Basel III and the Prediction of Financial Crises*

Simon van Norden[a,b,*], Marc Wildi[c]

[a] *Finance Department, HEC Montréal*
[b] *CIRANO and CIREQ*
[c] *Institute of Data Analysis and Process Design, Winterthur*

Abstract

Basel III allows national regulators to adjust the minimum level of reserves held by financial institutions in response to changing perceptions of the fragility of the financial system. This requires regulators to forecast potential systemic financial crises. The extent to which this is feasible is the subject of considerable controversy, not least because of its international importance to current banking regulation. This paper contributes to this debate by examining the extent to which regulators' preferred variables can usefully forecast systemic banking crises.

We make two original contributions. First, we study how aspects of existing real-time measures of credit cycles affect the predictability of systemic banking crises. We show that simple theory-based modifications to filtering measures recently mandated by regulators can have an important effect on predictive power. We then study the performance of alternatives based on optimal filter designs. We discuss the results in light of the current debate over the expected performance of credit cycles as guide for Countercyclical Capital Buffers.

Keywords:
Basel III, Countercyclical Capital Buffer, Banking Crises, Credit Cycles, ROC analysis, Bandpass filter, Hodrick Prescott filter, Macroprudential regulation,

1. Introduction

This is a paper about the real-time measurement of trends and cycles in aggregate measures of credit.

The Countercyclical Capital Buffer (CCB) is a novel feature of the Basel III regulatory reforms adopted in response to global financial crisis of 2007-08. It provides for national regulators to vary bank capital requirements in response to the perceived needs of the banking system. The aim is to raise capital requirements prior to periods of system-wide stress in the banking sector. The buffer is then released at the onset of a banking crisis to help supply additional liquidity to the banking system when it is most needed. [1]

National banking authorities have begun setting national levels for the CCB based largely upon a measure of the credit cycle proposed by the Bank of International Settlements (BIS) and subsequently endorsed by the European Systemic Risk Board (ESRB) as a guide.[2] However, both note that this measure is ad hoc, does not work well for all countries at all times, and that further work is required to provide more reliable guides for policy. In this paper, we explore alternatives that draw upon the theory of filter design to refine measures of the credit cycle. We provide extensive evidence comparing the performance of existing measures and new measures of credit cycles.

[⋆]The authors would like to thank Mathias Drehmann and Jan P.A.M. Jacobs for helpful comments (without implicating them in any errors or omissions), as well as seminar participants at East Carolina University, the 2012 ANPEC conference, the 2013 Meetings of the Society for Computational Economics, the 7th International Conference on Computational and Financial Econometrics, the 2014 Mathematical Finance Days and the 2015 Joint Statistical Meetings.

[∗]Corresponding author

Email addresses: simon.van-norden@hec.ca (Simon van Norden), wlmr@zhaw.ch (Marc Wildi)

URL: www.svannorden.org (Simon van Norden), blog.zhaw.ch/sef/ (Marc Wildi)

[1] "The countercyclical capital buffer is meant to provide the banking system with an additional buffer of capital to protect it against potential future losses, when excess credit growth in the financial system as a whole is associated with an increase in system-wide risk. The capital buffer can then be released when the credit cycle turns so that the released capital can be used to help absorb losses and reduce the risk of the supply of credit being constrained by regulatory capital requirements. A side benefit of operating the buffer in this fashion is that it may lean against the build-up of excess credit in the first place." BCBS (2010b), p. 3.

[2] BCBS (2010b), EU (2014)

In the next section, we briefly review the proposed functioning of the CCB and the measures of "excessive" credit growth proposed by the BIS and the ESRB. The key feature is the use of a recursive Hodrick-Prescott (HP) filter to estimate credit growth trends and thereby determine whether recent growth has been excessive. However, it is well known that the HP filter has some undesirable features when used in this way.

Subsequently, we provide new evidence on the properties of the HP filter when used as proposed by the BIS and the ESRB, focusing on its ability to accurately measure fluctuations associated with credit cycles and the resulting impact on their predictive power for systemic banking crises. After illustrating some of the problems using simple modifications of the HP filter, we investigate the use of both optimal mean-squared error (MSE) linear bandpass filters as well as McElroy and Wildi (2014)'s customised Direct Filter Analysis (DFA) approach for nonstationary processes. We compare their preditive performance to those of the HP filters presently in use.

We begin in the next section by reviewing the role of the credit cycle as a potential guide for the operation of the CCB and the role of the HP filter in measuring that cycle. Thereafter, we review the data sources and methodology that will be used to assess the effectiveness of operational guides for the CCB. We then more closely examine the suitability of the HP for measuring credit cycles when it is used as proposed by the BIS and the ESRB. We show that some (but not all) simple modifications suggested by filtering theory have predictable effects on its usefulness for operation of the CCB. We then compare the performance of the existing filter to alternatives based on MSE and DFA filters. A final section offers conclusions and suggestions for further research.

2. The Credit Cycle

2.1. The CounterCyclical Capital Buffer

The Basel III framework incorporated many reforms intended to prevent a repeat of the systemic banking crises that had affected many countries after 2007. One of the more novel features was the Countercyclical Capital Buffer (CCB), which gives national regulators the discretion to raise and lower capital requirements in response to the perceived needs of the banking system. The intent is to have a buffer-stock of bank capital which may be drawn down to provide additional liquidity during periods of systemic stress. The Basel Committee on Banking Supervision (BCBS) extensively cite recent empirical work on credit cycles and its role in systemic banking crises as a key part of their rationale for the CCB.[3]

This empirical literature documents cycles in the rate of credit growth and argues that periods of excessive credit growth commonly precede financial crises. In particular, it argues that variations in the ratio of credit to GDP are distinct from business cycles and occur at lower frequencies.[4] Several studies, including Drehmann, Borio and Tsatsaronis (2011), Edge and Meisenzahl (2011), Drehmann and Juselius (2014), Aikman et al. (2014) and Detkin et al. (2014), have examined the relationship between such credit cycles and banking crises in major economies.[5] They present extensive evidence that credit "gaps" (i.e. deviations from longer-term trends) have had economically important and statistically significant predictive power for systemic banking crises.

It is important to note that this forecasting ability is an essential component for the operation of CCBs. Without a means of anticipating when additional liquidity might be required, the alternative would be to simply have permanently higher capital requirements that might be temporarily relaxed during crises. However, higher levels of capital are costly to the extent that they reduce the lending capacity of the banking system. The challenge

[3]See BCBS (2010a). Aikman et al. (2014) rationalise this with a model of strategic complementarities between banks in which small changes in fundamentals generate large swings in aggregate credit, arguing that this is the central feature of the credit cycle. Brunnermeier (2009) discusses the role of credit and liquidity in the 2007-08 financial crisis.

[4]For example, Drehmann, Borio and Tsatsaronis (2011), p. 206 note that credit cycles "... are four times longer than standard business cycles... [and] occur on average once in twenty to twenty-five years..." Borio (2014), Graph 1 illustrates the cycle with a filter designed to capture frequencies between 8 and 30 years. However, Leeper and Nason (2015) take a more skeptical view, citing conflicting evidence.

[5]Laeven and Valencia (2012) and Dembiermont, Drehmann and Muksakunratana (2013) describe much of the data upon this research is based. Christensen and Li (2014) and Schwaab et al. (2014) provide evidence on alternative approaches. Detkin et al. (2014) and Galati and Moessner (2014) provide overviews of this burgeoning recent literature.

is therefore to use the CCB to raise capital requirement only in anticipation of periods of high stress.[6] Skeptics have questioned whether regulators will be able to do so, noting that banks require notice of changes in required capital up to 12 months in advance.[7]

2.2. Credit-GDP Gaps and Bandpass Filters

Recent empirical work on credit cycles and the prediction of banking crises has been remarkably consistent in the way credit gaps are measured. Starting with data on credit available to the private sector (expressed as a share of GDP), a Hodrick-Prescott (HP) filter with a smoothing parameter $\lambda = 400,000$ is used to separate trend and cycle, with the latter becoming the estimate of the credit gap. This particular value of λ is rationalised as roughly equal to $1,600 \cdot 4^4$; the former is the value commonly used with quarterly data to capture business cycle frequencies of 6-32 quarters while the latter is an adjustment factor to capture frequencies four times as long.[8] This specific HP filter, applied to a credit-to-GDP ratio, forms the basis for the credit gap measures examined in Drehmann, Borio and Tsatsaronis (2011), Drehmann and Juselius (2014), Aikman et al. (2014) and Detkin et al. (2014) among others, and is the only credit gap measure specifically suggested by BCBS (2010b) and EU (2014). It has also been far and away the most common reference measure adopted by national regulators.[9]

The HP filter was widely adopted by economists after the 1970s as a means of estimating business cycles. An importance difference between the use of HP filters in historical analysis and their role as an operational guide for policy makers, however, is that the latter require

[6] In introducing the CCB, BCBS (2010a), para. 137 state "It will be deployed by national jurisdictions when excess aggregate credit growth is judged to be associated with a build-up of system-wide risk ... This focus on excess aggregate credit growth means that jurisdictions are likely to only need to deploy the buffer on an infrequent basis."For more details on the intended operation of the CCB, see BCBS (2010a) and BCBS (2010b).

[7] As stated in BCBS (2010a), para. 141 "To give banks time to adjust to a buffer level, a jurisdiction will pre-announce its decision to raise the level of the countercyclical buffer by up to 12 months."

[8] Drehmann et al. (2010), Annex 1.

[9] As of this writing, the authors are not away of any national banking regulator that has publicly preferred the use of any other indicator in setting the CCB.

a gap that can be calculated in real time. This obviously requires a one-sided filter whereas the business cycle literature justifies the use of the two-sided HP filter for *ex post* use, well away from the ends of samples.[10] While the HP filter can be applied to any point in a data sample, its weights (and therefore its transfer function and the cycles it suppresses and emphasises) will vary as a result, a point we examine in detail, below. One result of these changing weights is that cycles estimated with HP filters at the end of sample tend to be substantially revised once more data is available, a point that has been emphasised by St-Amant and van Norden (1997) and Orphanides and van Norden (2002) and others in the context of business cycles, and Edge and Meisenzahl (2011) in the context of credit cycles.

Although the empirical research using the HP filter as a measure of credit gaps is careful to emphasise that only one-sided gap estimates are used, the justification above in terms of the frequencies captured is based on the properties of the two-sided filter. There has been no analysis of the extent to which the one-sided filter now used by regulators actually captures the fluctuations that researchers associate with credit cycles, a point to which we will return, below. Before doing so however, we first describe the data and the methodology that will be used to assess the usefulness for operation of the CCB of any given measure of credit cycles.

3. Evaluating Potential Guides for the CCB

This section sets out the evaluation framework that will be used to gauge the performance of the various measures of credit cycle that we consider below. We begin with a brief review of the data sources used before turning to consider various measures of performance. We conclude this section by showing how the HP-filter measure presently used by policymakers performs. These results will then serve as benchmarks against which we can assess alternative measures of the credit cycle.

[10]See Hodrick and Prescott (1997) and Baxter and King (1999).

3.1. Data Series

Our credit measures are quarterly data published by the BIS as their *long series on credit to private non-financial sectors*. They provide measures covering all sources of credit (which we will refer to as "Total Credit") as well as those covering only credit extended by domestic banks (which we will refer to as "Bank Credit".)[11] Further details on the construction of the series are provided by Dembiermont, Drehmann and Muksakunratana (2013). The series are available from 1940Q2 onwards for 40 advanced and emerging economies, although availability varies widely across nations and only roughly half have data prior to 1970. The series are measured in nominal local currency terms; we follow policymakers in dividing these series by nominal GDP using quarterly series from the IMF's *International Financial Statistics*.[12]

As an example, Figure 1 shows the Total Credit ratios for four major advanced economies: the US, the UK, Switzerland and Japan.[13] It is apparent from the figure that the series are nonstationary and irregular; trend growth rates appear to change over time, sometimes abruptly, and while the overall trend is towards higher credit to GDP ratios, the trend may be reversed for substantial periods of time. Such behaviour complicates the measurement of credit cycles, as we will further discuss below.

The other element that we require for our empirical analysis is a set of dates for the start of systemic banking crises. The most commonly used dates are those of Laeven and Valencia (2012), who identify 35 systemic banking crises for our 38 countries from 1948 onwards as shown in Table 1. Henceforth, we'll refer to these collectively as the LV2012 dates. For each one, they estimate the month in which the banking crisis was deemed to have become

[11] In all cases, we use the series adjusted for breaks by the BIS. Data may be downloaded from http://www.bis.org/statistics/credtopriv.htm. We used the 8 June 2015 update. We did not use the series given for the Euro Area as dates for the onset of banking crises differ by nation.

[12] Seasonally adjusted series were used wherever possible. We also follow Detkin et al. (2014) in using a 4Q moving average of GDP. The length of the IMF series were augmented by additional observations from national sources for BE, CN, DK, GR, IE, IN, MX, NL, PT and SE. We also dropped Saudi Arabia from our analysis as quarterly GDP data were not available.

[13] Results for the Bank Credit ratios were qualitatively similar.

systemic.[14] A few countries suffered more than one banking crisis, as the table below shows, while several saw no systemic crises. The relative frequency of crises also seemed to vary, with crises appearing more frequently in the latter part of the sample. As we will see below, one implication of this is that the overall predictive performance of the various measures we examine may be sensitive to sample period examined. Also, some of the crisis dates shown were effectively excluded from our empirical analysis due to a lack of suitable credit measures, particularly in the earlier years of our sample. We return to this issue, below.

We also consider the impact of an alternative set of crisis starting dates, which we'll henceforth refer to as "Alternative" dates. These modify the LV2012 dates along the lines suggested by Drehmann, Borio and Tsatsaronis (2011) and Drehmann and Juselius (2014).[15] As shown in Table 1, the two sets of dates agree on crisis dates for only about one third of the economies. For roughly another third, the two sets of dates even disagree on the number of crises that occurred. In some cases, these differences are likely to be innocuous (e.g. dating the Global Financial Crisis as 2008Q3 rather than Q4, or the Asian Crisis as 1997Q3 rather than Q4.) More substantial differences, however, include moving the onset of the Japanese banking crisis from 1997Q4 to 1992Q4, and increasing the number of banking crises for both the US and the UK. Given the evident room for differences of opinions among experts, we include these alternative dates simply as a means of showing the sensitivity or

[14]We use the monthly dates given in Sheet 5 "additional details" of the data file accompanying Laeven and Valencia (2012). When available, we used the date shown as line 4 as "Date when crisis became systemic." When this was not available (in the case of CH, FR, HU, IT, PT, RU and SE) we simply used the date shown on line 3 as "Crisis date."

[15]Specifically, our alternative dates are created as follows.

1. We began with the LV2012 dates.
2. In cases where the LV2012 crisis dates differed from those used by Drehmann, Borio and Tsatsaronis (2011), we used the latter.
3. In cases where the resulting crisis dates differed from those used by Drehmann and Juselius (2014), we used the latter. Note that we used their broad set of crisis dates, including crises caused by cross-border exposures.

robustness of our results to plausible differences in crises dates.

