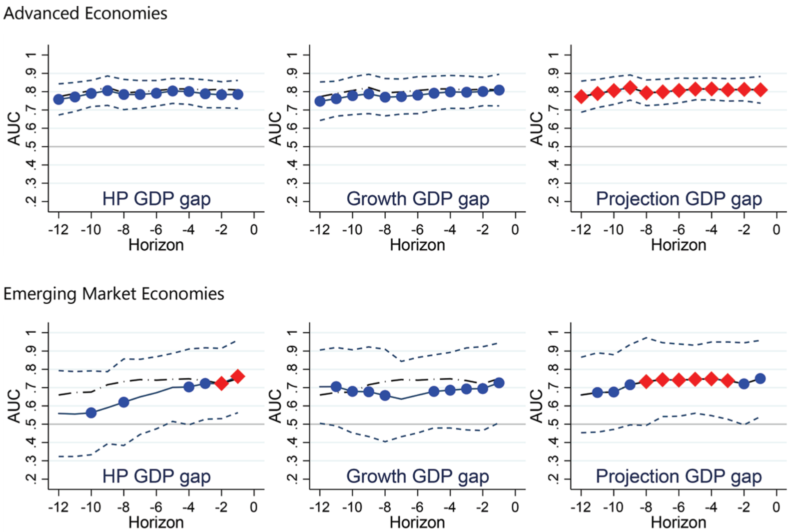
These results support our main takeaways above, that the projection GDP gap is the best-performing EWI overall in our sample, but differences are sometimes small and sample dependent.

8. Practical Implications

The analysis so far has several important practical implications for deriving indicators that signal "excessive" credit growth.

The core takeaway from our first set of results is that, when deriving projection gaps, it is crucial to use a panel approach rather

Figure 5. AUCs for Different Measures of the Credit Gap: Different Country Groups

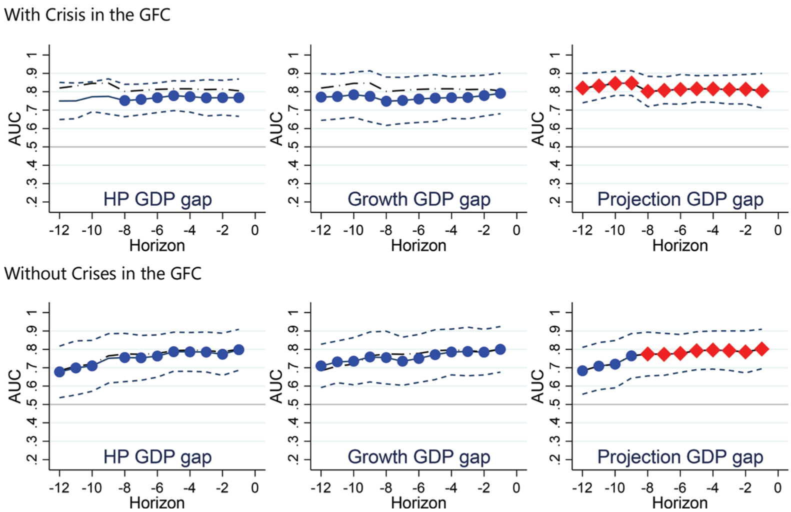


Notes: AUCs for different forecast horizons. A dot-dash line indicates the results for the projection GDP gap for the respective country group. Panels based on normalizing by population are excluded to save space, but are available in the online appendix (figures OA5.1 and OA5.2). Horizon: quarters before crises. Solid line: point estimates; dashed lines: 95 percent confidence intervals. Red diamond: highest AUC across the six specifications at that given forecast horizon for the respective country group. For horizons 9 to 12, the growth capita gap has the highest AUCs. Blue dot: AUC is not statistically different from the highest AUC at this horizon at 95 percent confidence level, based on bootstrapped critical values using 1,000 replications (for the respective country group). See the online appendix for the underlying statistics (tables OA5.1 and OA5.2).

than running country-by-country regressions.[23] It is also important to assess predictive performance by using real-time estimates, as this is what policymakers can do in practice and results can differ significantly from a full-sample analysis. The question of lag length, on the other hand, is second order as long as h is between five and nine years.

