Basel III Credit-to-GDP Gaps and the Origins of Their Unreliability: Introducing Historical Reliability Bands

Josefine Quast*
Deutsche Bundesbank, University of Würzburg

—This version: July 2022—

Abstract

Basel III credit-to-GDP gaps are used to assess whether aggregate credit is excessive or not and inform macroprudential policymaking. Yet, estimates from Basel III's prescribed detrending procedure are prone to continuous reevaluations that do not reflect changes in the data and exceed commonly discussed end-of-sample biases from converging one- and two-sided filtering procedures. To illustrate the extent of unreliability, I introduce historical reliability bands. Based on simulation and empirical evidence for 43 countries, I show that estimates do not converge to full sample estimates and each quarter is associated with a new trend history, which compromises the comparability of cyclical positions over time. This leads to relevant misalignment in countercyclical buffer decisions in both, direction and size, and may impair the regulator's credibility. Alternatively using a two-year difference filter would provide endpoint and historically reliable estimates, while yielding distinctly fewer and smaller decision misalignment and maintaining credit gap dynamics and amplitudes that the regulator deems important.

Keywords: Countercyclical capital buffers, detrending, stochastic trends, real-time analysis
JEL classification: C10, C32, E32, E44, G01

*This paper benefited greatly from extensive and fruitful discussions with James D. Hamilton, Yves S. Schüler, and Maik Wolters. I am further grateful to Christiane Baumeister, Graham Elliot, Stephanie Ettmeier, Isabel Gödl-Hanisch, Elmar Mertens, James Morley, Galina Potjagailo, Christian Proaño, Frank Schorfheide, Allan Timmermann as well as participants of the IAAE 2022 Annual Conference for their valuable comments. The views in this paper are solely mine and should not be interpreted as reflecting the views of the Deutsche Bundesbank or the Eurosystem.

1 Introduction

Since the Global Financial Crisis (GFC), excessive credit growth has been considered as an important early warning signal for financial and banking crises (see, e.g., Claessens et al., 2012; Drehmann et al., 2012; Schularick and Taylor, 2012; Jordà et al., 2013; Aikman et al., 2015; Drehmann and Yetman, 2021).[1] Further, financial-crisis recessions are more costly than normal recessions in terms of output losses (see, e.g., Kaminsky and Reinhart, 1999; Cerra and Saxena, 2008; Reinhart and Rogoff, 2009; Cecchetti et al., 2009; Schularick and Taylor, 2012; Jordà et al., 2013; Greenwood et al., 2020). But, aiming to prevent them through expanding capital requirements can be costly as lending capacity gets withdrawn from the banking system. Inaccurate early warning signals may, additionally, be costly in terms of foregone financial deepening (false-positive signals) or missed financial crises (false-negative signals).

The timely and reliable measurement of credit growth has, thus, become a crucial task for macroprudential regulation since the GFC (van Norden and Wildi, 2015). Countercyclical capital buffer (CCyB) decisions on raising or releasing bank capital requirements prior to or during periods of perceived distress need to be based on reliable indicators. If indicators are continuously revised instead, the regulator's credibility may get impaired. Not only may wrong buffer add-ons be prescribed in real time, but ongoing reevaluations may make it difficult to justify the initial buffer add-on decision. This may diminish the regulator's ability to effectively run CCyB regimes.

In this paper, I show that Basel III credit-to-GDP gaps remain *historically unreliable* since they continue to be revised when more data is added, even after endpoint biases from converging one- and two-sided filters have settled. Importantly, as the whole history of cyclical and trend positions gets continuously reevaluated, this does not reflect changes in the data, but filtering artifacts instead. While endpoint biases are typically associated with revisions within the first three years after initial estimation, continued wandering rather indicates lacking convergence to full sample estimates. I introduce *historical reliability bands* to illustrate the range with which a period t estimate varies between initial quasi real-time and full sample estimation, considering all values in between. I use simulation and empirical evidence for 43 countries to show that these unreliability patterns prevail because the Basel III prescribed detrending procedure does not sufficiently account for stochastic trends in the credit-to-GDP series. Each quarter is, further, associated with a new trend estimate, where changes in the whole trend history compromise the

[1] In addition, Greenwood et al. (2020) and Sufi and Taylor (2021) emphasize the predictive power of credit growth when it is accompanied by rapidly growing asset prices.

comparability of cyclical positions over time.[2]

This closely relates to the Nelson-Kang critique of inappropriately detrending random walks with procedures that resemble linear detrending techniques (Nelson and Kang, 1981; Watson, 1986; Cogley and Nason, 1995). I, finally, show that historical unreliability matters for CCyB decision making. There is distinct decision misalignment in both, direction and size. For example, 37% of the time, the initially indicated sign for the credit-to-GDP gap would not sustain so that the economic interpretation of the cyclical position would change often. The real-time indicated buffer add-on size would be reevaluated similarly often and these reevaluations would be sizable: across the 43 countries covered in the BIS database on long credit series, changes would amount to, on average, 0.73% of risky weighted assets per reevaluation.

This paper proposes to alternatively use a two-year difference filter, because doing so would cure the symptoms observed for the prescribed detrending procedure. The alternative accounts for stochastic trends and yields endpoint as well as historically reliable estimates. Moreover, it extracts a historical trend so that the comparability across cyclical positions over time is maintained. A two-year difference filter provides a similarly transparent and easy to use detrending tool, while preserving credit-to-GDP gap dynamics and amplitudes, the regulator deems important.[3] Buffer add-on decision misalignment is far less distinct. The signs of credit-to-GDP gaps do not switch as often and if add-on reevaluations occur, they remain small.

This paper, further, verifies that while the Basel III credit-to-GDP gap seems to work well for the United States as it signals buffer building with enough anticipation prior to US banking crises, it produces implausible answers for other economies. I use the country cases of Belgium and Brazil as examples to illustrate such implausibilities. Beyond these specific cases, I also discuss cross-country averages to illustrate that all countries are, on average, affected by such implausibilities. For these two economies, a prolonged and extensive need for buffer building would have been signaled in real time (for Belgium from the mid-1990s to 2005; for Brazil from 2008 to 2017/18), while later assessments would have indicated distinct underutilization of economy-wide leverage instead.[4] The ECB and the IMF document

[2]Hamilton (2021) and Baumeister and Guérin (2021) highlight similar issues when applying linear detrending in context of measuring global real economic activity as in Kilian (2009, 2019).

[3]Recursively demeaning so derived credit-to-GDP gaps would further ensure their zero-centering, so that CCyB add-ons could be operated. Moreover, while the Basel III credit-to-GDP gap remains indistinguishable from unit root processes for most countries (Karagedikli and Rummel, 2020), using a two-year difference filter would yield stationary gap estimates for the vast majority of countries.

[4]The following analysis contains multiple examples of false-positive signals. Emerging markets seem to be especially af-

similar economically implausible patterns for European economies after the credit cycle reverts (ECB, 2017; Baba et al., 2020). Against the very high degree of advised smoothing, both the one-sided quasi real-time and the ex post two-sided HP filter configuration, remain prone to the Nelson-Kang critique. Econometrically, no statement can be made about whether one of them provides the "right" assessment as both remain unreliable.

The main reason why the Basel III credit-to-GDP gap produces seemingly plausible US patterns is that it was partly designed to capture US crises signals (Borio and Drehmann, 2009). Basel III and its credit-to-GDP gap became effective *after* the GFC, building on ex post knowledge. But even for the United States, Edge and Meisenzahl (2011) estimate that false-positive signals could yield substantial costs in terms of capital shortfalls and reductions in lending. Those could have amounted to as much as 141 billion USD following a false-positive signal in 2001Q4, which would have slowed the economic recovery after the burst of the dot.com bubble. As discussed in ECB (2017), the credit-to-GDP gap's good early warning properties emphasized in the literature are associated with in-sample forecasting exercises. The picture remains more blurry for true out-of-sample projections beyond the available data.

Methodologically, the HP filter is considered to provide a simple stochastic detrending tool. Yet, it nests the special case of linear detrending if the smoothing parameter is set to $\lambda = \infty$. Given the Basel III prescription of setting $\lambda = 400,000$ to detrend quarterly credit-to-GDP series, it is a priori not clear whether this configuration is capable of absorbing potential unit roots or whether it de facto amounts to linear deterministic detrending instead. Based on simulation and empirical cross-country time-series evidence, I verify that the implementers of the prescribed procedure face very similar challenges and outcomes as if they had implemented linear detrending instead.

Highlighting the limitations of the Basel III credit-to-GDP gap is not new. Past criticism spans from GDP normalization and procyclicality issues (BCBS, 2010; Repullo and Salas, 2011; Jokipii et al., 2021), to the documentation of implausible country patterns (ECB, 2017; Baba et al., 2020), endpoint biases (Edge and Meisenzahl, 2011), the role of (spurious) cyclical dynamics (Hamilton, 2018; Schüler, 2020), and misconceptions between one- and two-sided cyclical properties (van Norden and Wildi, 2015; Wolf et al., 2020). Karagedikli and Rummel (2020) further remark that observations from the distant past distinctly impact current credit-to-GDP gap and trend estimates, which is at odds with the Basel Committees' intentions (BCBS, 2010, p. 13).

fected (see results Appendix B).

Linking Basel III's detrending procedure to the Nelson-Kang critique has, however, been overlooked. I establish this crucial link that reconciles many of the critiques raised in the literature: most notably the spurious cycles discussion, the documentation of implausible, seemingly random, country patterns, and Edge and Meisenzahl's (2011) endpoint bias considerations. More specifically, Nelson and Kang (1981) show that linearly detrending a random walk introduces spurious periodicity to the resulting cyclical component that is not characteristic of the underlying data generating process (DGP), but is a function of the sample length instead. The spurious medium-term cyclical dynamics discussed in Schüler (2018, 2020) are a direct reflection of this link. Thus, understanding the nature of the trend in credit-to-GDP is essential for deriving appropriate gaps. To the best of my knowledge, only Hamilton and Leff (2020) and Karagedikli and Rummel (2020) emphasize this need.

The remainder of this paper is organized as follows. Section 2 provides simulation evidence to discuss reliability patterns for known DGPs. Section 3 then compares these stylized patterns to those characterizing the empirical country series, where the true DGP remains unknown. Section 4 proposes to use a two-year difference as a simple alternative, because it circumvents the econometric issues associated with the Basel III credit-to-GDP gap. Section 5 shows how historical unreliability affects CCyB decision making and section 6 concludes.

2 Deriving Basel III Credit-to-GDP Gaps: A Simulation Exercise

Since the underlying DGP of credit-to-GDP remains unknown, it is a priori not clear whether its long-term trend is deterministic or stochastic. To inform macroprudential policymaking about whether aggregate credit is excessive or not, the Basel Committee's Macro Variables Task Force's (MVTF) proposed to derive country-specific credit-to-GDP gaps from a one-sided HP filter with a very large smoothing parameter: $\lambda = 400,000$. This procedure became an integral part of the Basel III regulations. Aiming to circumvent the well-established endpoint problem from converging one- and two-sided filtering techniques at both sample ends, the guidelines (BCBS, 2010; European Systemic Risk Board, 2014) explicitly advise implementing the one-sided version of the following HP filter (Hodrick and Prescott, 1981, 1997):

$$\min_{\{g_t\}_{t=-1}^{T}} \left\{ \sum_{t=1}^{T} (y_t - g_t)^2 + \lambda \sum_{t=1}^{T} \left[(g_t - g_{t-1}) - (g_{t-1} - g_{t-2}) \right]^2 \right\}. \quad (1)$$

Hamilton (2018) shows that the solution for equation (1) is given by:

$$g^* = (H'H + \lambda Q'Q)^{-1}H'y = A^*y, \tag{2}$$

where $\tilde{T} = T + 2$, $\underset{(Tx\tilde{T})}{H} = \begin{bmatrix} \underset{(TxT)}{I_T} & \underset{(Tx2)}{0} \end{bmatrix}$, and

$$\underset{(Tx\tilde{T})}{Q} = \begin{bmatrix} 1 & -2 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 1 & -2 & 1 \end{bmatrix}.$$

For the one-sided HP endpoint series and $s = 1 : t$, it then follows for the recursive quasi real-time credit-to-GDP trend: $g_s^* = A^*_{[1xs]}y_{[1xs]'}$ and the recursive quasi real-time credit-to-GDP gap estimate: $c_s^* = y_s - g_s^*$.

A friction for the regulator's potential intention to extract a stochastic trend may arise, because it is a priori not clear whether $\lambda = 400,000$ is large or small enough to be associated with stochastic or deterministic detrending. While the HP filter is generally considered to provide a simple stochastic detrending tool, recall that as the smoothing parameter λ approaches infinity, the one- and the two-sided HP trends approach linear deterministic trends (see, e.g., Hodrick and Prescott, 1981; Cogley and Nason, 1995; Hodrick and Prescott, 1997; Hamilton, 2018; Hamilton and Leff, 2020). However, the regulator's trend assumption remains unclear since the publicly accessible guidelines do not contain an explicit mentioning.

De Jong and Sakarya (2016) provide a formalization that the HP filter is generally capable of absorbing unit roots in the trend component. Moreover, they provide further justification for Ravn and Uhlig's (2002) power four rescaling of λ in context of frequency changes in the observations. They, however, maintain Hodrick and Prescott's (1981; 1997) recommendation to set $\lambda = 1,600$ for analyzing quarterly data, which is associated with a prior belief that the cycle (σ_c^2) is much more volatile than the trend (σ_t^2) ($\lambda = \frac{\sigma_c^2}{\sigma_t^2} = (\frac{5}{1/8})^2 = 1,600$). BCBS (2010) explicitly refers to Ravn and Uhlig (2002) when advising to set $\lambda = 400,000$. This setting is motivated by the desire to extract long cycles lasting approx. 30 years ($\lambda = 400,000 \sim 4^4 \cdot 1,600$). But in contrast to Ravn and Uhlig (2002), this is not accompanied by a change in observation frequency, which implies a very large cyclical variability with regular within-quarter changes in credit-to-GDP of up to 5% (Hamilton and Leff, 2020).

Nelson and Kang's Critique (1981) If credit-to-GDP were characterized by a unit root, while Basel III's detrending procedure closely resembles linear detrending, spurious periodicity would be introduced to the extracted cyclical component (Nelson and Kang, 1981). Fitting a linear trend to a random walk will, thus, yield the cyclical component at period t to contain ex post information from later observed values (Hamilton, 2021). Instead of characterizing genuine cyclical dynamics of the underlying DGP, the predominant peak in the cyclical component's spectral density function will depend on the sample length instead (Chan et al., 1977; Nelson and Kang, 1981; Watson, 1986; Cogley and Nason, 1995; Hamilton, 2018; Hamilton and Leff, 2020; Schüler, 2020).[5] As a result, so extracted cyclical components perform spuriously well in in-sample forecasting evaluations (Hamilton, 2021).

2.1 Outlining the Simulation Exercise

As the HP filter spans stochastic to linear deterministic detrending, I am interested in where the Basel III prescribed HP filter configuration aligns when it is applied to credit-to-GDP series. To assess this, I set up a simulation exercise where the procedure is used to detrend three known benchmark DGPs: (i) a random walk with drift, (ii) the implied HP model, the regulator prefers, and (iii) a trend-stationary process. To assess how much the quasi real-time estimate changes when more data is added but the procedure is maintained, I construct a measure that quantifies *historical reliability* in the estimated cyclical components. By historical reliability I mean the range with that a period t estimate varies between its initial quasi real-time value, its full sample estimate, and all values in between.

Recall that recursive quasi real-time estimates are based on information available up until t, while full sample estimates are obtained via running the same procedure on information until T. Instead of only focusing on these two estimates, which is usually done in the real-time literature, I also investigate all estimates in between. This allows me to assess historical reliability, which goes beyond endpoint considerations. As will be seen below, estimates can appear real-time reliable when initial quasi real-

[5] Long and Herrera (2020) further argue that doing so yields the OLS estimators themselves to become spurious.

time and full sample estimates align. However, the estimate is only historically reliable if all estimates in between align as well. If they do not, real-time reliability is illusive and the estimates remain historically unreliable. All analyses are conducted as quasi real-time exercises on an expanding sample since 'true' real-time vintages for credit-to-GDP are not attainable from the database introduced below.

While GDP is revised substantially up to several years after initial publication, the output gap literature shows that data revisions are relatively unimportant drivers of these revisions. Ex post reevaluations of cyclical positions are predominantly filter-induced, i.e. an artifact of the employed trend-cycle decomposition (see, e.g., Orphanides and van Norden, 2002; Hamilton, 2018; Quast and Wolters, 2022). Edge and Meisenzahl (2011) show that this carries over to the credit-to-GDP context. They construct 'real' credit-to-GDP real-time vintages for the US and show that the main source for revisions is related to filter-induced endpoint biases instead of revisions to the underlying data.[6]

Historical reliability is considered to be low if the quasi real-time estimate differs greatly from later estimates and high if the two differ little. I, further, consider historical unreliability to be excessive if the difference between the quasi real-time and later estimates continues to increase beyond three years. Borrowing from the output gap reliability literature and the associated credit-to-GDP gap discussion, endpoint measurement issues from converging one- and two-sided filtering techniques are typically associated with revisions up to two or three years (see, e.g., Baxter and King, 1999; Orphanides and van Norden, 2002; Bernhardsen et al., 2005; Cayen and van Norden, 2005; Marcellino and Musso, 2011; Edge and Meisenzahl, 2011; Quast and Wolters, 2022). By then, HP filtered weights should have settled. If the estimates continue to change and do not converge to full sample estimates, the resulting gap and trend measures remain historically unreliable beyond endpoint issues.

To assess historical reliability in the resulting gap and trend estimates, I simulate the implementation of the Basel III prescribed HP filter configuration alongside recursive linear detrending for the following three benchmark processes. For parameter calibration, I use US credit-to-GDP (1947Q4 to 2020Q4). In spirit of Hodrick (2020) and Schüler (2021), coefficients for the drift and error standard deviation align with the sample mean (0.428) and standard deviation (0.892) of the full-sample quarterly continuously compounded rate of US credit-to-GDP growth.

[6] They obtain series on US nominal GDP from Alfred and credit series "from Federal Reserve sources" (Edge and Meisenzahl, 2011, p. 268).

Random Walk Model with Drift: $y_t = 0.428 + y_{t-1} + 0.892\varepsilon_t$ with $\varepsilon_t \sim N(0,1)$.

Regulator's Preferred One-Sided HP Model: The regulator's preferred trend-cycle decomposition, $y_t = g_t + c_t$, implies for the trend component: $g_t = 2g_{t-1} - g_{t-2} + \sqrt{\frac{0.892}{\lambda}}\eta_t$ and the cyclical component: $c_t = 0.892\varepsilon_t$. $\eta_t \sim N(0,1)$ and $\varepsilon_t \sim N(0,1)$, which are uncorrelated. Imposing $\lambda = 400,000$ and calibrating σ_{c_t} to 0.892 then scales σ_{g_t} via $\lambda = \frac{\sigma_c^2}{\sigma_t^2} \rightarrow \sigma_t^2 = \frac{0.892}{400,000}$ (Hamilton, 2018). The first three trend observations are initialized using US credit-to-GDP.

Trend-Stationary Process: $y_t = 0.428t + 0.892\varepsilon_t$ with $\varepsilon_t \sim N(0,1)$.

I choose a sample length in correspondence to the length of the available US credit-to-GDP series, $T = 293$, and conduct 5000 simulations per DGP.[7]

While the trend-stationary process unlikely reflects the regulator's null hypothesis regarding the nature of the trend in credit-to-GDP, I include it to assess where the regulator's preferred model aligns: are reliability patterns in the country series more similar to detrending a random walk or a trend-stationary process with the prescribed HP filter configuration? As mentioned above, this is a priori not clear since the filter nests the special case of linear detrending as $\lambda \rightarrow \infty$. Beyond gap reliability, I am explicitly interested in trend reliability because if a procedure is prone to the Nelson-Kang critique, the trend characterizing coefficients will be continuously revised, i.e. slopes and intercepts will continue to change on an expanding sample (Nelson and Kang, 1981; Watson, 1986). With gap estimates being derived as trend deviations, this would compromise their comparability over time.

After discussing historical reliability for the three benchmark processes, I compare these stylized patterns to those obtained from analogously detrending the country credit-to-GDP series in section 3.