3.2. Receiver Operating Characteristics

To compare the usefulness of alternative predictors of banking crises, it is important that we keep in mind the nature of the problem facing banking regulators operating a CCB. First, they are trying to anticipate the *start* of systemic crisis. The two sets of crisis start dates discussed above therefore implicitly define two alternative sets of binary-valued variables which take on the value of 1.0 at the start of crises and 0.0 at all other times.[16] The problem of forecasting a binary variable of interest is commonly analysed as probability forecast using popular methods such as probit or logit analysis. One drawback of such methods is that, by minimising the mean squared error (MSE) of the forecast, they weight both crisis and non-crisis outcomes equally. Given the relative rarity of crises, this places relatively little weight on their correct prediction.

An alternative approach, described by Fawcett (2004) and Jordà and Taylor (2011) and applied in this context by IMF (2011), Drehmann and Juselius (2014) and Detkin et al. (2014) is to use ROC (Receiver Operating Characteristic) analysis. The underlying assumption is that the regulator operating the CCB faces a binary choice for its level: 0% or 2.5%. Together with our binary crisis indicators, this gives us only 4 possible outcomes as summarized in Table 2

Outcomes (a) and (d) are preferred over the other two, which represent type I (outcome c) and type II (outcome b) errors. However, assigning relative values to these four cases (as would be required if we sought to minimize an expected loss function, for example) is problematic.[17] Instead, ROC analysis examines a hit rate (defined here as the fraction

[16]This is the same approach used, *inter alia*, by Drehmann, Borio and Tsatsaronis (2011), IMF (2011), Drehmann and Juselius (2014) and Detkin et al. (2014). This implicitly treats all crises equally. While it might in principle be better to put more weight on the prediction of most costly crises, the lack of a concensus on the measured costs of banking crises has thus far stymied the application of such an approach.

[17]For example, it would require us to estimate the expected reduction in the overall economic costs of a systemic banking crisis conditional on having a higher value of the CCB as well as gauge the expected economic costs of additional restrictions on bank capital in the absence of a systemic crisis.

$a/(a+c)$) and a false alarm rate (defined here as the fraction $b/(b+d)$.)[18]

Given a real-valued indicator X_t and the decision rule $X_t > \alpha \Rightarrow CCB = 2.5\%$ for some threshold value α, we can trace out all feasible combinations of hit rates and false alarm rates by varying α over the full range of X_t. The resulting curve, shown in Figure 2, is called an ROC curve. High values of α imply that the CCB will only rarely be set above 0, giving both a low hit rate and a low false alarm rate (in the lower left of the figure.) Conversely, low values of α imply that the CCB will commonly be at 2.5%, resulting in both higher hit rates and false alarm rates (in the upper right of the figure.)

If X_t contains no useful information about the onset of banking crises, then the CCB would in effect be set randomly and we would expect ROC curve to lie along the main diagonal from $(0,0)$ to $(1,1)$. Indicators with ROC curves that lie above and the left of those of alternative are to be strictly perfered as they permit superior combinations of hit and false alarm rates. In general, however, ROC curves of alternative indicators may cross, which prevents us from ranking their desirability without additional information. Despite this, a popular characterisation of ROC curves is the Area Under the Curve (AUC), which simply measures the surface area of the unit square which lies below the ROC. Curves with higher AUCs are associated with more desirable outcomes, while an AUC of 0.5 implies an indicator with no useful information.

Detkin et al. (2014) argue that banking regulators care at least as much about Type I errors as about Type II.[19] This in turn leads them to suggest the use of the partial AUC (pAUC), which measures only the area under the ROC curve above a given threshold (again, standardised so that an uninformative X_t would be expected to have a $pAUC = 0.5$ while the maximum possible value of the $pAUC = 1$.) Accordingly, we will report such pAUCs below, using Robin et al. (2011) to calculate the area under the ROC curve corresponding to hit rates $\geq 50\%$.[20] Mason and Graham (2002) show that tests of the null hypothesis

[18]The hit rate is also referred to as the True Positive (TP) rate or the sensitivity, while $1-$ the false alarm rate is also referred to as the selectivity or the True Negative (TN) rate.

[19]Detkin et al. (2014), section 2.2.3.

[20]A sufficient, but not necessary, condition to rule out the optimality of hit rates $\leq 50\%$ is that missed

of equal AUCs are equivalent to Mann-Whitney U-statistics, but Robin et al. (2011) note that similar tests are not available for pAUCs. Instead, we follow their suggestion and use bootstrap confidence intervals based on Carpenter and Bithell (2000)'s stratified bootstrap.[21]

An alternative characterisation of the tradeoff facing regulators for a given measure X_t may be constructed by showing the average value that the CCB would have had historically at the onset of systemic banking crises if the regulator had been setting the CCB using $X_t \geq \alpha$ as a guide. In other words, this simply shows how $E(CCB(X_t \geq \alpha)|$ Crisis Starts) varies as a function of α.[22] The same could be done for all other periods (i.e. $E(CCB(X_t \geq \alpha)|$ No Crisis Starts)) Comparing these two functions shows more clearly the expected costs of higher CCB hit rates in terms of capital requirements during non-crisis periods.

3.3. Prediction Timing and Scoring

The ROC analysis described above relies on the classification of actions into "successful" and "unsuccessful" outcomes. In so doing, it abstracts from two important aspects of timing related to the operation of the CCB. In the remainder of this section, we consider each of these two aspects in turn.

First, our goal is to gauge the degree to which a given measure X_t may help regulators in deciding when to *increase* the CCB. As mentioned above, once a systemic crisis begins, the CCB should be lowered. We leave the important question of judging when to lower the CCB in a systemic banking crisis to other research. However, this means that there is no further need to consider increases in the level of the buffer until the crisis has abated. The implication is that the performance of X_t *during* a banking crisis is of no consequence. Accordingly, in our analysis below, we omit the 8 quarters following the start of a banking

alarms are at least twice as costly as false alarms.

[21] To the extent that observations across time and economies are positively correlated, these bootstrap p-values may overstate the true significance of test results. We investigated this by allowing for contagion effects across economies in the bootstrap; the resulting changes in p-values were typically small. Furthermore, as we detail below, we are typically unable to reject the null hypothesis, in which case overstating the significance of the results is relatively innocuous.

[22] Using Basic scoring as described below, this is simply 2.5% times the Hit Rate $a/(a + c)$.

crisis from all our calculations.[23]

Second, because a lag of several quarters is required before institutions can comply with changes in the CCB, we must take this lag into account when evaluating the usefulness of a given measure X_t. We consider three alternative approaches.

1. Our simplest approach assumes that a fixed lag of four quarters is required. Therefore, $X_{t-4} > c$ is counted as a true positive prediction (i.e. a "hit") if and only if t is the first quarter of a systemic banking crisis. Below, we refer to this as "Basic" scoring.
2. We also follow the approach of Drehmann and Juselius (2014) who vary the above lag from 1 to 20 quarters. Results are summarised by showing the pAUC as a function of the prediction horizon. An unusual feature of their approach is that, for a chosen lag h, all values of $\{X_{t-20}, \ldots, X_{t-1}\}$ other than X_{t-h} are ignored and contribute to neither the false alarm nor the hit rate. Below, we refer to this as "DJ2014" scoring.
3. It has been argued that increases in CCB during periods of increased systemic fragility may be beneficial in that they may, in and of themselves, reduce systemic risk.[24] We therefore also consider the entire range of dates from $t - 20$ to $t - 4$ as part of the target for the CCB, scoring each value of X over that period as either a "hit" (a) or a missed signal (c). This has the effect of making positive outcomes much less rare, while simultaneously requiring sustained predictions over longer horizons. Below, we refer to this as "Comprehensive" scoring.

[23]This is the same approach used by Drehmann and Juselius (2014), among others. Results are relatively robust to the precise number of quarters omitted. This reflects the facts that (1) crises are relatively rare, so this has little effect on the overall false alarm rate, and (2) the minimum time span between crisis starts is considerably more than 8 quarters.

[24]For example, BCBS (2010b), Section 2 state that "In addressing the aim of protecting the banking sector from the credit cycle, the countercyclical capital buffer regime may also help to lean against the build-up phase of the cycle in the first place. ... This potential moderating effect on the build-up phase of the credit cycle should be viewed as a positive side benefit, rather than the primary aim of the countercyclical capital buffer regime."

3.4. Base Case Results

Figures 6 to 8 and Table 3 describe the simulated historical efficacy of HP-filtered credit gaps. The 1-sided HP filter used here is the same as that mandated by BCBS (2010b) and EU (2014); a recursively-calculated one-sided filter with a smoothing parameter of $\lambda = 400,000$.[25] Results in the Figures use the LV2012 crisis dates and the Total Credit series; the Table provides additional information on the robustness of these results.[26] Following Drehmann, Borio and Tsatsaronis (2011) we also require the availability of at least 40 quarters of data prior to using the filter. The resulting need for credit series which begin 10 years in advance effectively limits the sample period that we can examine. After pooling results across economies, we are left with 3,613 quarterly observations and 24 of the 36 LV2012 crisis dates or 34 of the 54 Alternative crisis dates listed in Table 1. Put another way, quarters in which crises start represent a bit less than 1% of our sample, suggesting that on average crises are observed no more than about once every twenty-five years.

As shown by the blue curve in the upper panel of 6, the ROC curve for the 1-sided HP(400K) filter lies well above the 45-degree, suggesting that it contains useful predictive information for systemic crises. The jagged appearance of the line simply reflects the limited number of systemic crises in our historical sample; the smallest vertical step simply reflects a change of a single crisis in the overall number (out of 24) successfully predicted. The ROC curve shows that in order to have successfully predicted about half of the crises in the sample, our HP-filtered credit gap would have generated false alarms 24% of the time; an 80% hit rate would have raised the false-alarm rate to 59%. The lower panel of the figure shows the implications of this for the operation of the CCB. For the CCB to have averaged 2.0% at the onset of crises, the average value of the CCB at all other times would have been only about 50 basis points lower. The blue curve in Figure 7 shows the ROC curve using Comprehensive scoring. We can see that the overall position of the curve is little changed,

[25]The Figures and Table also provide information on the performance of a number of alternative gap measures; we defer our discussion of those results to the next section.

[26]Prior to filtering, the Credit/GDP series for each economy were scaled to have mean 0 and a standard deviation of 1 to allow for a meaningful cross-sectional comparison of gaps.

although its shape at high values of the Hit Rate now appears somewhat flatter. Finally, the blue curve in Figure 8 shows us how the pAUC varies as a function of the forecast horizon using the DJ2014 scoring. Values decline slightly from roughly 70% at the shortest horizons to about 60% beyond 6 quarters and slowly recover towards their starting value after 10 quarters.

Table 3 provides estimates of the pAUCs for the 1-sided HP filter in the fifth column and shows the robustness and sensitivity of these results to the choice of data series. Comparing results for Bank Credit and Total Credit, for example, we see that the pAUCs are always similar and that often (but not always) the latter gives better slightly better results. The choice of crisis dates has a larger impact, however, with the Alternative dates always providing pAUCs that are typically higher by 0.06 to 0.10. The various scoring methods also tended to give broadly similar results in this case.

These results summarise the historical performance of conventionally HP-filtered credit cycles as operational guides for the CCB. To understand how representative this performance is and how it might be improved, we turn in the next section to a consideration of the properties of the HP filter.

4. HP Filters for Credit Cycles

The *de facto* adoption of HP-filtered credit cycles as guides for the operation of CCBs has also seen the expression of substantial skepticism about their usefulness. One important criticism has the been the lack of a rigorous structural aggregate economic model justifying their use.[27] In this paper, however, we will instead focus on two related potential statistical criticisms of these estimated credit cycles. First, the historical performance of existing filters may be misleading to the extent that it has been exaggerated by data mining. Second, it may also understate the importance of credit cycles to the extent that these HP filters do a poor job of capturing any "true" credit cycles.

[27]For example, see the discussions in Borio (2014), Detkin et al. (2014), Galati and Moessner (2014) and Leeper and Nason (2015).

The data mining critique is one that is most relevant when looking at the development of the empirical literature as a whole since the onset of the Global Financial Crisis. The research community has mobilised substantial resources to identify variables which might serve as predictors of systemic banking crises. These efforts have focused in particular on data from the most recent several decades for higher and some middle-income economies and a roughly coherent set of crisis dates to be predicted. While useful and important, such a search across indicators (and their transformations) should be expected to produce results that are somewhat over-optimistic due to selection bias and repeated testing. Ideally, forward-looking assessments of the usefulness of the indicators selected by such a process would correct for such bias. However, such corrections are extremely problematic in practice. Another possible check on the results, however, would be to see how theory-based modifications to such indicators affect their performance. With this perspective in mind, we turn below to consider the ability of the HP filter to capture the behaviour of "true" credit cycles. Before doing so, however, we must consider how we might identify such cycles.

In the work that follows, we use a frequency-based characterisation of credit cycles as being those regular fluctuations that occur with a period of between 10 and 30 years. This is consistent with the non-structural spirit of the recent empirical literature on credit cycles, while leaving open the question of whether such cycles are necessarily useful for the prediction of banking crises. It is also consistent with the characterisation of credit cycles as having longer durations than business cycles and a maximum duration of twenty to thirty years.[28]

4.1. Detrending and the HP(400,000) filter

The justification for the use of the HP filter with a smoothing parameter of 400,000 may be traced back to Drehmann et al. (2010) who argued that

- it seemed to work better than other values that they examined, and

[28]For example, see the discussion in Borio (2014), section 2, who stresses the longer duration of credit cycles and includes fluctuations of up to 30 years in his statistical evidence. The maximum length of business cycles is typically taken to be between 8 years (as popularised by Baxter and King (1999)) and 10 to 12 years, as suggested by Burns and Mitchell (1946).

- it fit the characterisation of credit cycles as lasting three to four times as long as business cycles.[29]

The latter conclusion was based on the reasoning that

1. "Using frequency analysis, it can be shown that [a smoothing parameter $\lambda = 1600$] implicitly assumes a business cycle frequency of around 7.5 years."[30]
2. Ravn and Uhlig (2002) show that changing the target frequency by a factor of k requires that we change λ by a factor of k^4.
3. $\lambda = 400,000 \approx 4^4 \cdot 1,600$

The problem with this logic is that relies on the *asymptotic* properties of the *two-sided* HP filter. It also confuses the role of a high-pass filter with that of a band-pass filter.

First, the HP filter functions as an approximate high-pass filter.[31] This means that it eliminates cycles exceeding a certain duration while preserving those with shorter durations. In the case of the asymptotic two-sided HP filter, that critical frequency is given (in radians) by

$$\phi = \pi / \sinh(0.5 \cdot \lambda^{0.25}) \tag{1}$$

For quarterly data, $\lambda = 1650$ gives us a cutoff frequency of 10 years while $\lambda = 133,000$ gives 30 years and $\lambda = 400,000$ gives 39.5 years. The HP($\lambda = 400,000$) filter is therefore designed to capture all regular fluctuations lasting less than 40 years, including those at business cycle frequencies. Of course, the inclusion of the latter may be innocuous if those fluctuations are trivial when compared to those at lower frequencies, a question which we will investigate below.

The properties of the one-sided HP filter differ from those described by the above formulae, and such differences are more pronounced when we apply the filter in small samples. To

[29] Drehmann et al. (2010), Annex 1.

[30] Ibid.