[23]While our analysis is clear that a panel approach is important, the optimal panel of countries may differ for specific practical purposes.

Figure 6. AUCs for Different Measures of the Credit Gap: Countries with or without Crises in the GFC

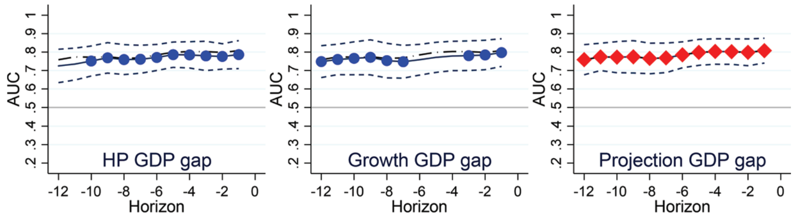


Notes: AUCs for different forecast horizons. A dot-dash line indicates the results for the projection GDP gap for the respective country group. Panels based on normalizing by population are excluded to save space, but are available in the online appendix (figures OA6.1 and OA6.2). Horizon: quarters before crises. Solid line: point estimates; dashed lines: 95 percent confidence intervals. Red diamond: highest AUC across the six specifications at that given forecast horizon for the respective country group. For horizons 9 to 12, the growth capita gap has the highest AUCs for countries without crises in the GFC. Blue dot: AUC is not statistically different from the highest AUC at this horizon at 95 percent confidence level, based on bootstrapped critical values using 1,000 replications (for the respective country group). See the online appendix for the underlying statistics (tables OA6.1 and OA6.2).

Across all the results, it also stands out that normalizing credit by GDP results in superior forecast performance than normalizing by the population.

Our analysis is, however, less clear cut on the best approach to derive gaps. The statistical results (figures 3–7) suggest that the projection GDP gap is marginally better than the HP GDP gap, which in turn somewhat outperforms the growth GDP gap. But despite the statistically significant differences in forecast performance between the different gaps, they are not meaningful from a practical perspective. The main uncertainty policymakers face is that indicators give

Figure 7. AUCs for Different Measures of the Credit Gap, Projection GDP Gap Based on Lags 20–23



Notes: AUCs for different forecast horizons. A dot-dash line indicates the results for the projection GDP gap. Panels based on normalizing by population are excluded to save space, but are available in the online appendix (figure OA7). Horizon: quarters before crises. Solid line: point estimates; dashed lines: 95 percent confidence intervals. Red diamond: highest AUC across the six specifications at that given forecast horizon. Blue dot: AUC is not statistically different from the highest AUC at this horizon at 95 percent confidence level, based on bootstrapped critical values using 1,000 replications. See the online appendix for the underlying statistics (table OA7).

wrong signals: they may miss crises or may issue wrong crises calls in calm times.

To illustrate this, we take the HP and the projection gaps and undertake a simplified analysis where we do not look at 12 different forecast horizons but instead differentiate between no forthcoming crisis (labeled "normal") and pre-crisis periods. The pre-crisis periods are the 12 quarters in the run-up to crises. As before, we drop the observations during actual crises. In this analysis, the AUC of the projection GDP gap (0.80) is higher than the AUC of the HP GDP gap (0.77) but the difference is not statistically significant at the 5 percent level. We then pick, for each of the GDP gaps, one particular threshold which, if breached, is seen as a crisis signal. This threshold is the one with the lowest noise-to-signal ratio that signals at least a 66 percent probability of a crisis in the pre-crisis periods.[24]

[24]This assumes that policymakers are more worried about missing crises than false alarms, and follows some of our earlier work (e.g., Borio and Drehmann 2009; and Drehmann, Borio, and Tsatsaronis 2011). However, the exact specification is arbitrary and many different approaches are possible, and sophisticated policy analysis often uses a range of different rule for robustness (see, e.g., Alessi and Detken 2018).

The identified thresholds are 6.0 for the HP GDP gap and 14.9 for the projection GDP gap.