2.2 Measuring Historical Reliability

How much does a period t value change between initial quasi real-time and full sample estimation? To assess this historical reliability, I consider reference observation 150, which roughly marks the middle of each simulated series, and construct the following cross-sectional statistic: first, I run the filtering procedures through observation 150 to obtain its initial quasi real-time value: $Gap(150, 150)$. Subsequently, I

[7]Comparing the average trend levels that result from the implied endpoint weights of the prescribed procedure: $g_s^* = A^*_{[1xs]}y_{[1xs]'}$ with those from recursive linear detrending: $\hat{\alpha}_{qrt}t$ with $\hat{\alpha}_{qrt} = \frac{\sum_{s=1}^{t} sy_s}{\sum_{s=1}^{t} s^2}$ across simulations, yields the same average linear trend estimate, irrespective of the underlying benchmark process. In single runs, results from both procedures can, however, differ.

run the same procedure on an expanding sample up to $T = 293$ across all N simulations. Through that, I obtain estimates for observation 150 from later vintages runs: $Gap(150, i)$, where $150 < i \leq T$. I then compute the absolute mean difference between $Gap(150, 150)$ and $Gap(150, i)$ to measure how much they deviate. To assess whether this deviation is small or large, I standardize the absolute difference with the typical gap size that is observed for observation 150, measured as the median absolute value across all N simulations.[8] This cross-sectional statistic can be expressed as:

$$Median(HistoricalReliability_{Gap}) = \left(\frac{|Gap(150, i) - Gap(150, 150)|}{|Gap(150, 150)|_{[\frac{N}{2}]}} \right)_{[\frac{N}{2}]}, \text{where } 150 < i \leq T. \quad (3)$$

Standardizing by the typical gap size yields the following interpretation: a value of zero reflects perfect historical reliability. A value of one indicates that historical unreliability is as large as the gap itself. If the statistic exceeds one, historical unreliability is larger than the gap itself.[9]

I want to explicitly distinguish end-of-sample unreliability from historical unreliability associated with the Nelson-Kang critique. To do so, I report statistics for one ($i = 154$), two ($i = 158$), and three years ($i = 162$) after initial quasi real-time estimation alongside statistics from later vintages that are approx. equally apart: $i = 200$, $i = 250$, and for $T = 293$. The first three choices are motivated by the output gap reliability literature and the associated credit-to-GDP gap discussion (see, e.g., Orphanides and van Norden, 2002; Bernhardsen et al., 2005; Cayen and van Norden, 2005; Marcellino and Musso, 2011; Edge and Meisenzahl, 2011; Quast and Wolters, 2022). Therein, revisions during the first two years after initial estimation are typically associated with end-of-sample measurement biases from converging one- and two-sided filtering procedures. Moreover, Baxter and King (1999) discuss choices for HP parameters and their consequences for the resulting filter weights. According to them, the HP filter's endpoint bias vanishes within the first three years after initial estimation.[10] The remaining columns outline the evolution of historical reliability for the observation 150 estimate across later vintages up to $i = T$. If the

[8]Considering the mean instead yields similar results.

[9]For the random walk with drift process, the absolute median gap size amounts to 1.75 and 2.69 for the preferred HP model and recursive linear detrending, respectively. Analogously it is given with 0.58 and 0.62 for the regulator's preferred HP model; and 0.58 and 0.59 for the trend-stationary process.

[10]Similar insights are discussed in the literature on improving the HP filter's real-time reliability, e.g. via forecast-augmentation (see, e.g., Kaiser and Maravall, 1999; Mise et al., 2005; Garratt et al., 2008; Kaiser and Maravall, 2012). Further recall that compared to other bandpass filter such as the Christiano-Fitzgerald bandpass (CF BP) filter, HP filter weights are independent of the available sample so that end-of-sample measurement error is high initially. With additionally incoming observations, it then declines relatively quickly. On the other hand, one-sided CF BP filter weights are chosen optimally based on the available sample, yielding its end-of-sample biases to vanish more slowly with additionally incoming observations (Christiano and Fitzgerald, 2003; Quast and Wolters, 2022).

statistic continues to increase beyond two to three years after initial estimation, revisions are associated with historical instead of endpoint unreliability.

In table 1, I report the evolution of these statistics for detrending the three benchmark DGPs with the Basel III prescripted HP filter configuration as well as recursive linear detrending. I do so to evaluate whether both procedures produce distinct outcomes or whether they yield similar reliability patterns.

Table 1: Historical Reliability Statistics for Known Benchmark DGPs

Simulated Process / Observation Detrending Procedure	$i = 154$	$i = 158$	$i = 162$	$i = 200$	$i = 250$	$i = 293$
DGP: Random Walk						
One-sided HP ($\lambda = 400,000$)	0.208	0.381	0.520	0.837	0.866	0.880
Recursive Linear Detrending	0.104	0.195	0.284	0.860	1.178	1.277
DGP: Implied HP Model						
One-sided HP ($\lambda = 400,000$)	0.109	0.149	0.169	0.217	0.218	0.220
Recursive Linear Detrending	0.069	0.108	0.143	0.410	0.653	0.788
DGP: Trend-Stationary Process						
One-sided HP ($\lambda = 400,000$)	0.108	0.138	0.155	0.163	0.164	0.166
Recursive Linear Detrending	0.053	0.074	0.087	0.136	0.155	0.155

Notes: Historical reliability statistics for detrending the three known benchmark DGPs with the Basel III prescribed detrending procedure and recursive linear detrending, respectively. $i = 154/158/162$ indicate observations one, two, and three years after initial estimation. $i = 293 = T$ marks the sample size. A value of zero reflects perfect historical reliability. A value of one indicates that historical unreliability is as large as the gap itself.

First, it becomes apparent that applying the two detrending procedures to a trend-stationary process would yield highly historically reliable credit-to-GDP gap estimates. While linear detrending is unsurprisingly more efficient in recognizing the linear trend and associated with smaller end-of-sample measurement error (5% to 9% of a typical gap size compared to 11% to 16% for one-sided HP filtering), differences between both procedures remain relatively small. Further, differences between early follow-up vintages that are related to endpoint issues and later vintages that are related with Nelson-Kang considerations remain marginal for Basel III's prescribed detrending procedure.

Second, if credit-to-GDP were to follow the regulator's preferred HP process, the prescribed detrending procedure would, again unsurprisingly, effectively derive reliable credit-to-GDP gaps. Unreliability does not increase much beyond a small endpoint measurement error. Recursive linear detrending, on the other hand, would produce a smaller end-of-sample bias, but would subsequently be associated with increasingly higher historical unreliability.

Third, if credit-to-GDP were to follow a random walk with drift instead, using Basel III's detrending procedure yields historically unreliable gap estimates. This unreliability goes well beyond endpoint biases. Note that endpoint biases are themselves more distinct compared to the two cases discussed before. With continuously increasing historical reliability statistics, revisions become almost as large as the gap itself for the common sample length of $T = 293$ that is studied here. Overall historical unreliability in $T = 293$ exceeds endpoint measurement biases by 69% (0.52% compared to 0.88% of a typical gap size). Again, recursive linear detrending is associated with a smaller endpoint bias, but higher historical unreliability, amounting up to 1.28 times of a typical gap size, which reflects the essence of the Nelson-Kang critique.

While Basel III's detrending procedure does not amount to pure linear detrending, it generates similar (un)reliability patterns as recursive linear detrending, irrespective of the underlying DGP. This implies that it is similarly prone to the Nelson-Kang critique when one applies it to a random walk. Thus, its implementers face similar challenges as when they recursively linearly detrend a random walk.

3 Introducing Historical Reliability Bands for Country-specific Credit-to-GDP Series

Next, I analyze the historical reliability of country-specific credit-to-GDP gap and trend estimates, where the true underlying DGP remains unknown. While doing so, I discuss three symptoms that emerge from applying the prescribed detrending procedure to the credit-to-GDP country series. The resulting gap and trend estimates (i) remain historically unreliable because they (ii) do not converge to full sample estimates, but continue to wander as more data is added. Moreover, (iii) each quarterly trend estimate is associated with changing slopes and intercepts, which compromises the comparability of derived cyclical positions over time.

3.1 Data

The Basel III guideline advises to derive credit-to-GDP gaps from aggregate private sector credit-to-GDP ratios (BCBS, 2010, p.12). These series are obtained from the BIS's "Long series on credit-to-GDP to the non-financial sector" database. I consider their break-adjusted ratio series at market value, expressed in percent. 43 countries are covered in the database.[11] Sample starts vary considerably across countries with the earliest start for the US (1947Q4) and the latest for Colombia (1996Q4). For 21/43 countries, quarterly information is available at least since 1970Q1. 2020Q4 is the last observation considered.

3.2 Symptom I: Historically Unreliable Period t Estimates

To gauge whether country-specific credit-to-GDP gap and trend estimates get revised beyond endpoint revisions, I report and discuss the results in analogy to table 1 for the United States, Belgium, and Brazil in table 2.[12] To go beyond the three specific country cases, I also report simple averages for the respective i across all 43 countries at the bottom of the table. As in the simulation exercise before, observation 150 is considered as the reference observation for the US. For this country case, observation 150 reflects 1985Q1 during the run-up to the 1988 systemic US banking crisis. For all other countries, I consider $T/2$ if T is even. If T is uneven, I round up to the next integer. While some country series are too short to report all historical reliability statistics, I always include the statistic associated with the country-specific sample size T.

Revisions between initial quasi real-time and full sample estimations are larger compared to the results from the simulation evidence. When the Basel III prescribed one-sided HP filter configuration is implemented, this amounts to the a typical gap size for the US, 1.5 times of a typical gap size for Belgium and approx. 2 times of a typical gap for the Brazil. While recursive linear detrending generates comparatively higher historical unreliability for Belgium, Brazil, and most other country cases, it does not do so for the United States. Using linear recursive detrending for the US would produce historically reliable estimates. However, note that this is also a reflection of the spuriousness that characterizes cyclical components in context of the Nelson-Kang critique. As the evidence suggests, the extent to which end-of-sample biases contribute to overall reliability patterns varies across countries. Yet, there are revisions beyond end-point measurement error for the vast majority of countries. This pattern also emerges from the cross-country averages.[13] Similar to the simulation evidence, there is an increase in historical

[11] Basel III is implemented in 100 jurisdictions (Hohl et al., 2018), whereof 73 operate CCyB regimes (Schüler, 2020). Data on long credit-to-GDP series is available for 43 countries: Argentina, Australia, Austria, Belgium, Brazil, Canada, Chile, China, Colombia, Czech Republic, Denmark, Finland, France, Germany, Greece, Hong Kong, Hungary, India, Indonesia, Ireland, Israel, Italy, Japan, Korea, Luxembourg, Malaysia, Mexico, the Netherlands, Norway, New Zealand, Poland, Portugal, Russia, Saudi Arabia, Singapore, South Africa, Spain, Sweden, Switzerland, Thailand, Turkey, the United Kingdom, and the United States.

[12] Now, historical reliability statistics are standardized using a series' median of the absolute gap. This allows me to gauge a typical gap size while circumventing using a standardization that is very close to zero, which would inflate the historical reliability statistic. Results for all other countries are displayed in Appendix A.

[13] Since the underlying data panel is unbalanced, the statistics are not representative at $i = T$, where they reflect US results.

Table 2: Historical Reliability Statistics for the United States, Belgium, and Brazil

Country / Observation Detrending Procedure	$i = 154$	$i = 158$	$i = 162$	$i = 200$	$i = 250$	$i = T$
United States						
One-sided HP ($\lambda = 400,000$)	0.095	0.352	0.700	1.088	0.837	1.023
Recursive Linear Detrending	0.055	0.032	0.057	0.021	0.370	0.111
Belgium						
One-sided HP ($\lambda = 400,000$)	0.101	0.266	0.437	1.369	1.452	1.452
Recursive Linear Detrending	0.107	0.265	0.438	2.183	3.241	3.241
Brazil						
One-sided HP ($\lambda = 400,000$)	0.405	0.733	1.004	2.152	–	2.152
Recursive Linear Detrending	0.671	1.226	1.701	4.449	–	4.449
Cross-Country Average						
One-sided HP ($\lambda = 400,000$)	0.381	0.737	0.980	1.177	1.052	1.023
Recursive Linear Detrending	0.312	0.598	0.792	1.575	2.196	0.111

Notes: $i = 154/158/162$ indicate observations one, two, and three years after initial estimation. $i = T$ marks the respective sample size. A value of zero reflects perfect historical reliability. A value of one indicates that historical unreliability is as large as the gap itself.

unreliability amounting to a typical gap size when the Basel III detrending procedure is implemented on the country series. Average historical unreliability is larger when recursive linear detrending is used.

Compared to the simulation evidence where revisions associated with the Nelson-Kang critique and historical unreliability increase monotonically up to T, country-specific patterns can be non-monotonic. For example, there is a decrease from $i = 200$ to $i = 250$ for the United States, whereas the statistic subsequently increases again. Thus, with initially increasing historical unreliability, the observation 150 estimate seems to revert back closer to its initial quasi real-time value, before either surpassing that initial estimate in the opposite direction or tending outward again. The country series, thus, seem to additionally exhibit excessive movement. Between $i = 200$ and $i = 250$ there is also a little excess movement when the cross-country average is considered. This symptom will be illustrated in more detail in the next section.

Before turning to that, I illustrate the range with which credit-to-GDP gap and trend estimates vary between their initial quasi-real time and full sample estimation. To do so, I compute *historical reliability bands* that show the largest and smallest value obtained for gap_t or $trend_t$ across all historical vintages. These bands are included as gray shades in figure 1 and 2. Dashed and solid lines show the quasi-real time and full sample estimates. Gray bars indicate systemic banking crises based on the updated classification provided in Laeven and Valencia (2018). Dash-dotted lines for 2pp and 10pp highlight the range when national authorities should consider activating CCyB add-ons (BCBS, 2010). Gap estimates for the United States, Belgium, and Brazil are shown in figure 1, while the associated trend estimates are shown in figure 2. The trend plots also contain credit-to-GDP level series as dotted light gray lines.[14]

The range with which Basel III credit-to-GDP gap and trend estimates vary is distinct. At first glance, the indicator seems to work fine for the United States, signaling buffer building prior to US systemic banking crises. However, recall that it was partly designed for that purpose (Borio and Drehmann, 2009), which makes the relatively good US fit less surprising. This impression changes if other economies such as Belgium or Brazil are considered. Here, quasi real-time assessment would have suggested extensive need for buffer building for both (Belgium: mid-1990s-2005) and Brazil (2008-2017/18). Yet, later assessments would have signaled distinct underutilization of economy-wide leverage instead. Compared to these, the initial signals, thus, turned out as distinct and persistent false-positive signals. Considering cross-country averages hits a similar vein. I show results since 1960 so that at least three countries are included in the averages.[15] Gap signals are, on average across countries, not robust in terms of sign and size since the historical reliability bands tend to span a wide range from negative to positive values that also affect the potential activation and extent of CCyB add-ons. Only during the financial crisis, zero is not included. Importantly, the bands to not get tighter over time, but stay wide instead. For periods currently associated with the current end they will widen again as more data is added.

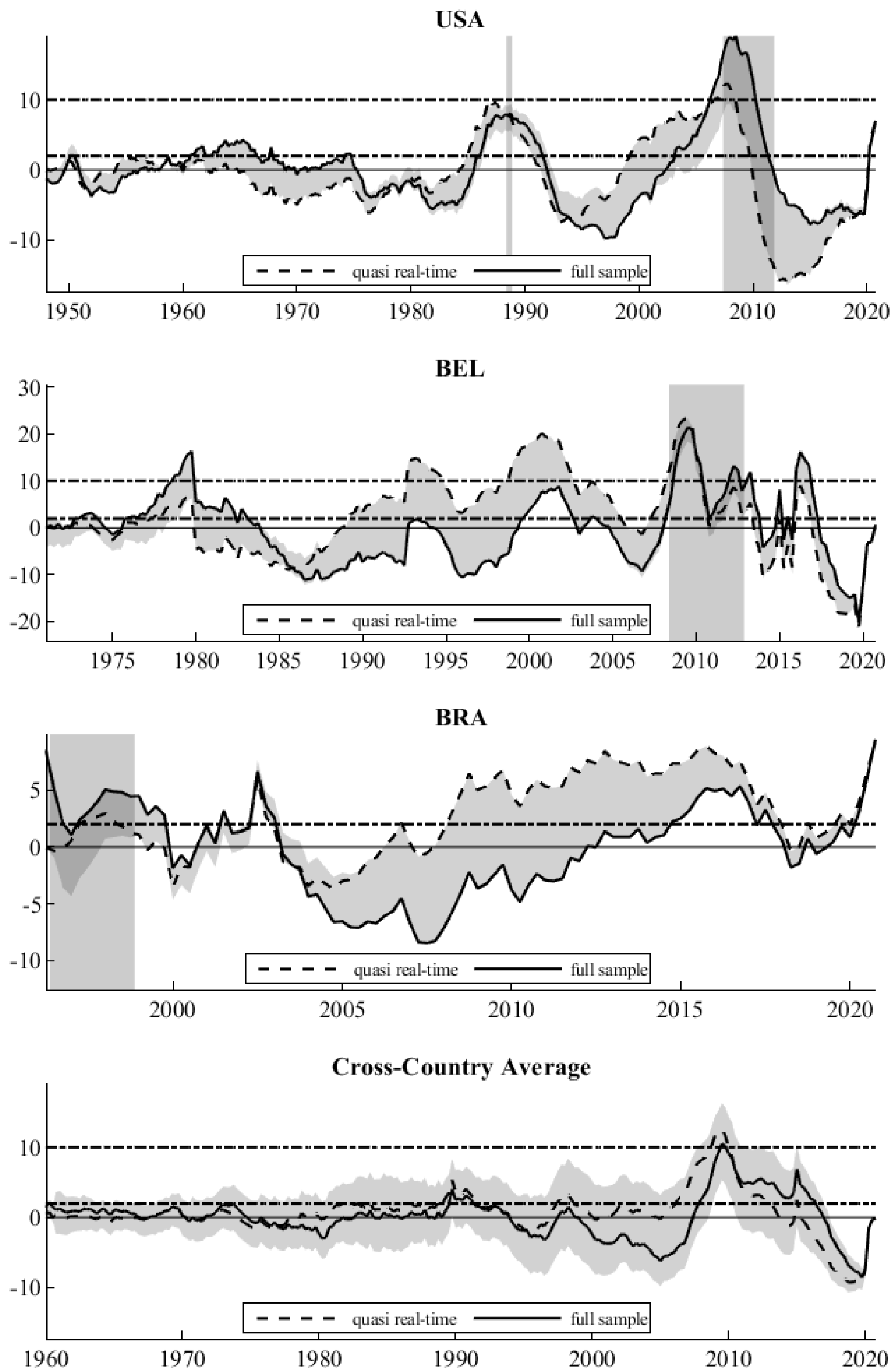
Edge and Meisenzahl (2011) discuss possibly costly false-positive signals for the United States stemming from end-of-sample biases. Their evidence suggests that costs in terms of capital shortfalls and reductions in lending can be substantial, amounting to as much as 141 billion USD following a false-positive signal in 2001Q4, which would have slowed the economic recovery after the burst of the dot.com bubble. The analysis here emphasizes that costs might be even larger than projected for endpoint biases since historical unreliability exceeds endpoint issues for Basel III credit-to-GDP gaps.

This can become costly for society when it undermines the regulator's credibility. Initial false-positive signals that suggest tightening and buffer building, but are corrected downwards in subsequent quarters are hard to maintain and communicate. As noted in Edge and Meisenzahl (2011), banks are given one year to accumulate additional capital to meet the new CCyB requirements. However, as credit-to-

[14]Plots for all other countries are displayed in Appendix B.

[15]All other countries enter successively as the sample extends and their results become available.

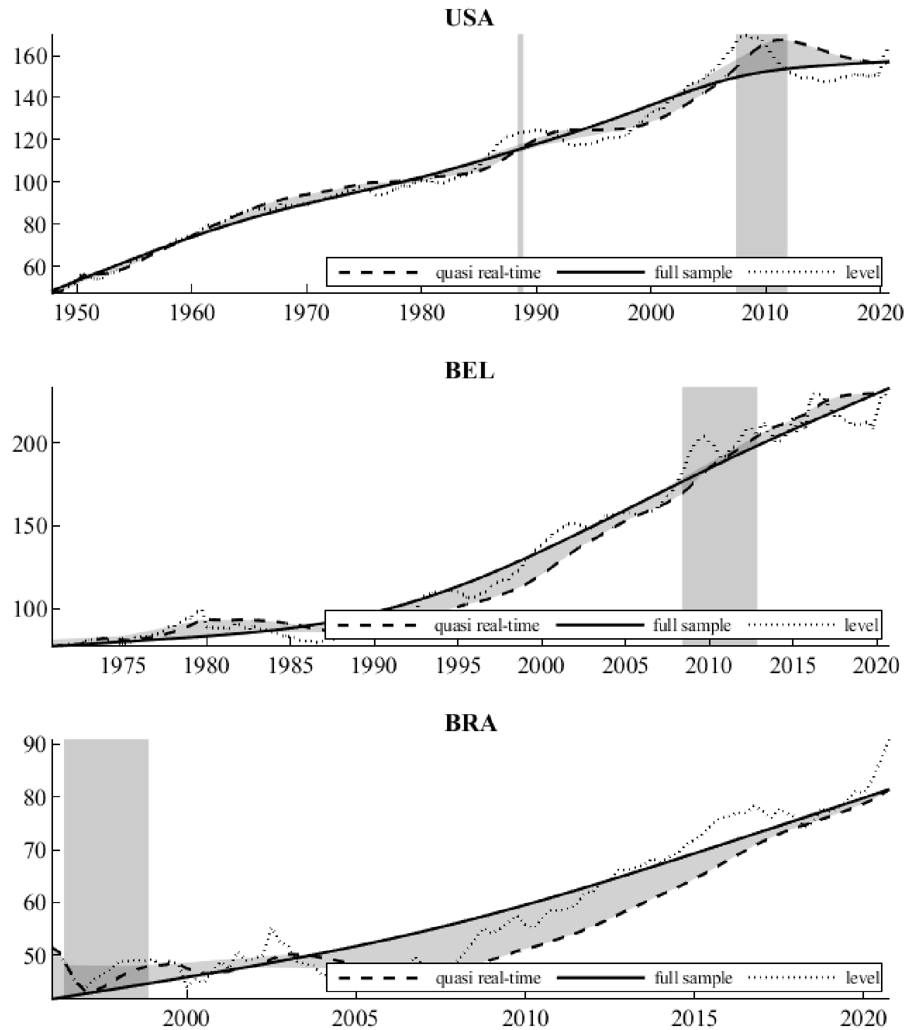
Figure 1: Historical Reliability Bands for Selected Countries and Cross-Country Average: Gap Estimates



Notes: Gap estimates in terms of percentage deviations. Dashed lines show the quasi real-time and solid lines the full sample estimates. Gray shades indicate historical reliability bands, i.e. the smallest and largest estimated value for period *t* across all vintages. Dash-dotted lines for 2pp and 10pp indicate the range when national authorities should consider activating CCyB add-ons (BCBS, 2010). Gray bars indicate systemic banking crises (Laeven and Valencia, 2018).

