DEUTSCHE BUNDESBANK
EUROSYSTEM

Discussion Paper

Deutsche Bundesbank
No 33/2020

Identifying indicators of systemic risk

Benny Hartwig
(Goethe University Frankfurt and Deutsche Bundesbank)

Christoph Meinerding
(Deutsche Bundesbank)

Yves Schüler
(Deutsche Bundesbank)

Deutsche Bundesbank, Wilhelm-Epstein-Straße 14, 60431 Frankfurt am Main,
Postfach 10 06 02, 60006 Frankfurt am Main

Tel +49 69 9566-0

Please address all orders in writing to: Deutsche Bundesbank,
Press and Public Relations Division, at the above address or via fax +49 69 9566-3077

Internet http://www.bundesbank.de

ISBN 978–3–95729–724–2 (Printversion)
ISBN 978–3–95729–725–9 (Internetversion)

Non-technical summary

Research question

One of the main goals of the newly-established macroprudential policies is to mitigate systemic risk. However, there is still an open debate on which indicators best measure systemic risk. In this paper, we propose an objective way to identify indicators of systemic risk and therefore to consolidate the large number of candidates proposed in the literature.

Contribution

We develop a hierarchical testing framework that operationalizes the definition of systemic risk provided by the International Monetary Fund, the Bank of International Settlements, and the Financial Stability Board. We apply our hierarchical testing framework to a set of popular indicators that are claimed to measure systemic risk.

Results

We determine that the Basel III credit-to-GDP gap – arguably our most prominent candidate variable – does not indicate systemic risk coherently across G7 countries. This is because elevated systemic risk is signaled by a high value of the credit-to-GDP gap in some countries and a low value in other countries. A composite financial cycle indicator, which combines information from the growth of credit and asset prices, signals systemic risk consistently for all G7 countries except Canada.

Overall, our results suggest that countercyclical macroprudential policy may address vulnerability episodes ahead of financial crises, rather than systemic risk itself according to its definition. It may smooth the financial cycle in boom phases, which then indirectly mitigates the amount of systemic risk that can build up in the distant future.

Nichttechnische Zusammenfassung

Fragestellung

Eines der Hauptziele der neuen makroprudenziellen Regulierungsmaßnahmen ist die Minderung von systemischem Risiko. Die Debatte darüber, welche Indikatoren systemisches Risiko messen können, wird jedoch nach wie vor geführt. In diesem Aufsatz stellen wir ein objektives Verfahren vor, solche Indikatoren zu identifizieren, mit dem die große Anzahl der in der Literatur vorgeschlagenen Indikatoren konsolidiert werden kann.

Beitrag

Wir entwickeln einen hierarchischen Testrahmen, der die Definition von systemischem Risiko operationalisiert, die vom Internationalen Währungsfonds, der Bank für Internationalen Zahlungsausgleich und dem Financial Stability Board festgelegt wurde. Diesen hierarchischen Testrahmen wenden wir auf eine Reihe gängiger Indikatoren an, die zur Messung von systemischem Risiko vorgeschlagen werden.

Ergebnisse

Wir zeigen, dass die Kredit/BIP-Lücke gemäß Basel III – die gebräuchlichste der von uns betrachteten Variablen – systemisches Risiko nicht kohärent über die G7-Länder anzeigt. In einigen Ländern wird nämlich durch ein hohes, in anderen Ländern hingegen durch ein niedriges Niveau der Kredit-BIP-Lücke signalisiert, dass systemisches Risiko besteht. Ein zusammengesetzter Indikator für den Finanzzyklus, der Informationen aus dem Wachstum von Kreditvolumen und Wertpapierpreisen kombiniert, signalisiert indes systemisches Risiko konsistent für alle G7-Länder mit Ausnahme Kanadas.