To highlight the real-time uncertainty, table 2 shows the fraction of correct and incorrect signals in the normal and pre-crisis (i.e., in the 12 quarters before crisis) periods for the full sample and also the average across the robustness checks run in the previous section. Numbers in italics show the fraction of correct/incorrect signals for the individual indicators, while the other numbers provide the percentage of observations where both signals are giving the same or different messages.

The results have important implications from a policy perspective. First, independent of the indicator, around 30 percent of signals are wrong. Second, there is disagreement between the indicators in around 10 percent of the cases. Third, both indicators perform exactly equal in pre-crisis periods. Fourth, the projection gap performs marginally better in normal times by issuing 2–3 percentage points fewer wrong calls.

If policymakers would mechanically follow this rule, this would imply that, over a 10-year period, they could expect that the indicators would give wrong signals for around 3 years with either of the two gaps. Over the same period, the 2–3 percentage points difference of fewer wrong calls in normal times for the projection GDP gap relative to the HP GDP gap amounts to a single quarter. As such, dealing with the inherent uncertainty in identifying credit booms is an order of magnitude more important in practice than the choice between the different credit-to-GDP gaps.[25] Note, however, that despite this inherent uncertainty, all these gaps perform better than simple coin tosses. Thus, using them to calibrate prudential policies improves welfare, the more so if we consider the high typical costs of systemic crisis—100 percent of GDP or more (e.g., BCBS 2010a; Fender and Lewrick 2016).

9. Conclusions

The credit gap, defined as the deviation of the credit-to-GDP ratio from a one-sided HP-filtered trend with a smoothing parameter of

[25] This also holds true if we add the growth GDP gap into the comparison.

Table 2. The Fraction of Correct and Wrong Signals of the HP and Projection Gap in Percent

			Projection GDP Gap				
			Normal		Pre-crisis		
			Correct	**Incorrect**	**Incorrect**	**Correct**	**Total**
Full Sample							
HP GDP Gap	**Normal**	**Correct**	66	5			*71*
		Incorrect	7	22			*29*
	Pre-crisis	**Incorrect**			25	9	*34*
		Correct			9	57	*66*
Total			*73*	*27*	*34*	*66*	
Robustness Checks (Average)							
HP GDP Gap	**Normal**	**Correct**	64	5			*69*
		Incorrect	8	23			*31*
	Pre-crisis	**Incorrect**			27	7	*33*
		Correct			7	60	*67*
Total			*71*	*29*	*33*	*67*	

Notes: Fraction of correct and incorrect signals in normal and pre-crisis periods. A crisis signal is issued if the gap breaches the critical threshold. For the projection gap, the threshold is 14.9. For the HP GDP gap, the threshold is 6.0. If a crisis signal is issued in a pre-crisis period, it is counted as correct. If it is issued in normal times, it is counted as incorrect. Pre-crisis periods are the 12 quarters in the run-up to a crisis. For robustness, we also show the average fraction in each cell for the sample splits shown in figures 4–6 (pre-2001, post-2001, advanced economies, emerging markets, countries that had a crisis during the GFC, countries that did not have a crisis during the GFC).

400,000 (for quarterly data), has been suggested as a useful measure for predicting crises. Two criticisms leveled at this measure are that (i) the normalization may be problematic because of the positive correlation between credit and GDP, and (ii) the HP filter has undesirable properties.

In this paper, we examine alternative measures of the credit gap that have been advocated by others to address these concerns.

We find that credit gaps based on linear projections in real time perform poorly in real time when based on country-by-country estimation. But when we estimate as a panel, and impose the same coefficients on all economies, linear projections perform marginally better than the baseline credit-to-GDP gap, with larger improvements concentrated in the post-2000 period and for emerging market economies, although the differences across the measures are often statistically small. The improvement in performance between the linear projection using a panel instead of applied to individual economies points to the importance of considering international evidence when calculating credit gaps for individual economies. That said, the practical relevance of the improvement is limited. Over a 10-year horizon, policymakers could expect one less wrong call on average.