GDP gap estimates continue to be revised, the initial tightening may be questioned. This weakens the regulator's ability to effectively run CCyB regimes.

Figure 2: Historical Reliability Bands for Selected Countries: Trend Estimates



Notes: *Level* refers to the country-specific credit-to-GDP series. Dashed lines shows the quasi real-time and solid lines full sample estimates. Gray shades indicate historical reliability bands, i.e. the smallest and largest estimated value for period t across all vintages. Gray bars indicate systemic banking crises (Laeven and Valencia, 2018).

Figure 2 illustrates the historical unreliability around the trend estimates for the three country cases. Especially, for Belgium and Brazil it becomes apparent that the quasi-real time trend is often distinctly misaligned with the level series. Moreover, for no country does the trend build consecutively. This is problematic, because it compromises the comparison of cyclical positions over time. This symptom will be discussed in more detail below.

3.3 Symptom II: On The Illusion of Convergence

This section investigates the excessive movement in the credit-to-GDP gap estimates observed above in more detail. Convergence to full sample estimates remains elusive even though the range depicted in figure 1 is predominately spanned by the quasi real-time and the full sample estimates.

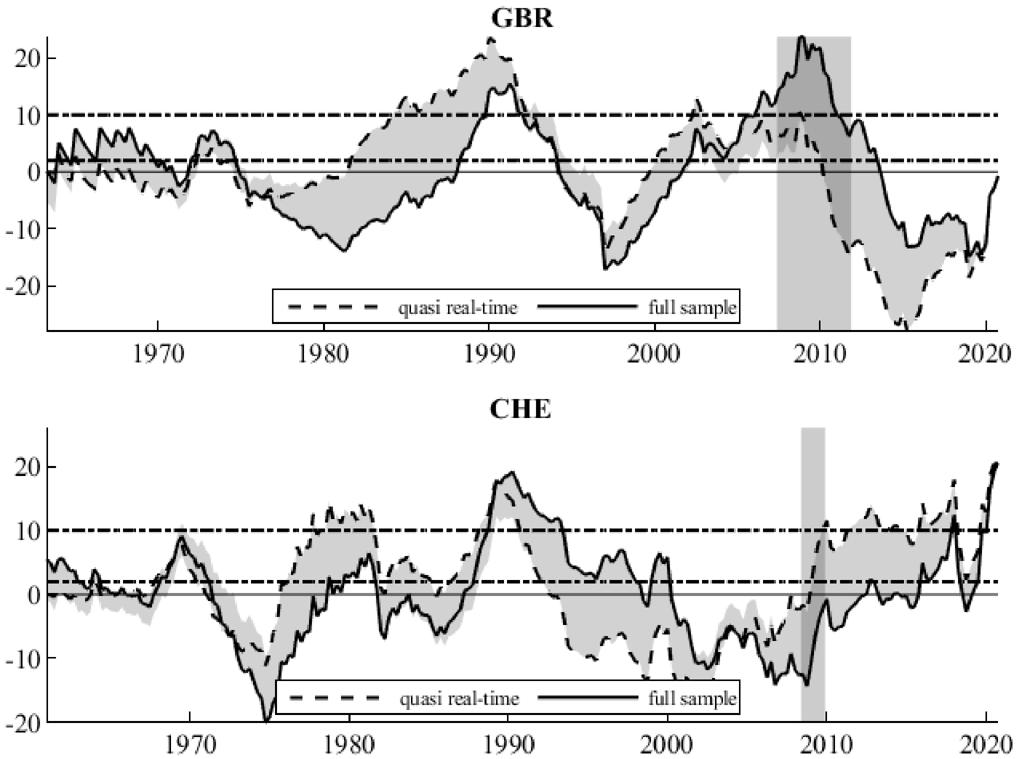
Figure 3 shows credit-to-GDP gap estimates for the United Kingdom (top panel) and Switzerland (bottom panel). For both countries there are episodes where the quasi real-time and full estimates are almost identical. Consider, for example, the 1990s for the United Kingdom or between 1965-1975 for Switzerland. Just considering these two estimates would suggest high real-time reliability, because both estimates align almost perfectly. Considering, however, estimates from all vintages in between, changes the inference on historical reliability. During these episodes, the range of period t estimates spans beyond the quasi real-time and full samples estimates. This implies that the estimated gap values first moved away from the initial estimate, before reverting afterwards. The initial real-time reliable impression is, thus, not robust if the cross-section of all later period t estimates is considered.

To illustrate further, consider the three cross-sectional panels shown in figure 4. Here, I show the cross-sectional evolution of specific period t estimates from their initial quasi real-time to full sample estimates and all vintages in between. The left panel shows the US credit-to-GDP gap estimate for 2006Q4, on the onset of the GFC. The middle and right panel show the credit-to-GDP gap estimate evolution for the UK 1995Q4 and the Swiss 1995Q4 value. I choose 1995Q4 for the latter two, because it roughly lays in the middle of the sample where the HP filter is supposed to work well. The choice, further, allows me to evaluate the estimate's evolution for 25 years.

The US estimate is revised downwards during the first three years after initial estimation. Then, it reverts over time, even past its initial value. The estimate for the 2006Q4 US credit-to-GDP gap, thus, ranges from about 6pp to 12pp, which amounts to a doubling in absolute gap size. This is well in line with the simulation evidence when the prescribed procedure is applied to a random walk. Beyond the pure magnitude of unreliability, this example also illustrates the excessive movement that can occur between initial and full sample assessments. The estimate does not settle for about 10 years and remains historically unreliable, well beyond endpoint issues.

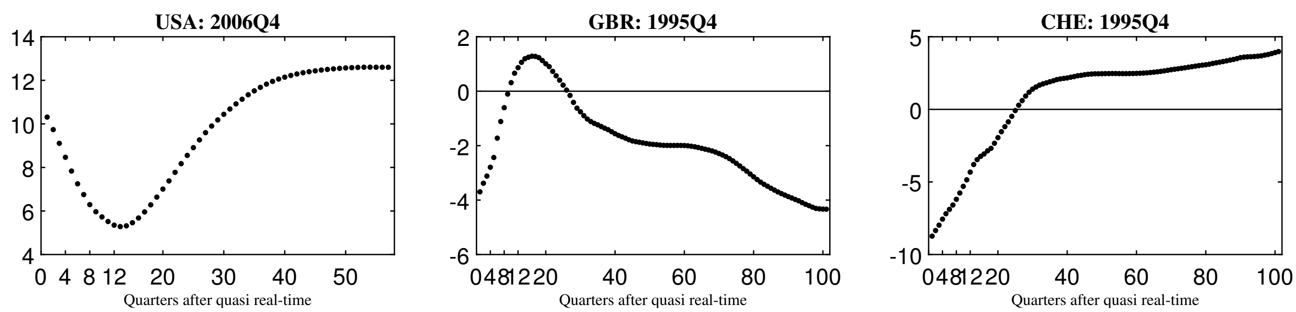
The other two examples reveal that even such late estimate settling is not guaranteed. The UK 1995Q4 estimate starts at about -4pp, surpasses 0pp up to about 1pp before it continuously moves back toward

Figure 3: Illusion of Convergence



Notes: Gap estimates in terms of percentage deviations. Dashed lines show the quasi real-time and solid lines the full sample estimates. Gray shades indicate historical reliability bands, i.e. the smallest and largest estimated value for period t across all vintages. Dash-dotted lines for 2pp and 10pp indicate the range when national authorities should consider activating CCyB add-ons (BCBS, 2010). Gray bars indicate systemic banking crises (Laeven and Valencia, 2018).

Figure 4: Illustrating Lacking Convergence to Full Sample Estimates: Selected Examples



Notes: The dots show the cross-sectional evolution of period t estimates for three selected examples.

-4pp over two decades. This case provides an example where just observing the quasi real-time and full sample estimate would give the impression of a very reliable UK 1995Q4 estimate. However, the cross-sectional evolution shows that this is not the case. The estimate remains historically unreliable instead. Lastly, the Swiss 1995Q4 credit-to-GDP gap estimate has been continuously revised upwards for 25 years.

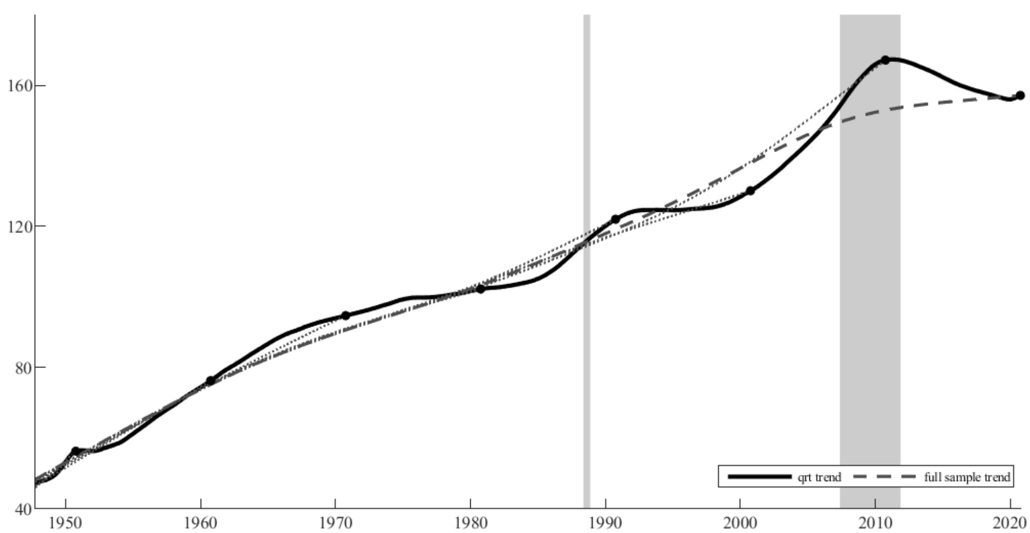
Similar historical unreliability patterns can be observed for the associated trend estimates.

3.4 Symptom III: A New Trend Estimate for Each Quarter

The regulator is interested in credit-to-GDP deviations from "*its long-term trend*" (see, e.g., BCBS, 2010, p. 9) in order to evaluate whether credit-to-GDP is excessively high compared to historical episodes. However, figure 5 suggests that this is rendered unattainable if credit-to-GDP is detrended as prescribed. One does not obtain ***one*** historic trend to derive cyclical positions from. Instead, the whole trend history is continuously reevaluated as the sample expands.

Figure 5 shows recursive US trend estimates that are obtained from the prescribed HP filter configuration. The solid black line represents the quasi real-time estimate and the dotted lines represent selected underlying recursive trend estimates, where endpoints are marked as filled circles. The dashed gray line shows the most currently available full sample trend estimate.

Figure 5: Recursive Trend Estimates for the United States



Notes: Recursive US credit-to-GDP trend estimates as extracted from the prescribed procedure. The solid black line represents the quasi real-time estimate and the dotted lines represent selected underlying recursive trend estimates, where endpoints are marked as filled circles. Gray bars indicate systemic banking crises (Laeven and Valencia, 2018).

It becomes apparent that with each new quarter, the trend estimate changes in slope and intercept. Consider, for example, trend values for 2000Q4. The quasi real-time estimate is marked as a filled circle. Estimates for 2000Q4 from later vintages indicate distinctly higher trend levels. Further illustrations are provided in figure 2 and Appendix B, which show the range of historical trend unreliability for all other countries. While the quasi real-time endpoint series indicates some degree of curvature, note that the underlying recursive trend estimates closely resemble linear lines.

The trend base changes with each new observation instead of building consecutively on an expanding sample. This compromises the comparability of cyclical positions that are derived as trend deviations over time. For example, a 2pp deviation relative to one underlying trend differs from a 2pp deviation associated with a trend estimate from another period. Crucially, *the* historic long-term trend emphasized in the guideline (see, e.g., BCBS, 2010, p. 9) is not extracted. The trend keeps on changing because the prescribed HP filter configuration does not sufficiently recognize the stochastic trend in credit-to-GDP. Instead, it is set to extract a deterministic trend component that is not characteristic for credit-to-GDP.

3.5 Two Complementary Perspectives

To complement the analysis, I reconcile my findings with evidence from formal unit root testing (Schüler, 2018; Hamilton and Leff, 2020; Karagedikli and Rummel, 2020) and assess whether fluctuations in credit-to-GDP are rather driven by transitory or persistent shocks in spirit of Aguiar and Gopinath (2007).

Formal Unit Root Testing In this paper, I gain insights on the nature of the trend in credit-to-GDP by detrending stylized known DGPs and country series as prescribed in Basel III and comparing their reliability patterns. Traditional, formal unit root testing provides a natural complementary perspective, but is less agnostic. I also discuss it in a supplementary manner because testing power is limited, e.g. in light of local alternatives or trend breaks (Schüler, 2018).

Augmented-Dicky-Fuller Tests: First, I conduct augmented ADF tests (Dickey and Fuller, 1979). To test for a unit root, the null hypothesis $H_0 : \rho = 1$ is evaluated in the following regression:[16]

$$y_t = \alpha + \rho y_{t-1} + \delta t + \sum_{i=1}^{p} \gamma_i \Delta y_{t-i} + \varepsilon_t. \tag{4}$$

Kwiatkowski–Phillips–Schmidt–Shin (KPSS) Tests: The testing strategy is complemented by KPSS tests to check whether trend-stationarity would provide a more appropriate trend characterization.[17]

[16]Evaluation is based on the OLS t-statistics. I follow Hamilton (1994) and Hamilton and Leff (2020) and use $p = 4$.

[17]The null hypothesis for the KPSS test is for the series to be stationary around a deterministic time trend.

Detailed results for all country series are provided in table 9, Appendix C. Large negative ADF test statistics indicate evidence to reject the null hypothesis of a unit root, which would be rejected for only 4 of 43 countries at common significance levels. There is, thus, little evidence to rule out unit root characterizations for the vast majority of countries. This is supported by the KPSS test results, which indicate that trend-stationary would only be preferred for 2 of 43 countries.[18] Compared to other financial series such as house, bond, or equity prices, unit root characterizations are found hardest to be rejected for credit-to-GDP (Schüler, 2018). These results, thus, support stochastic credit-to-GDP trend characterizations. As discussed before, applying the Basel III prescribed detrending procedure to series that are characterized by unit roots rather than linear deterministic trends yields historically unreliable estimates.

Fluctuations in Credit-to-GDP: Transitory versus Persistent Shocks Stochastic trend characterizations would imply that quarterly fluctuations in credit-to-GDP are primarily driven by permanent rather than transitory shocks. In the following, I explore empirical credit-to-GDP cycle regularities in spirit of Aguiar and Gopinath (2007) who find that business cycle volatility in emerging markets stems primarily from shocks to trend growth.

Table 3 summarizes key statistics on credit-to-GDP growth, averaged for the three following country groups: G7 economies, advanced small open economies, and emerging markets.[19] The coefficients for early lags of the sample autocorrelation function remain positive for several lags. This indicates that shocks to trend growth that persist beyond business cycle horizons mainly drive quarterly credit-to-GDP fluctuations. In contrast, negative ρ's would indicate that fluctuations are mainly driven by less persistent shocks (Aguiar and Gopinath, 2007).[20]

[18]Note that these updated results are in line with Schüler (2018) and Hamilton and Leff (2020).

[19]Emerging markets are classified as in the IMF's country classification: Argentina, Brazil, Chile, China, Colombia, Hungary, India, Indonesia, Malaysia, Mexico, Poland, Russia, Saudi Arabia, South Africa, Thailand, and Turkey. All remaining economies that do not fall within the G7 group are subsumed as advanced small open economies.

[20]Patterns are largely similar across country groups, which does not necessarily suggest that they face the same combination of persistent and transitory shocks (Aguiar and Gopinath, 2007). Estimating shock histories requires clear shock identification via structural analysis and is beyond the scope of this paper. Table 3 also reveals that advanced small economies are currently, on average, most leveraged and characterized by the highest average credit-to-GDP growth rates. Emerging markets are considerably less leveraged while they are exposed to similarly high volatility. With ongoing financial deepening in advanced small open economies, aggregate financial market exposure is less distinct in emerging markets, which, however, does not guard them from relatively high volatility.

Table 3: Characterizing Credit-to-GDP Growth (ΔY)

	G7 Economies	Adv. Small Open Economies	Emerging Markets
Number of Economies	7	20	16
$\rho_1(\Delta Y_t, \Delta Y_{t-1})$	0.22	0.25	0.24
$\rho_2(\Delta Y_t, \Delta Y_{t-2})$	0.28	0.26	0.16
$\rho_3(\Delta Y_t, \Delta Y_{t-3})$	0.08	0.13	0.07
$\rho_4(\Delta Y_t, \Delta Y_{t-4})$	0.36	0.20	0.14
$\rho_5(\Delta Y_t, \Delta Y_{t-5})$	0.01	0.06	0.06
$\rho_6(\Delta Y_t, \Delta Y_{t-6})$	0.13	0.10	0.02
$\rho_7(\Delta Y_t, \Delta Y_{t-7})$	-0.02	0.04	0.00
$\rho_8(\Delta Y_t, \Delta Y_{t-8})$	0.27	0.15	0.07
$\rho_9(\Delta Y_t, \Delta Y_{t-9})$	-0.03	0.00	-0.06
$\rho_{10}(\Delta Y_t, \Delta Y_{t-10})$	0.12	0.04	-0.05
$\mu(\Delta Y)$	0.44	0.74	0.40
$\sigma(\Delta Y)$	1.63	3.14	2.58
Mean Leverage Level in 2020Q4	178.9	216.3	97.3

Notes: Sample autocorrelation functions: average reported across G7 economies, advanced small open economies, and emerging markets. μ and σ represent respective means and standard variations across groups. Additionally, the leverage level, averaged across groups, as of 2020Q4 is reported.

4 Proposing an Alternative

This section proposes an alternative that circumvents the conceptual econometric concerns associated with the prescribed HP filter configuration discussed so far. Two considerations guide the suggestion: first, I take as given that there is merit in deriving credit-to-GDP gaps from simple univariate detrending procedures to obtain transparently derived and internationally comparable credit-to-GDP gaps to base CCyB decisions on.

Second, I assume that the Basel Committee's Macro Variables Task Force has deemed gap dynmaics as obtained from their prescribed detrending procedure important and policy relevant from the regulator's perspective. While the guideline notes that *"The guide does not always work well in all jurisdictions at all times."* (BCBS, 2010, p. 3), Borio and Drehmann (2009) pay explicit attention to US credit-to-GDP gap dynamics. Based on that, I propose an alternative that aligns with their preferred US credit-to-GDP gap dynamics and amplitudes. I am not proposing an alternative that necessarily optimizes the indicator's predictive ability for financial crises. Such considerations are, for example, discussed in Beltran et al. (2021).

With the empirical evidence suggesting stochastic rather than deterministic credit-to-GDP trend characterizations, difference filtering provides a natural simple detrending alternative:

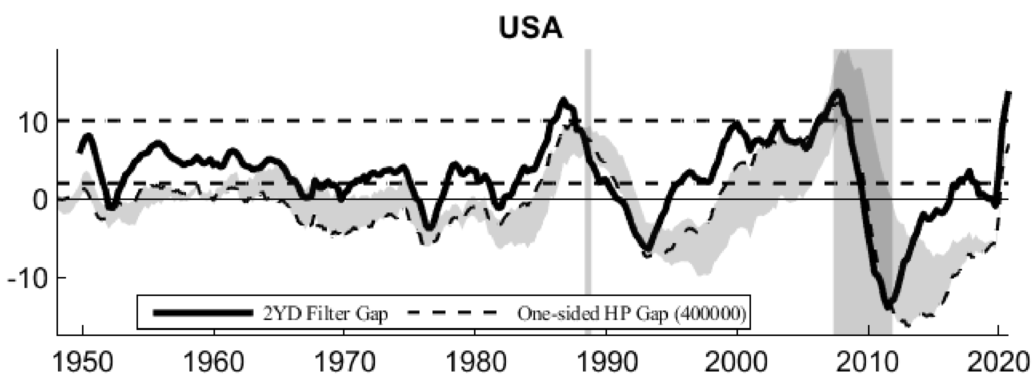
$$\varepsilon_{t+h} = y_{t+h} - y_t. \tag{5}$$

The cyclical component is derived as the difference between y_{t+h} and the actual realization, y_t, and reflects how much the series changes over horizon h. Any change over horizon h is entirely attributed to cyclical factors and does not affect the trend estimate. As in Hamilton's (2018) proposed business cycle argument, a two-year difference (2YD) filtered credit-to-GDP gap then captures unexpected changes such as the exact timing of important turning points over a two year horizon.

Hamilton (2018) proposes to use $h = 20$ to analyze debt cycles. I, nevertheless, stick to a 2YD filter, i.e. $h = 8$, and his business cycle argument because choosing a high h in small samples is problematic since *"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."* (Hamilton, 2018, p. 838). Further, the assumption that all unexpected changes between t and $t + h$ only pass through to the cyclical component becomes harder to maintain as h increases. Note also that the residual series becomes increasingly autocorrelated for larger h, so that the gap estimate becomes increasingly hard to distinguish from a unit root process.

Figure 6 compares US credit-to-GDP gap estimates from the prescribed HP filter configuration (dashed line and historical reliability bands as shaded area) with those from the 2YD filter (solid black line). Historical reliability as measured in this paper is perfect if the 2YD filter is applied, because there are no subsequent revisions to the initial quasi real-time estimates.

