Measuring the Credit Gap

James D. Hamilton (jhamilton@ucsd.edu)
Daniel Leff (dleffyaf@ucsd.edu)
University of California at San Diego

March 6, 2020
Revised: June 25, 2020

Abstract

We revisit the analysis by Drehmann and Yetman (2018) and conclude that measuring the credit gap based on the 5-year growth rate of the credit-to-GDP ratio produces a more reliable and robust predictor of financial crises than does the Hodrick-Prescott filtered series. We also conclude that estimating the credit gap based on the forecast error of a 5-year-ahead regression can be even more useful, provided a sufficiently long sample is available to estimate coefficients of the regression.

1 Introduction.

A recent paper by Drehmann and Yetman (2018) defended the use of the Hodrick-Prescott filter as a tool for identifying deviations from trend in the credit-to-GDP ratio for purposes of setting countercyclical capital buffers under Basel Banking Supervision guidelines. Their conclusions have been cited by over a dozen studies, including Beltran, Jahan-Parvar, and Paine (2019), Reichlin, Ricco and Hasenzagl (2019), Sprincean (2019), and De Jong and Sakarya (2020). In this paper we revisit their findings and reach a different conclusion. Recent papers raising complementary concerns to ours include Schüler (2020) and Karagedikli and Rummel (2020).

2 Summary of earlier results.

Hamilton (2018) proposed interpreting the cyclical component of a time series as the component that could not be predicted two years in advance based on the variable's own lagged values. He showed that for a broad class of nonstationary processes, this component can be estimated from the residuals of a simple linear regression, though he noted that for many economic and financial time series we could accomplish something very similar simply by taking the 2-year change in the variable. He recommended a 2-year horizon on the grounds that "the primary reason that we would be wrong in predicting the value of most macro and financial variables at a horizon of $h = 8$ quarters ahead is cyclical factors such as whether a recession occurs over the next two years and the timing of recovery from any downturn." He further raised a practical reason for not using a larger value of h: "a bigger sample size T will be needed the bigger is h. The information in a finite data set about very long-horizon forecasts is quite limited."

Credit cycles are typically viewed as evolving much slower than business cycles. For analyzing credit cycles, Hamilton recommended, "For such an application I would use $h = 5$ years, with the regression-free implementation $(y_{t+5} - y_t)$ having particular appeal given the length of datasets available."

Drehmann and Yetman (2018) compared Hamilton's two approaches with the popular Hodrick-Prescott characterization of trends for purposes of identifying the cyclical component of the credit-to-GDP ratio. They used quarterly observations on y_{it}, the ratio of credit to GDP for country i in quarter t for $i = 1, ..., 42$ different countries[1]. The first observations on y_{it} come as early as 1970:Q1 for some countries while data for others like Peru only begin in 2001:Q4. They used data through date T_i to estimate a linear 5-year-ahead predictive

[1] This is measured as credit to the private non-financial sector, capturing total borrowing from all domestic and foreign sources. See Dembiermont, Drehmann and Muksakunratana (2013).

regression for country i,

$$y_{it} = \beta_{i0} + \beta_{i1} y_{i,t-20} + \beta_{i2} y_{i,t-21} + \beta_{i3} y_{i,t-22} + \beta_{i4} y_{i,t-23} + v_{i,t} \quad \text{for } t = 1, 2, ..., T_i \tag{1}$$

for T_i the observation 10 years after the data for country i begins. They were interested in the usefulness of the estimated residual $\hat{v}_{i,T_i}$ for purposes of predicting a financial crisis in that country at date $T_i + s$ for different forecasting horizons $s = 1, ..., 12$ quarters. They then re-estimated the regression using data now through date $T_i + 1$ in order to predict a financial crisis at $T_i + 1 + s$, and so on through the end of the sample. The average area under the Receiver Operating Characteristic curve (denoted AUC) for $\hat{v}_{i,T_i}$ as a predictor of financial crisis at $T_i + s$ is then calculated across T_i and across countries. Panel A of Figure 1 reproduces some of their key results. The dot-dashed blue line plots the value of AUC when the predictor is the residual from (1) as a function of the horizon s. Drehmann and Yetman compared this with the signal provided by v^*_{i,T_i}, the terminal residual from an HP filter applied to country i through date T_i using a smoothing parameter of $\lambda = 400,000$ (shown in solid black in Panel A) and with the simple five-year growth rate $v^{\dagger}_{i,T_i} = \log(y_{i,T_i}) - \log(y_{i,T_i-20})$ (dashed red).