References

Aikman, D., A. Haldane, and B. Nelson. 2015. "Curbing the Credit Cycle." *Economic Journal* 125 (585): 1072–1109.

Aldasoro, I., C. Borio, and M. Drehmann. 2018. "Early Warning Indicators of Banking Crises: Expanding the Family." *BIS Quarterly Review* (March): 29–45.

Alessi, L., and C. Detken. 2018. "Identifying Excessive Credit Growth and Leverage." *Journal of Financial Stability* 35 (April): 215–25.

Basel Committee on Banking Supervision (BCBS). 2010a. "An Assessment of the Long-term Economic Impact of Stronger Capital and Liquidity Requirements." Bank for International Settlements (August).

———. 2010b. "Guidance for National Authorities Operating the Countercyclical Capital Buffer." Bank for International Settlements (December).

Berge, T., and O. Jordà. 2011. “Evaluating the Classification of Economic Activity into Recessions and Expansions.” *American Economic Journal: Macroeconomics* 3 (2): 246–77.

Bernanke, B. 2004. “Gradualism.” Remarks at an economics luncheon co-sponsored by the Federal Reserve Bank of San Francisco (Seattle Branch) and the University of Washington, Seattle, May 20.

Borio, C. 2014. “The Financial Cycle and Macroeconomics: What Have We Learnt?” *Journal of Banking and Finance* 45 (August): 182–98.

Borio, C., and M. Drehmann. 2009. “Assessing the Risk of Banking Crises — Revisited.” *BIS Quarterly Review* (March): 29–46.

Borio, C., and P. Lowe. 2002. “Asset Prices, Financial and Monetary Stability: Exploring the Nexus.” BIS Working Paper No. 114.

Buncic, D., and M. Melecky. 2014. “Equilibrium Credit: The Reference Point for Macroprudential Supervisors.” *Journal of Banking and Finance* 41 (April): 135–54.

Bussiere, M., and M. Fratzscher. 2006. “Towards a New Early Warning System of Financial Crises.” *Journal of International Money and Finance* 25 (6): 953–73.

Caruana, J. 2010. “The Challenge of Taking Macroprudential Decisions: Who Will Press Which Button(s)?” Speech at the 13th Annual International Banking Conference, Federal Reserve Bank of Chicago, in cooperation with the International Monetary Fund, Chicago, September 24.

Claessens, S., M. Kose, and M. Terrones. 2012. ”How Do Business and Financial Cycles Interact?” *Journal of International Economics* 87 (1): 178–90.

Committee on the Global Financial System (CGFS). 2012. “Operationalising the Selection and Application of Macroprudential Instruments.” CGFS Papers No. 48.

Dembiermont, C., M. Drehmann, and S. Muksakunratana. 2013. “How Much Does the Private Sector Really Borrow? A New Database for Total Credit to the Private Non-financial Sector.” *BIS Quarterly Review* (March): 65–81.

Drehmann, M., C. Borio, L. Gambacorta, G. Jiménez, and C. Trucharte. 2010. “Countercyclical Capital Buffers: Exploring Options.” BIS Working Paper No. 317.

Drehmann, M., C. Borio, and K. Tsatsaronis. 2011. “Anchoring Countercyclical Capital Buffers: The Role of Credit Aggregates.” *International Journal of Central Banking* 7 (4, December): 189–240.

———. 2012. “Characterising the Financial Cycle: Don’t Lose Sight of the Medium Term!” BIS Working Paper No. 380.

Drehmann, M., and M. Juselius. 2014. “Evaluating Early Warning Indicators of Banking Crises: Satisfying Policy Requirements.” *International Journal of Forecasting* 30 (3): 759–80.

Drehmann, M., and K. Tsatsaronis. 2014. “The Credit-to-GDP Gap and Countercyclical Capital Buffers: Questions and Answers.” *BIS Quarterly Review* (March): 55–73.

Drehmann, M., and J. Yetman. 2018. “Why You Should Use the Hodrick-Prescott Filter — At Least to Generate Credit Gaps.” BIS Working Paper No. 744.