Figure 6: US Credit-to-GDP Gap Estimates: One-sided HP and 2YD Filter



Notes: Gap estimates in terms of percentage deviations. The dashed line shows the one-sided HP filtered gap estimate, the solid line the two-year difference filtered gap estimate. Gray shades indicate historical reliability bands, i.e. the smallest and largest estimated value for period t across all vintages. Dash-dotted lines for 2pp and 10pp indicate the range when national authorities should consider activating CCyB add-ons (BCBS, 2010). Gray bars indicate systemic banking crises (Laeven and Valencia, 2018).

Gap amplitudes from a 2YD filter are very similar to those obtained from the prescribed procedure.[21] Note further that both US systemic banking crises would have been signaled in advance. While overall dynamics are similar[22], an important difference remains. In the aftermath of the GFC, the 2YD filter indicates much quicker reversion to trend, whereas the Basel III credit-to-GDP gap indicates distinct and persistent underutilization of US economy-wide leverage.

Gap and trend estimates from a 2YD filter are one-sided per construction. From the regulator's perspective, one potential challenge might be that the resulting credit-to-GDP gap estimate is not centered around zero. Operating buffer add-ons as prescribed in the Basel III presumes a zero-centered gap. I, therefore, additionally consider a recursively demeaned 2YD filter gap.[23] Appendix D illustrates. In what follows, I check if the suggested alternative circumvents the econometric issues associated with the prescribed procedure.

4.1 Difference Filtering in Light of the Simulation Exercise

Historical reliability as measured and discussed here is perfect for pure difference filtering approaches, irrespective of the choice for h. This is because the initial quasi real-time estimates do not get revised subsequently. Zero-centering difference filtered gaps can introduce unreliability due to the estimation uncertainty that surrounds the recursively estimated mean. Table 4 reports results for the two difference filtering approaches, I consider here.

For all three benchmark processes, the recursively demeaned 2YD filtered gap measure is distinctly more historically reliable compared to the Basel III prescribed filtering procedure discussed above. Endpoint measurement issues and Nelson-Kang considerations are practically absent.

4.2 On Curing the Symptoms

Next, I check whether or to what extent the proposed alternative circumvents the three symptoms discussed earlier: (i) historical unreliability beyond endpoint biases, (ii) lacking convergence to full sample

[21] For large h, credit-to-GDP gap amplitudes increase, whereas they decrease for smaller h.

[22] The cross-country average correlation between both quasi real-time gap series is 0.84.

[23] Alternatively, equation 5 could be extended by a constant that is estimated recursively. The results are equivalent and numerically very similar. One could also consider implementing Hamilton's (2018) proposed regression filter. This would be associated with introducing some estimation uncertainty around the intercept and the four AR-coefficients. Given short samples for some countries and the evidence for unit root characterizations, the regression-free implementation seems to be well suited for the credit-to-GDP context.

Table 4: Historical Reliability Statistics: 2YD Filter

Simulated Process / Observation Detrending Procedure	$i = 154$	$i = 158$	$i = 162$	$i = 200$	$i = 250$	$i = 293$
Random Walk						
2YD Filter	0	0	0	0	0	0
2YD Filter: Recursively Demeaned	0.025	0.046	0.059	0.120	0.152	0.168
Implied HP Model						
2YD Filter:	0	0	0	0	0	0
2YD Filter: Recursively Demeaned	0.014	0.020	0.020	0.027	0.039	0.049
Trend-Stationary Process						
2YD Filter:	0	0	0	0	0	0
2YD Filter: Recursively Demeaned	0.014	0.020	0.020	0.018	0.017	0.017

Notes: Historical reliability statistics for detrending the three known benchmark DGPs two-year difference filters. $i = 154/158/162$ indicate observations one, two, and three years after initial estimation. $i = 293 = T$ marks the sample size. A value of zero reflects perfect historical reliability. A value of one indicates that historical unreliability is as large as the gap itself.

estimates, and (iii) new trend estimates as the sample expands. Table 5 reports the historical reliability statistic for the United States, Belgium, Brazil, and the cross-country average and reveals that historical reliability is very high when the recursively demeaned 2YD filter is used.

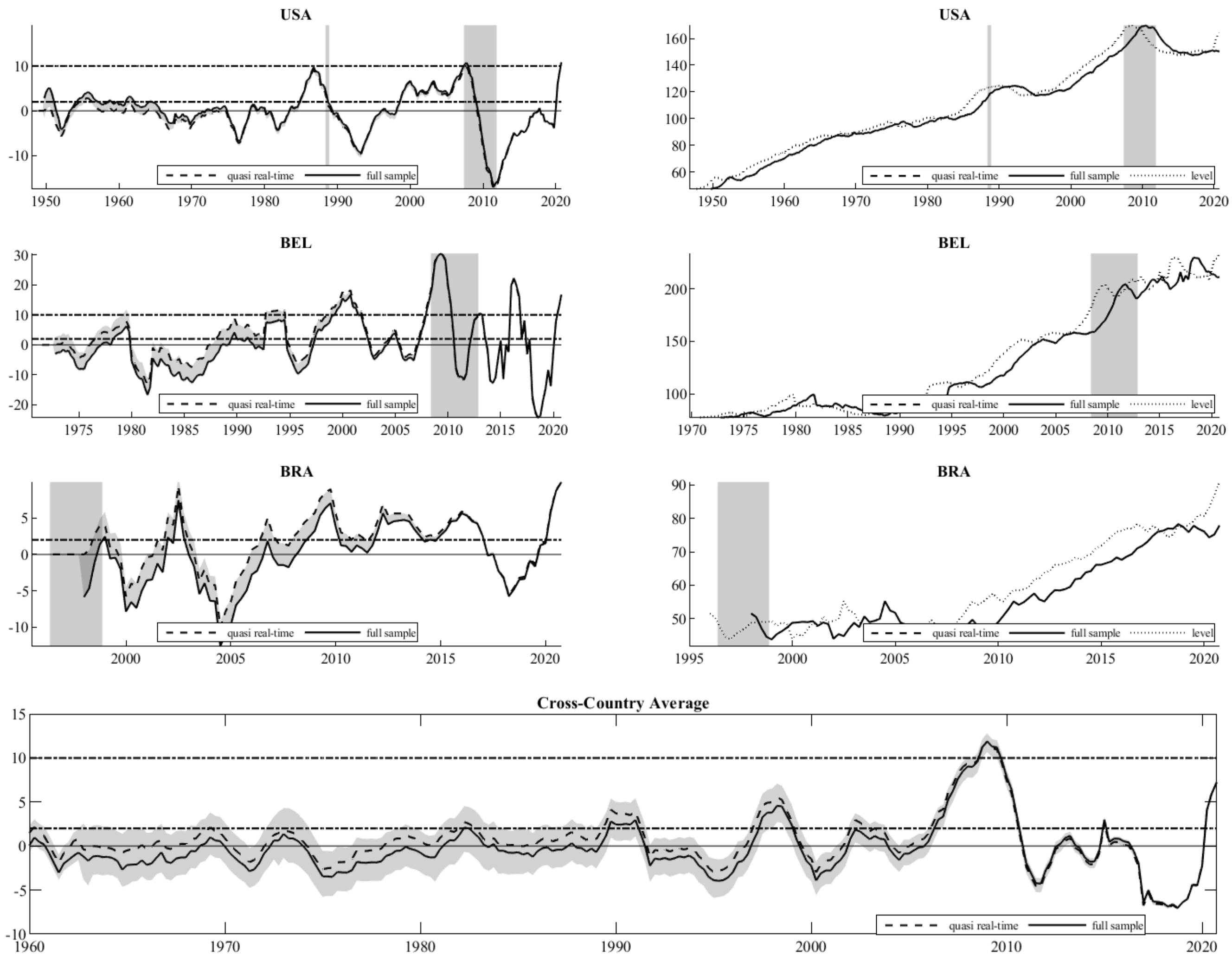
Table 5: Historical Reliability Statistics for the United States, Belgium, and Brazil: 2YD Filter

Country / Observation Detrending Procedure	$i = 154$	$i = 158$	$i = 162$	$i = 200$	$i = 250$	$i = 293$
United States						
2YD Filter: Recursively Demeaned	0.001	0.002	0.005	0.001	0.009	0.001
Belgium						
2YD Filter: Recursively Demeaned	0.002	0.000	0.003	0.021	0.029	0.029
Brazil						
2YD Filter: Recursively Demeaned	0.006	0.012	0.014	0.032	–	0.032
Cross-Country Average						
2YD Filter: Recursively Demeaned	0.007	0.014	0.018	0.030	0.020	0.001

Notes: $i = 154/158/162$ indicate observations one, two, and three years after initial estimation. $i = T$ marks the sample size. A value of zero reflects perfect historical reliability. A value of one indicates that historical unreliability is as large as the gap itself.

Figure 7 illustrates. The range between quasi-real time and full sample estimates remains very small compared to the distinct unreliability bounds associated with the prescribed procedure shown in figures 1 and 2. As expected, mean estimation uncertainty is limited to the earlier part of the respective samples. As the estimates converge to full sample estimates, unreliability vanishes. Importantly, even during the slight early unreliability, there are almost no sign changes, which indicates that the initial assessment of whether the gap is positive or negative sustains. These patterns are also reflected in the cross-country gap averages. Overall, unreliability is smaller compared to the official procedure and limited to the first part of the sample. Importantly, buffer add-on decisions are affected distinctly less often. And if they are, there would be almost no decision misalignment. Showing the cross-country average for the gap estimates, thus, illustrates that there is improvement that is not limited to specific country cases, but is relevant for all countries. Trend estimates are not associated with any historical unreliability in context of difference filtering.

Figure 7: Historical Reliability Bands for Selected Countries: 2YD Filter (Recursively Demeaned)

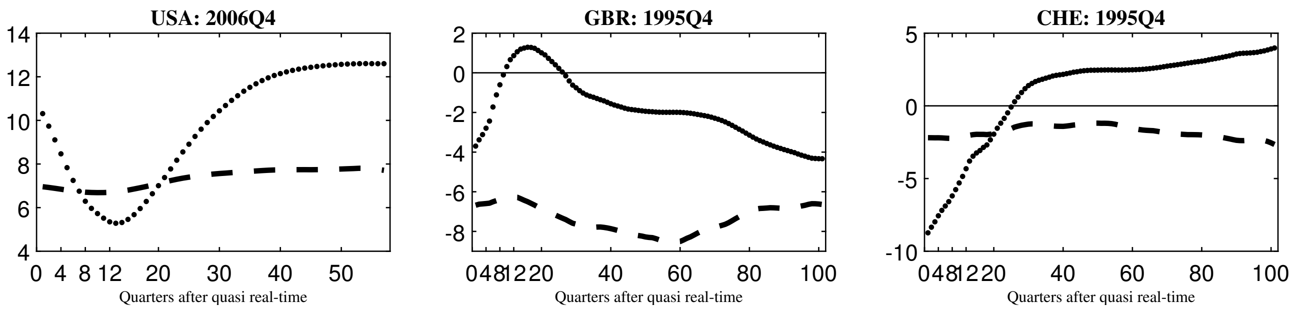


Notes: Gap estimates in terms of percentage deviations on the left, country-specific trend estimates and level series on the right. The dashed line shows the quasi real-time, the solid the full sample two-year difference filtered gap estimate. The bottom panel shows the cross-country average gap estimates. Gray shades indicate historical reliability bands, i.e. the smallest and largest estimated value for period t across all vintages. Dash-dotted lines for 2pp and 10pp indicate the range when national authorities should consider activating CCyB add-ons (BCBS, 2010). Gray bars indicate systemic banking crises (Laeven and Valencia, 2018).

Compared to before, episodes signaling the need or no need for buffer building in real time alternate between 1990 and 2015 for Belgium. Importantly, shortly before the onset of the GFC, the need for buffer building would have been indicated. Similarly for Brazil, signals alternate between 2008 and 2017/18 instead of signaling continuous need for buffer building. Most notably, the ex post signals for distinct underutilization do not distort the communication of initial buffer decisions.

Figure 8 extends figure 3 by including results from the recursively demeaned 2YD filter (dashed lines) to contrast the estimates' cross-sectional evolution. Compared to before, estimates from the recursively demeaned 2YD filter do converge to full sample estimates. Here, estimates change very little and remain highly historically reliable. Moreover, signs are maintained so that cyclical positions are not subject to continuous economic reinterpretation.

Figure 8: Illustrating Convergence to Full Sample Estimates: 2YD Filter



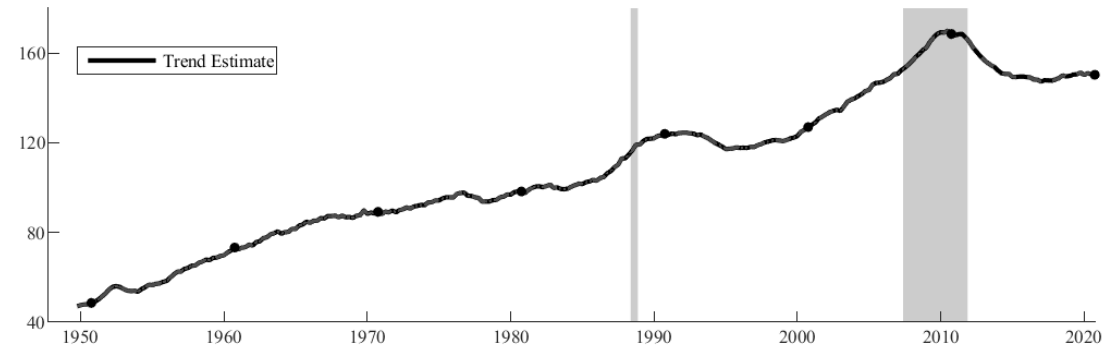
Notes: The dots show the cross-sectional evolution of period t estimates for three selected examples as derived from the Basel III prescribed detrending procedure, the dashed line as derived from the recursively demeaned 2YD filter, respectively.

Figure 9 shows the historic long-term trend as derived from a 2YD filter, which builds consecutively on an expanding sample. Comparability across cyclical positions, thus, remains intact.

5 How are CCyB Decisions Affected? Is it About the Sign or the Size?

It is about both. To illustrate, I ask the following four questions: (i) how often does the sign switch for an initially estimated period t value compared to all later estimates. This gives a sense of how often the cyclical position is economically reinterpreted. (ii) How do later estimates differ in terms of size? (iii) How often would have initial buffer add-on (de)activation decisions been reevaluated? And (iv) by how much would the buffer add-on have been reevaluated in terms of % of risky weighted assets? I account for all 43 countries and report respective averages for the first three questions in table 6 and 7. Following

Figure 9: Recursive Trend Estimates for the US: 2YD Filter



Notes: Recursive US credit-to-GDP trend estimates as extracted from a two-year difference filter. Filled circles represent selected endpoint estimates. Gray bars indicate systemic banking crises (Laeven and Valencia, 2018).

this, I will then discuss question (iv). To illustrate the variability around the cross-country mean, I include the smallest and largest value that is obtained across countries in brackets below.

To start with, I look at a direct comparison between Basel III's one-sided HP filter configuration and the recursively demeaned 2YD. Results for the first three questions are summarized in table 6. First, different signs and therewith distinct economic interpretations of the cyclical position are indicated 17% of the time. Second, I compute the pairwise absolute differences between the two series and assort them to the following four brackets: small differences [0pp-2pp], medium differences [2pp-4pp] as well as larger [4pp-6pp] and very large differences [>6pp].[24] While absolute size differences tend to be small (0pp-2pp: 39% of the time), for more than 1/4 of all periods they become rather large ($> 6pp$). Third, I determine whether the initial HP filter based assessment was too tight compared the 2YD filter, i.e. when did the official procedure assessed need for buffer activation while the alternative would have suggested no add-on activation (HP assessment: $> 2pp$, 2YD filter assessment: $< 2pp$). I proceed analogously for too loose add-on decisions. Compared to signals derived from a recursively demeaned 2YD filter, initial assessment from the prescribed credit-to-GDP gap measures were, on average, too tight 9% and too loose 7% of the time. On average, thus, both procedures yield different buffer add-on decisions to a non-negligible extent.

Next, I am interested in the frequency of decision reevaluations for the procedures themselves, i.e. to what extent would initial buffer add-on decision deviate from later assessments when more data is

[24]Keep in mind that buffer add-ons are suggested to be adjusted for positive gaps between 2pp and 10pp, from 0% to 2.5% of risk weighted assets (BCBS, 2010; European Systemic Risk Board, 2014).

Table 6: Effects on CCyB Decisions: Comparing Filters

Sign Switch	Size: Abs. Difference				Initial Buffer Decision	
	0pp-2pp	2pp-4pp	4pp-6pp	>6pp	too tight	too loose
			Comparing HP1 and 2YD (QRT Demeaned)			
0.17	0.39	0.22	0.13	0.26	0.09	0.07
[0.08;0.27]	[0.07;0.65]	[0.07;0.46]	[0.07;0.22]	[0.00;0.74]	[0.02;0.20]	[0.03;0.16]

Notes: Rec.Dem. 2YD refers to the recursively demeaned two-year difference filter. Column *Sign Switch* reports the fraction of time periods when the signs misalign. Four columns report fractions for the associated absolute size differences. Columns *too tight* and *too loose* report the fractions of time periods when the initial buffer add-on decision was comparatively too tight or too loose, respectively.

added. In context of output gap estimates, Orphanides and van Norden (2002) initially looked at sign switches to determine whether the initially estimated direction of the cyclical position would have been reevaluated ex post. I extend their concept and consider not only potential decision misalignment between the initial and the ex post evaluation, but all potential reevaluations in between as well. For Basel III credit-to-GDP gaps, sign switches occur, on average, 37% of the time. Hence, very often the initial directional assessment of the cyclical position would not sustain. Communication-wise this is a problem for the regulator, because banks would question the initial decision as it would get continuously reevaluated. Beyond directional misalignment, differences can be become large in terms of size, i.e. later assessments would have suggested distinctly different buffer add-ons. Note that the absolute gap size exceeds 4pp, on average, in about half of all associated time periods. Moreover, initial buffer add-on (de)activation was, on average, too tight 17% and too loose 13% of the time. Hence, as more data is added, implementing the official detrending technique would yield continuous reevaluations of past CCyB decisions that are relevant in both, direction and size.

When signals are based on a recursively demeaned 2YD filter instead, there is distinctly smaller decision misalignment across all indicators. Sign switches are fewer, absolute gap differences remain small, and reevaluations of initial buffer add-on (de)activation decisions would be seldom. To assess how much of the small remaining misalignment is due to estimation uncertainty around the recursively estimated mean, I report the same statistics for the second half of available vintages only in the bottom panel of table 7. As estimation uncertainty vanishes with increasing sample size, practically relevant decision misalignment vanishes as well. Differences between the quasi real-time estimate and later estimates remain small, i.e. between 0pp-2pp, 77% of the time.

In contrast, decision misalignment does not vanish for Basel III credit gaps when only the second half of vintages is considered. There are fewer sign switches, because the weight of observations closer to the current end where quasi-real time and ex post estimates converge increases. But the fraction of large absolute differences in gap sizes even increases. Misalignment for initial buffer add-on decisions also remains relevant.

Table 7: Effects on CCyB Decisions: Within Filters

Detrending Procedure	Sign Switch	Size: Abs. Difference 0pp-2pp	2pp-4pp	4pp-6pp	>6pp	Initial Buffer Decision too tight	too loose
				All Vintages			
One-sided HP	0.37	0.31	0.20	0.14	0.35	0.17	0.13
	[0.23;0.51]	[0.06;0.70]	[0.08;0.41]	[0.03;0.23]	[0.00;0.80]	[0.04;0.37]	[0.03;0.31]
Rec. Dem. 2YD	0.14	0.59	0.24	0.09	0.06	0.08	0.04
	[0.06;0.30]	[0.15;0.95]	[0.04;0.39]	[0.00;0.28]	[0.00;0.50]	[0.00;0.23]	[0.00;0.19]
				Second Half of Vintages			
One-sided HP	0.28	0.21	0.19	0.14	0.46	0.21	0.08
	[0.07;0.54]	[0.00;0.68]	[0.04;0.49]	[0.04;0.30]	[0.00;0.90]	[0.00;0.60]	[0.00;0.40]
Rec. Dem. 2YD	0.05	0.77	0.15	0.04	0.02	0.04	0.01
	[0.00;0.20]	[0.12;0.99]	[0.00;0.4]	[0.00;0.28]	[0.00;0.54]	[0.00;0.21]	[0.00;0.08]

Notes: Rec.Dem. 2YD refers to the recursively demeaned two-year difference filter. Column *Sign Switch* reports the fraction of time periods when the signs misalign. Four columns report fractions for the associated absolute size differences. Columns *too tight* and *too loose* report the fractions of time periods when the initial buffer add-on decision was comparatively too tight or too loose, respectively.

Lastly, I compute the average add-on revision size in terms of % of risk weight assets. To do so, I associate the output gap values with the corresponding buffer add-ons. For gaps $\leq 2pp$, the buffer add-on is zero. For gaps $\geq 10pp$, the buffer add-on is 2.5% of risky weighted assets. In between, it increases linearly (BCBS, 2010). I then compute the absolute difference between the initially suggested and the later reevaluated add-on. Subsequently, I compute the country-specific mean across all periods where reevaluation would have taken place. Finally, I average across all 43 countries and report the mean reevaluation size. For the Basel III credit-to-GDP gap, later estimates would indicate add-on revisions often, on average, in 39% of all associated time periods. These reevaluations would be sizable with, on average, 0.73% of risky weighted assets. To put this into perspective, 0.73% amounts to 29% of the maximal buffer add-on of 2.5%.