The estimated regression does significantly worse than either HP or the 5-year growth. HP does slightly better than the 5-year growth, though the differences are not statistically significant, and the ordering is sensitive to the sample period used. For example, if we start the evaluation with T_i = 1995:Q1 instead of the earliest possible date T_i = 1980:Q1, the 5-year growth does slightly better than HP (see Panel B).

3 Interpreting previous results.

These results raise two interesting questions. First, why does the estimating regression (1) do so poorly, and second, why does HP do so well?

The answer to the first question is that this exercise uses very short horizons to estimate the parameters in regression (1). For example, the first regression for the United States uses T_i = 1980:Q1. Because the regression is estimating a 5-year-ahead forecast and includes 4 lags beyond those 5 years, the dependent variable in regression (1) only runs from t = 1975:Q4 through 1980:Q1, 18 quarterly observations to estimate 5 coefficients. Moreover, the residuals $v_{i,t}$ from this regression are very highly serially correlated due to the 5-year overlapping structure. One would expect regression estimates to perform extremely poorly in this setting, and indeed Hamilton (2018) recommended using the full sample to estimate the trend parameters rather than expanding windows as in the Drehmann and Yetman analysis for exactly this reason. Panel C shows the AUC for the three approaches when the parameters of regression (1) are estimated using the full sample of observations on each country.[2] When implemented this way, the regression approach turns out to do a little better than either of the other two.

To answer the second question, note that the smoothing parameter $\lambda = 400,000$ that Drehmann and Yetman use in their exercise is two orders of magnitude bigger than the value $\lambda = 1600$ that is typically used for quarterly data. Indeed if one were to use $\lambda = 1600$ the performance of HP turns out to be far worse, as seen in Panel D.

When applied to quarterly data of these sample sizes, HP with $\lambda = 400,000$ is almost equivalent to just estimating a time trend. Figure 2 plots credit/GDP for the U.S. along with full-sample estimates resulting from HP with $\lambda = 400,000$ and the estimate of a deterministic time trend. HP and the linear trend differ little. Indeed, if we estimate a deterministic time trend for each country with expanding samples as in regression (1),

$$y_{it} = \alpha_i + \delta_i t + \tilde{v}_{i,t} \quad \text{for } t = 1, 2, ..., T_i, \tag{2}$$

the deterministic time trend if anything performs slightly better than does HP (see Figure 3). Perhaps the title for their paper should have been, "Why You Should Use a Linear Time Trend – At Least to Assess Credit Gaps."

4 Do either linear- or HP-detrending succeed in producing a stationary series?

It is straightforward to investigate whether a deterministic time trend offers a plausible description of these data. An augmented Dickey-Fuller test of the null hypothesis that y_{it} contains a unit root is based on the OLS t-statistic for testing $H_0 : \rho = 1$ in the regression

$$y_t = \alpha + \rho y_{t-1} + \delta t + \textstyle\sum_{i=1}^{p} \zeta_i \Delta y_{t-i} + \varepsilon_t.$$

A large negative value for the test statistic would lead us to reject the null hypothesis of a unit root. These test statistics are reported in column 1 of Table 1.[3] The null hypothesis that the series is characterized by a unit-root process rather than a simple time trend is rejected at the 5% level for only one and at the 10% level for another 2 of the 42 countries investigated by Drehmann and Yetman. On the basis of these tests, it appears that a unit root rather than deterministic time trend might be a better description of most of these series.