Insgesamt deuten unsere Ergebnisse darauf hin, dass sich mit einer antizyklischen makroprudenziellen Politik zwar Verwundbarkeiten im Vorfeld von Finanzkrisen eindämmen lassen, diese Politik jedoch nicht das eigentliche systemische Risiko im Sinne der Definition bekämpfen kann. Somit ist antizyklische makroprudenzielle Politik eher dafür geeignet, den Finanzzyklus in Aufschwungphasen zu glätten. Hierdurch verringert sich indirekt das Ausmaß von in der Zukunft entstehendem systemischem Risiko.

DEUTSCHE BUNDESBANK DISCUSSION PAPER NO 33/2020

IDENTIFYING INDICATORS OF SYSTEMIC RISK

Benny Hartwig*   Christoph Meinerding**   Yves Schüler**

May 26, 2020

This paper contains charts in color. Use color printer for best results.

Abstract: We operationalize the definition of systemic risk provided by the IMF, BIS, and FSB and derive testable hypotheses to identify indicators of systemic risk. We map these hypotheses into a two-stage hierarchical testing framework, combining insights from the early-warning literature on financial crises with recent advances on growth-at-risk. Applying this framework to a set of candidate variables, we find that the Basel III credit-to-GDP gap does not indicate systemic risk coherently across G7 countries. Credit growth and house price growth also do not pass our test in many cases. By contrast, a composite financial cycle signals systemic risk consistently for all countries except Canada. Overall, our results suggest that systemic risk may be consistently measured only once the turning points of indicators have been observed. Therefore, pre-emptive countercyclical macroprudential policy may smooth the financial cycle in boom phases, which then indirectly mitigates the amount of systemic risk in the future.

Keywords: Systemic risk, macroprudential regulation, forecasting, growth-at-risk, financial cycles

JEL classification: E37, E44, G17

*Goethe University Frankfurt and Deutsche Bundesbank, e-mail: benny.hartwig@bundesbank.de.
**Deutsche Bundesbank, e-mails: christoph.meinerding|yves.schueler@bundesbank.de

We gratefully acknowledge comments from seminar and conference participants at AFA 2020, Bundesbank Spring Conference 2019, Bank of Lithuania, and University of Freiburg. Special thanks go to Klaus Adam, David Aikman, Geert Bekaert, Johannes Beutel, Paul Hiebert, Manfred Kremer, Andreas Lehnert, Thomas Lux, Ralf Meisenzahl, Norbert Metiu, José-Luis Peydró, and colleagues at Goethe University and Deutsche Bundesbank. Moreover, we thank Birte Lohmann for her excellent research assistance. This work represents the authors' personal opinions and does not necessarily reflect the views of the Deutsche Bundesbank or the Eurosystem.

1 Introduction

In response to the global financial crisis, new macroprudential policies have been implemented. One of their main goals is to mitigate systemic risk. Despite these policy reforms, however, there is still no consensus on which tools best measure systemic risk. Rather than a unique variable or a narrow set, a wide variety of systemic risk indicators has been proposed (see Bisias, Flood, Lo, and Valavanis (2012); Aikman, Kiley, Lee, Palumbo, and Warusawitharana (2017)). Consolidating this large number of candidate indicators and selecting the relevant indicator(s) for systemic risk assessment therefore remains a formidable challenge.

The purpose of this paper is to provide objective guidance on which indicators actually qualify for measuring systemic risk. To this end, we operationalize the definition of systemic risk that the International Monetary Fund (IMF), the Bank of International Settlements (BIS), and the Financial Stability Board (FSB) have agreed upon and derive testable hypotheses that allow us to identify indicators of systemic risk. The resulting statistical hypotheses test is straightforward to implement and easy to interpret, as we demonstrate for a set of candidate indicators. The outcomes enhance our understanding of systemic risk and of the indicators used to measure it, representing a step towards the thorough calibration of macroprudential policies.