That historical unreliability exceeds commonly discussed endpoint considerations for Basel III credit-to-GDP gaps becomes clear when I evaluate the same statistics for the subset of endpoint-associated time periods only. Decision misalignment is considerably less distinct for revisions within the first three years.

Buffer add-on reevaluations would only be suggested in 4% of endpoint-related time periods. Moreover, they would be smaller, amounting, on average, to 0.54% of risky weighted assets.[25]

6 Conclusion

This paper illustrates that implementers of the Basel III procedure to detrend credit-to-GDP series face very similar challenges as when they implement recursive linear detrending instead. Credit-to-GDP gap and trend estimates remain historically unreliable, beyond endpoint considerations, as they lack convergence to full sample estimates. Moreover, the trend history changes continuously with each incoming observation, which compromises the comparability of cyclical positions over time. This is because the prescribed HP filter configuration does not sufficiently absorb stochastic trends in credit-to-GDP and remains prone to the Nelson-Kang critique of inappropriately detrending random walks with techniques that resemble linear detrending techniques. This historical unreliability matters for CCyB decision making, both direction- and size-wise. Buffer add-on reevaluations would occur often and are sizable.

I propose to alternatively use a two-year difference filter. Doing so (i) accounts for stochastic trends, (ii) continues to provide a transparent and easy to implement detrending tool, (iii) yields historically reliable credit-to-GDP gap and trend estimates in terms of endpoint and Nelson-Kang considerations, while (iv) it maintains credit-to-GDP gap dynamics and amplitudes that the regulator deems important. Recursively demeaning such gaps ensures their zero-centering, which enables the implementation of CCyB add-ons as intended by the Basel Committee and would be associated with distinctly fewer and smaller CCyB decision misalignments.

References

Aguiar, M. and G. Gopinath (2007). Emerging market business cycles: The cycle is the trend. *Journal of Political Economy 115*(1), 69–102.

Aikman, D., A. G. Haldane, and B. D. Nelson (2015). Curbing the credit cycle. *The Economic Journal 125*(585), 1072–1109.

[25]For the recursively demeaned 2YD filter, reevaluations would be, on average, only half the size and would be indicated less often. Disregarding the earlier sample half, where estimation uncertainty around the recursively estimated mean is larger, would yield very small reevaluation sizes and very few reevaluation periods. Endpoint-related decision misalignment is small, occurs rarely, and further diminishes when only the second half of vintages is considered.

Baba, C., S. Dell'Erba, E. Detragiache, O. Harrison, A. Mineshima, A. Musayev, and A. Shahmoradi (2020). How should credit gaps be measured? An application to european countries. *IMF Working Paper* (WP/20/6).

Baumeister, C. and P. Guérin (2021). A comparison of monthly global indicators for forecasting growth. *International Journal of Forecasting 37*(3), 1276–1295.

Baxter, M. and R. G. King (1999). Measuring business cycles: Approximate band-pass filters for economic time series. *Review of Economics and Statistics 81*(4), 575–593.

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

Beltran, D. O., M. R. Jahan-Parvar, and F. A. Paine (2021). Optimizing credit gaps for predicting financial crises: Modelling choices and tradeoffs. *FRB International Finance Discussion Paper 1307*.

Bernhardsen, T., Ø. Eitrheim, A. S. Jore, and Ø. Røisland (2005). Real-time data for Norway: Challenges for monetary policy. *North American Journal of Economics and Finance 16*(3), 333–349.

Borio, C. E. V. and M. Drehmann (2009, March). Assessing the risk of banking crises - revisited. *BIS Quarterly Review*.

Cayen, J.-P. and S. van Norden (2005). The reliability of Canadian output-gap estimates. *North American Journal of Economics and Finance 16*(3), 373–393.

Cecchetti, S. G., M. Kohler, and C. Upper (2009). Financial crises and economic activity. *NBER Working Paper* (w15379).

Cerra, V. and S. C. Saxena (2008). Growth dynamics: the myth of economic recovery. *American Economic Review 98*(1), 439–457.

Chan, K. H., J. C. Hayya, and J. K. Ord (1977). A note on trend removal methods: The case of polynomial regression versus variate differencing. *Econometrica 45*(3), 737–744.

Christiano, L. J. and T. J. Fitzgerald (2003). The band pass filter. *International Economic Review 44*(2), 435–465.

Claessens, S., M. A. Kose, and M. E. Terrones (2012). How do business and financial cycles interact? *Journal of International Economics 87*(1), 178–190.

Cogley, T. and J. M. Nason (1995). Effects of the Hodrick-Prescott filter on trend and difference stationary time series implications for business cycle research. *Journal of Economic Dynamics and Control 19*(1-2), 253–278.

De Jong, R. M. and N. Sakarya (2016). The econometrics of the Hodrick-Prescott filter. *Review of Economics and Statistics 98*(2), 310–317.

Dickey, D. A. and W. A. Fuller (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association 74*(366a), 427–431.

Drehmann, M., C. E. V. Borio, and K. Tsatsaronis (2012). Characterising the financial cycle: Don't lose sight of the medium term! *BIS Working Paper* (No. 380).

Drehmann, M. and J. Yetman (2021). Which credit gap is better at predicting financial crises? A comparison of univariate filters. *International Journal of Central Banking 17*(4), 225–256.

ECB (2017, May). Financial stability review.

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

European Systemic Risk Board (2014). Recommendation of the european systemic risk board of 18 june 2014 on guidance for setting countercyclical buffer rates. *Official Journal of the European Union* (2014/C 293/01).

Garratt, A., K. Lee, E. Mise, and K. Shields (2008). Real-time representations of the output gap. *The Review of Economics and Statistics 90*(4), 792–804.

Greenwood, R., S. G. Hanson, A. Shleifer, and J. A. Sørensen (2020). Predictable financial crises. *NBER Working Paper* (w27396).

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

Hamilton, J. D. (2018). Why you should never use the Hodrick-Prescott filter. *Review of Economics and Statistics 100*(5), 831–843.

Hamilton, J. D. (2021). Measuring global economic activity. *Journal of Applied Econometrics 36*(6), 293–303.

Hamilton, J. D. and D. Leff (2020). Measuring the credit gap. mimeo.

Hodrick, R. J. (2020). An exploration of trend-cycle decomposition methodologies in simulated data. *NBER Working Paper 26750*.

Hodrick, R. J. and E. C. Prescott (1981). Postwar US business cycles: An empirical investigation. *Northwestern University Working Paper*.

Hodrick, R. J. and E. C. Prescott (1997). Postwar US business cycles: An empirical investigation. *Journal of Money, Credit, and Banking 29*(1), 1–16.

Hohl, S., M. C. Sison, T. Stastny, and R. Zamil (2018). The basel framework in 100 jurisdictions: Implementation status and proportionality practices. Technical report, Bank for International Settlements, Financial Stability Institute.

Jokipii, T., R. Nyffeler, and S. Riederer (2021). Exploring BIS credit-to-GDP gap critiques: The Swiss case. *Swiss Journal of Economics and Statistics 157*(7).

Jordà, O., M. Schularick, and A. M. Taylor (2013). When credit bites back. *Journal of Money, Credit and Banking 45*(s2), 3–28.

Kaiser, R. and A. Maravall (1999). Estimation of the business cycle: A modified Hodrick-Prescott filter. *Spanish Economic Review 1*(2), 175–206.

Kaiser, R. and A. Maravall (2012). *Measuring business cycles in economic time series*. Springer Science & Business Media.

Kaminsky, G. L. and C. M. Reinhart (1999). The twin crises: The causes of banking and balance-of-payments problems. *American Economic Review 89*(3), 473–500.

Karagedikli, O. and O. J. Rummel (2020). Weighing up the Credit-to-GDP gap: A cautionary note. *Joint Discussion Paper Series in Economics* (No. 22-2020).

Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review 99*(3), 1053–1069.

Kilian, L. (2019). Measuring global real economic activity: Do recent critiques hold up to scrutiny? *Economics Letters 178*, 106–110.

Laeven, L. and F. Valencia (2018). Systemic banking crises revisited. *International Monetary Fund*.

Long, Z. and R. Herrera (2020). Spurious OLS estimators of detrending method by adding a linear trend in difference-stationary processes - a mathematical proof and its verification by simulation. *Mathematics 8*(11), 1931.

Marcellino, M. and A. Musso (2011). The reliability of real-time estimates of the Euro Area output gap. *Economic Modelling 28*(4), 1842–1856.

Mise, E., T.-H. Kim, and P. Newbold (2005). On suboptimality of the Hodrick-Prescott filter at time series endpoints. *Journal of Macroeconomics 27*(1), 53–67.

Nelson, C. R. and H. Kang (1981). Spurious periodicity in inappropriately detrended time series. *Econometrica 49*(3), 741–751.

Orphanides, A. and S. van Norden (2002). The unreliability of output-gap estimates in real time. *Review of Economics and Statistics 84*(4), 569–583.

Quast, J. and M. H. Wolters (2022). Reliable real-time output gap estimates based on a modified Hamilton filter. *Journal of Business & Economic Statistics 40*(1), 152–168.

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.

Reinhart, C. M. and K. S. Rogoff (2009). The aftermath of financial crises. *American Economic Review 99*(466-472).

Repullo, R. and J. S. Salas (2011). The countercyclical capital buffer of basel III: A critical assessment. *CEPR Discussion Paper* (No. DP8304).

Schularick, M. and A. M. Taylor (2012). Credit booms gone bust: Monetary policy, leverage cycles, and financial crises, 1870-2008. *American Economic Review 102*(2), 1029–61.

Schüler, Y. S. (2018). Detrending and financial cycle facts across G7 countries: Mind a spurious medium term! *ECB Working Paper Series* (No. 2138).

Schüler, Y. S. (2020). On the credit-to-GDP gap and spurious medium-term cycles. *Economics Letters 192*(109245).

Schüler, Y. S. (2021). On the cyclical properties of Hamilton's regression filter. mimeo.

Sufi, A. and A. M. Taylor (2021). Financial crises: A survey. *NBER Working Paper* (w29155).

van Norden, S. and M. Wildi (2015). Basel III and the prediction of financial crises. *Prepared for Swiss National Bank Conference*. mimeo.

Watson, M. W. (1986). Univariate detrending methods with stochastic trends. *Journal of Monetary Economics 18*(1), 49–75.

Wolf, E., F. Mokinski, and Y. S. Schüler (2020). On adjusting the one-sided Hodrick-Prescott filter. *Deutsche Bundesbank Discussion Paper*.

Appendix A: Historical Reliability Statistics for Remaining Countries

Table 8: Historical Reliability Statistics for Remaining Countries

Country / Observation Detrending Procedure	$i = 154$	$i = 158$	$i = 162$	$i = 200$	$i = 250$	$i = T$
Argentina						
One-sided HP ($\lambda = 400,000$)	0.014	0.093	0.186	0.432	–	0.443
Recursive Linear Detrending	0.001	0.054	0.127	0.445	–	0.423
2YD Filter: Recursively Demeaned	0.006	0.043	0.062	0.074	–	0.058
Australia						
One-sided HP ($\lambda = 400,000$)	0.103	0.113	0.017	0.347	0.329	0.299
Recursive Linear Detrending	0.188	0.306	0.326	0.399	1.257	1.293
2YD Filter: Recursively Demeaned	0.000	0.002	0.005	0.002	0.004	0.001
Austria						
One-sided HP ($\lambda = 400,000$)	0.083	0.094	0.063	0.242	0.109	0.005
Recursive Linear Detrending	0.061	0.079	0.068	0.150	0.934	0.873
2YD Filter: Recursively Demeaned	0.000	0.000	0.000	0.001	0.005	0.008
Canada						
One-sided HP ($\lambda = 400,000$)	0.103	0.313	0.543	0.960	0.863	0.743
Recursive Linear Detrending	0.047	0.169	0.313	0.901	1.070	1.070
2YD Filter: Recursively Demeaned	0.002	0.005	0.007	0.006	0.011	0.020
Chile						
One-sided HP ($\lambda = 400,000$)	0.156	0.163	0.089	0.212	–	0.235
Recursive Linear Detrending	0.938	1.180	1.006	0.177	–	0.960
2YD Filter: Recursively Demeaned	0.002	0.004	0.012	0.009	–	0.009
China						
One-sided HP ($\lambda = 400,000$)	0.263	0.303	0.246	0.357	–	0.354
Recursive Linear Detrending	1.227	1.514	1.369	3.922	–	5.882
2YD Filter: Recursively Demeaned	0.009	0.006	0.003	0.030	–	0.035
Colombia						
One-sided HP ($\lambda = 400,000$)	0.387	0.784	1.181	–	–	2.576
Recursive Linear Detrending	0.831	1.698	2.590	–	–	6.752
2YD Filter: Recursively Demeaned	0.004	0.009	0.014	–	–	0.032
Czech Republic						
One-sided HP ($\lambda = 400,000$)	0.286	0.636	0.976	1.923	–	1.922
Recursive Linear Detrending	0.584	1.318	2.056	4.962	–	5.025
2YD Filter: Recursively Demeaned	0.010	0.022	0.033	0.055	–	0.055
Denmark						
One-sided HP ($\lambda = 400,000$)	1.416	3.116	4.100	2.663	1.836	1.498
Recursive Linear Detrending	0.504	1.231	1.720	0.908	7.879	7.812
2YD Filter: Recursively Demeaned	0.002	0.005	0.006	0.013	0.014	0.014

Country / Observation Detrending Procedure	$i = 154$	$i = 158$	$i = 162$	$i = 200$	$i = 250$	$i = T$
Finland						
One-sided HP ($\lambda = 400,000$)	1.041	1.831	2.649	3.632	3.711	3.711
Recursive Linear Detrending	0.500	0.922	1.426	2.648	1.681	1.681
2YD Filter: Recursively Demeaned	0.008	0.011	0.014	0.007	0.000	0.000
France						
One-sided HP ($\lambda = 400,000$)	0.180	0.446	0.717	0.773	1.055	1.104
Recursive Linear Detrending	0.002	0.062	0.155	0.298	2.154	2.254
2YD Filter: Recursively Demeaned	0.002	0.004	0.005	0.000	0.012	0.014
Germany						
One-sided HP ($\lambda = 400,000$)	0.230	0.372	0.379	0.072	0.179	0.180
Recursive Linear Detrending	0.202	0.362	0.437	0.271	1.253	1.624
2YD Filter: Recursively Demeaned	0.001	0.002	0.000	0.005	0.012	0.012
Greece						
One-sided HP ($\lambda = 400,000$)	0.014	0.010	0.076	0.863	0.999	0.999
Recursive Linear Detrending	0.052	0.070	0.043	1.541	2.916	2.916
2YD Filter: Recursively Demeaned	0.002	0.006	0.011	0.103	0.098	0.098
Hong Kong						
One-sided HP ($\lambda = 400,000$)	2.203	4.284	6.241	6.577	–	7.873
Recursive Linear Detrending	0.350	0.703	1.063	1.072	–	0.855
2YD Filter: Recursively Demeaned	0.011	0.019	0.024	0.003	–	0.034
Hungary						
One-sided HP ($\lambda = 400,000$)	0.144	0.198	0.211	0.213	0.398	0.398
Recursive Linear Detrending	0.193	0.314	0.397	0.202	0.694	0.694
2YD Filter: Recursively Demeaned	0.003	0.002	0.001	0.059	0.025	0.025
India						
One-sided HP ($\lambda = 400,000$)	0.049	0.050	0.025	0.390	0.656	0.570
Recursive Linear Detrending	0.124	0.222	0.300	0.248	1.023	1.295
2YD Filter: Recursively Demeaned	0.010	0.000	0.000	0.009	0.024	0.009
Indonesia						
One-sided HP ($\lambda = 400,000$)	0.024	0.106	0.171	0.315	–	0.324
Recursive Linear Detrending	0.128	0.728	1.268	3.603	–	3.697
2YD Filter: Recursively Demeaned	0.000	0.132	0.194	0.217	–	0.191
Ireland						
One-sided HP ($\lambda = 400,000$)	0.048	0.165	0.317	1.247	–	1.555
Recursive Linear Detrending	0.139	0.452	0.866	6.773	–	12.022
2YD Filter: Recursively Demeaned	0.003	0.005	0.012	0.091	–	0.069
Israel						
One-sided HP ($\lambda = 400,000$)	0.211	0.261	0.472	1.816	–	1.820
Recursive Linear Detrending	0.176	0.224	0.411	2.336	–	2.544
2YD Filter: Recursively Demeaned	0.003	0.003	0.006	0.048	–	0.051
Italy						
One-sided HP ($\lambda = 400,000$)	0.590	1.216	1.749	2.410	2.240	2.458
Recursive Linear Detrending	0.268	0.625	0.992	2.605	5.895	6.034
2YD Filter: Recursively Demeaned	0.000	0.006	0.009	0.015	0.034	0.025

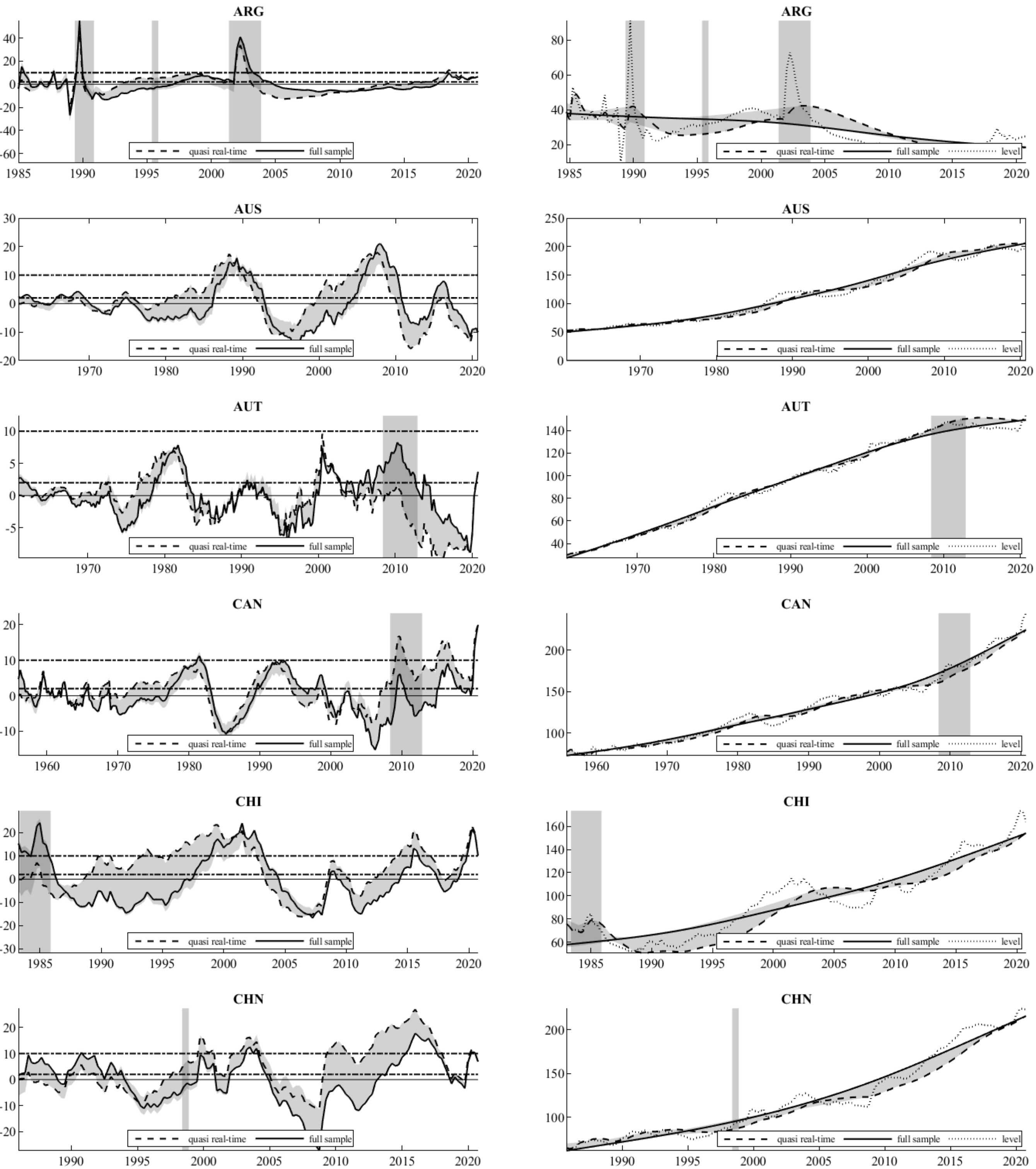
Country / Observation Detrending Procedure	$i = 154$	$i = 158$	$i = 162$	$i = 200$	$i = 250$	$i = T$
Japan						
One-sided HP ($\lambda = 400,000$)	0.058	0.055	0.004	0.650	0.662	0.686
Recursive Linear Detrending	0.104	0.162	0.177	0.850	1.830	1.904
2YD Filter: Recursively Demeaned	0.002	0.000	0.002	0.021	0.026	0.024
Korea						
One-sided HP ($\lambda = 400,000$)	0.368	0.799	1.111	1.129	0.880	0.817
Recursive Linear Detrending	0.165	0.363	0.536	0.500	0.551	0.630
2YD Filter: Recursively Demeaned	0.134	0.003	0.004	0.007	0.007	0.004
Luxembourg						
One-sided HP ($\lambda = 400,000$)	0.314	0.583	0.658	0.473	–	0.471
Recursive Linear Detrending	2.429	4.561	5.235	3.259	–	3.137
2YD Filter: Recursively Demeaned	0.002	0.184	0.170	0.000	–	0.007
Malaysia						
One-sided HP ($\lambda = 400,000$)	1.488	3.194	3.723	1.119	1.416	1.075
Recursive Linear Detrending	0.009	0.028	0.012	0.665	2.113	2.382
2YD Filter: Recursively Demeaned	0.003	0.006	0.005	0.011	0.025	0.026
Mexico						
One-sided HP ($\lambda = 400,000$)	0.445	0.745	0.931	1.010	–	1.063
Recursive Linear Detrending	0.417	0.727	0.949	1.264	–	0.863
2YD Filter: Recursively Demeaned	0.007	0.008	0.006	0.019	–	0.040
Netherlands						
One-sided HP ($\lambda = 400,000$)	0.212	0.386	0.472	0.707	0.582	0.977
Recursive Linear Detrending	0.463	0.893	1.230	3.583	4.974	3.672
2YD Filter: Recursively Demeaned	0.000	0.000	0.000	0.002	0.006	0.006
Norway						
One-sided HP ($\lambda = 400,000$)	0.054	0.146	0.263	0.794	0.881	0.882
Recursive Linear Detrending	0.060	0.076	0.049	0.463	0.176	0.365
2YD Filter: Recursively Demeaned	0.003	0.006	0.008	0.009	0.001	0.003
New Zealand						
One-sided HP ($\lambda = 400,000$)	0.457	0.520	0.222	0.476	1.454	2.194
Recursive Linear Detrending	0.398	0.714	0.940	2.946	4.026	3.404
2YD Filter: Recursively Demeaned	0.002	0.004	0.004	0.011	0.008	0.003
Poland						
One-sided HP ($\lambda = 400,000$)	0.005	0.125	0.384	0.666	–	0.660
Recursive Linear Detrending	0.005	0.118	0.369	0.711	–	0.612
2YD Filter: Recursively Demeaned	0.010	0.033	0.064	0.056	–	0.039
Portugal						
One-sided HP ($\lambda = 400,000$)	0.150	0.241	0.254	0.058	0.048	0.129
Recursive Linear Detrending	0.135	0.246	0.316	0.213	0.331	0.266
2YD Filter: Recursively Demeaned	0.000	0.000	0.002	0.039	0.047	0.026