Another popular approach is the KPSS test of Kwiatkowski et al. (1992), which takes the

[2] For most countries this goes through 2017:Q3.
[3] We used the Case 4 test described in Hamilton (1994, p. 529) with $p = 4$ lags.

null hypothesis to be that deviations from a linear time trend are stationary:

$$y_t = \alpha + \delta t + v_t \tag{3}$$

$$v_t = \xi_t + u_t$$

$$\xi_t = \xi_{t-1} + \gamma a_t.$$

The null hypothesis that the nonstationarity comes solely from the linear time trend (H_0 : $\gamma = 0$) is tested based on the statistic

$$k_\ell = \frac{\sum_{t=1}^{T} S_t^2}{T^2 \hat{\lambda}_\ell} \tag{4}$$

$$\hat{\lambda}_\ell = \hat{\gamma}_0 + 2\sum_{r=1}^{\ell} \left(1 - \frac{r}{\ell+1}\right) \hat{\gamma}_r$$

$$\hat{\gamma}_r = T^{-1} \sum_{t=r+1}^{T} \hat{v}_t \hat{v}_{t-r}$$

for $\hat{v}_t$ the residuals from OLS estimation of (3). When the lag order $\ell = 1$, the statistic (4) would be proportional to the Lagrange multiplier statistic for testing the null hypothesis $\gamma = 0$ assuming that a_t and u_t are independent Gaussian white noise. Kwiatkowski et al. (1992) derived the properties of the statistic k_ℓ under more general null hypotheses such as $u_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ for a_t and ε_t independent martingale difference sequences with $\sum_{j=0}^{\infty} |\psi_j| < \infty$ so that v_t is stationary. A larger value for the test statistic (4) is stronger evidence against the null hypothesis that y_t is stationary around a deterministic time trend. Column 2 of Table 1 report the test statistic k_ℓ based on $\ell = 8$ lags. The null hypothesis is rejected at the 5% level for 34 out of the 42 countries and at the 10% level for 3 more.

For fixed smoothing parameter λ, a sufficiently large sample size T, and observations near the middle of the sample (t near $T/2$), HP filtering approximately amounts to taking fourth differences of the original series and then resmoothing (e.g., King and Rebelo, 1993), which one might have expected to make most series stationary. This is not the case for observations near the beginning or end of the sample (t near 1 or T), and for finite T and large λ, the full sample for $t = 1, ..., T$ of an HP-filtered series may not behave like a stationary process even if the original series is only $I(1)$, as shown by De Jong and Sakarya (2016) and Phillips and Jin (2015). Phillips and Shi (2019) recommended performing an augmented Dickey-Fuller test on the cyclical component resulting from HP filtering, and if the cyclical component appears to be nonstationary using standard p-values, "boosting" the filter by HP-filtering a second time the residuals that resulted from the first pass of the HP filter. Given the close similarity between the dotted and dashed lines in Figure 2, and given the fact that removing a linear trend does not appear produce a stationary series for most countries, it seems quite possible that HP filtering for these sample sizes and this smoothing parameter does not produce a stationary time series.

It is simple enough to repeat the above exercises with y_{it} now replaced by c_{it}, the cyclical component of the credit-to-GDP ratio that results from applying a two-sided HP filter with $\lambda = 400,000$. Column 3 of Table 1 repeats the augmented Dickey Fuller test with y_{it} replaced by c_{it}. We'd now reject the unit-root hypothesis at the 5% level for two countries and at the 10% level for 9 more; for most countries we'd conclude that the HP-filtered series still exhibit a unit root. Column 4 repeats the KPSS test with y_{it} replaced by c_{it}. This test provides evidence against stationarity at the 10% level for 22 countries, about half of the total..

The conclusion we draw from Table 1 is that either linear detrending or applying the HP filter with $\lambda = 400,000$ often fail to capture the trend in the credit-to-GDP ratio.

The main problem with using a deterministic time trend, or with using an HP filter very similar to a time trend, to try describe a unit-root process is that patterns in the resulting cyclical component will be an artifact of the filtering procedure itself and not a feature of the data-generating process. If a random run-up in credit turned out to be followed by a decline, taking residuals from a time trend or applying HP with $\lambda = 400,000$ will make it artificially appear ex post that one could somehow have known ex ante that a crash was certain to follow after the run-up.