Based on the definition of systemic risk, we argue that a variable qualifies as an indicator of systemic risk if it predicts the probability of a disruption to financial services several periods ahead and, subsequently, the predicted probability is positively correlated with the amount of downside risk for the real economy. To check these conditions, we develop a two-stage hierarchical testing framework that combines methodologies from the early-warning literature on financial crises (e.g. Demirgüç-Kunt and Detragiache (1998)) and from the growth-at-risk literature (Adrian, Boyarchenko, and Giannone (2019)).

For the G7 countries, we document the following major empirical results. First, the credit-to-GDP gap, which plays a prominent role in Basel III regulations, does not indicate systemic risk coherently across G7 countries. This is because elevated systemic risk is signaled by a high value of the credit-to-GDP gap in some countries and a low value in other countries. Statistically, the null hypothesis of not indicating systemic risk can be rejected in many of the cases considered. However, the test delivers inconclusive signs for the slope coefficients across countries. In particular, our findings for Canada, Japan, and Germany contradict the interpretation of the credit-to-GDP gap given under Basel III (high value = high level of systemic risk). Altogether, this impedes a coherent interpretation of the signals sent to policymakers.[1]

Second, a composite measure of the financial cycle put forward by Schüler, Hiebert, and Peltonen (2015, 2020), which combines information from the growth of credit and asset prices, performs better than the credit-to-GDP gap. It signals systemic risk consistently for all G7 countries except Canada. Notably, we observe that low (and not high) values of the financial cycle indicate high systemic risk. We argue that this makes intuitive sense. On the one hand, it is in line with empirical observations: declines in the growth rates of equity and house prices preceded the onset of the global financial crisis by several years. On the other hand, turning points are unpredictable. That is, observing a high current growth rate of credit and/or asset prices cannot lead to the conclusion that systemic risk is elevated at a fixed period ahead.[2] This is also in line

[1]Basel III regulations advise to contemplate an activation of the countercyclical capital buffer in case a country's credit-to-GDP gap exceeds two percentage points. This would require a positive sign in our testing framework.

[2]Many studies artificially generate the result of a positive sign. These studies do not predict financial crisis themselves but rather "vulnerability periods" ahead of them, i.e. periods that, by some quarters, precede the turning point by definition. Thus, in these exercises the turning point is assumed to be known (e.g. Anundsen, Gerdrup, Hansen, and Kragh-Sørensen (2016)). Due to their yearly sampling frequency, the seminal paper by Schularick and Taylor (2012) also predicts vulnerability periods, rather than crisis periods. The authors report that a rise in credit growth predicts financial crises.

with theoretical mechanisms underpinning financial cycles in the spirit of Geanakoplos (2010) or Brunnermeier and Sannikov (2014).

Third, the individual components of the composite financial cycle (credit growth as well as equity, house and bond price growth) do not pass our test in many of the cases considered. Therefore, they should not be used as indicators of systemic risk on a stand-alone basis. If anything, house price growth performs best among these individual components, in line with the evidence in Reinhart and Rogoff (2009).

Fourth, financial conditions indices capturing contagion and spillover effects that are prominently featured in Adrian et al. (2019) also do not indicate systemic risk. They fail in the early-warning stage of our test, i.e. they do not predict the probability of a disruption to financial services several periods ahead. Given their input variables (e.g. implied/realized volatilities, various credit spreads, or large downward spikes in asset prices), they can be used to assess the contemporaneous state of the financial system, but not to calibrate countercyclical macroprudential policies.

Fifth, all these results survive a battery of robustness checks concerning both the structure of our test and the choice of left-hand side and right-hand side variables.

Importantly for policy, the results deepen our understanding of systemic risk, the target of countercyclical macroprudential policy. We find that systemic risk may be consistently measured only once the turning points of indicators – from high to low, possibly leading into financial crisis – have been observed. Given the impossibility of predicting turning points, we argue that pre-emptive countercyclical macroprudential policy may address vulnerability episodes ahead of financial crises, rather than systemic risk itself according to the definition provided by the IMF, BIS, and FSB. It may smooth the financial cycle in boom phases, which then indirectly mitigates the amount of systemic risk that can build up in the distant future.