[31] To be precise, the HP filter acts a low-pass filter in estimating a trend. The "cycle" is then simply the deviation of the raw data from that trend. The HP cycle is therefore a high-pass filter.

understand the importance of this problem, we can simply recover the weights used by the 1-sided HP filter in constructing its estimate of the deviation from trend. These are jointly determined by both λ and the sample length T and are shown in Figure 3 for $\lambda = 400,000$ and the range $T = 20, \ldots, 240$ which corresponds to the sample lengths available in applied studies such as Drehmann, Borio and Tsatsaronis (2011) and Drehmann and Juselius (2014).[32] The definition of the HP cycle as the actual value X_t minus the HP trend is responsible for the peak shown at lag 0; the remaining (mostly negative) weights reflect the estimate of the HP trend. We can see that for $T \geq 100$ the weighting functions are almost identical, implying that weights on lags of more than 25 years are negligible. In shorter samples, however, differences in weights are more apparent.

Figure 4 shows the spectral gain function implied by the above HP cycle weights as a function of T. The vertical line at the left marks the frequency corresponding to cycles lasting 30 years; an ideal high-pass filter would have a gain of zero for values to the left of its cutoff frequency and a gain of one for all values to the right. For comparison, the heavy black line shows the gain for a 2-sided HP(400,000) filter with $T = 240$. We see that the gain at frequencies in the pass band fluctuates close to the desired value of 1. The gain falls sharply as we pass durations of 30 years and more, with a gain of 0.5 at roughly the frequency corresponding to 40 years as predicted by equation (1). For the 1-sided HP filter, we again see very similar behaviour for all $T \geq 100$, with a still-lower cutoff frequency than for the 2-sided filter and a gain in the pass-band that is slightly but consistently < 1. For smaller values of T however, we see increasingly large deviations from the ideal filter, with increasing degrees of amplification or suppression of arbitrary ranges in the pass-band.

To help assess the practical importance of these problems for the measurement of credit cycles, Figure 5 compares the various estimates of the credit cycle that we would find for the US if we used the weighting functions shown in Figure 3. This is precisely what we would find if we applied the 1-sided HP(400,000) filter to a rolling window of length T. Again, we see that cycles estimated for values of $T \geq 100$ are virtually identical. However, estimates

[32] Those studies typically limit their analysis to cases where $T \geq 40$, although the latter uses some samples as small as $T = 24$.

from shorter samples often deviate substantially. The figure also shows the estimates of the conventional (i.e. 2-sided, full-sample) HP filter: these also deviate from the one-sided estimates, often by substantial amounts.[33]

The problems of the HP filter in shorter samples should not come as a surprise in this context. The filter is being used to identify trends with a duration of 30-40 years or longer, but it is only provided with much shorter spans of data with which to identify those trends. Unsurprisingly, it performs poorly under such conditions. The fact that relatively short data samples have been included in previous studies may therefore have tended to produce estimates of credit cycles that were less reliable than what should be expected with longer samples. For the Credit/GDP data we analyse below, 11 of the 38 economies we examine have series of only 80 observations or less, implying that a substantial proportion of the HP-filtered credit gaps analysed in previous empirical studies have likely lacked precision.

4.2. Modified HP Filters

Given the above properties of the HP filter, we can now try to understand how changing the properties of the filter should affect its predictive ability for systemic banking crises. We do so under the maintained hypothesis that a "better" measure of the credit cycle should have improved predictive performance. We can then see whether the data are consistent with this hypothesis. If so, this should improve our confidence that the relationship between credit cycles and banking crises is not primarily an artifact of data mining.

We investigate the performance of three alternatives to the 1-sided HP(400,000) filter.

1. a 2-sided HP(400,000) filter
2. a rolling 1-sided HP(400,000) of length $T = 40$
3. a 1-sided HP BandPass filter with $\lambda = (400000, 1600)$

[33] Note that, by construction, the 2-sided estimates must converge to the large-T 1-sided estimates as we approach the end of the sample, which explains the striking similarity of these estimates immediately following the 2008 financial crisis. Edge and Meisenzahl (2011) have previously emphasised the substantial difference between 1- and 2-sided estimates of US credit gaps.

The 2-sided filter is clearly not a filter that is feasible for policy purposes. However, because it better approximates the low-frequency cutoff of the ideal filter, it can be used to better understand the effects of measurement errors stemming from the inclusion of excessively long cycles. Drehmann, Borio and Tsatsaronis (2011) previously compared the performance of these 1- and 2-sided filters and, based on their full data sample, found that the 2-sided estimates gave slightly noisier forecasts of banking crises.[34]

The fixed-length 1-sided filter is simply constructed using the corresponding weights shown in Figure 3. Its forecast performance indicates whether the problems caused by small values values of T shown above are harmful.

Finally, the HP BandPass filter is constructed by passing the raw data through two distinct 1-sided HP filters. First the data are detrended using the $\lambda = 400,000$ filter; the resulting cycles are then passed through the $\lambda = 1600$ and the resulting HP trends are saved, thereby purging the effects of cycles at or above business cycle frequencies.[35] Its forecast performance indicates whether measurement errors stemming from the inclusion of excessively short cycles are harmful.

The results of these alternative measures of the credit cycle are shown in Figures 6 to 8 and Table 3 alongside the results for the 1-sided HP(400,000) considered previously. For the 2-sided filter, Figure 8 shows that the results are sensitive to the forecast horizon, with the 2-sided filter substantially improving on the 1-sided's filters pAUC at the very shortest horizon, but then rapidly deteriorating as the forecast horizon increases. As a result, Table 3 shows that the 1- and 2-sided filters perform similarly when using Basic scoring, but the 2-sided filter performs much worse when using Comprehensive scoring. This implies that part of the good forecasting performance of the 1-sided HP filter at longer horizons is the result of including cycles longer than those normally associated with credit cycles.

Results for the 40Q fixed-length filter are much less ambiguous, with a performance that

[34] Drehmann, Borio and Tsatsaronis (2011), Table 8.

[35] The bandpass filter may also be of interest for more prosaic reasons: regulators are likely to be wary of any indicator which implies continual and volatile adjustment of the CCB. By suppressing high-frequency movements, a bandpass filter should produce a smoother signal.

is worse than that of the recursive 1-sided filter in every case, and often substantially so. This suggests that poor measurement of credit cycles in short samples has probably biased estimates of the predictability of banking crises somewhat downwards. It also seems to imply that accurate measurement of longer-term trends in credit/GDP ratios plays an important role in correctly identifying periods of financial stress.

Finally, results for the HP bandpass filter contrast with those of the 2-sided filter; although the bandpass filter gives lower pAUCs than the high-pass filter at short forecast horizons, its performance improves as the horizon increases and gives higher pAUCs for horizons of 7Q and more. As a result, while its ROC curve is often slightly inferior to that of the standard 1-sided filter using Basic scoring, Comprehensive scoring shows that it produces a marked improvement for hit rates in the range of $60 - 90\%$. These results are not surprising when we keep in mind that the one-sided band-pass filter will tend to lag the one-sided high-pass filter because it replaces the current value X_t with a 1-sided moving average. This lag becomes less serious as the forecast horizon increases. The results also suggest that, at least at longer horizons, fluctuations at business cycle frequencies impair the measurement of credit cycles in a substantive way.

To summarise, there is good evidence that problems in measuring credit cycles in shorter samples tends to attenuate the predictive ability of credit gaps at all forecast horizons. The evidence also suggests that including excessively long fluctuations in credit actually *helps* predict systemic banking crises at multi-year forecast horizons, while including excessively short fluctuations hurts long-horizon forecasts but helps at short horizons.

To understand whether these results are specific to HP filters, in the next section we extend our analysis to consider "optimal" estimates of credit cycles and their predictive performance.

5. Optimal Bandpass Filters for Credit Cycles

A bandpass filter is a weighted moving average of a series that precisely captures cycles of a particular range of durations (called the "pass band") while eliminating both shorter fluctuations and longer trends. The use of approximate linear bandpass filters has a long history in both economics and finance as a way of separating short-lived movements in series from longer-lived underlying tendencies. In the 1970s Hodrick and Prescott (1997) advocated the use of splines that approximated high-pass filters while Baxter and King (1999) suggested using truncated versions of the symmetric doubly-infinite linear filters that produced ideal band-pass filters. However, the latter noted that such symmetric filters could not be used near the end of a data sample (or "on-line" in engineering parlance.) Koopmans (1974)) showed how to construct asymmetric band-pass filters suitable for on-line use based only on finite samples. Wildi (2005) showed that similar solutions applied in the case of non-stationary processes.

5.1. Minimum Mean Squared Error Filters

In the case of credit cycles, we could precisely capture fluctuations lasting 10-30 years with an ideal bandpass filter that eliminates all cycles with periods > 30 years or < 10 years. To do so, however, we would need a doubly infinite filter of the form

$$y_t \equiv \sum_{j=-\infty}^{\infty} \gamma_j \cdot x_{t-j} \tag{2}$$

where

- y_t is the estimated cyclic component
- x_t is the series we wish to filter
- $\gamma_j = (\sin j \cdot u - \sin j \cdot l)/\pi \cdot j$ are the required filter weights
- u and l are the upper and lower cutoff frequencies of the pass band

Note that this ideal filter requires observations infinitely far into the future and into the past. However, given that we have only T observations and that we must use a 1-sided filter, we could approximate y_t using

$$\widehat{y}_T \equiv \sum_{j=0}^{T} b_j \cdot x_{T-j} \tag{3}$$

for some set of weights $\{b_j\}$. We can then define the 1-sided filter that is optimal in a minimum mean-squared error sense as the set $\{b_j\}$ which solves

$$\min_{b_1 \ldots b_T} E\left([y_t - \widehat{y}_t]^2\right) = \int_{-\pi}^{\pi} \left|\Gamma(\omega) - \widehat{\Gamma}(\omega)\right|^2 \cdot S_x(\omega) \cdot d\omega \tag{4}$$

where

- $S_x(\omega) \equiv$ Spectral Density of x
- $\Gamma(\omega) \equiv$ target transfer function ($= 1 \quad \forall \omega \in (l, u)$, 0 elsewhere)
- $\widehat{\Gamma}(\omega) \equiv$ transfer function of $\{b_1 \ldots b_T\}$

We can approximate this by solving

$$\min_{b_1 \ldots b_T} E\left([y_t - \widehat{y}_t]^2\right) \approx \frac{2\pi}{T} \cdot \sum_{k=\frac{-(T-1)}{2}}^{\frac{T-1}{2}} \left|\Gamma(\omega_k) - \widehat{\Gamma}(\omega_k)\right|^2 \cdot I_x(\omega_k) \tag{5}$$

where

- $I_x(\omega_k)$ is the discrete-valued periodigram of x
- $\omega_k \equiv 2\pi \cdot \frac{k}{T}$

We call the linear filter with the weights $\{b_1 \ldots b_T\}$ which solve (5) the *MSE filter* since it minimises the expected mean squared error of our estimate $[y_t - \widehat{y}_t]$. Unlike the weights of the HP or the ideal band-pass filter, these weights depend on the spectral properties of the series to be filtered. [36] In particular, the MSE filter will try harder to match the

[36] Koopmans (1974) notes that the MSE filter can be thought of as the result of an two-step process

1. Use $I_x(\omega_k)$ to pad the available series $\{y_1 \ldots y_T\}$ with forecasts $\{\hat{y}_{T+1} \ldots \hat{y}_\infty\}$ and backcasts $\{\hat{y}_{-\infty} \ldots \hat{y}_0\}$
2. Apply the ideal band-pass filter weights $\{\gamma_j\}$ to the resulting doubly-infinite series.

We can therefore interpret the MSE filter weights as the convolution of the ideal band-pass filter weights with those needed to forecast and backcast the "missing" values in the series. See also Christiano and Fitzgerald (2003).

ideal band-pass filter around those frequencies where the spectral density is highest. For series with the typical Granger shape, this means matching the low-frequency cutoff will be emphasised more than the high-frequency cutoff. In the extreme case where there is no spectral density near or above the high-frequency stop band, the MSE bandpass filter will simply try to approximate a high-pass filter using the low-frequency cutoff.

5.2. Customized Direct Filter Analysis (DFA)

One of the potential shortcomings of the MSE filter is that, while optimal in a MSE sense, this may not be a good reflection of the loss function facing the user. McElroy and Wildi (2014) therefore extend the analysis to allow for a more flexible approach. They begin by noting that we can always decompose

$$\Gamma(\omega) \equiv A(\omega) \cdot e^{i \cdot \phi(\omega)} \tag{6}$$

and

$$\begin{aligned} \left|\Gamma(\omega) - \widehat{\Gamma}(\omega)\right|^2 &= A(\omega)^2 + \widehat{A}(\omega)^2 - 2 \cdot A(\omega) \cdot \widehat{A}(\omega) \cdot \cos\left(\phi(\omega) - \widehat{\phi}(\omega)\right) & (7) \\ &= \left(A(\omega) - \widehat{A}(\omega)\right)^2 + 4 \cdot A(\omega) \cdot \widehat{A}(\omega) \cdot \sin\left(\frac{\phi(\omega) - \widehat{\phi}(\omega)}{2}\right) & (8) \end{aligned}$$

where

- $A(\omega) \equiv |\Gamma(\omega_k)| \equiv$ Amplitude (real-valued)
- $\phi(\omega) \equiv$ phase (real-valued)

We can use this to similarly decompose the MSE of (5) as

$$\sum_{k=\frac{-(T-1)}{2}}^{\frac{T-1}{2}} \left|\Gamma(\omega_k) - \widehat{\Gamma}(\omega_k)\right|^2 \cdot I_x(\omega_k) = \left[\sum_{k=\frac{-(T-1)}{2}}^{\frac{T-1}{2}} \left(A(\omega_k) - \widehat{A}(\omega_k)\right)^2 \cdot I_x(\omega_k)\right] \tag{9}$$

$$+ 4 \cdot \left[\sum_{k=\frac{-(T-1)}{2}}^{\frac{T-1}{2}} A(\omega_k) \cdot \widehat{A}(\omega_k) \cdot \sin\left(\frac{\left(\widehat{\phi}(\omega) - \phi(\omega)\right)}{2}\right)^2 \cdot I_x(\omega_k)\right]$$

$$= [\text{ Leakage and Compression }] + [\text{ Phase Shift }]$$

McElroy and Wildi (2014) then propose modifying the MSE criterion by selectively reweighting the above components of (9) in the pass band and in the stop band to obtain a customised direct filter approach (DFA)

$$\begin{aligned}
\min_{b_1 \ldots b_T} & \sum_{\omega_k \in Passband} \left(A(\omega_k) - \widehat{A}(\omega_k)\right)^2 \cdot I_x(\omega_k) \\
+(1+\lambda_S)\cdot & \sum_{\omega_k \in Stopband} \left(A(\omega_k) - \widehat{A}(\omega_k)\right)^2 \cdot I_x(\omega_k) \\
+(1+\lambda_T)\cdot & \sum_{\omega_k \in Passband} A(\omega_k) \cdot \widehat{A}(\omega_k) \cdot \sin\left(\frac{\widehat{\phi}(\omega)}{2}\right)^2 \cdot I_x(\omega_k)
\end{aligned} \tag{10}$$

where $-1 < \lambda_S, \lambda_T$. Note that $0 = \lambda_S = \lambda_T$ gives us (5) as a special case. Values $\lambda_S > 0$ provide "smoother" estimates by putting greater weight on noise suppression in the stopband (thereby reducing leakage), while values $\lambda_T > 0$ reduce phase shift and so emphasize the "timeliness" of filtered signals.[37] Of course, these advantages do not come without a cost; putting greater weight on timeliness or smoothness implies that the filter will have a greater expected MSE. Below we examine how such a customised filter performs relative to the filters discussed above.