Complementary concerns to ours have also been raised by two other recent studies. Schüler (2020) argued that HP detrending with $\lambda = 400,000$ introduces spurious cycles in the credit-GDP ratio at precisely the medium-term frequencies in which researchers and regulators are interested and which have no basis in the underlying data-generating process. Karagedikli and Rummel (2020) expressed concern that the procedure puts high weights on observations from the distant past, citing the experiences of Malaysia in 1997 and Argentina in 2002 as examples where this inference may be problematic.

What then is the justification for using $\lambda = 400,000$? Drehmann and Yetman cited Borio and Lowe (2002), though that study appears to use annual data. Ravn and Uhlig (2002) argued that the value of $\lambda = 1600$ that is used in almost all applications of the HP filter to quarterly data is analogous to using $\lambda = 129,600$ on monthly or $\lambda = 6.25$ on annual data. The primary justification for using $\lambda = 400,000$ for quarterly data appears to come from Annex 1 in Drehmann et al. (2010), who argued that the choice of $\lambda = 1600$ for quarterly data is motivated by the desire to match business cycle periodicities whereas credit may exhibit longer cycles of 5 to 20 years. In point of fact, the justification that Hodrick and Prescott gave for $\lambda = 1600$ was not the expected length of the business cycle, but rather their expectation that the trend would typically not change by more than 1/8% in a single quarter, whereas the cyclical component might change 5%. The square of the ratio between these, $(5 \times 8)^2 = 1600$ is the basis for the standard recommendation of $\lambda = 1600$. Adapting that argument in the present context, the implicit assumption would be that if the cyclical component of the credit-to-GDP ratio often changes by as much as 5% within a quarter, the trend in the credit-to-GDP ratio that we are seeking to identify typically does not change by more than 0.008% in a single quarter.[4] This seems a hard case to make on its own merits, particularly given its inconsistency with the observed facts documented in Table 1.

5 Conclusion.

The claim that one can identify the trend component in the credit-to-GDP ratio on the basis of a linear time trend or something very close to it is not what would have been expected a priori and receives no support from the data. Moreover, its use is motivated at least in part by the fact that the resulting estimate of the gap was found to be correlated with subsequent financial crises by other researchers using earlier data sets. To the extent this is the case, it calls into question whether exercises like those in Drehmann and Yetman (2018) can truly be characterized as a clean out-of-sample forecast evaluation.

The simple 5-year growth is a robust, model-free and estimation-free alternative that performs almost as well or better than HP with $\lambda = 400,000$ and far better than HP with conventional smoothing weights. Furthermore, allowing for the possibility of multiple unit roots and estimating parameters with a linear autoregression as in (1) also appears to be a useful approach in this setting, though we would probably want 40 years of data before trusting those estimates for purposes of identifying the cyclical component in the credit-to-GDP ratio. Greenwood et al. (2020) provided a nice illustration of model-free use of credit-to-GDP ratios.

[4] That is, $(5/0.0079)^2 = 400,000$.

References

Beltran, Daniel O., Mohammad R. Jahan-Parvar, and Fiona A. Paine (2019). "Designing Gaps as Early Warning Indicators: Accuracy, Predictive Power, and Robustness," working paper, Federal Reserve Board.

Borio, Claudio EV, and Philip William Lowe (2002). "Asset Prices, Financial and Monetary Stability: Exploring the Nexus," BIS Working Paper 114.

De Jong, Robert M., and Neslihan Sakarya (2016). "The Econometrics of the Hodrick-Prescott Filter," *Review of Economics and Statistics* 98: 310–317.

De Jong, Robert M., and Neslihan Sakarya (2020). "The one-sided Hodrick-Prescott filter," working paper, Ohio State University.

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.

Drehmann, Mathias, Claudio E.V. Borio, Leonardo Gambacorta, Gabriel Jimenez, and Carlos Trucharte (2010). "Countercyclical Capital Buffers: Exploring Options," BIS Working Paper 317.