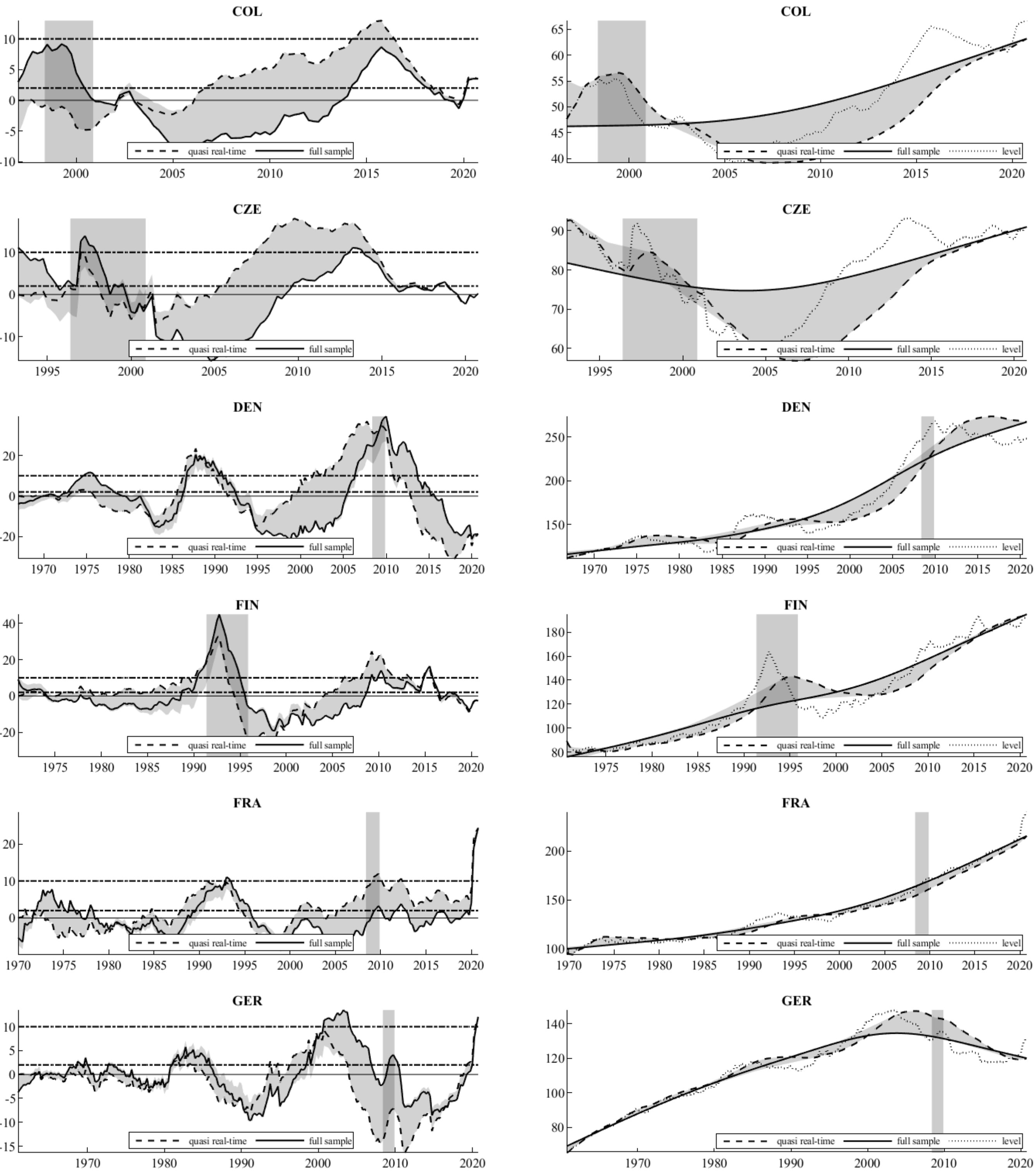
Country / Observation Detrending Procedure	$i = 154$	$i = 158$	$i = 162$	$i = 200$	$i = 250$	$i = T$
Russia						
One-sided HP ($\lambda = 400,000$)	0.528	0.850	0.799	1.051	–	1.053
Recursive Linear Detrending	0.192	0.314	0.303	0.466	–	0.471
2YD Filter: Recursively Demeaned	0.011	0.012	0.005	0.007	–	0.008
Saudi Arabia						
One-sided HP ($\lambda = 400,000$)	0.707	1.045	2.025	1.553	–	1.553
Recursive Linear Detrending	0.068	0.102	0.201	0.196	–	0.213
2YD Filter: Recursively Demeaned	0.002	0.004	0.009	0.005	–	0.007
Singapore						
One-sided HP ($\lambda = 400,000$)	2.379	5.325	6.824	6.952	4.895	4.895
Recursive Linear Detrending	0.375	1.052	1.369	0.443	1.098	1.098
2YD Filter: Recursively Demeaned	0.007	0.012	0.012	0.006	0.013	0.013
South Africa						
One-sided HP ($\lambda = 400,000$)	0.324	0.572	0.623	0.051	0.108	0.125
Recursive Linear Detrending	0.177	0.334	0.389	0.270	1.386	1.453
2YD Filter: Recursively Demeaned	0.003	0.005	0.004	0.004	0.013	0.014
Spain						
One-sided HP ($\lambda = 400,000$)	0.061	0.106	0.185	1.500	1.948	1.961
Recursive Linear Detrending	0.072	0.133	0.251	5.044	6.814	6.772
2YD Filter: Recursively Demeaned	0.000	0.001	0.003	0.081	0.036	0.037
Sweden						
One-sided HP ($\lambda = 400,000$)	0.151	0.247	0.321	0.134	0.051	0.058
Recursive Linear Detrending	0.225	0.409	0.582	0.832	1.741	1.961
2YD Filter: Recursively Demeaned	0.000	0.001	0.000	0.003	0.013	0.016
Switzerland						
One-sided HP ($\lambda = 400,000$)	0.090	0.103	0.067	0.337	0.398	0.439
Recursive Linear Detrending	0.109	0.177	0.213	0.070	0.271	0.185
2YD Filter: Recursively Demeaned	0.000	0.002	0.003	0.011	0.007	0.003
Thailand						
One-sided HP ($\lambda = 400,000$)	0.163	0.306	0.377	0.204	0.242	0.242
Recursive Linear Detrending	0.225	0.441	0.590	0.523	1.012	1.012
2YD Filter: Recursively Demeaned	0.007	0.012	0.013	0.048	0.046	0.046
Turkey						
One-sided HP ($\lambda = 400,000$)	0.107	0.159	0.105	0.470	–	0.475
Recursive Linear Detrending	0.162	0.247	0.169	1.437	–	1.791
2YD Filter: Recursively Demeaned	0.014	0.012	0.004	0.113	–	0.111
United Kingdom						
One-sided HP ($\lambda = 400,000$)	0.175	0.267	0.225	0.062	0.187	0.234
Recursive Linear Detrending	0.477	0.853	1.078	2.298	3.134	2.866
2YD Filter: Recursively Demeaned	0.000	0.002	0.004	0.004	0.005	0.007

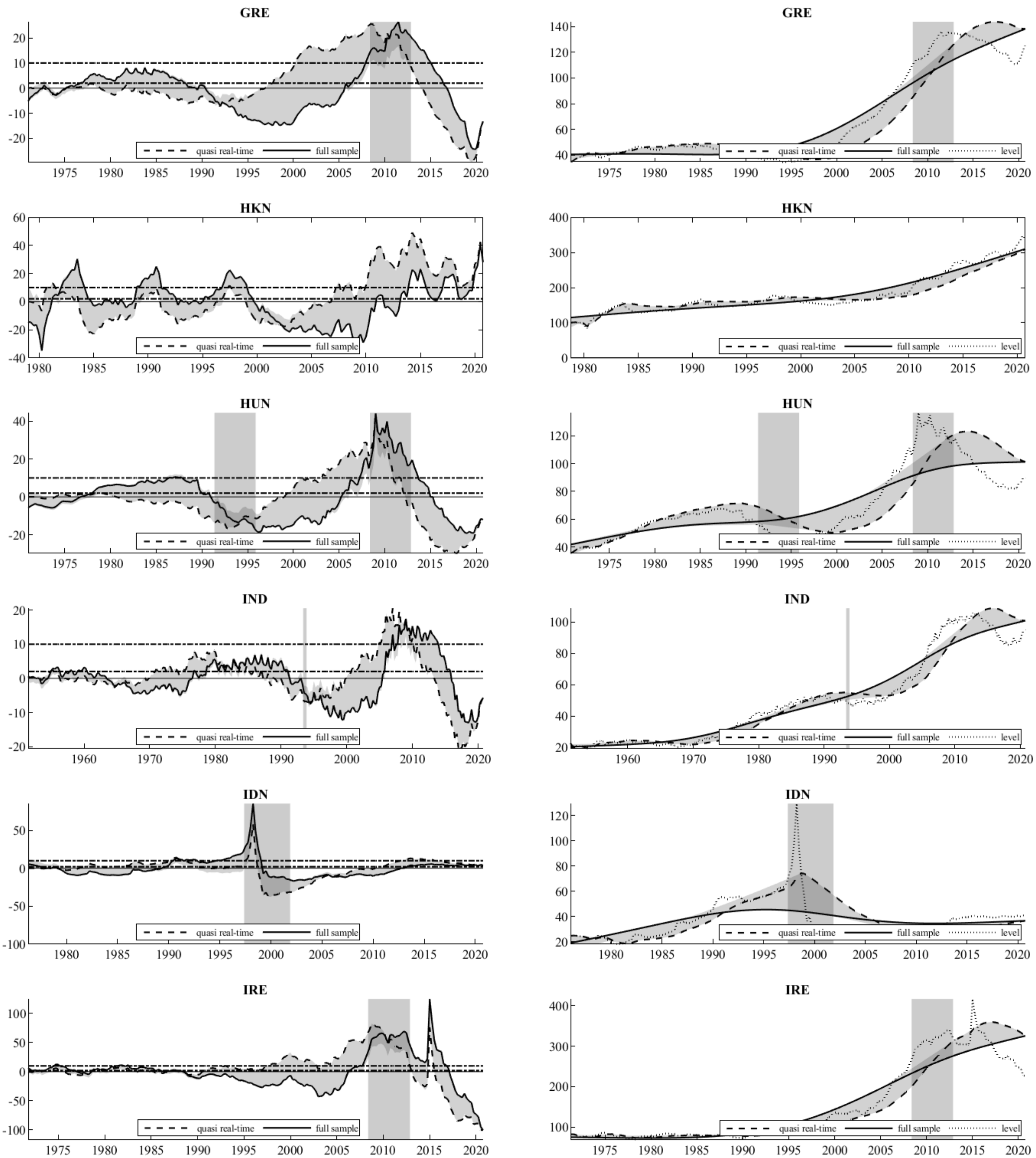
Notes: $i = 154/158/162$ indicate observations one, two, and three years after initial estimation. $i = T$ marks the respective sample size. A value of zero reflects perfect historical reliability. A value of one indicates that historical unreliability is as large as the gap itself.

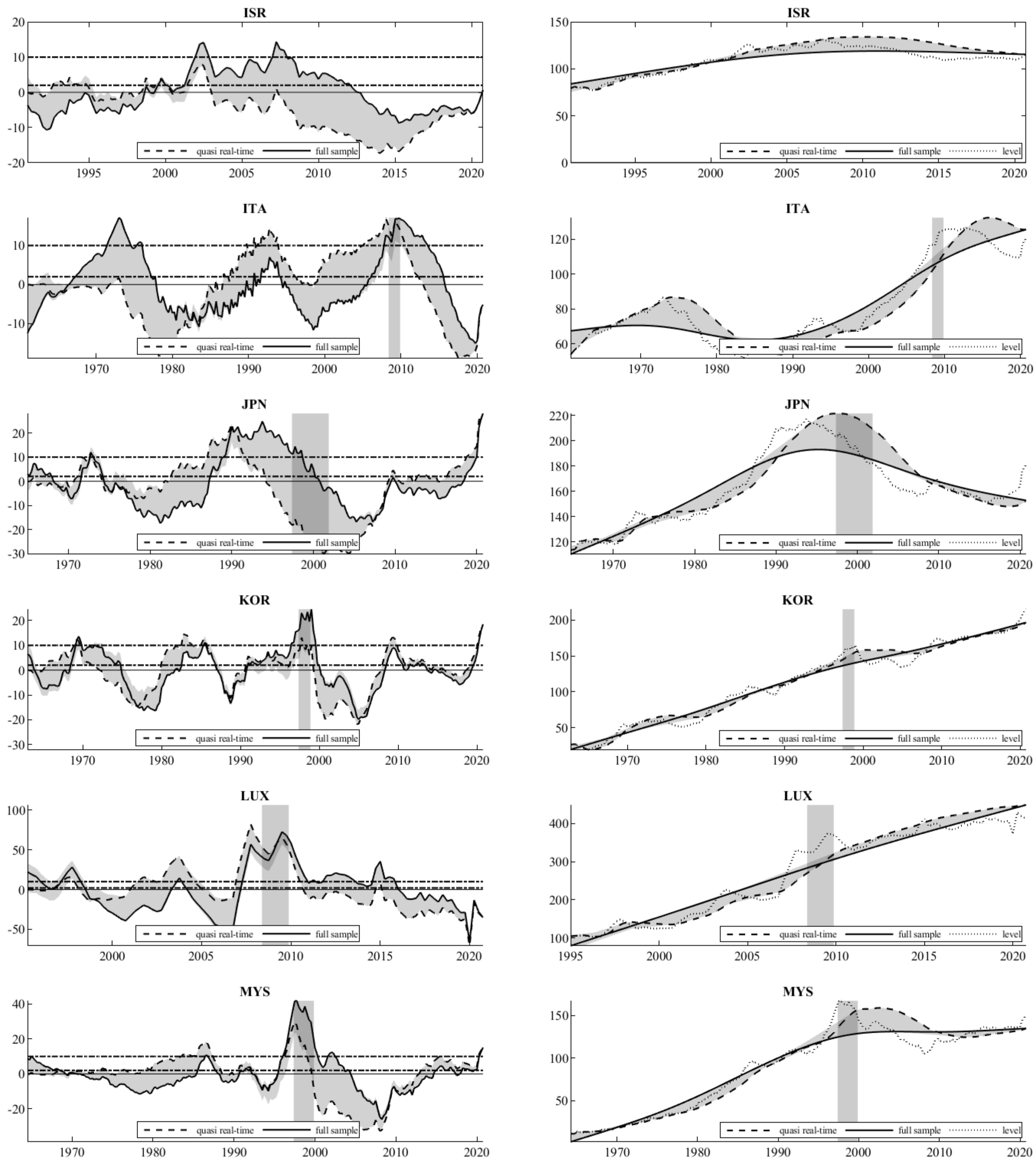
Appendix B: Historical Reliability Bands for Remaining Countries

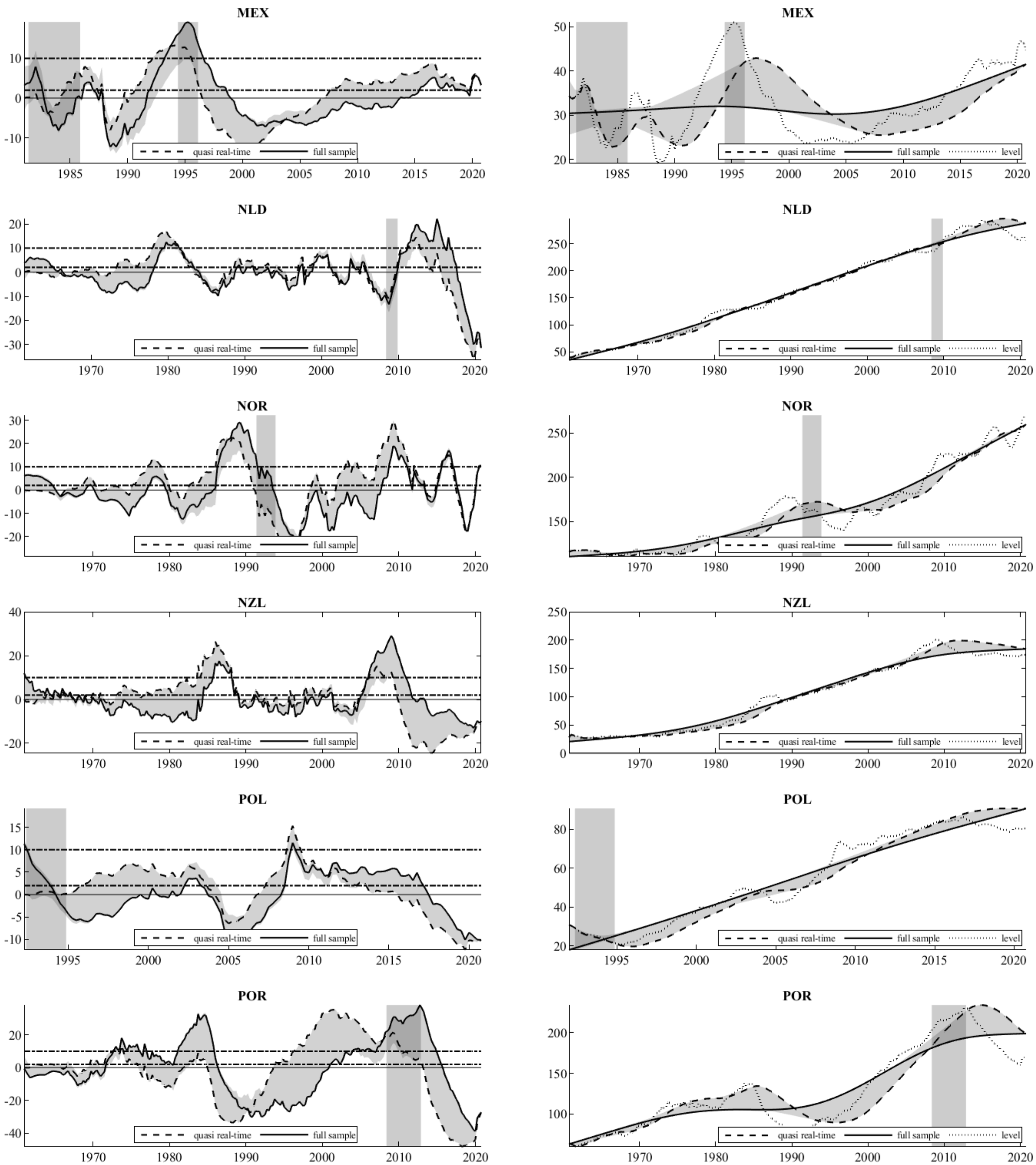
Figure 10: Historical Reliability Bands for Gap (left) and Trend (right) Estimates for Remaining Countries: Basel III Prescribed HP Filter Configuration