[37] The terms "smoothness" and "timeliness" are those used by McElroy and Wildi (2014). Whether the resulting filtered series becomes less variable than or leads that from the MSE filter will depend on the specifics of the application.

6. Crisis Prediction with Optimal Filters

We now examine whether credit cycles estimated by the MSE or DFA filters can improve the prediction of systemic banking crises. We compare their performance to two benchmarks

- the 1-sided HP(400,000) mandated by the BIS and the ESRB
- the quarterly change in the Credit/GDP ratio (previously studied by IMF (2011)) which we refer to below as "Change."

6.1. Construction of the MSE and DFA Estimates

Both the MSE and the DFA filters require estimates of the periodigram of the Credit/GDP series. We used consistent OLS estimates of a common autocorrelation function across economies, having first standardised the raw series across economies.[38] For the DFA filter, we followed McElroy and Wildi (2014) in penalising high-frequency deviations more heavily by setting $\lambda_S = 0$ in the low-frequency stop-band and

$$1 + \lambda_S = (1 + |\omega_k| - \omega^u)^\eta \tag{11}$$

in the high-frequency stop-band, where ω^u is the high-frequency cutoff. We experimented with a small (9 value) grid of weights (λ_T, η) and found results tended to improve monotonically with increases in both parameters. To avoid data mining, we made no further attempt to optimise these parameters; the results we report below use $\lambda_T = 40$ and $\eta = 1$.

We imposed a constant filter length of $T = 80$ for the MSE and DFA filters; the 1-sided HP filter was calculated recursively as before. However, comparing the performance of these different filters requires that we compare them over the same sample period. This means that no HP filter estimates on samples shorter than 80 quarters were used to generate the results shown below. Based on the results in the previous section, we would expect that this should

[38] To reflect the real-time nature of the filter estimation, series were adjusted to have mean zero and a standard deviation of one over the first 20 years of available data. However, using the full data sample for standardisation caused only trivial changes.

improve the performance of the HP filter somewhat. However, as we also noted above, 11 of the 38 economies with Credit/GDP series have ≤ 80 observations. Therefore this sample length constraint significantly reduces both the number of economies and banking crises in our sample.[39]

We again calculated pAUCs for standardised hit rates in the range $[0.5, 1.0]$ for all four measures of credit gaps, for all three of our scoring schemas, and for both the LV2012 and the Alternative sets of banking crisis start dates. We also repeated the calculation of MSE and DFA gaps for the change in (rather than the level of) the credit/GDP ratio. All these calculations were then repeated using both our credit series. Figures 9 to 14 compare results for all six gap estimates using the Total Credit/GDP series and the LV2012 crisis dates. Table 4 provides additional information on the robustness of these results.[40] Missing values again reflect cases where the pAUC is less than that of the uninformative case; in these situation, the standardised pAUC is not defined.

We also tested for equality of pAUCs of the HP gaps with those of each other gap method. All null hypotheses of equal pAUCs were tested against a two-sided alternative allowing for matched outcome measures.[41] Cases where the null hypothesis could be rejected at a 5% significance level are marked with a * in Table 4. Rejections were never found using Basic scoring; rejections using DJ2014 scoring were restricted to horizons of 10Q or longer and implied that the 1-sided HP filter provided statistically superior forecasts.[42] Rejections were much more frequent using Comprehensive scoring. We attribute much of this difference to

[39]This constraint eliminates AR, BR, CL, CZ, GR, HU IN, LU, PL, RU, SG and TH from our sample, leaving only 26 of our 36 economies. Using the LV2012 crisis dates, it leaves only 18 of the 36 crises; using the Alternative dates, it leaves 27 of the 54; in both cases this effectively excludes all crises prior to 1990.

[40]For DJ2014 scoring, results are for a 4 quarter forecast horizon.

[41]Significance levels were determined using a stratified bootstrap with 2000 replications. Carpenter and Bithell (2000) find 2000 replications is typically adequate for two significant figures of accuracy.

[42]Of the 300 two-sided tests that we performed at the 5% significance level, in 40 cases we rejected H_0 in favour of the alternative that the HP filter produced a higher pAUC and in 3 cases we rejected in favour of the alternative that it produced a lower pAUC. 39 of the 40 rejections occurred using the Alternative crisis dates and of those 33 occurred using the Total Credit series, where the HP filter was statistically superior to all three alternatives at the longest horizons.

much lower number of events observed using the former scoring schemas which appears to greatly reduce the power of the tests.[43]

6.2. Prediction Performance Comparisons

Figures 9 and 10 show that with Basic scoring, the ROC curve of HP gaps is dominated by that of simple Changes in Credit/GDP and (when used on data in Levels) by the gaps of the MSE and DFA filters on changes through much of relevant hit rate range. While differences in the pAUCs were not statistically significant, the implied differences in performance of the CCB are potentially economically important. For example, to achieve an 80% hit rate, the expected value of the CCB in normal times would be 1.64% for the HP filtered gaps vs 1.67, 1.52 and 1.10% for the DFA, MFA and Change measures of the credit gap respectively. This suggests potentially important savings in bank capital from using the Change measure in place of the HP filter.

Figures 11 and 12 show that this result is completely reversed with Comprehensive scoring, where the ROC curve of the HP gaps dominates all alternative measures (with the sole exception of the DFA gap from changes in the credit ratio for some hit rates.) Furthermore, Table 4 shows that most of these differences are statistically significant. This suggests that the HP gap forecasts significantly better than the alternative measures at longer forecast horizons, a result that we see confirmed in Figures 13 and 14. They show that the HP gap seems to be slightly inferior to the simple Change in the credit ratio at very short horizons, and inferior to the DFA (and, in levels, also the MSE gap) at intermediate horizons, while the HP gap has the highest pAUCs at forecast horizons beyond three or four years.

This sensitivity to forecast horizon is consistent with the evidence we saw above from the various HP filters, where the filters that passed more high-frequency fluctuations (such as the 1-sided HP filter) performed better at short horizons but worse at longer horizons. The relatively good performance of simple Changes in the Credit/GDP ratio at short horizons

[43]Further evidence is shown in the Appendix, where Table A.5 and Figure A.15 compare the 95% confidence intervals around the ROC curves for the 1-sided HP filter using Basic and Comprehensive scoring.

suggests that this extends to fluctuations of even shorter durations than business cycles. Similarly, we again see that the 1-sided HP filter's inclusion of excessively long cycles is associated with superior forecast performance at the longest horizons.

Table 4 provides further evidence on the robustness of these results to changes in the credit series and the crisis start dates used.[44] When using the Alternative crisis dates, the MSE and DFA filters almost always performed better when applied to changes in Credit/GDP rather than the level of the series, which suggests that *cycles* in the growth rate of Credit/GDP have useful predictive information. The use of Bank Credit/GDP causes gap measures based on first differences (the Change as well as the MSE and DFA measures based on changes) to perform better than those based on levels (the HP as well as the MSE and DFA measures based on levels) in most cases, whereas when using Total Credit/GDP the performance of gap measures based on levels relative to those based on changes is more ambiguous.

Finally, Table 4 and Figures 9 through 14 also show that DFA filter's emphasis on speed and smoothness often, but not always, improved the filter's predictive performance relative to that of the MSE filter; this result was more robust using the data on changes in credit ratios than the corresponding data in levels.

In summary, these results give mixed support to the hypothesis that better measurement of credit cycle frequencies will give better prediction of systemic banking crises. When using Total Credit/GDP data and the LV2012 crisis dates, the optimal filters often gave better predictions of systemic crises at shorter forecast horizons than the existing 1-sided HP filter measures used by regulators. However, the differences were neither statistically significant, nor robust to changes in crisis dates or credit measures. Furthermore, measures of changes in Credit/GDP ratios frequently did as well or better than any of the above at short horizon forecasts, and MSE and DFA measures of longer term movements in *changes* also frequently showed strong forecasting power. Finally, the relatively good performance of the HP filter at

[44] Note that, due to the change in the effective sample period, Table 4 also shows that pAUCs for the HP filter are now always lower than in Table 3, despite the expectation that the higher values of T used here should lead to better predictive performance.

longer forecasting horizons is consistent with the findings of Drehmann and Juselius (2014) who reported that it compared favourably to a variety of alternative indicators at longer horizons.

7. Conclusions

We have examined whether better measurement of credit cycles improves the prediction of systemic banking crises. Our definition of "better" has been a narrow one: we focus on the ability of an estimator to isolate regular fluctuations in the Credit/GDP ratio at the frequencies previously identified with credit cycles – 10 to 30 years. We showed that the 1-sided recursive HP filter commonly used both by regulators and researchers has some flaws by this criterion; it captures movements at frequencies outside this band and has difficultly capturing long-term trends on short samples. This may have weakened the empirical evidence linking credit cycles and the prediction of banking crises. We also investigated the forecast performance of alternative credit cycle measures based on "optimal" measures of fluctuations at credit cycle frequencies.

We found that in some cases better measures of the credit cycle improved forecast performance at shorter policy-relevant horizons. However, such improvements were not statistically significant, nor were the results robust. More robust and statistically significant was the superior forecast performance at multi-year horizons of simple HP-filter measures of cycles currently used by many regulators. This improved performance seems to be the result of mismeasuring the credit cycle, particularly the inclusion of cycles longer than those associated with credit cycles. The results also showed that measures based on changes in the Credit/GDP ratio often performed better than the simple HP filter at short forecast horizons, evidence that is again at odds with the view of credit cycles as slow-moving phenomena.

References

Aikman, David, Andrew G. Haldane and Benjamin D. Nelson (2014) "Curbing the Credit Cycle" *The Economic Journal*, doi:10.1111/ecoj.12113.

Basel Committee on Banking Supervision (2010a) *Basel III: A global regulatory framework for more resilient banks and banking systems.* Bank of International Settlements, December 2010.

Basel Committee on Banking Supervision (2010b) *Guidance for national authorities operating the countercyclical capital buffer.* Bank of International Settlements, December 2010.

Baxter, Marianne and Robert G. King (1999) "Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series" *The Review of Economics and Statistics*, Vol. 81, No. 4, Pages 575-593, doi:10.1162/003465399558454

Borio, C. (2014) "The Financial Cycle and Macroeconomics: What Have We Learnt?" *Journal of Banking and Finance*, 45(August), 182-198.

Brunnermeier, Marcus (2009) "Deciphering the Liquidity and Credit Crunch 2007-08," *Journal of Economic Perspectives,* 23(1), 77-100.

Burns, A.F. and W.C. Mitchell (1946) *Measuring Business Cycles*, NBER, New York.

Carpenter, James and John Bithell (2000)"Bootstrap condence intervals: when, which, what? A practical guide for medical statisticians". *Statistics in Medicine* 19, 11411164.

Christiano,Lawrence J. and Terry J. Fitzgerald (2003) "The Band Pass Filter" *International Economic Review*, 44(2), 435465.

Christensen, Ian and Fuchun Li (2014) "Predicting Financial Stress Events: A Signal Extraction Approach" *Journal of Financial Stability*, *http://dx.doi.org/10.1016/j.jfs.2014.08.005*

Dembiermont, Christian, Mathias Drehmann and Siriporn 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 2013, 65-81.

Detken, Carsten, Olaf Weeken, Lucia Alessi, Diana Bonfim, Miguel M. Boucinha, Christian Castro, Sebastian Frontczak, Gaston Giordana, Julia Giese, Nadya Jahn, Jan Kakes, Benjamin Klaus, Jan Hannes Lang, Natalia Puzanova and Peter Welz (2014) "Operationalising the countercyclical capital buffer: indicator selection, threshold identification and calibration options" *European Systemic Risk Board Occasional Paper Series*, No. 5

Drehmann, Mathias, Claudio Borio, Leonardo Gambacorta, Gabriel Jimnez and Carlos Trucharte (2010) "Countercyclical capital buffers: exploring options" *BIS Working Paper*, No. 317.

Drehmann, Mathias, Claudio Borio and Kostas Tsatsaronis (2011) "Anchoring countercyclical capital buffers: the role of credit aggregates" *International Journal of Central Banking*, 7(4), December 2011, p. 189-240.

Drehmann, Mathias and Mikael Juselius (2013) "Evaluating early warning indicators of banking crises: Satisfying policy requirements" *BIS Working Papers*, No. 421.

Drehmann, Mathias and Mikael Juselius (2014) "Evaluating early warning indicators of banking crises: Satisfying policy requirements" *International Journal of Forecasting* Volume 30, Issue 3, JulySeptember 2014, Pages 759780.

Edge, R. and Meisenzahl, R.R. (2011), "The unreliability of the credit-to-GDP ratio gaps in real-time: Implications for countercyclical capital buffers," *International Journal of Central Banking*, 7(4), 261-298.

European Systemic Risk Board (2014) "Recommendation of the European Systemic Risk Board of 18 June 2014 on guidance for setting countercyclical buffers." *Official Journal of the European Union*, C293.

Fawcett, Tom (2004) "ROC Graphs: Notes and Practical Considerations for Researchers" *Machine learning*, 31, 1-38.

Hodrick, Robert J. and Edward C. Prescott (1997) "Postwar U.S. Business Cycles: An Empirical Investigation" *Journal of Money, Credit and Banking*, 29(1), 1-16.

Galati, Gabriele and Richhild Moessner (2014) "What Do We Know About the Effects of Macroprudential Policy?" *De Nederlandsche Bank Working Paper* No. 440

International Monetary Fund (2011) *Global Financial Stability Report*, September, Chapter 3.

Jordà, Ò., and Taylor, A. M. (2011). "Performance evaluation of zero net-investment strategies." *National Bureau of Economic Research*, WP 17150.

Koopmans, L.H. (1974), *The Spectral Analysis of Time Series*, San Diego, Academic Press.

Laeven, Luc and Fabin Valencia (2012) "Systemic Banking Crises Database: An Update" *IMF Working Paper* No. 12/163.

Leeper, Eric and James Nason (2015) *Bringing Financial Stability into Monetary Policy* Sveriges Riksbank Working Paper 305.

Mason, S. J., and Graham, N. E. (2002) "Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation." *Quarterly Journal of the Royal Meteorological Society,* 128(584), 2145-2166.

McElroy, Tucker and Marc Wildi (2014) *The Trilemma Between Accuracy, Timeliness and Smoothness in Real-Time Signal Extraction*, manuscript.

Orphanides, Athanasios and Simon van Norden (2002) "The Unreliability of Output-Gap Estimates in Real Time," *The Review of Economics and Statistics* 84(4), 569-583.

Ravn, M. O. and H. Uhlig, (2002). "On adjusting the Hodrick-Prescott filter for the frequency of observations." *Review of economics and statistics*, 84(2), 371-376.

Xavier Robin, Natacha Turck, Alexandre Hainard, Natalia Tiberti, Frdrique Lisacek, Jean-Charles Sanchez and Markus Mller (2011). "ROC: an open-source package for R and S+ to analyze and compare ROC curves." *BMC Bioinformatics*, 12, p. 77. *DOI: 10.1186/1471-2105-12-77*

Schwaab, Bernd, Siem Jan Koopman and André Lucas (2014) "Nowcasting and forecasting global financial sector stress and credit market dislocation" *International Journal of Forecasting* Volume 30, Issue 3, JulySeptember 2014, Pages 741758.

St-Amant, Pierre and Simon van Norden (1997) "Measurement of the Output Gap: A Discussion of Research at the Bank of Canada" *Bank of Canada Technical Report,* 79.