Drehmann, Mathias, and James Yetman (2018). "Why You *Should* Use the Hodrick-Prescott Filter– At Least to Assess Credit Gaps," BIS Working Paper 744.

Predictable Financial Crises

Greenwood, Robin, Samuel G. Hanson, Andrei Shleifer, and Jakob Ahm Sørensen (2020). "Predictable Financial Crises," NBER Working Paper No. 27396.

Hamilton, James D. (1994). *Time Series Analysis,* Princeton University Press.

Hamilton, James D. (2018). "Why You Should Never Use the Hodrick-Prescott Filter," *Review of Economics and Statistics,* 100: 831-843.

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

Karagedikli, Özer and Ole Rummel (2020). "Weighing Up the Credit-to-GDP Gap: A Cautionary Note," working paper, SEACEN Research and Training Centre.

King, Robert and Sergio T. Rebelo (1993). "Low Frequency Filtering and Real Business Cycles," *Journal of Economic Dynamics and Control* 17: 207-231.

Kwiatkowski, Denis, Peter C.B. Phillips, Peter Schmidt, and Yongcheol Shin (1992). "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root," *Journal of Econometrics* 54: 159-178.

Phillips, Peter C.B., and Sainan Jin (2015). "Business Cycles, Trend Elimination, and the HP Filter," working paper, Yale University.

Phillips, Peter C.B. Phillips and Zhentao Shi (2019). "Boosting: Why you Can Use the HP Filter," working paper, Yale University.

Ravn, Morten O., and Harald Uhlig (2002). "On Adjusting the Hodrick-Prescott Filter for the Frequency of Observations," *Review of Economics and Statistics* 84: 371-376.

Reichlin, Lucrezia, Giovanni Ricco, and Thomas Hasenzagl (2019). "Financial Variables as Predictors of Real Growth Vulnerability," working paper, London Business School.

Schüler, Yves Stephan (2020). "On the Credit-to-GDP Gap and Spurious Medium-term Cycles," *Economics Letters*, 192: 109245.

Sprincean, Nicu (2019). "Early warning indicators for macrofinancial activity in Romania," *Review of Economic and Business Studies* 12: 137-162.

Table 1. Unit-root tests on original data and HP-filtered data.

	Original data		HP-filtered	
	(1) ADF	(2) KPSS	(3) ADF	(4) KPSS
Argentina	-2.99	0.17**	-3.37*	0.11
Australia	-3.18*	0.19**	-3.30*	0.08
Austria	-1.01	0.21**	-2.34	0.09
Belgium	-1.79	0.51***	-3.27*	0.17**
Brazil	-1.64	0.23***	-1.78	0.21**
Canada	-1.68	0.17**	-2.65	0.09
Chile	-2.62	0.10	-2.64	0.10
China	-1.57	0.26***	-2.54	0.16**
Colombia	-2.12	0.25***	-2.23	0.24***
Czech Republic	-1.55	0.28***	-1.66	0.25***
Denmark	-1.57	0.35***	-0.88	0.11
Finland	-2.34	0.15**	-2.61	0.10
France	-1.01	0.44***	-3.23*	0.11
Germany	-1.48	0.36***	-2.77	0.15**
Greece	-1.87	0.48***	-1.43	0.20**
Hong Kong	-1.07	0.28***	-2.66	0.17**
Hungary	-0.75	0.16**	-0.70	0.16**
India	-1.75	0.36***	-1.43	0.16**
Indonesia	-1.97	0.26***	-2.46	0.13*
Ireland	-1.68	0.46***	-1.57	0.14*
Italy	-2.06	0.46***	-2.06	0.14*
Japan	-1.65	0.49***	-2.38	0.19**
Korea (South)	-3.08	0.17**	-3.58**	0.08
Malaysia	-1.47	0.43***	-2.96	0.13*
Mexico	-2.94	0.12*	-3.20*	0.11
Netherlands	0.27	0.37***	-1.99	0.10
Norway	-2.60	0.25***	-3.40*	0.11
Peru	-3.11	0.15**	-3.08	0.15**
Philippines	-2.91	0.13*	-3.00	0.11
Poland	-2.75	0.12*	-2.66	0.09
Portugal	-1.92	0.39***	-2.18	0.16**
Russia	-3.07	0.09	-3.10	0.08
Saudi Arabia	-3.30*	0.07	-3.41*	0.06
Singapore	-2.23	0.12	-2.63	0.08
South Africa	-2.64	0.29***	-3.31*	0.07
Spain	-2.40	0.39***	-2.16	0.15**
Sweden	-2.40	0.33***	-3.26*	0.09
Switzerland	-2.08	0.27***	-2.87	0.13*
Thailand	-2.02	0.34***	-2.79	0.14*
Turkey	-0.93	0.34***	-1.61	0.25***
United Kingdom	-1.81	0.17**	-2.18	0.12*
USA	-3.51**	0.12	-3.78**	0.10