2 Methodology

In this section we describe our approach to identifying indicators of systemic risk. We first present the definition of systemic risk, which we map into two key principles. Based on these principles we then derive a two-stage hierarchical testing framework with testable hypotheses.

2.1 The definition of systemic risk and two principles

In their report to the G20 finance ministers in 2009, the IMF, BIS, and FSB define systemic risk as a

> *"risk of disruption to financial services that is (i) caused by an impairment of all or parts of the financial system and (ii) has the potential to have serious negative consequences for the real economy"* (IMF, BIS, and FSB (2009)).[3]

From this definition, we can deduce two key principles for indicators of systemic risk.

Principle 1 *An indicator of systemic risk has to measure, as of today, the probability of a future event that qualifies as a disruption to financial services caused by an impairment of the financial system.*

[3] This definition is very similar to the definition of systemic risk used by other institutions, such as the European Central Bank.

To arrive at this principle, we interpret the word "risk" as today's *probability* of an event in the future. Clearly, the event of interest is the disruption to financial services. Importantly, the fact that this event lies in the future adds a time dimension to the concept and measurement of systemic risk.

Principle 2 *The probability of a future disruption must be negatively related to the left tail of the conditional distribution of real economic variables.*

The second part of the definition indicates that not all probabilities of disruption qualify as systemic risk. They only do so if they relate to real outcomes. The words "potential ... consequences" imply that the probabilities must relate to the future conditional distribution of real economic variables. The term "serious negative" indicates that we are particularly interested in the *left tail* of such conditional distributions. Specifically, a rise in the probability of a future disruption should lower the left tail of the conditional distribution (or in other words, increase downside risk).

2.2 A two-stage hierarchical testing framework

We map these two principles into a two-stage hierarchical testing framework to identify indicators of systemic risk. In the first stage we test whether a candidate indicator measures the probability of future disruptions to financial services. In the second stage we check for a link between the estimated probability of disruption and the left tail of real economic variables. We conclude that a candidate variable serves as an indicator of systemic risk if it passes both stages of our test.

2.2.1 Stage 1

In Stage 1, we borrow from methods that have been established in the literature on early warning models of financial crises, which was initiated, amongst others, by Demirgüç-Kunt and Detragiache (1998) and Kaminsky and Reinhart (1999). Within a logit framework, we predict a "financial disruption" dummy variable in time period $t+h$ using a candidate indicator in time period t in order to receive an "early warning" signal of a financial disruption lying ahead.

More formally, let $\pi_{t,t+h} = P(d_{t+h} = 1|\mathcal{F}_t)$ denote the conditional probability of a disruption, where d_{t+h} refers to the dummy at time period $t+h$. $\mathcal{F}_t$ is the information set containing information up to time period t. We model the h-period ahead probabilities using the lagged relation

$$\text{logit}(\pi_{t,t+h}) = \alpha + \sum_{k=0}^{K} \beta_k x_{t-k}, \qquad (1)$$

where $\text{logit}(\pi_{t,t+h}) = \ln(\pi_{t,t+h}/(1-\pi_{t,t+h}))$ is the log of the odds ratio, α is the intercept, the β_k are lag coefficients, and x_{t-k} is the candidate indicator of systemic risk at time period $t-k$. We select the number of lags K by minimizing the Bayes information criterion (BIC), as it consistently selects the true lag length and favors a parsimonious model (Greene (2012)). The maximum number of lags is 6 for half-yearly and 12 for quarterly data.

A candidate indicator *passes* Stage 1 if $\beta_k \neq 0$ for at least one k. For this, we require the null hypothesis of all β_k being equal to zero to be rejected by a likelihood ratio test. Furthermore, if a candidate indicator passes both stages of our hierarchical testing framework, we will also interpret the sum of all coefficients, $\sum_{k=0}^{K} \beta_k$. More precisely, we say that a high (low) level of a candidate indicator predicts a high probability of a disruption if $\sum_{k=0}^{K} \beta_k > 0 \quad (\sum_{k=0}^{K} \beta_k < 0)$.