Wildi, Marc (2005) "Signal Extraction: Efficient Estimation, 'Unit Root'-Tests and Early Detection of Turning Point" in *Lecture Notes in Economics and Mathematical Systems*, Springer, Vol. 547.

8. Tables

Table 1: Starting Dates for Systemic Banking Crises

Country	LV2012	Alternative	Country	LV2012	Alternative
AR	1980Q2,89Q4,1995Q1,2001Q4	1980Q1,89Q4,1995Q1,2001Q4	IN	none	1993Q3
AT	20008Q4	20008Q3	IT	2008Q3	1992Q3,2008Q4
AU	none	1989Q4,2008Q4	JP	1997Q4	1992Q4
BE	20008Q4	2008Q4	KR	1997Q4	1997Q3
BR	1990Q1,1994Q4	1990Q1,1994Q4	LU	2008Q3	2008Q3
CA	none	none	MX	1995Q1	1981Q3,1994Q4
CH	2008Q3	1991Q3,2007Q3	MY	1998Q1	1997Q3
CN	none	1998Q1	NL	2008Q4	2008Q4
CZ	1996Q2	1996Q2	NO	1991Q4	1990Q4
DE	2009Q4	2007Q3	PL	none	none
DK	2009Q1	1987Q4,2008Q4	PT	2008Q3	2008Q4
ES	2011Q2	1977Q3,1993Q4,2008Q4	RU	1999Q1	1999Q1
FI	1993Q1	1991Q3	SE	1992Q3,2008Q3	1991Q3,2008Q4
FR	2008Q3	1994Q1,2008Q4	SG	none	1982Q4
GR	2009Q2	2008Q4	TH	1997Q4	1997Q3
HK	none	none	TR	2000Q4	1982Q2,2000Q4
HU	2008Q3	2008Q3	UK	2008Q4	1973Q3,1990Q2,2007Q3
ID	1997Q4	1997Q4	US	2008Q4	1990Q2,2007Q3
IE	2009Q1	2008Q4	ZA	none	1989Q4

Table 2: Calculations underlying ROC curves

	Crisis Occurs	**No Crisis Occurs**
Crisis Signalled	(a) Correct	(b) Type II error
No Crisis Signalled	(c) Type I error	(d) Correct

Notes:

See Figure 2 for example of ROC curve.

Table 3: Partial AUCs for Credit Gaps from Various HP Filters

Scoring	Crisis Dates	Measure	2-Sided	1-Sided	1-Sided 40Q	1-Sided BP
Basic	LV2012	Bank Credit	0.681	0.662	0.574	0.643
Basic	LV2012	Total Credit	0.660	0.661	0.552	0.636
Basic	Alternative	Bank Credit	0.730	0.726	0.627	0.708
Basic	Alternative	Total Credit	0.689	0.730	0.609	0.704
DJ2014	LV2012	Bank Credit	0700	0.642	0.557	0.609
DJ2014	LV2012	Total Credit	0.679	0.651	0.531	0.611
DJ2014	Alternative	Bank Credit	0.775	0.745	0.631	0.719
DJ2014	Alternative	Total Credit	0.735	0.755	0.614	0.725
Comprehensive	LV2012	Bank Credit	0.562	0.639	0.574	0.652
Comprehensive	LV2012	Total Credit	0.566	0.674	0.530	0.712
Comprehensive	Alternative	Bank Credit	0.590	0.712	0.636	0.712
Comprehensive	Alternative	Total Credit	0.597	0.735	0.616	0.752

Notes:

Partial AUCs are calculated for hit rates in the range $[0.5, 1.0]$.

Figures for DJ2014 scoring reflect a 4 quarter forecast horizon.

Table 4: Partial AUCs for MSE and other filters

Scoring	Crisis Dates	Measure	Change	HP	MSE Level	MSE Change	DFA Level	DFA Change
Basic	LV2012	Bank Credit	0.660	0.612	0.572	0.571	0.590	0.574
Basic	LV2012	Total Credit	0.670	0.636	0.641	0.572	0.650	0.592
Basic	Alternative	Bank Credit	0.726	0.683	0.590	0.669	0.573	0.651
Basic	Alternative	Total Credit	0.656	0.700	0.632	0.658	0.612	0.655
DJ2014	LV2012	Bank Credit	0.665	0.620	0.571	0.581	0.587	0.587
DJ2014	LV2012	Total Credit	0.662	0.638	0.641	0.570	0.666	0.612
DJ2014	Alternative	Bank Credit	0.773	0.707	0.631	0.709	0.615	0.716
DJ2014	Alternative	Total Credit	0.637	0.714	0.692	0.686	0.673	0.719
Comprehensive	LV2012	Bank Credit	0.550	0.576	NA	0.567	NA	0.600
Comprehensive	LV2012	Total Credit	0.546*	0.624	0.552*	0.562*	0.508*	0.601
Comprehensive	Alternative	Bank Credit	0.592*	0.675	0.506*	0.677	0.526*	0.683
Comprehensive	Alternative	Total Credit	0.589*	0.714	0.559*	0.671*	0.501*	0.671*

Notes:

Partial AUCs are calculated for hit rates in the range $[0.5, 1.0]$.

Figures for DJ2014 scoring reflect a 4 quarter forecast horizon.

* indicates a rejection of the null hypothesis of a pAUC equal to that of the 1-sided HP filter at a 5% significance level based on a stratified bootstrap with 2,000 replications.

9. Figures

Figure 1: Total Private Credit / GDP

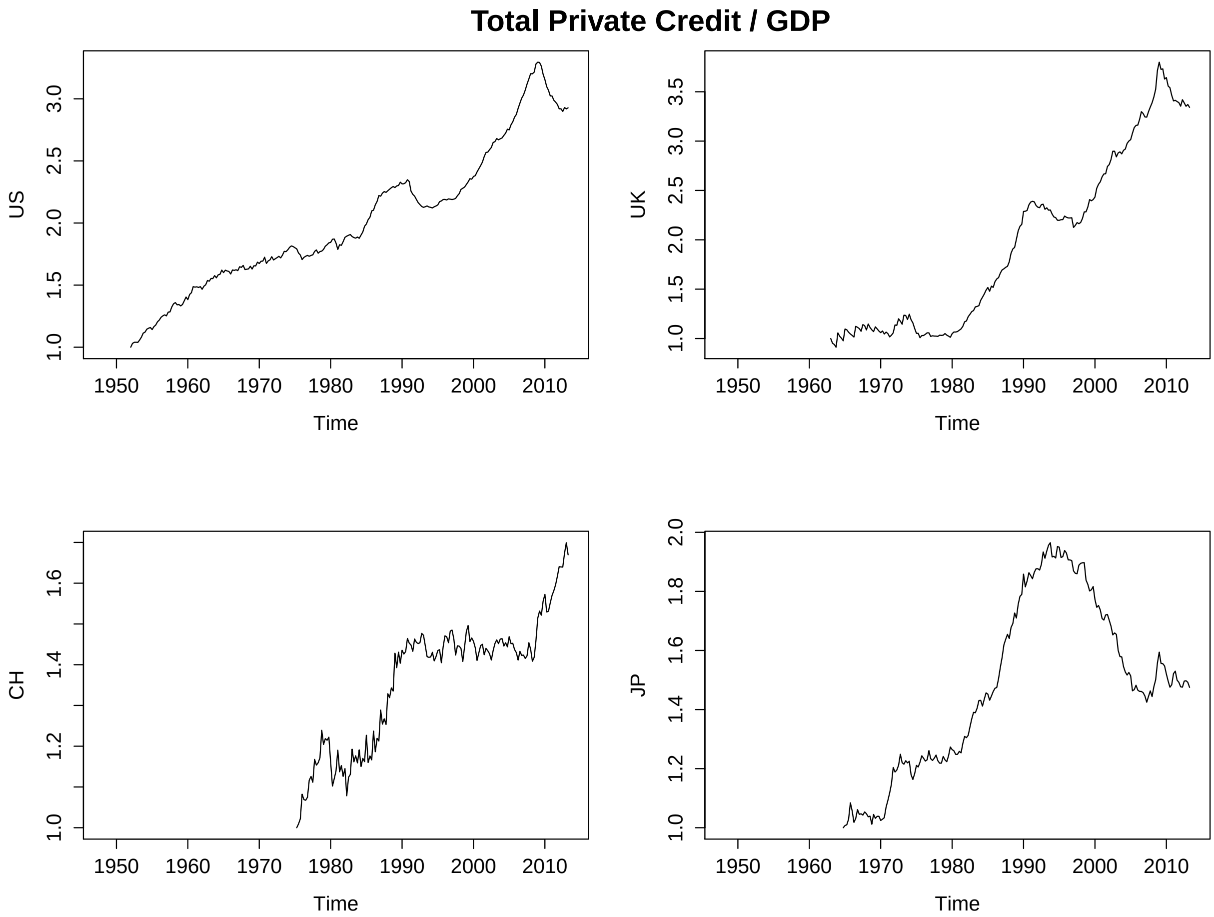
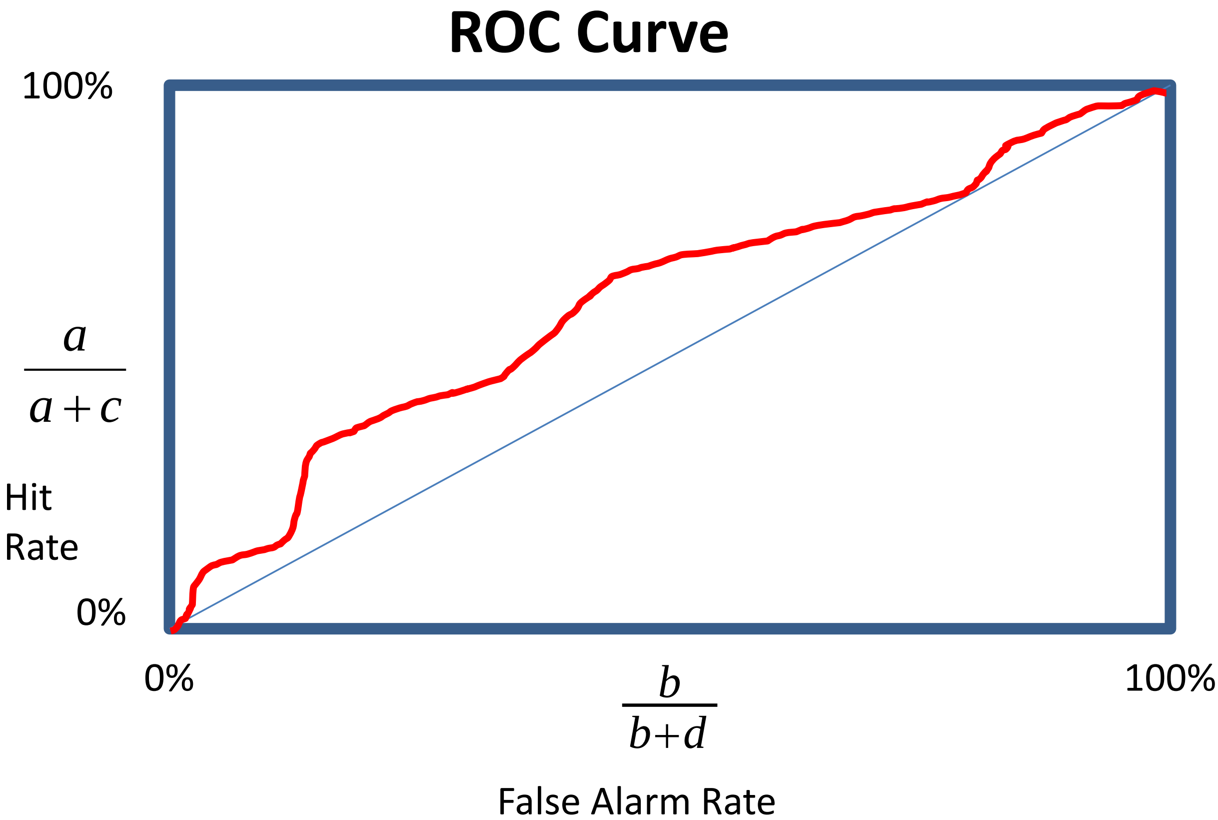


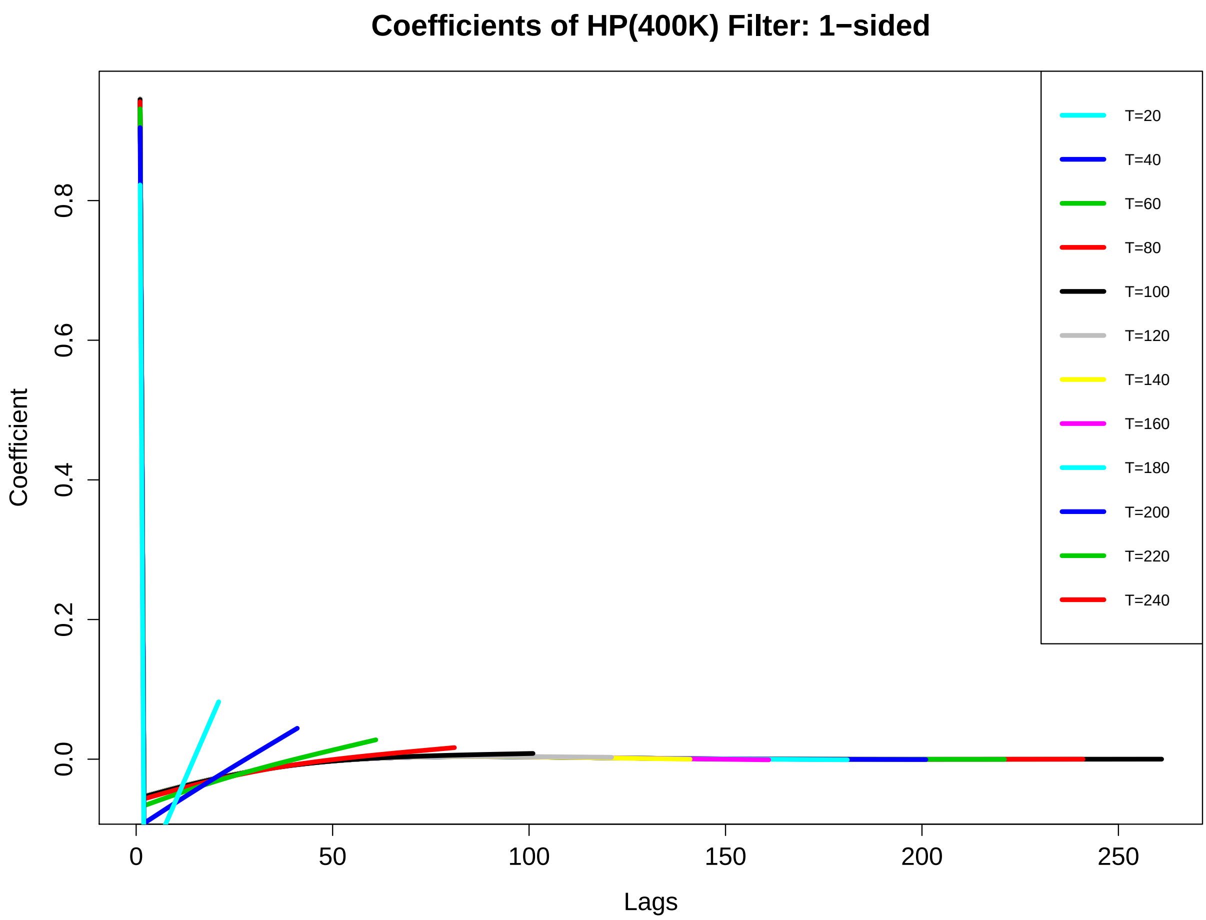
Figure 2: An ROC Curve



Notes:

See Table 2 for definitions of (a)-(d).