Notes to Table 1: * denotes statistically significant at 10% level; ** 5%; *** 1%.

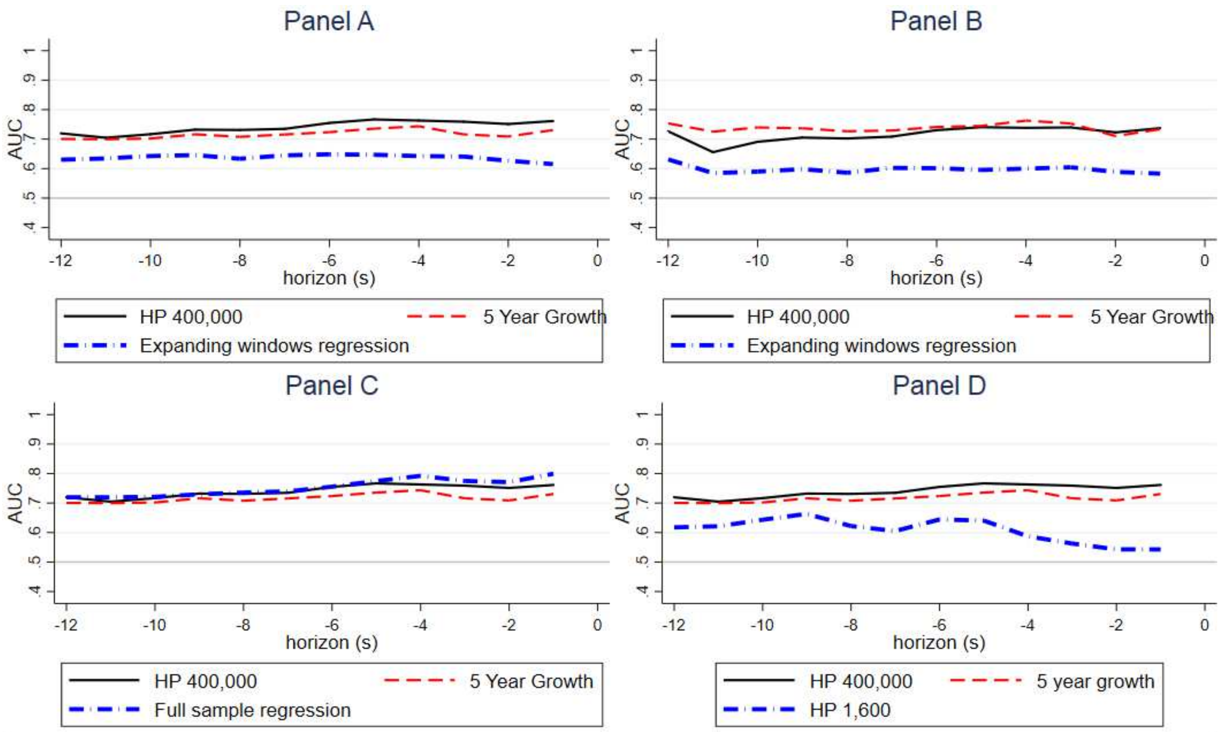

Figure 1. AUC as a function of horizon s based on cyclical component of credit/GDP estimated in different ways.