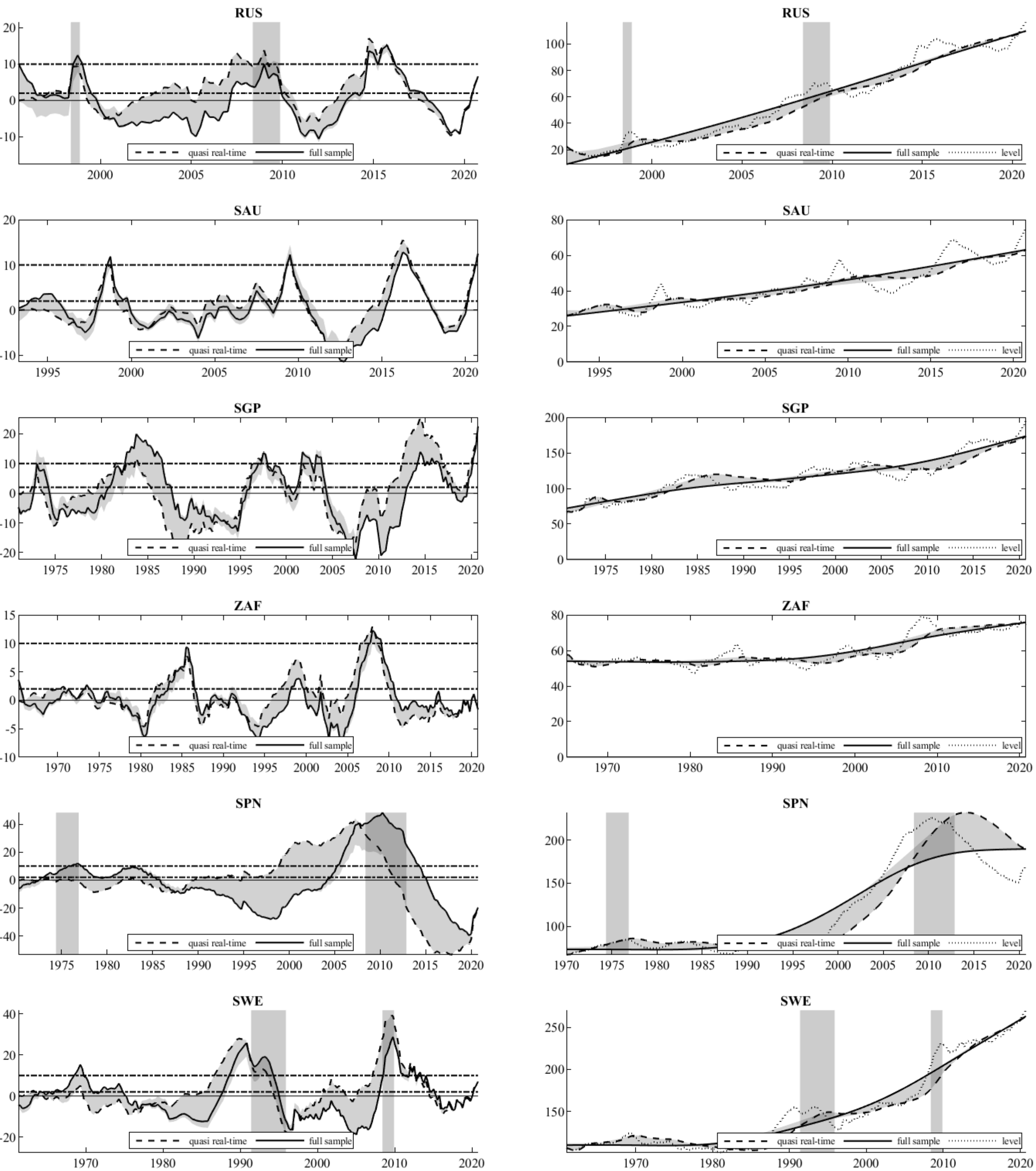


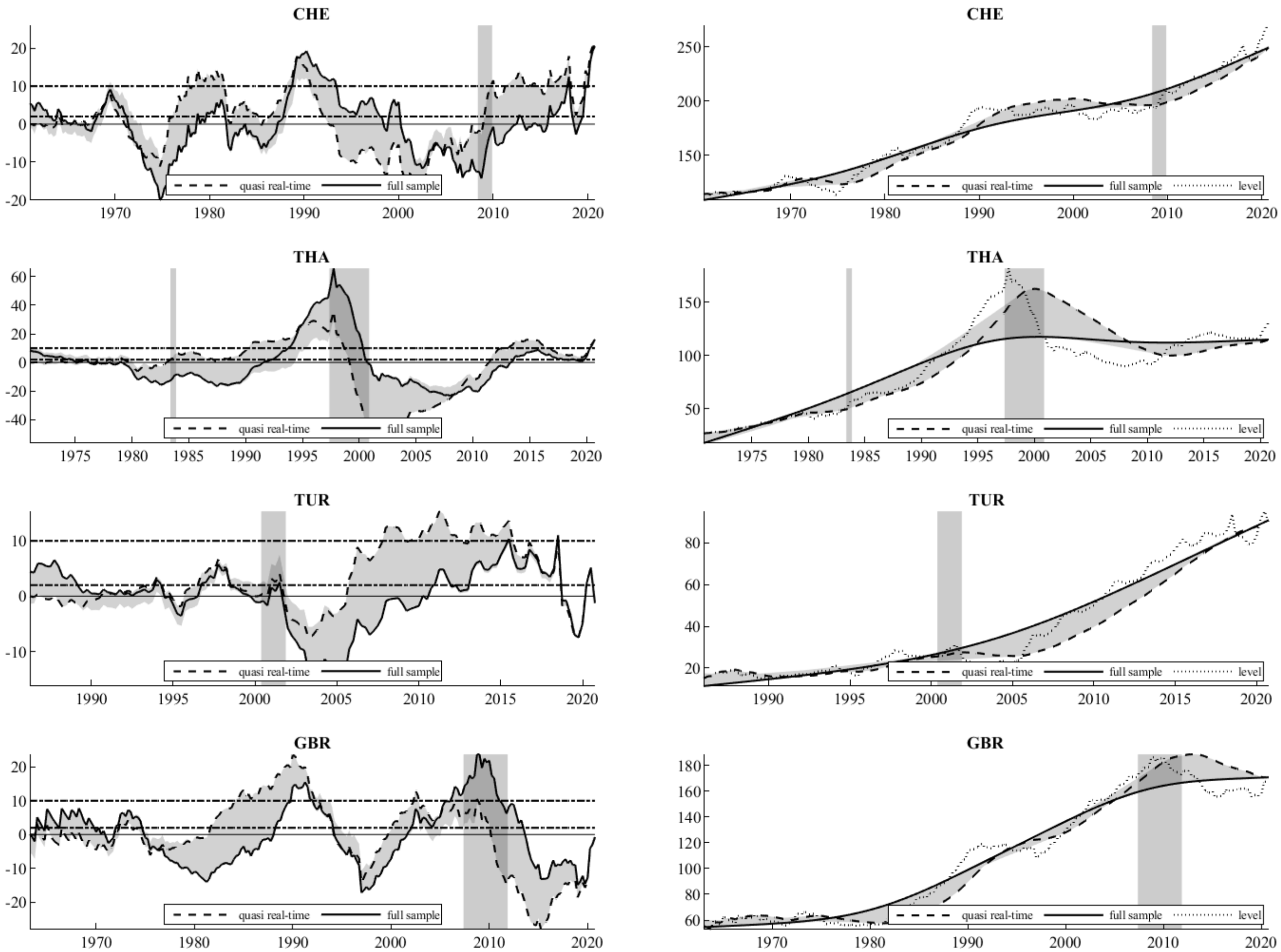

COL
quasi real-time
full sample
level
CZE
DEN
FIN
FRA
GER


GRE
quasi real-time
full sample
GRE
quasi real-time
full sample
level
HKN
quasi real-time
full sample
HKN
quasi real-time
full sample
level
HUN
quasi real-time
full sample
HUN
quasi real-time
full sample
level
IND
quasi real-time
full sample
IND
quasi real-time
full sample
level
IDN
quasi real-time
full sample
IDN
quasi real-time
full sample
level
IRE
quasi real-time
full sample
IRE
quasi real-time
full sample
level


ISR
quasi real-time
full sample
level
ITA
JPN
KOR
LUX
MYS

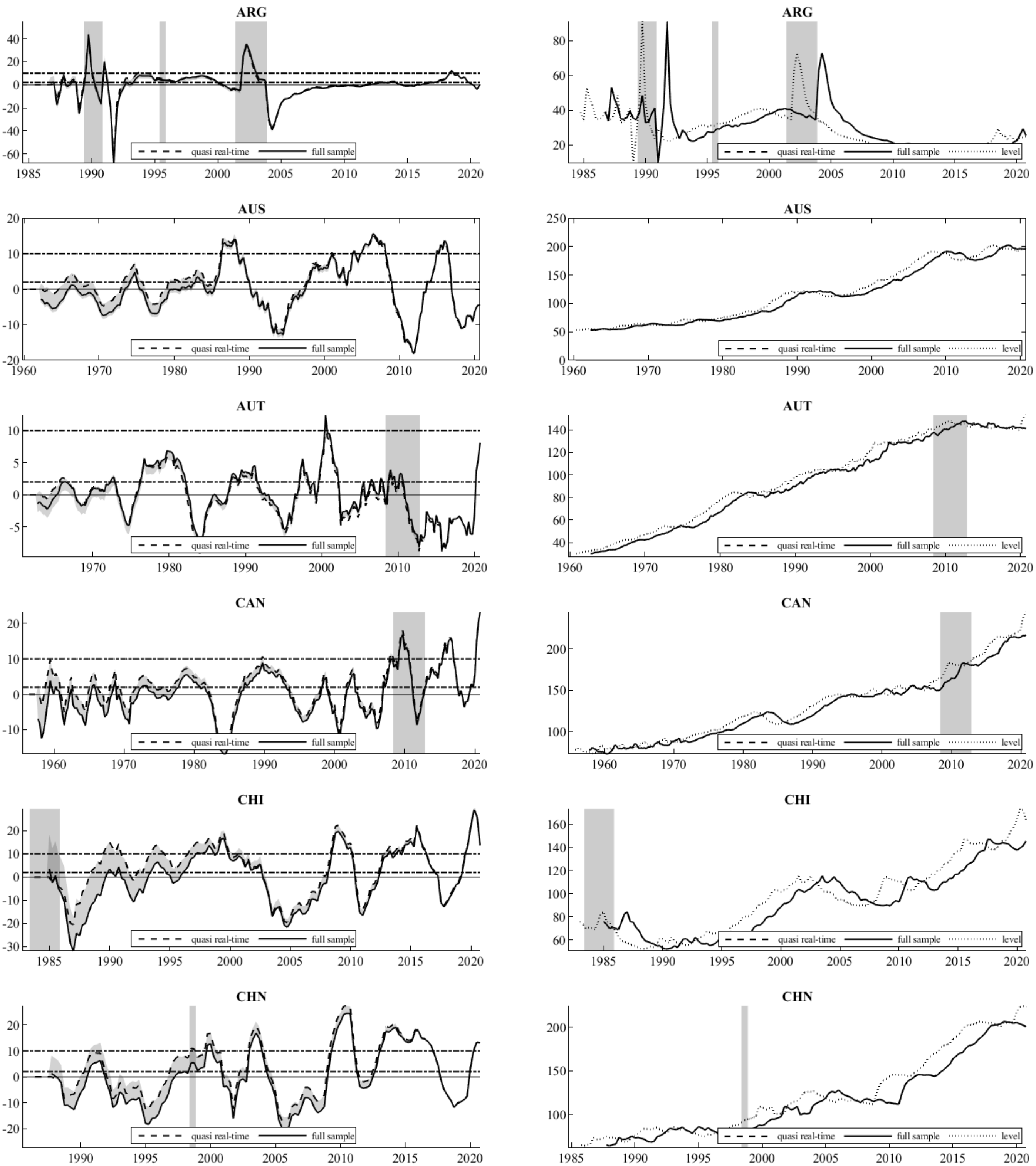

MEX
quasi real-time
full sample
MEX
quasi real-time
full sample
level
NLD
quasi real-time
full sample
NLD
quasi real-time
full sample
level
NOR
quasi real-time
full sample
NOR
quasi real-time
full sample
level
NZL
quasi real-time
full sample
NZL
quasi real-time
full sample
level
POL
quasi real-time
full sample
POL
quasi real-time
full sample
level
POR
quasi real-time
full sample
POR
quasi real-time
full sample
level

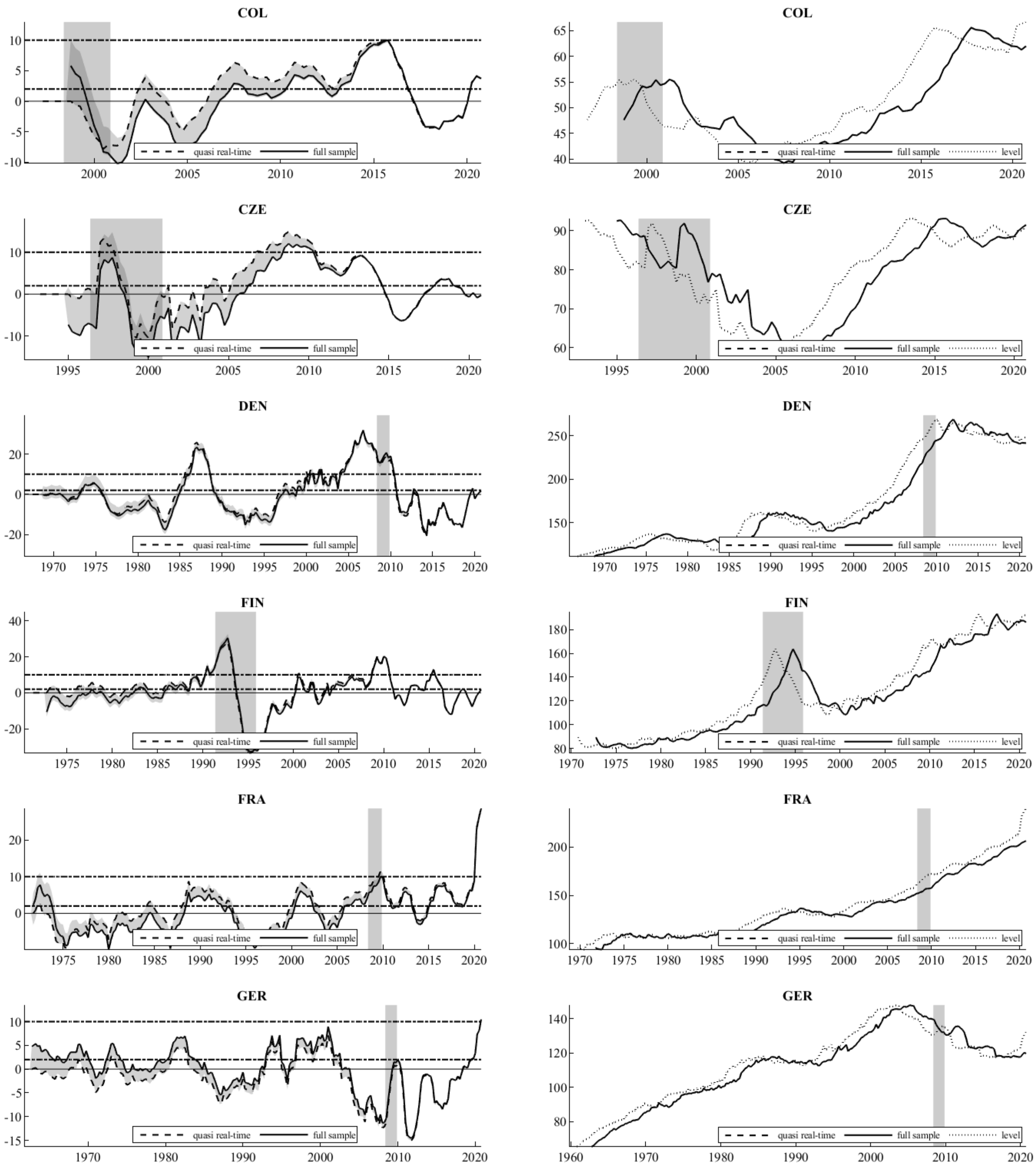

RUS
quasi real-time
full sample
RUS
quasi real-time
full sample
level
SAU
quasi real-time
full sample
SAU
quasi real-time
full sample
level
SGP
quasi real-time
full sample
SGP
quasi real-time
full sample
level
ZAF
quasi real-time
full sample
ZAF
quasi real-time
full sample
level
SPN
quasi real-time
full sample
SPN
quasi real-time
full sample
level
SWE
quasi real-time
full sample
SWE
quasi real-time
full sample
level

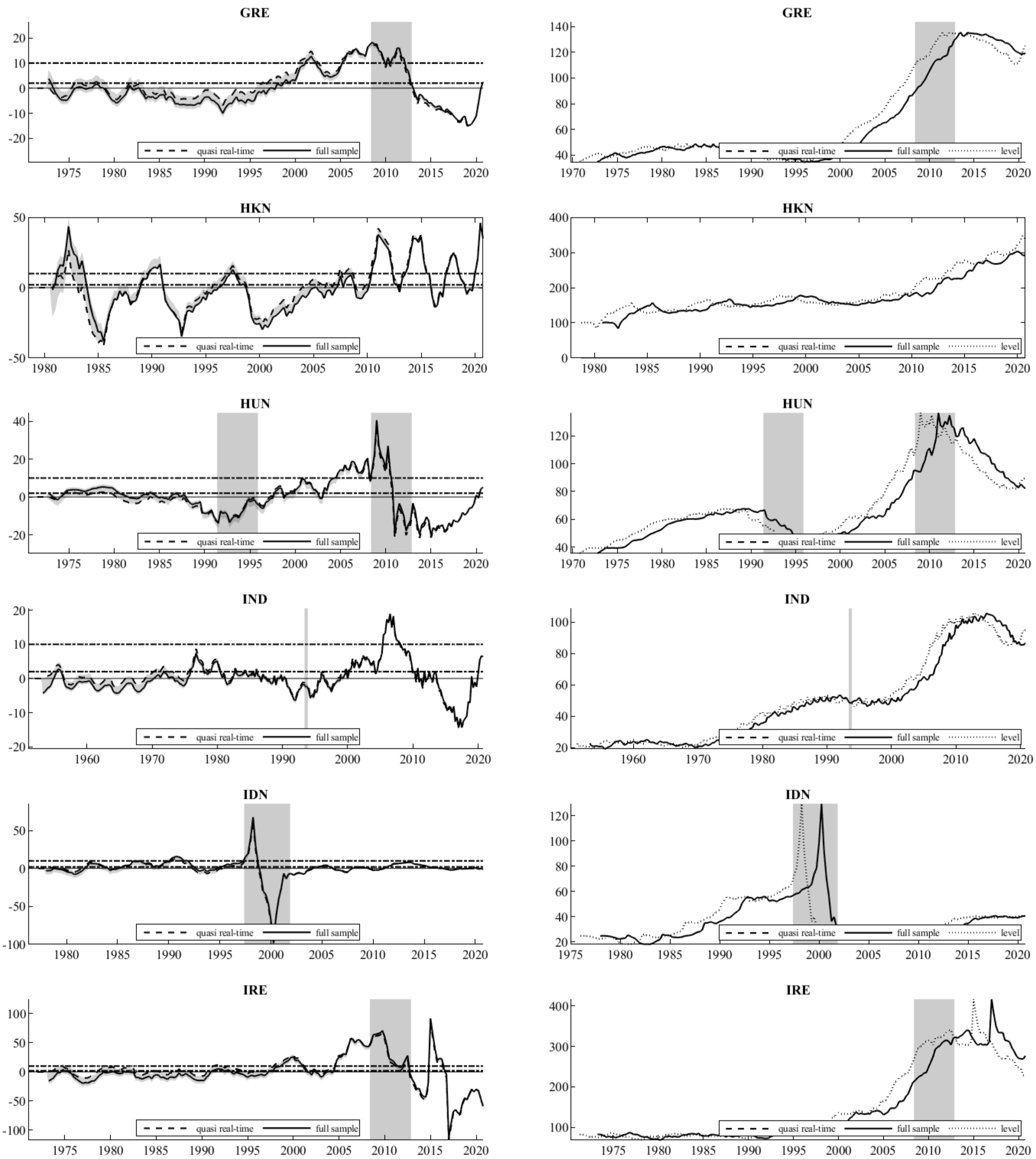


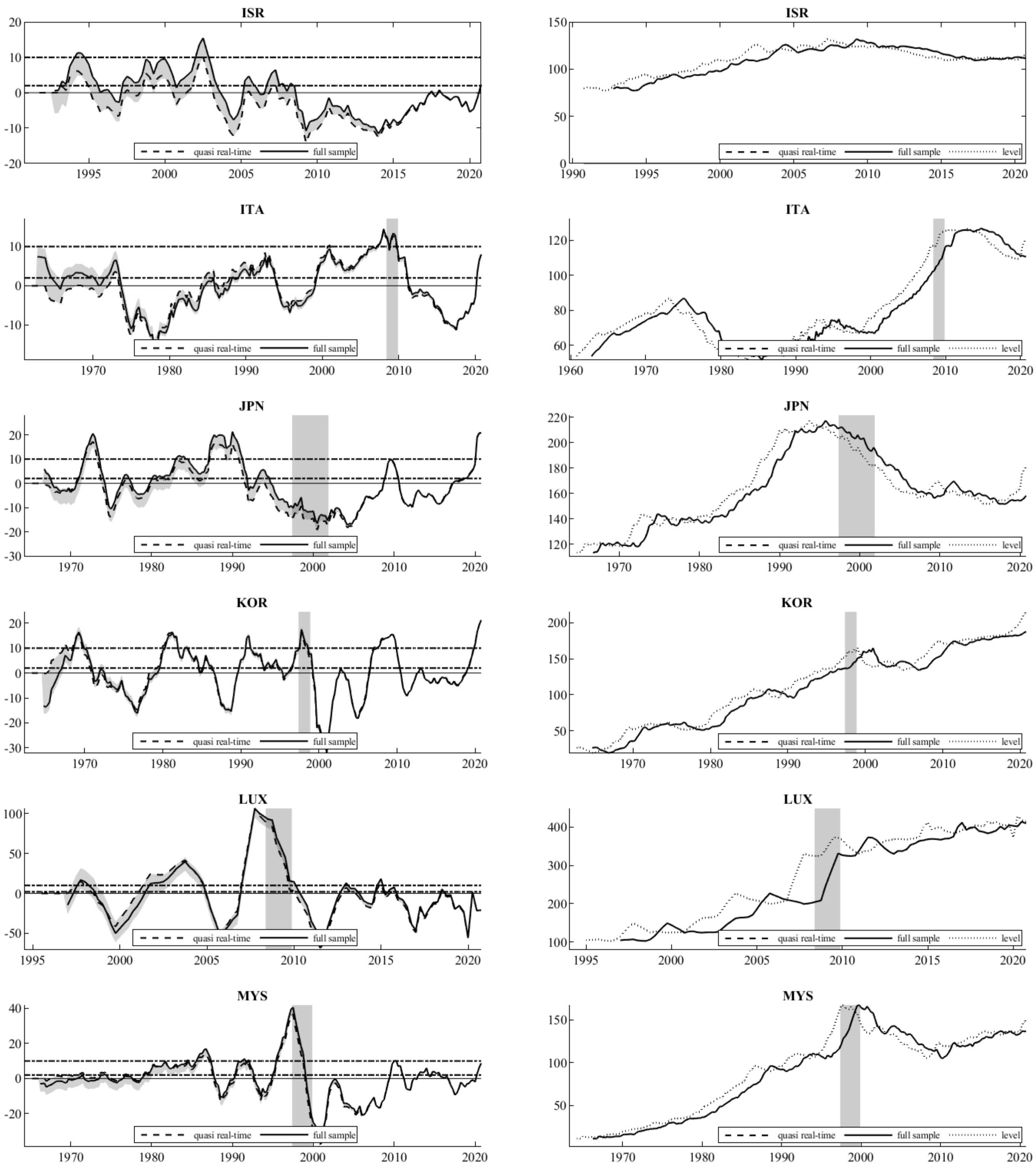
Notes: Gap estimates in terms of percentage deviations. Dashed lines show the quasi real-time and solid lines the full sample estimates. Gray shades indicate historical reliability bands, i.e. the smallest and largest estimated value for period *t* across all vintages. Dash-dotted lines for 2pp and 10pp deviations indicate the range when national authorities should consider activating CCyB add-ons (BCBS, 2010). Grey bars indicate systemic banking crisis from Laeven and Valencia (2018).