2.2.2 Stage 2

In Stage 2, we relate the probability of disruption from Stage 1 to the left tail of real economic variables. We rely on two distinct frameworks: regular linear regressions (henceforth labeled "mean regressions") and quantile regressions (Koenker and Bassett (1978)). This allows us to model both the center of the conditional distribution and a specific quantile of the conditional left tail.[4] Following Adrian et al. (2019), we choose the 5% conditional quantile.[5]

More precisely, we regress real GDP growth y_{t+h} in period $t+h$ on the predicted probabilities $\widehat{\pi}_{t,t+h}$ from Stage 1:

$$y_{t+h} = \gamma + \delta\widehat{\pi}_{t,t+h} + \boldsymbol{\omega}'\mathbf{z}_t + \varepsilon_{t+h}. \tag{2}$$

γ is a scalar intercept and δ is a scalar coefficient. As controls, we include a $d \times 1$ vector of lags of GDP growth, $\mathbf{z}_t$, with the corresponding coefficient vector $\boldsymbol{\omega}$.

In a quantile regression framework, the error term ε_{t+h} is assumed to satisfy $Q_\tau(\varepsilon_{t+h}) = 0$, where Q_τ denotes the τ-th quantile (here: $\tau = 5\%$). Furthermore, the regression coefficients minimize the objective function $\sum_{t=1}^{T} \rho_\tau(\varepsilon_{t+h})$, where

$$\rho_\tau(\varepsilon_{t+h}) = \begin{cases} \varepsilon_{t+h} \cdot \tau & \text{, if } \varepsilon_{t+h} > 0 \\ \varepsilon_{t+h} \cdot (\tau - 1) & \text{, if } \varepsilon_{t+h} < 0. \end{cases} \tag{3}$$

We obtain two estimates of δ in Stage 2, one from the mean regression framework (henceforth $\hat{\delta}$) and one from the quantile regression framework ($\hat{\delta}_\tau$). A candidate indicator passes Stage 2 if either $\delta < 0$ or $\delta_\tau < 0$. For this, we require that either the null hypothesis $\delta \geq 0$ or $\delta_\tau \geq 0$ is rejected by a one-sided t-test. As we discuss below in Section 2.3, we adjust the standard errors of the coefficients for two potential biases: (i) for the fact that $\widehat{\pi}_{t,t+h}$ is estimated with uncertainty and (ii) for possible heteroskedasticity.

Economically, we argue that an indicator which passes Stage 2 comes closer to the spirit of the definition of systemic risk if $\delta_\tau < \delta$. This implies that an increase in the disruption probability has a larger effect on the downside risk for the real economy than on the mean, in line with the notion of "severe negative consequences". We determine $\delta_\tau < \delta$ by checking that the confidence bands as defined by the t-statistics for the two parameters δ and δ_τ do not overlap. If we cannot reject that $\delta \geq 0$, we argue that it suffices to reject $\delta_\tau \geq 0$ to fulfil $\delta_\tau < \delta$. Overall, this procedure fosters the detection of non-linear effects.[6]

2.3 Adjusting the standard errors

Since our framework involves a generated regressor in the second stage estimation, the precision of our parameter estimates may be reduced by estimation uncertainty.[7] Traditional formulas for standard errors do not account for this type of uncertainty and therefore deliver an inconsistent

[4] We also ran tests with a median regression framework instead of the mean regression framework. We opted for the mean regression as our benchmark because if we were to use median regressions, fewer indicators would pass our test. This is because our candidate variables are especially informative for the tails of the distribution, and this is also reflected in the marginal effects at the mean, but less so at the median. Detailed results for the median framework are available upon request.