Edge, R., and R. Meisenzahl. 2011. “The Unreliability of Credit-to-GDP Ratio Gaps in Real-Time: Implications for Countercyclical Capital Buffers.” *International Journal of Central Banking* 7 (4, December): 261–98.

Fender, I., and U. Lewrick. 2016. “Adding It All Up: The Macroeconomic Impact of Basel III and Outstanding Reform Issues.” BIS Working Paper No. 591.

Galati, G., I. Hindrayanto, S. Koopman, and M. Vlekke. 2016. “Measuring Financial Cycles in a Model-based Analysis: Empirical Evidence for the United States and the Euro Area.” *Economics Letters* 145 (August): 83–87.

Hamilton, J. 2018. “Why You Should Never Use the Hodrick-Prescott Filter.” *Review of Economics and Statistics* 100 (5): 831–43.

Hodrick, R., and E. Prescott. 1997. “Postwar U.S. Business Cycles: An Empirical Investigation.” *Journal of Money, Credit and Banking* 29 (1): 1–16.

Jordà, O. 2011. “Discussion of ‘Anchoring Countercyclical Capital Buffers: The Role of Credit Aggregates’.” *International Journal of Central Banking* 7 (4, December): 241–59.

Jordà, O., B. Richter, M. Schularick, and A. Taylor. 2017. “Bank Capital Redux: Solvency, Liquidity, and Crisis.” NBER Working Paper No. 23287.

Jordà, O., M. Schularick, and A. Taylor. 2011. "Financial Crises, Credit Booms, and External Imbalances: 140 Years of Lessons." *IMF Economic Review* 59 (2): 340–78.

———. 2017. "Macrofinancial History and the New Business Cycle Facts." *NBER Macroeconomics Annual 2016,* Vol. 31, ed. M. Eichenbaum, E. Hurst, and J. A. Parker, 213–63 (chapter 3). Chicago: University of Chicago Press.

Juselius, M., and M. Drehmann. 2020. "Leverage Dynamics and the Burden of Debt." *Oxford Bulletin of Economics and Statistics* 82 (2): 347–64.

Kindleberger, C. 2000. *Maniacs, Panics and Crashes.* Cambridge: Cambridge University Press.

Laeven, L., and F. Valencia. 2012. "Systemic Banking Crises Database: An Update." IMF Working Paper No. 12/163.

Lang, J., C. Izzo, S. Fahr, and J. Ruzicka. 2019. "Anticipating the Bust: A New Cyclical Systemic Risk Indicator to Assess the Likelihood and Severity of Financial Crises." ECB Occasional Paper No. 219.

Lo Duca, M., A. Koban, M. Basten, E. Bengtsson, B. Klaus, P. Kusmierczyk, J. Lang, C. Detken, and T. Peltonen. 2017. "A New Database for Financial Crises in European Countries: ECB/ESRB EU Crises Database." European Systemic Risk Board/ECB Occasional Paper No. 194.

Minsky, H. 1982. "Can It Happen Again? Essays on Instability and Finance." Armonk, NY: M. E. Sharpe.

Ravn, M., and H. Uhlig. 2002. "On Adjusting the Hodrick-Prescott Filter for the Frequency of Observations." *Review of Economics and Statistics* 84 (2): 371–76.

Repullo, R., and J. Saurina. 2011. "The Countercyclical Capital Buffer of Basel III: A Critical Assessment." CEPR Discussion Paper No. 8304.

Richter, B., M. Schularick, and P. Wachtel. 2017. "When to Lean against the Wind." CEPR Discussion Paper No. 12188.

Rünstler, G., and M. Vlekke. 2018. "Business, Housing and Credit Cycles." *Journal of Applied Econometrics* 33 (2): 212–26.

Schüler, Y., P. Hiebert, and T. Peltonen. 2015. "Characterising the Financial Cycle: A Multivariate and Time-varying Approach." ECB Working Paper No. 1846.

Swets, J., and R. Picket. 1982. *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory.* New York: Academic Press.