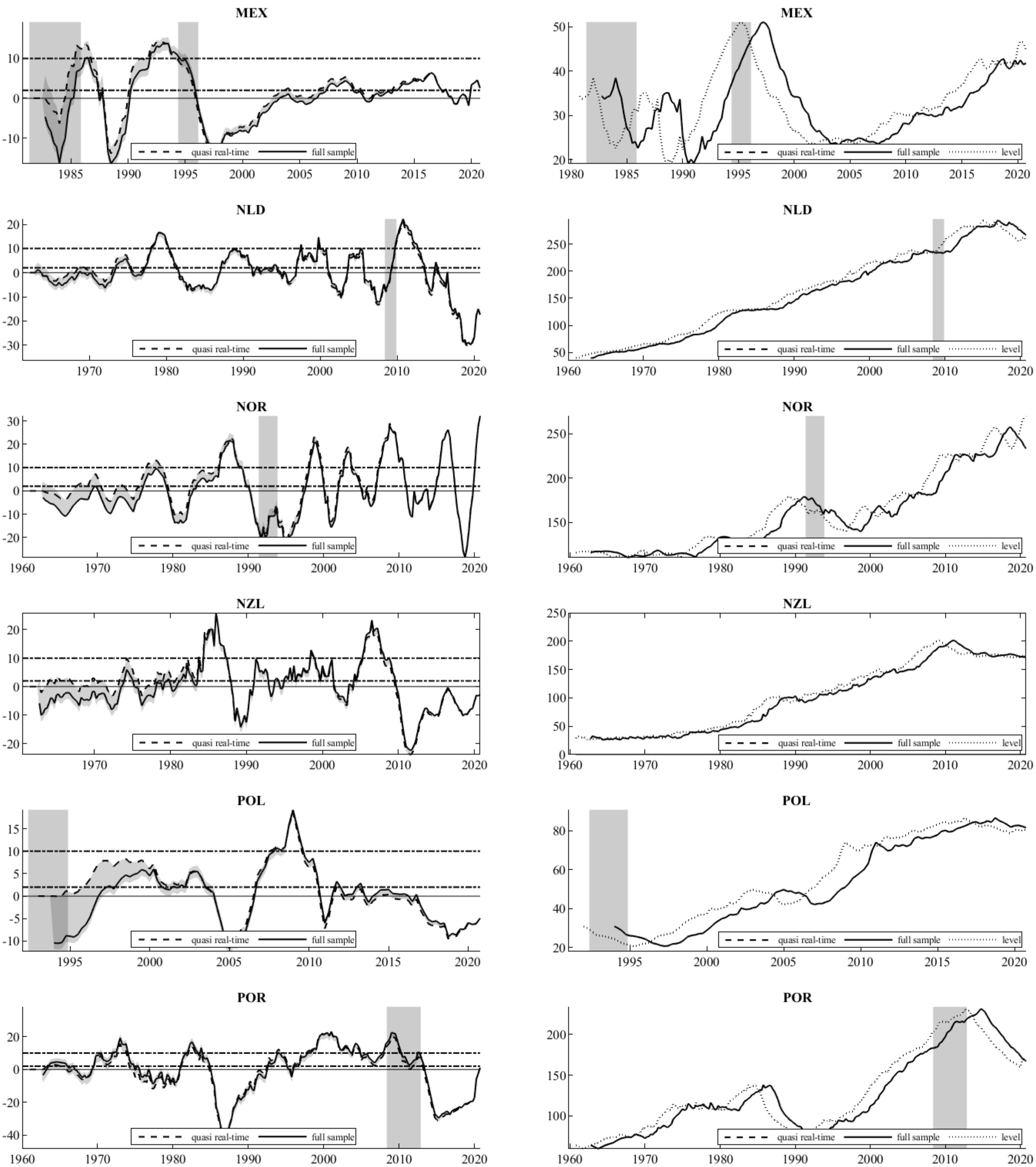
Figure 11: Historical Reliability Bands for Gap (left) and Trend (right) Estimates for Remaining Countries: Recursively Demeaned 2YD Filter

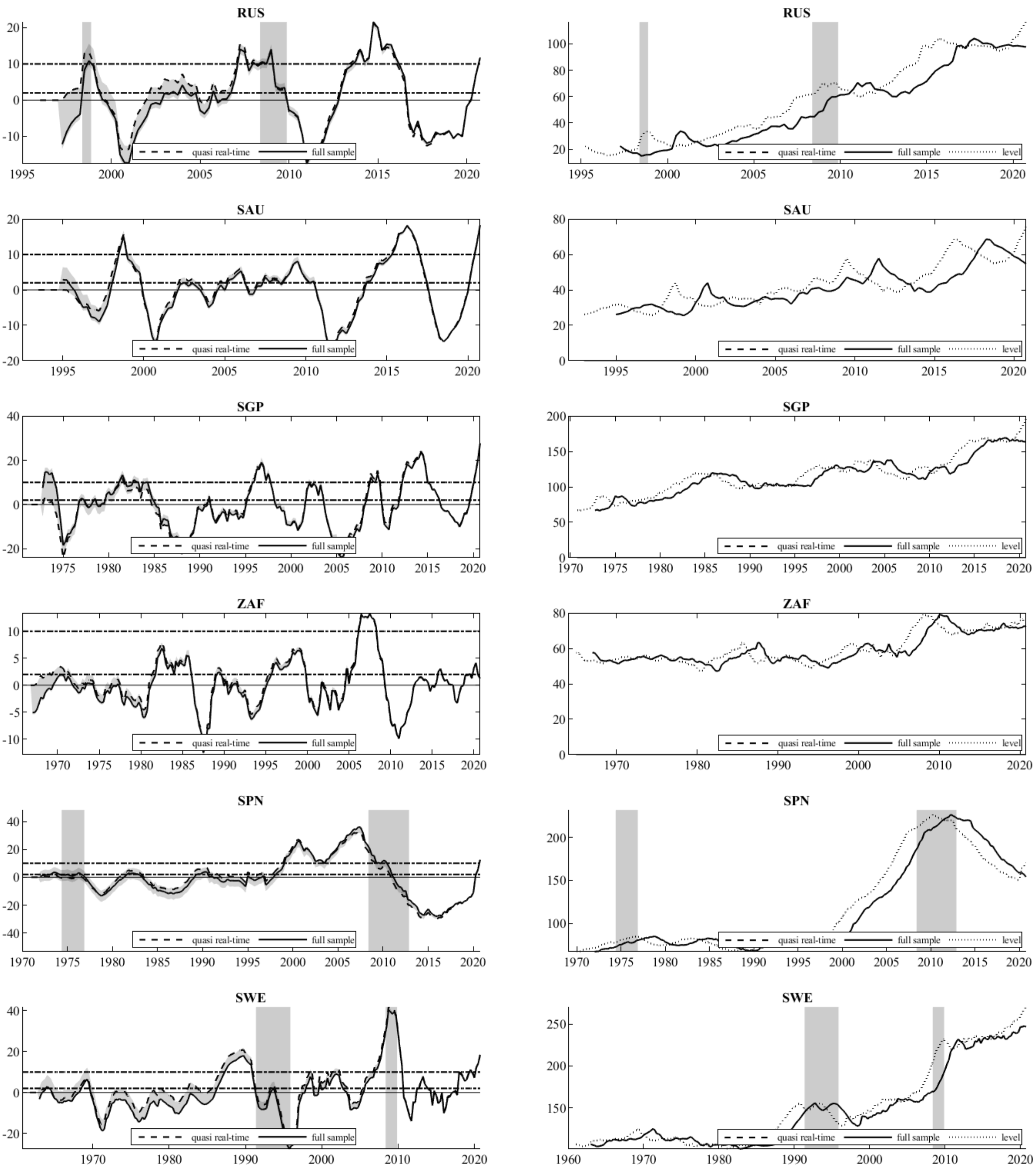


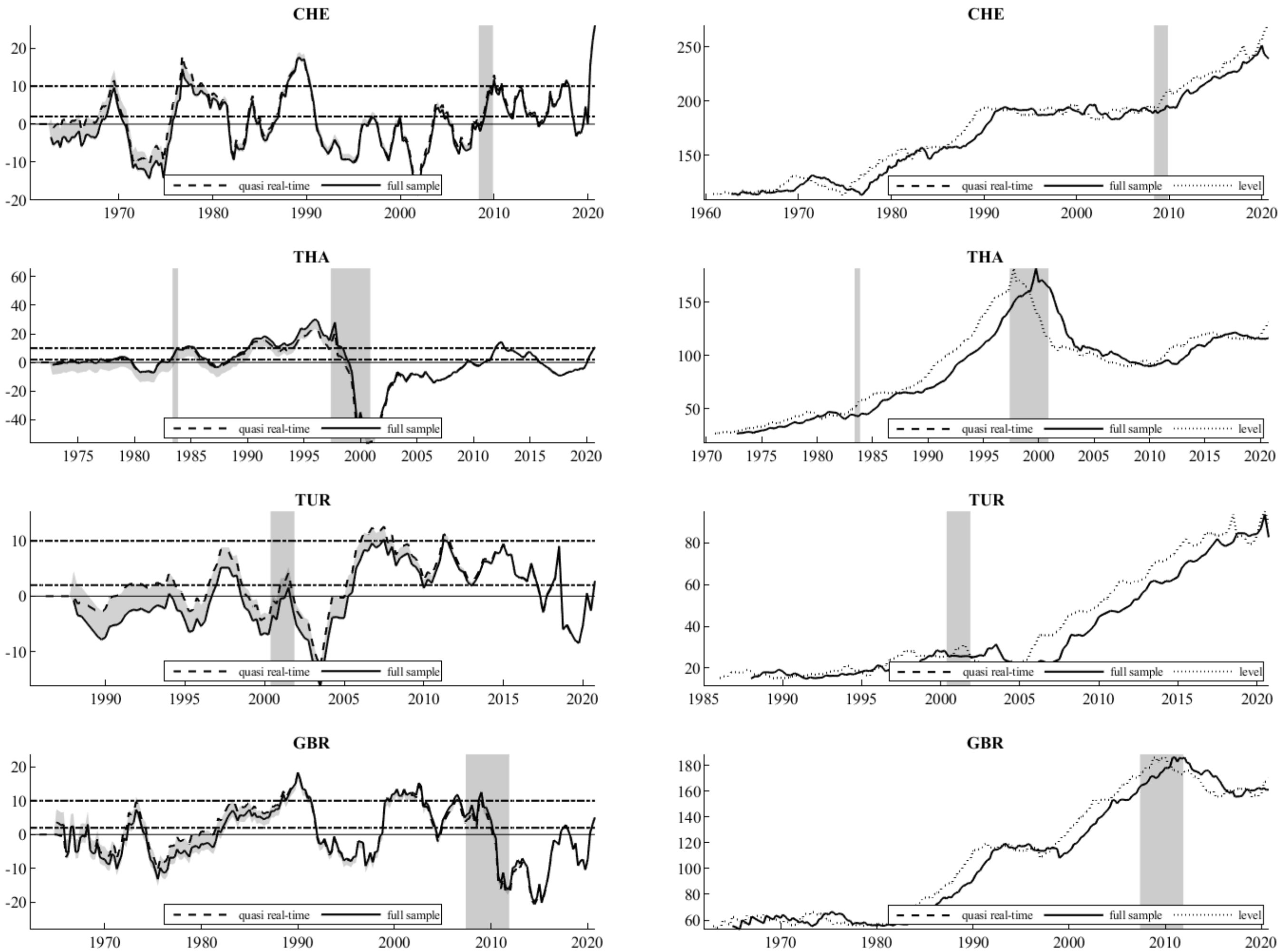

COL
quasi real-time
full sample
COL
quasi real-time
full sample
level
CZE
quasi real-time
full sample
CZE
quasi real-time
full sample
level
DEN
quasi real-time
full sample
DEN
quasi real-time
full sample
level
FIN
quasi real-time
full sample
FIN
quasi real-time
full sample
level
FRA
quasi real-time
full sample
FRA
quasi real-time
full sample
level
GER
quasi real-time
full sample
GER
quasi real-time
full sample
level


GRE
quasi real-time
full sample
GRE
quasi real-time
full sample
level
HKN
quasi real-time
full sample
HKN
quasi real-time
full sample
level
HUN
quasi real-time
full sample
HUN
quasi real-time
full sample
level
IND
quasi real-time
full sample
IND
quasi real-time
full sample
level
IDN
quasi real-time
full sample
IDN
quasi real-time
full sample
level
IRE
quasi real-time
full sample
IRE
quasi real-time
full sample
level


ISR
quasi real-time
full sample
ISR
quasi real-time
full sample
level
ITA
quasi real-time
full sample
ITA
quasi real-time
full sample
level
JPN
quasi real-time
full sample
JPN
quasi real-time
full sample
level
KOR
quasi real-time
full sample
KOR
quasi real-time
full sample
level
LUX
quasi real-time
full sample
LUX
quasi real-time
full sample
level
MYS
quasi real-time
full sample
MYS
quasi real-time
full sample
level


MEX
quasi real-time
full sample
MEX
quasi real-time
full sample
level
NLD
quasi real-time
full sample
NLD
quasi real-time
full sample
level
NOR
quasi real-time
full sample
NOR
quasi real-time
full sample
level
NZL
quasi real-time
full sample
NZL
quasi real-time
full sample
level
POL
quasi real-time
full sample
POL
quasi real-time
full sample
level
POR
quasi real-time
full sample
POR
quasi real-time
full sample
level


RUS
quasi real-time
full sample
RUS
quasi real-time
full sample
level
SAU
quasi real-time
full sample
SAU
quasi real-time
full sample
level
SGP
quasi real-time
full sample
SGP
quasi real-time
full sample
level
ZAF
quasi real-time
full sample
ZAF
quasi real-time
full sample
level
SPN
quasi real-time
full sample
SPN
quasi real-time
full sample
level
SWE
quasi real-time
full sample
SWE
quasi real-time
full sample
level



Notes: Gap estimates in terms of percentage deviations. Dashed lines show the quasi real-time and solid lines the full sample estimates. Gray shades indicate historical reliability bands, i.e. the smallest and largest estimated value for period *t* across all vintages. Dash-dotted lines for 2pp and 10pp indicate the range when national authorities should consider activating CCyB add-ons (BCBS, 2010). Gray bars indicate systemic banking crises (Laeven and Valencia, 2018).

Appendix C: Unit Root Test Results for Level Credit-to-GDP Series

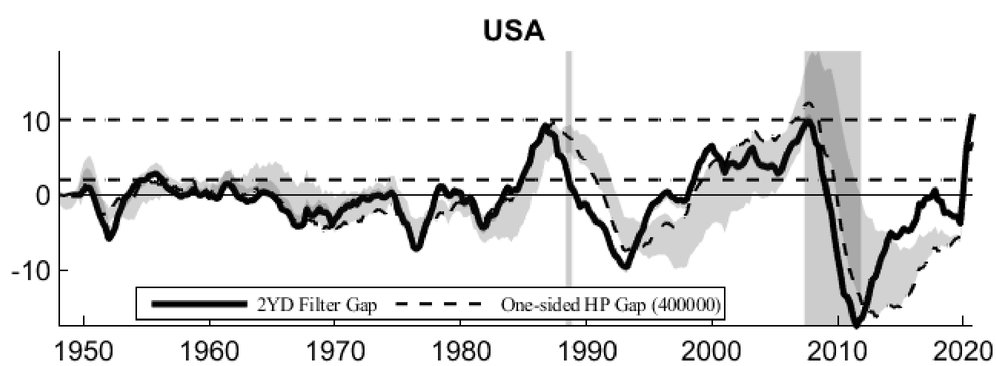
Table 9: Unit Root Test Results for Remaining Country Level Credit-to-GDP Series

Country	ADF	KPSS	Country	ADF	KPSS
Argentina	-3.10 (0.11)	0.13 (0.07)	Japan	-1.77 (0.63)	0.54 (0.01)
Australia	-2.89 (0.17)	0.35 (0.01)	Korea	-3.27 (0.07)	0.20 (0.02)
Austria	-2.71 (0.24)	0.33 (0.01)	Luxembourg	-1.66 (0.75)	0.15 (0.04)
Belgium	-1.87 (0.66)	0.50 (0.01)	Malaysia	-1.80 (0.69)	0.47 (0.01)
Brazil	-0.72 (0.97)	0.28 (0.01)	Mexico	-2.86 (0.18)	0.15 (0.04)
Canada	-0.10 (0.99)	0.41 (0.01)	Netherlands	-1.62 (0.78)	0.12 (>0.10)
Chile	-2.65 (0.27)	0.15 (0.05)	Norway	-2.39 (0.40)	0.37 (0.01)
China	-2.06 (0.56)	0.33 (0.01)	New Zealand	-2.03 (0.58)	0.25 (0.01)
Colombia	-2.51 (0.34)	0.26 (0.01)	Poland	-1.84 (0.67)	0.12 (>0.10)
Czech Republic	-1.71 (0.73)	0.27 (0.01)	Portugal	-2.54 (0.32)	0.27 (0.01)
Denmark	-1.92 (0.63)	0.37 (0.01)	Russia	-3.19 (0.09)	0.12 (0.01)
Finland	-2.26 (0.46)	0.17 (0.03)	Saudi Arabia	-3.41 (0.06)	0.09 (>0.10)
France	1.14 (1.00)	0.49 (0.01)	Singapore	-1.64 (0.78)	0.16 (0.01)
Germany	-2.15 (0.52)	0.49 (0.01)	South Africa	-2.67 (0.26)	0.37 (0.04)
Greece	-1.74 (0.72)	0.46 (0.01)	Spain	-2.80 (0.20)	0.30 (0.01)
Hong Kong	-1.75 (0.97)	0.38 (0.01)	Sweden	-1.39 (0.86)	0.55 (0.01)
Hungary	-1.31 (0.88)	0.20 (0.02)	Switzerland	-1.51 (0.82)	0.16 (0.04)
India	-2.32 (0.43)	0.39 (0.01)	Thailand	-2.29 (0.45)	0.36 (0.01)
Indonesia	-2.05 (0.57)	0.24 (0.01)	Turkey	-1.60 (0.79)	0.37 (0.01)
Ireland	-1.35 (0.87)	0.39 (0.01)	United Kingdom	-2.55 (0.32)	0.26 (0.01)
Israel	-1.38 (0.86)	0.35 (0.01)	United States	-4.02 (0.01)	0.11 (>0.10)
Italy	-1.92 (0.86)	0.51 (0.01)			

Notes: ADF and KPSS test results for original level credit-to-GDP series with associated p-values given in parenthesis.

Appendix D: US Credit-to-GDP Gap Estimates: One-sided HP and Recursively Demeaned 2YD Filter

Figure 12: US Credit-to-GDP Gap Estimates: One-sided HP and Recursively Demeaned 2YD Filter



Notes: Gap estimates in terms of percentage deviations. The dashed line shows the one-sided HP filtered gap estimate, the solid line the two-year difference filtered gap estimate. Gray shades indicate historical reliability bands, i.e. the smallest and largest estimated value for period *t* across all vintages. Dash-dotted lines for 2pp and 10pp indicate the range when national authorities should consider activating CCyB add-ons (BCBS, 2010). Gray bars indicate systemic banking crises (Laeven and Valencia, 2018).