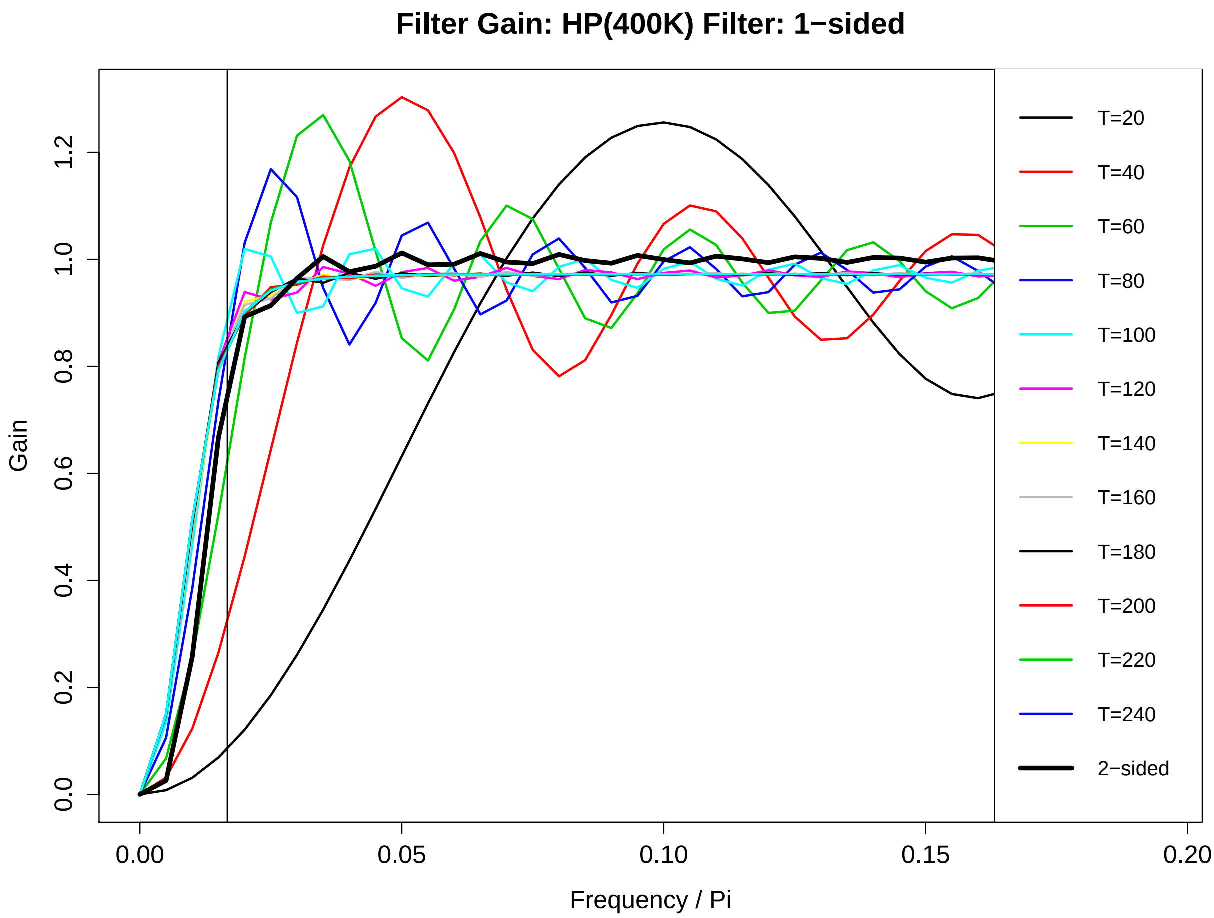
Figure 3: Filter Coefficients: HP Filter



Notes:

Curves show the weights assigned to each lag for 1-sided HP($\lambda = 400,000$) high-pass filters (i.e. for the cycle, not the trend) for samples sizes T given in the Legend.

Figure 4: Transfer Functions: HP Filter



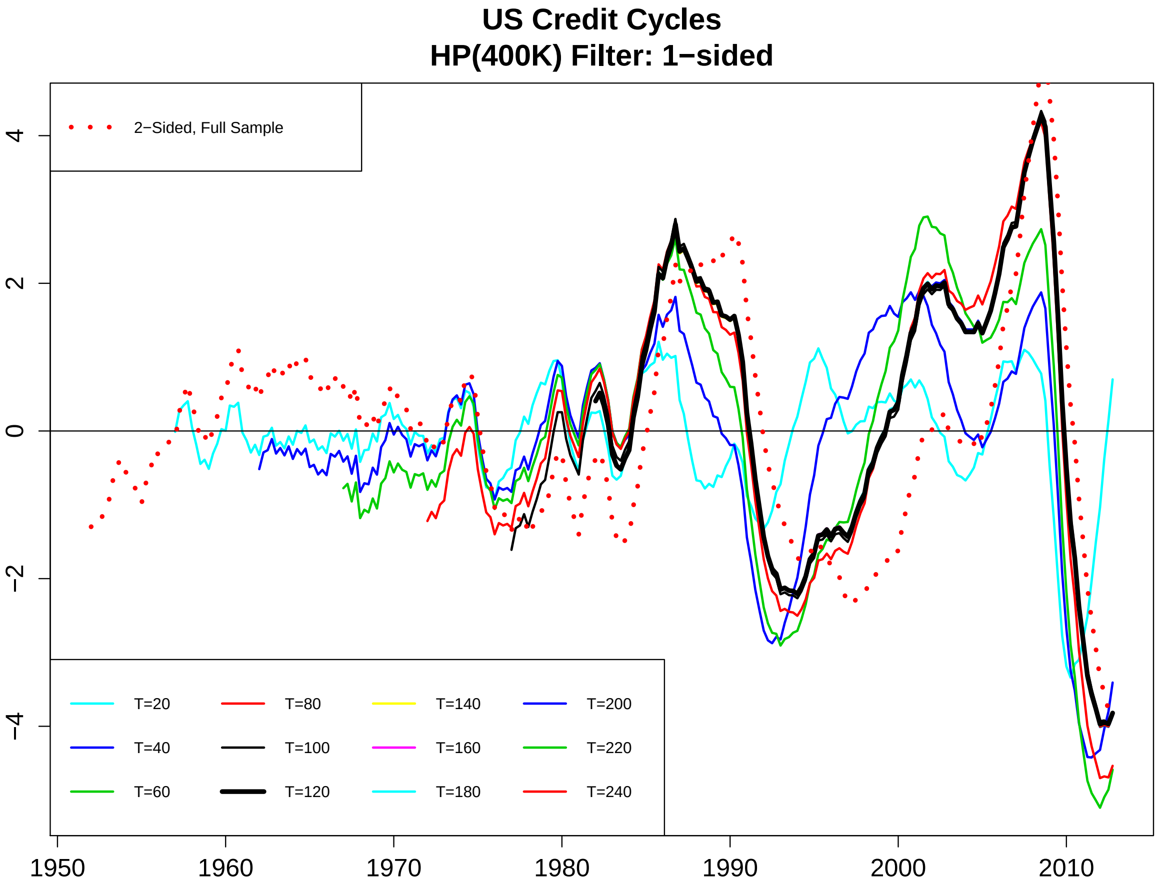
Notes:

Curves show the filter gain by frequency for 1-sided HP($\lambda = 400,000$) high-pass filters for samples sizes given by T.

The heavy black line shows the curve for the 2-sided HP($\lambda = 400,000$) high-pass filter.

The vertical line shows the frequency corresponding to cycles of 30 years.

Figure 5: US Credit Gaps: HP Filter

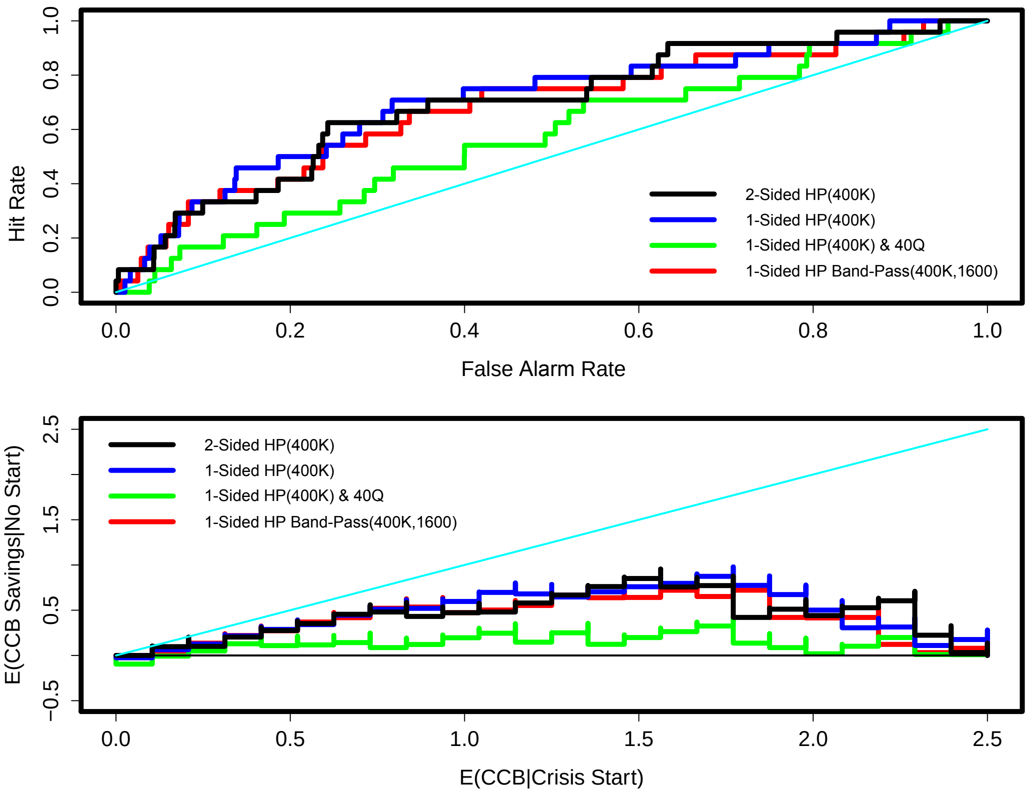


Notes:

Curves show the credit cycle estimated using the US Total Credit/GDP series detrended using the coefficients of the 1-sided HP($\lambda = 400,000$) high-pass filters for indicated sample size.

The heavy black line shows the curve for the 2-sided HP($\lambda = 400,000$) high-pass filter.

Figure 6: ROC Curves for Total Credit/GDP with HP Filters and Basic Scoring



Notes:

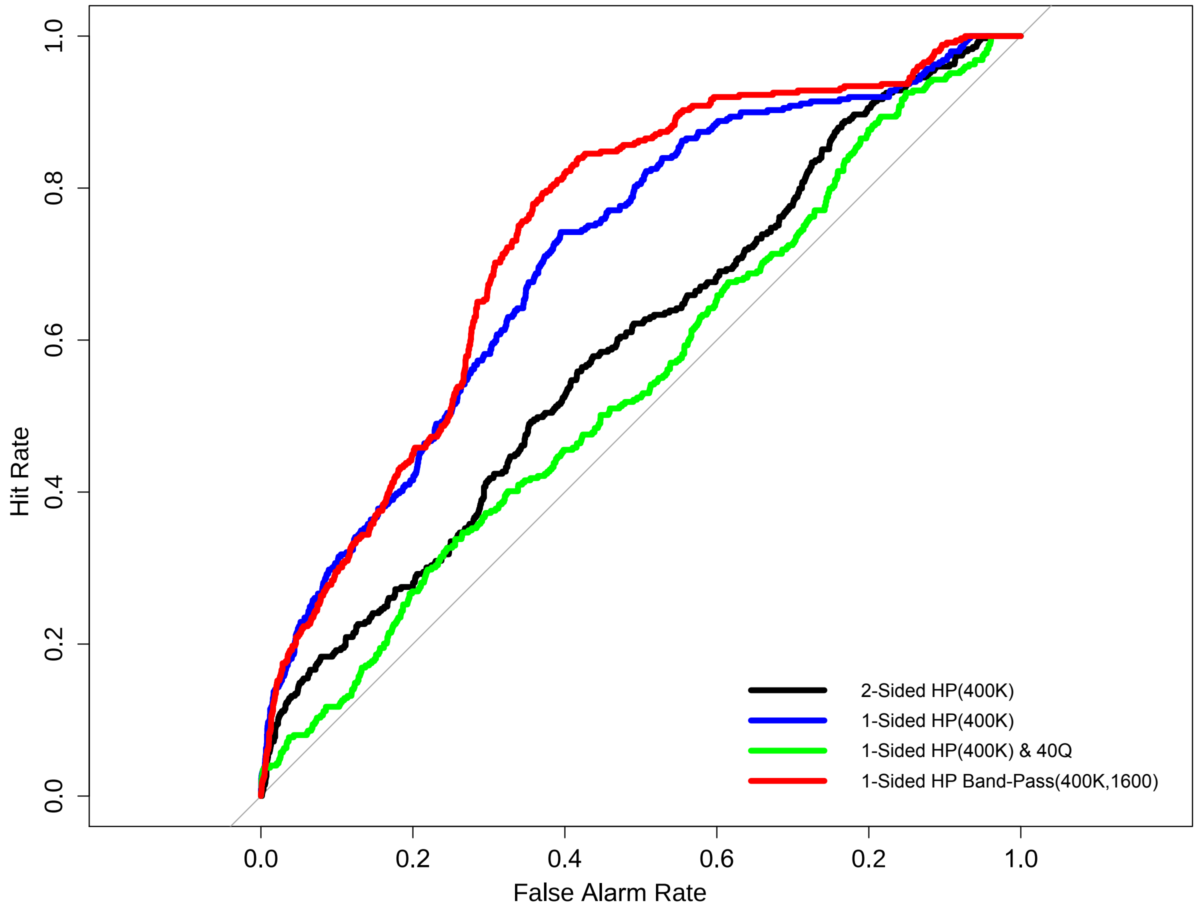
The upper panel shows ROC Curves using the LV2012 crisis dates and a minimum sample length $T = 40$.

The lower panel shows the implied tradeoff between Hit and False Alarm rates.