Notes to Figure 1. Panel A: credit gap as assessed using (1) HP with λ = 400,000; (2) 5-year change in the log of credit/GDP, and (3) regression estimated from credit/GDP data through date T_i to predict crisis s quarters after T_i, with evaluations beginning 10 years after the first observation of credit/GDP for each country. Panel B: same as panel A except evaluations begin 1995:Q1. Panel C: same as Panel A except regression parameters estimated using full sample of observations for each country. Panel D: (1)-(2) as in Panel A with (3) replaced by HP with λ = 1600.

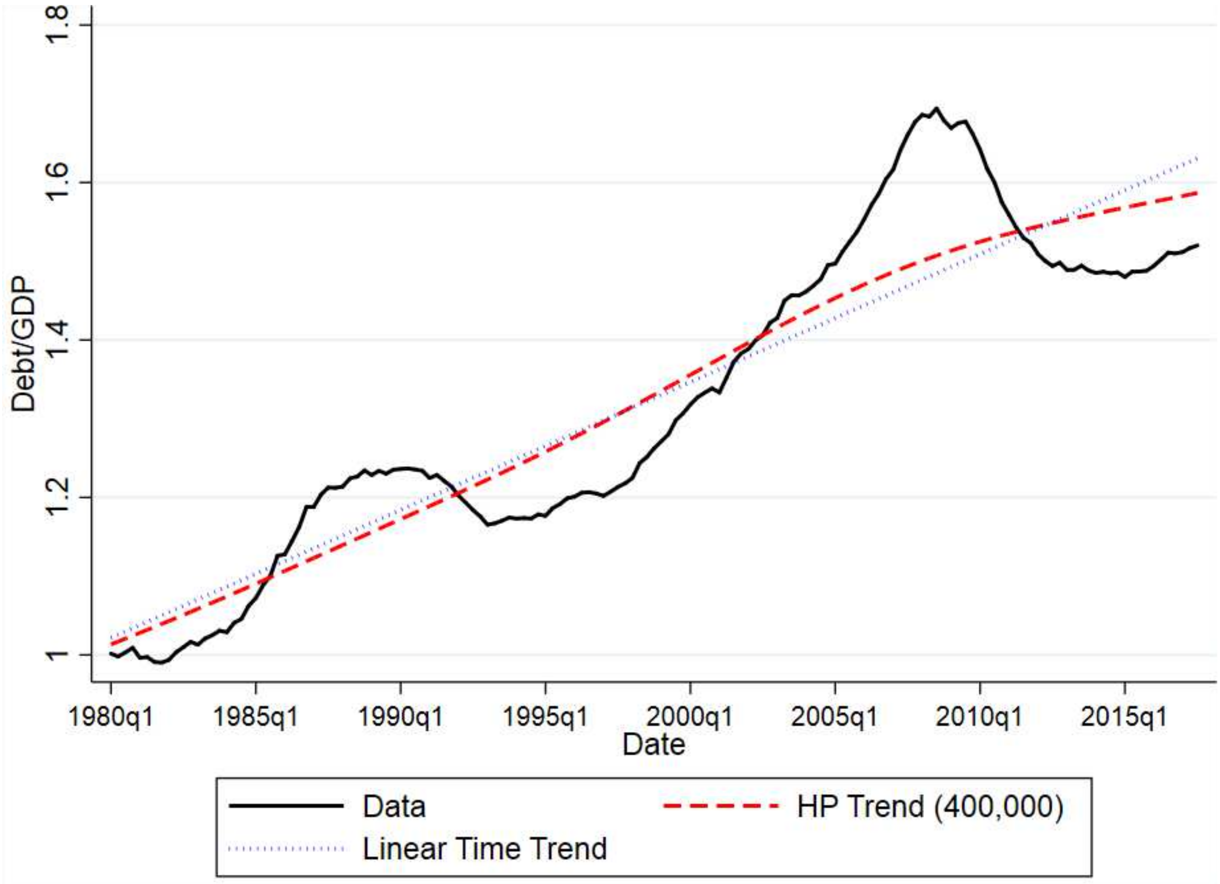

Figure 2. U.S. credit/GDP ratio 1980:Q1 to 2017:Q4 and full-sample estimates of (1) HP trend with λ = 400,000; (2) deterministic time trend.

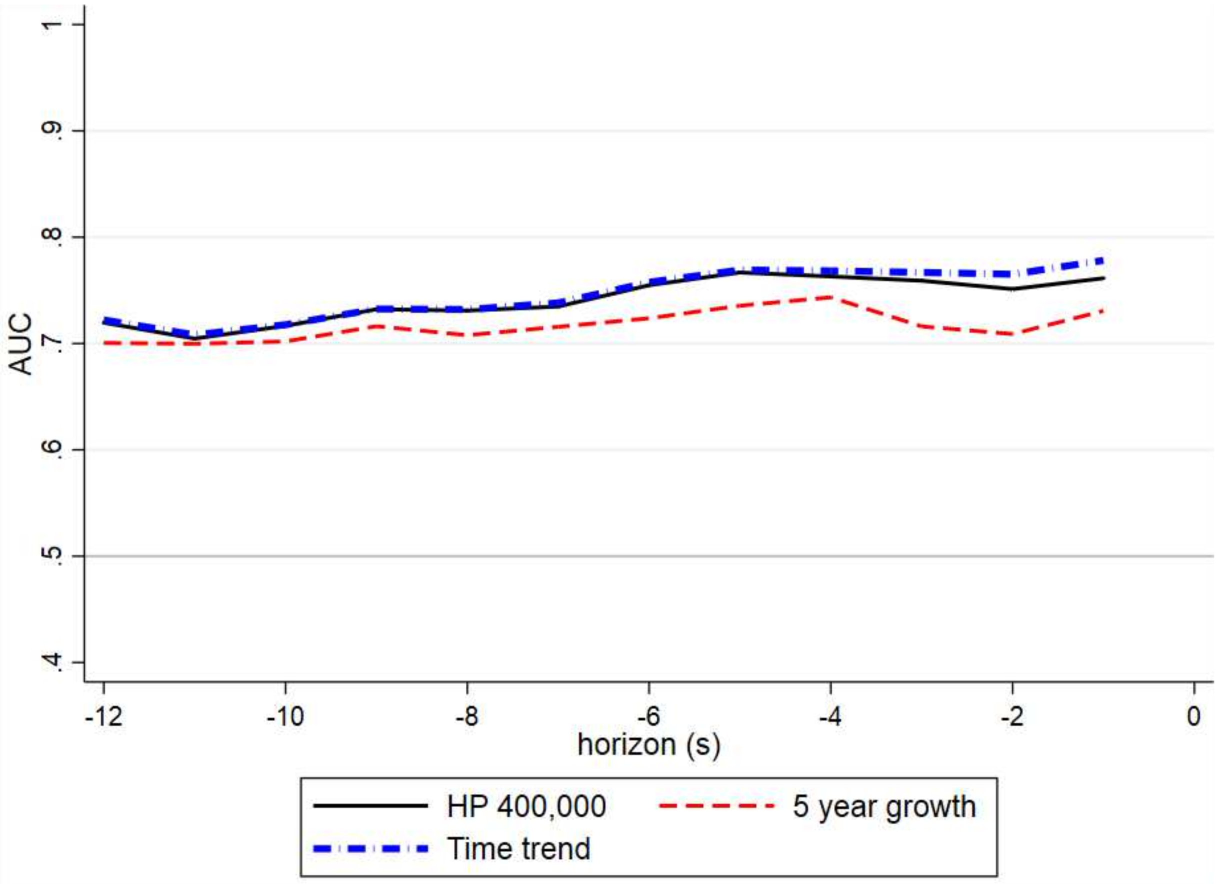

Figure 3. AUC as a function of horizon *s* based on cyclical component of credit/GDP estimated from (1) HP with λ = 400,000; (2) 5-year change in the log of credit/GDP, and (3) residuals from deterministic time trend estimated from regressions of expanding sample size.