$E(CBB|$ Crisis Start $) \equiv 2.5*$ Hit Rate

$E(CBB$ Savings | No Start $) \equiv 2.5 *$ (Hit Rate - False Alarm Rate)

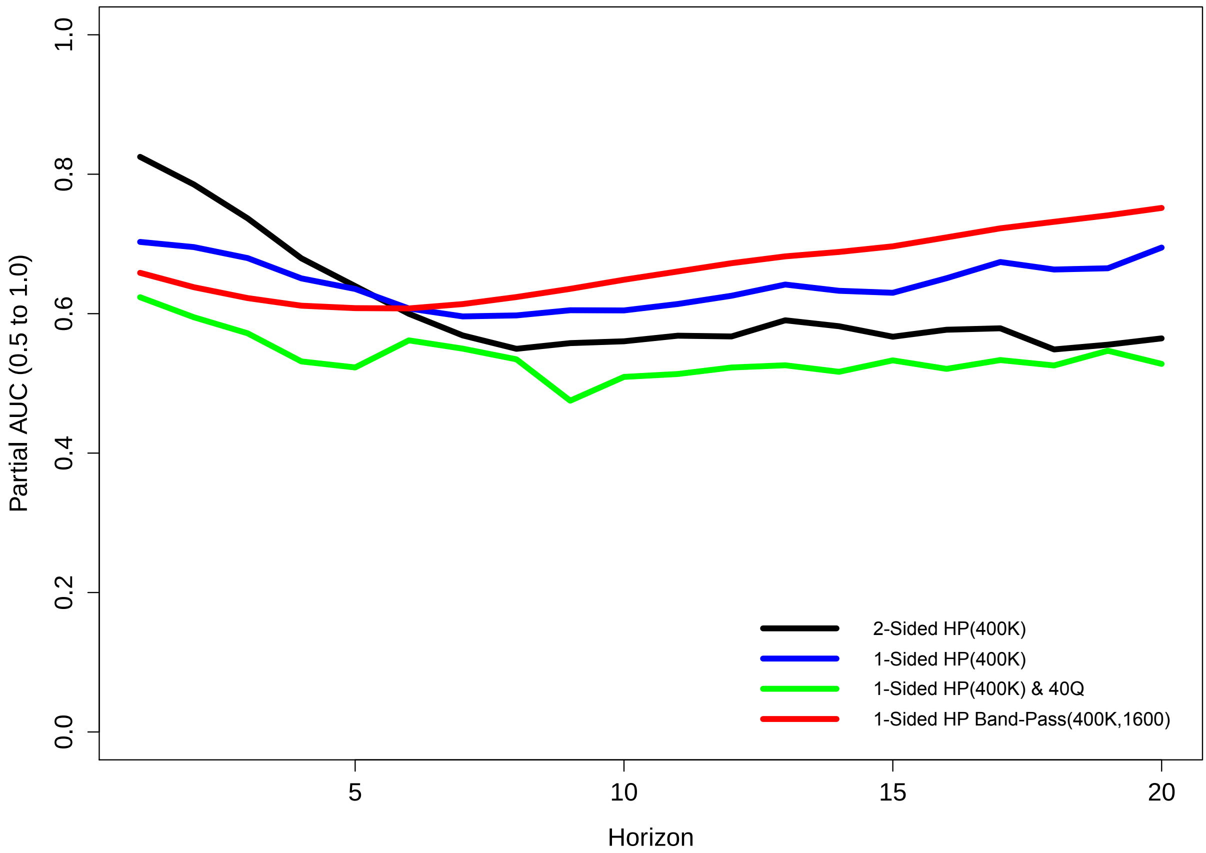
Figure 7: ROC Curve for Total Credit/GDP with HP Filters and Comprehensive Scoring



Notes:

ROC Curves are calculated using the LV2012 crisis dates and a minimum sample length $T = 40$.

Figure 8: pAUC for Total Credit/GDP with HP Filters and DJ2014 Scoring

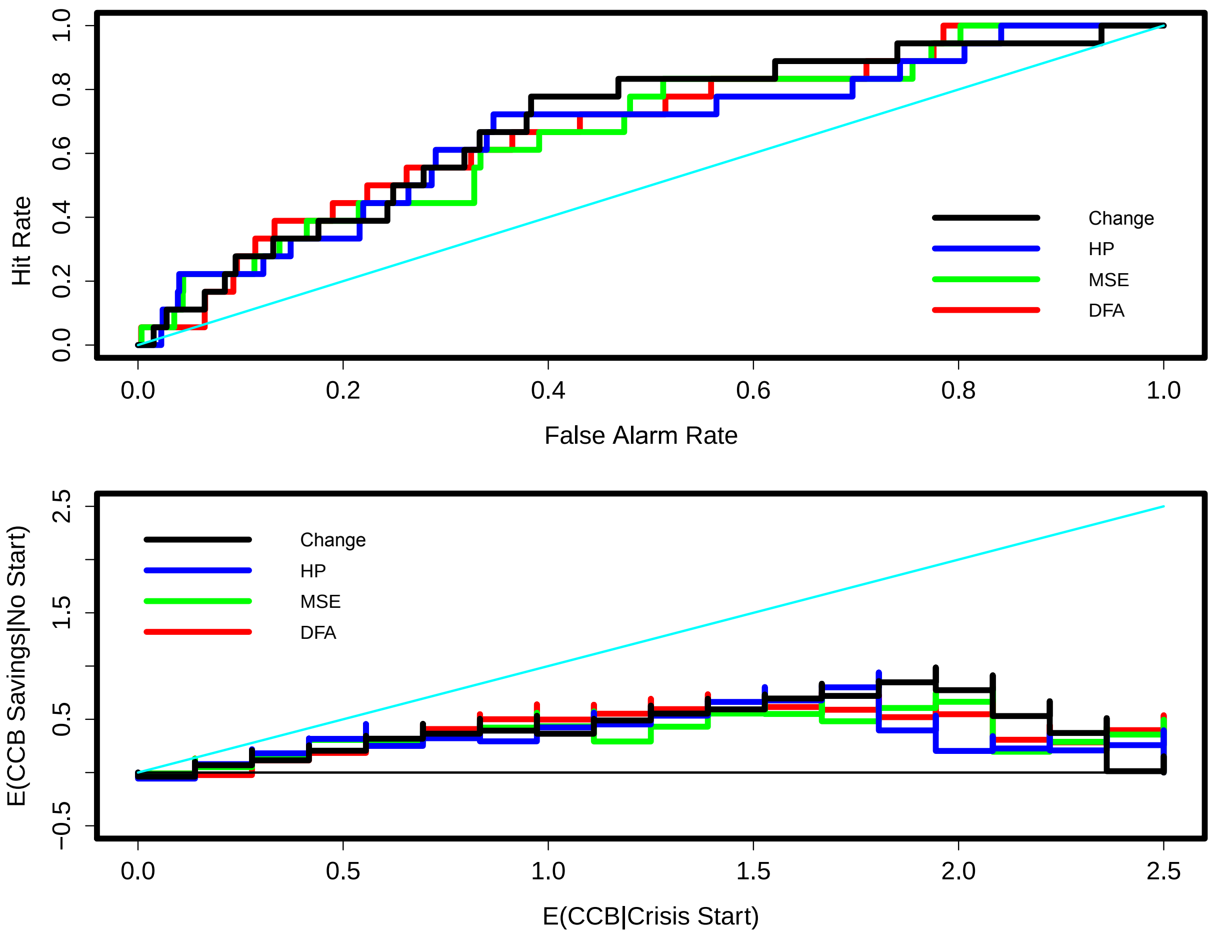


Notes:

ROC Curves are calculated using the LV2012 crisis dates and a minimum sample length $T = 40$.

Missing values indicate partial AUC < 0.5

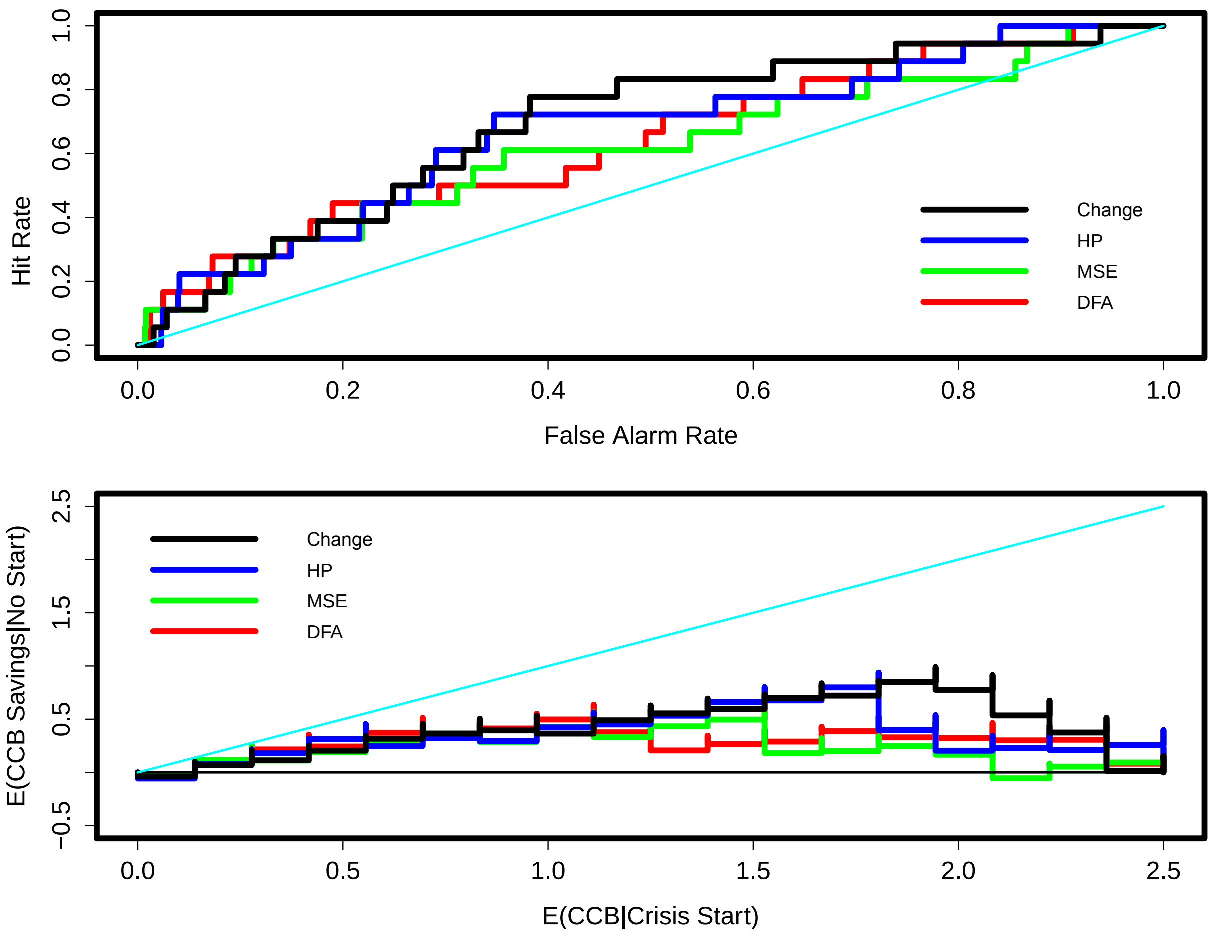
Figure 9: ROC Curve for Total Credit/GDP with Basic Scoring



Notes:

ROC Curves for the four measures shown in the Legend are calculated using the LV2012 crisis dates and a minimum sample length $T = 80$.

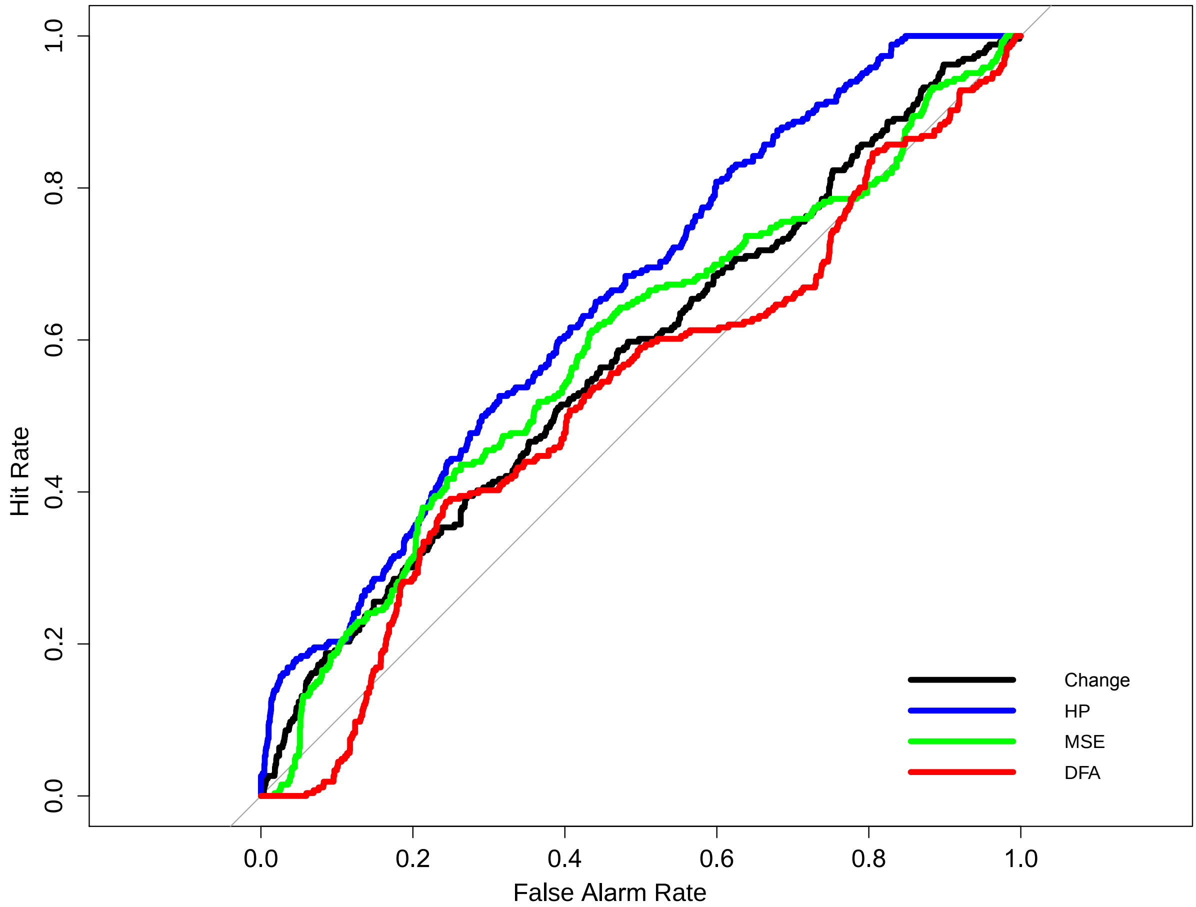
Figure 10: ROC Curve for Changes in Total Credit/GDP with Basic Scoring



Notes:

ROC Curves for the four measures shown in the Legend are calculated using the LV2012 crisis dates and a minimum sample length $T = 80$.

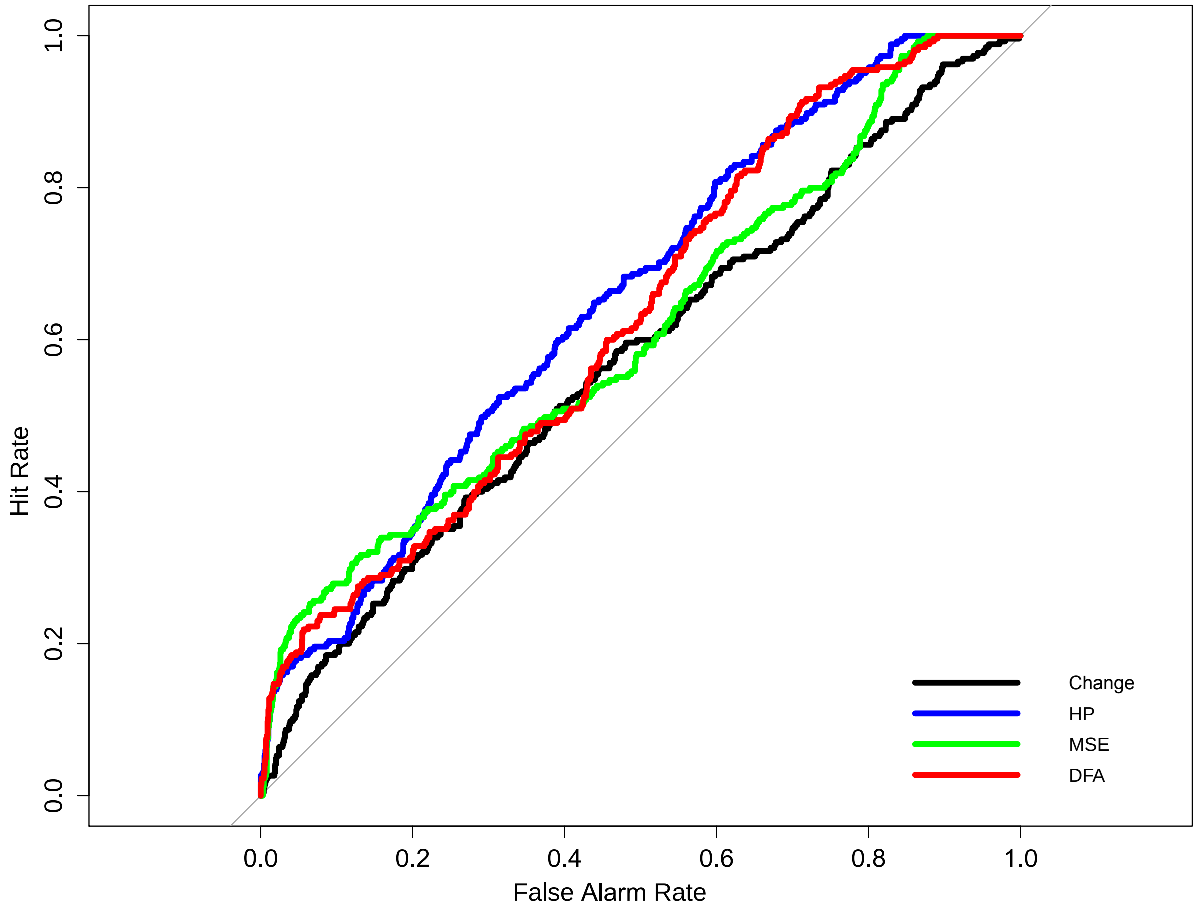
Figure 11: ROC Curve for Total Credit/GDP with Comprehensive Scoring



Notes:

ROC Curves for the four measures shown in the Legend are calculated using the LV2012 crisis dates and a minimum sample length $T = 80$.

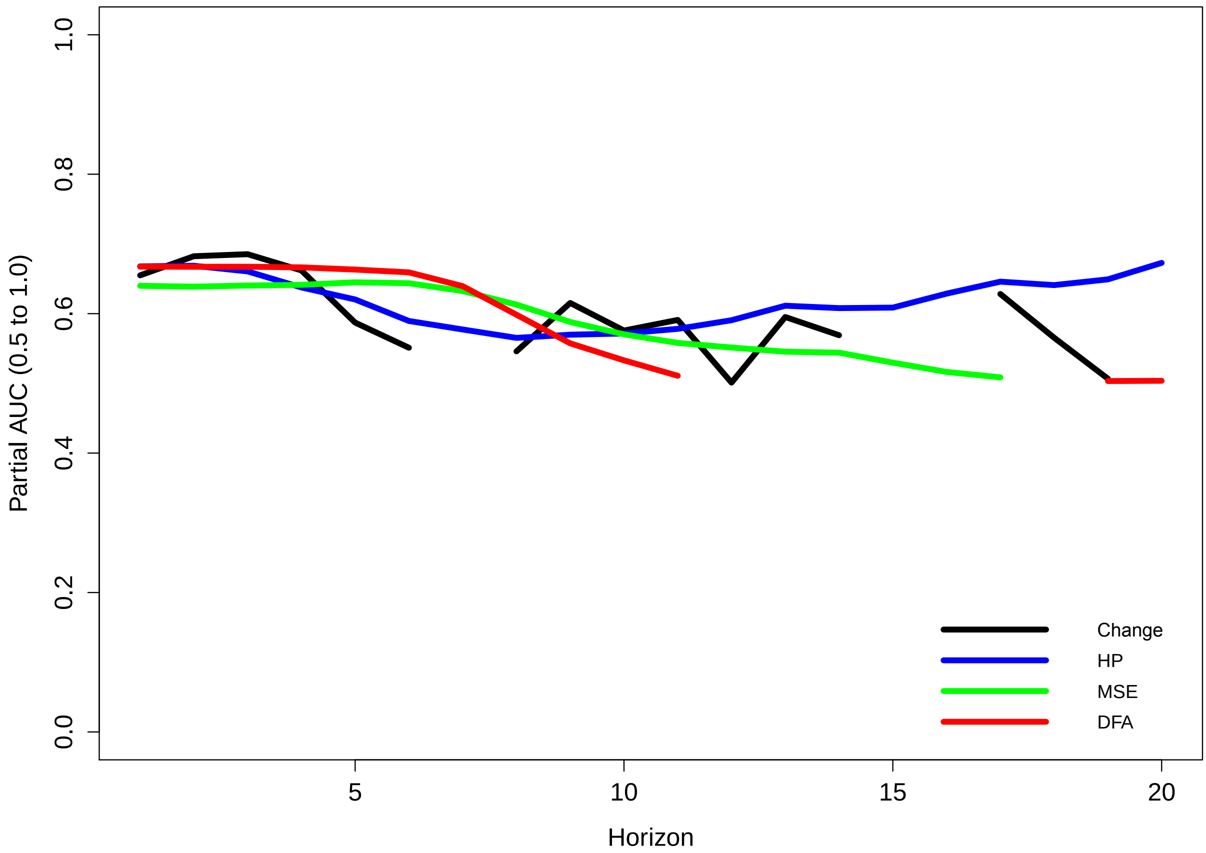
Figure 12: ROC Curve for Changes in Total Credit/GDP with Comprehensive Scoring



Notes:

ROC Curves for the four measures shown in the Legend are calculated using the LV2012 crisis dates and a minimum sample length $T = 80$.

Figure 13: pAUC for Total Credit/GDP with DJ2014 Scoring

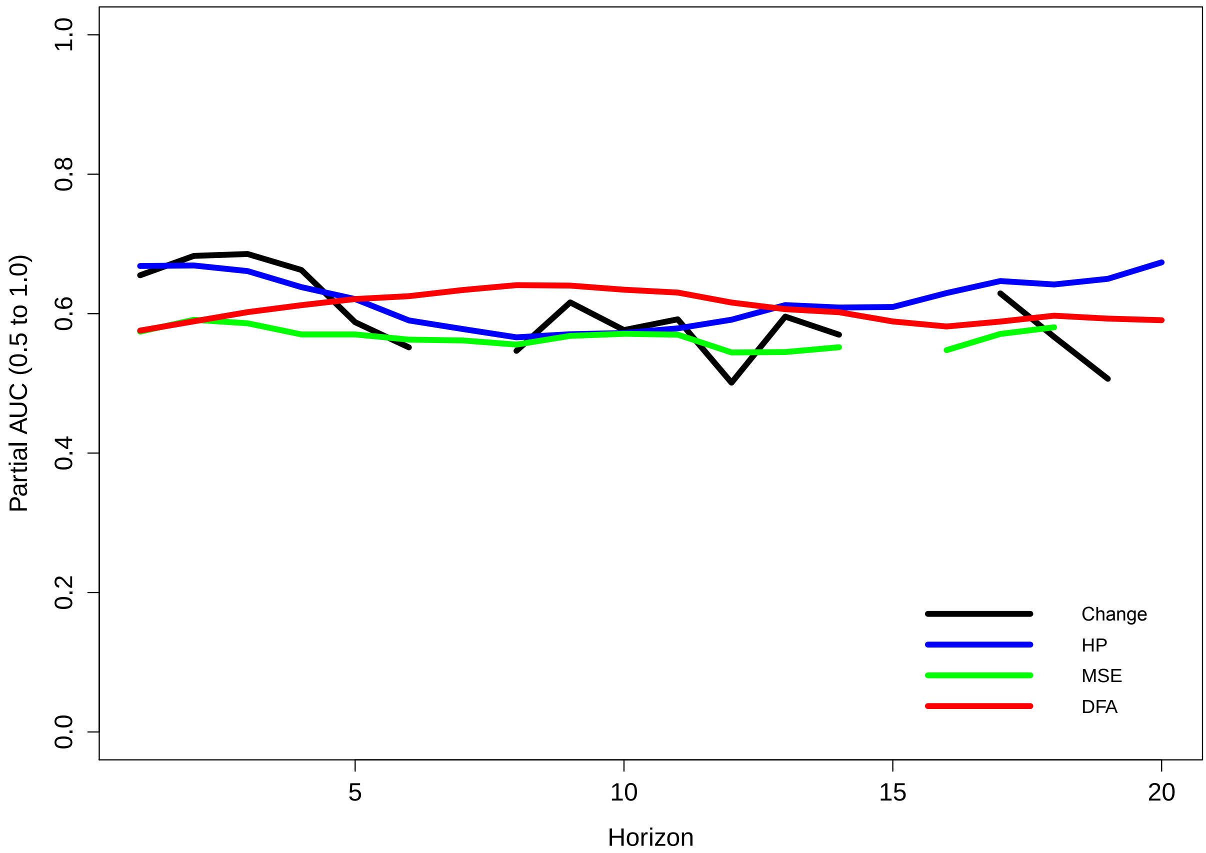


Notes:

ROC Curves for the four measures shown in the Legend are calculated using the LV2012 crisis dates and a minimum sample length $T = 80$.

Missing values indicate partial AUC < 0.5

Figure 14: pAUC for Changes in Total Credit/GDP with DJ2014 Scoring



Notes:

ROC Curves for the four measures shown in the Legend are calculated using the LV2012 crisis dates and a minimum sample length $T = 80$.

Missing values indicate partial AUC < 0.5

Appendix A. The Power of ROC Tests

Table A.5: Confidence Intervals for ROC Curves

	False Alarm Rate			
	Basic Scoring		Comprehensive Scoring	
Hit Rate	2.50%	97.50%	2.50%	97.50%
0.0	0	0	0	0
0.1	0.1262	0.0169	0.0212	0.0056
0.2	0.2154	0.0201	0.1357	0.0262
0.3	0.2725	0.0321	0.2094	0.1223
0.4	0.3202	0.0407	0.2747	0.1904
0.5	0.3342	0.1159	0.3719	0.2485
0.6	0.6449	0.1944	0.4595	0.3434
0.7	0.7069	0.2421	0.5768	0.4366
0.8	0.7711	0.2733	0.6577	0.5583
0.9	0.8031	0.3501	0.7739	0.6639
1.0	0.8068	0.6551	0.8632	0.8241

Notes:

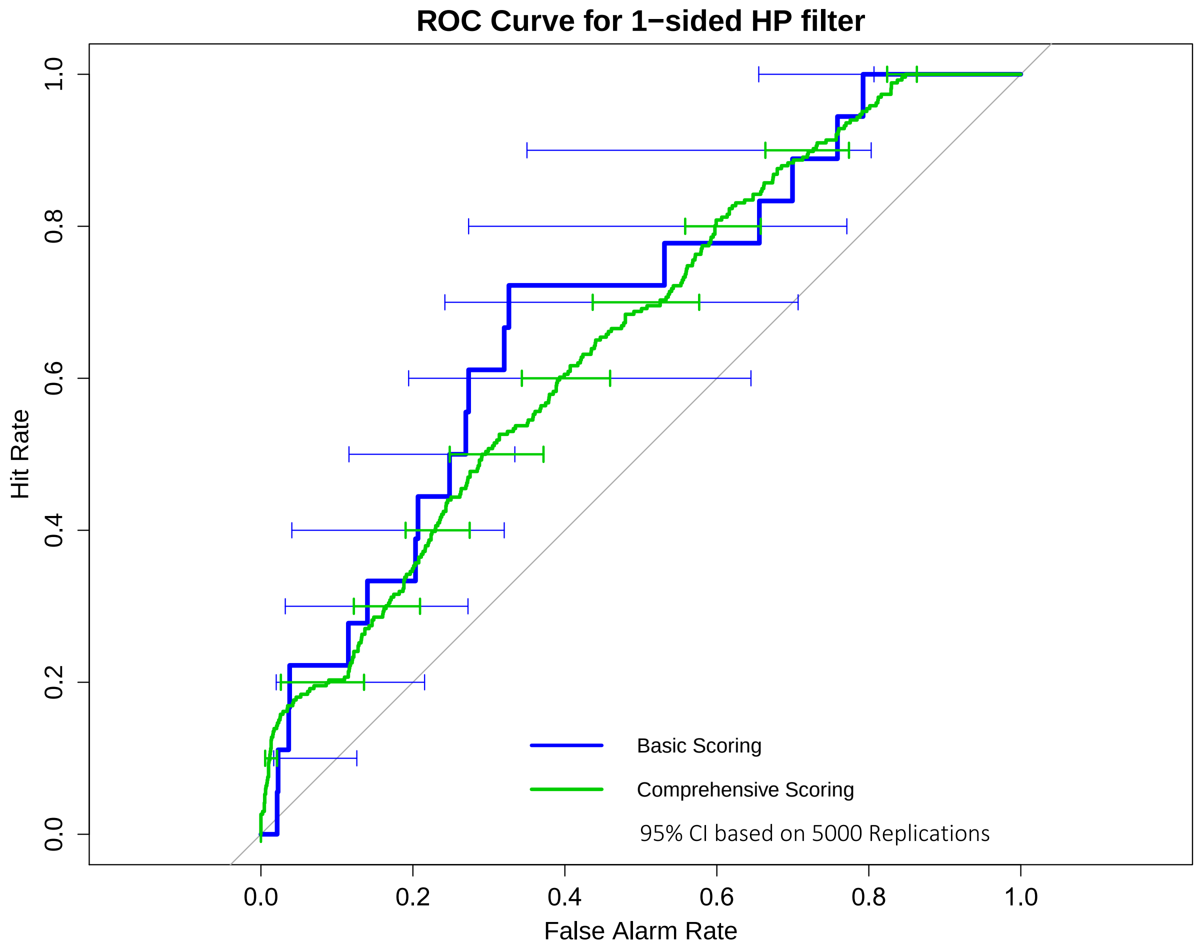
The Table compares the 95% bootstrap confidence intervals of the ROC Curves for both Basic and Comprehensive Scoring for the 1-sided recursive HP($\lambda = 400,000$) filter at selected Hit Rates.

Confidence intervals are calculated using a stratified bootstrap with 5,000 replications.

The ROC curves are calculated using Total Credit/GDP with LV2012 crisis dates and a minimum sample size of $T = 80$.

Also see Figure A.15

Figure A.15: Comparison of Confidence Intervals across Scoring Schemas



Notes:

ROC Curves shown are for the 1-sided recursive HP filter using Total Credit/GDP and LV2012 dates.

95% Confidence Intervals calculated via stratified bootstrap with 5000 replications.

Also see Table A.5