[5] Adrian et al. (2019) show that today's financial conditions provide valuable information for the 5% quantile of U.S. real GDP growth. The 5% quantile is also the quantile that is typically analyzed in the literature on value-at-risk. In robustness checks (results not reported for brevity), we also set $\tau = 10\%$ or $\tau = 25\%$. The results remain largely unchanged, but tend to become more similar to those from the mean regressions.

[6] We do not test whether the two coefficients are statistically different from each other because our framework does not allow for such a test. Our framework does not give us an estimate of the covariance of the two coefficients.

[7] Murphy and Topel (1985) show that this is the case if the error terms of the first and second stage estimation are independent.

estimate of the asymptotic covariance matrix of the parameters. To account for the estimation uncertainty, we derive a standard error correction from the general result of Murphy and Topel (1985) on two-stage maximum likelihood estimation with a generated regressor.

The maximum likelihood analysis of a logit model and of a mean regression model is well known and straightforward.[8] However, the analysis of a quantile regression is less common and more technically involved, in particular because the objective function ρ_τ is not twice continuously differentiable. Here we adopt the framework of Komunjer (2005), who provides generalized expressions for the first and second order conditions of the objective function, which can then be plugged into the formula for the standard error correction.

Furthermore, the forecast errors of the second stage may not be identically distributed, but rather exhibit some unknown form of conditional heteroskedasticity. This conjecture is motivated by the empirical evidence in Adrian et al. (2019), who argue that the conditional distribution of U.S. real GDP growth widens when financial conditions worsen. For this reason, we derive the standard error correction under the assumption that the error term is not identically distributed, i.e. the information matrix equality fails to hold. The derivation is based on the two-stage maximum likelihood framework discussed by Greene (2012). Under the assumption that the error term is misspecified, the parameter estimate of the second stage becomes a quasi maximum likelihood estimator (QMLE). In general, the parameter estimate of a QMLE is inconsistent. However, it is well known that for a mean regression model the QMLE is consistent under heteroskedasticity, and Komunjer (2005) proves consistency of the quantile estimator when the error term is misspecified.

Altogether, our standard error correction is summarized in the following theorem:

Theorem 1 (Asymptotic distribution of two-step quasi maximum likelihood) *Let the model consist of the two marginal distributions $f_1(y_1|x_1, \theta_1)$ and $f_2(y_2|x_1, x_2, \theta_1, \theta_2)$. The estimation proceeds in two steps:*

1. *Estimate θ_1 of model 1 by maximum likelihood: $L_1(\theta_1) = \Pi_{t=1}^T f_1(y_{1t}|x_{1t}, \theta_1)$.*
2. *Estimate θ_2 of model 2 by maximum likelihood, treating θ_1 as known by setting it to its estimate $\hat{\theta}_1$ from the first step: $L_2(\theta_1, \theta_2) = \Pi_{t=1}^T f_2(y_{2t}|x_{1t}, x_{2t}, \theta_1, \theta_2)$.*

If the standard regularity conditions for both log-likelihood functions hold and if the quasi maximum likelihood estimate of θ_2 is consistent, then the maximum likelihood estimate of θ_2 is asymptotically normally distributed with asymptotic covariance matrix

$$
\begin{aligned}
V_2 = {} & \frac{1}{T}(-H_{22}^{(2)})^{-1}\Sigma_{22}(-H_{22}^{(2)})^{-1} \\
& + \frac{1}{T}(-H_{22}^{(2)})^{-1}\Big(H_{21}^{(2)}(-H_{11}^{(1)})^{-1}H_{21}^{(2)\prime} + \Sigma_{21}(-H_{11}^{(1)})^{-1}H_{21}^{(2)\prime} + H_{21}^{(2)}(-H_{11}^{(1)})^{-1}\Sigma_{12}\Big)(-H_{22}^{(2)})^{-1}
\end{aligned}
\tag{4}
$$

where

$$
\begin{aligned}
\Sigma_{22} &= E\left[\frac{1}{T}\frac{\partial \ln L_2(\theta_1,\theta_2)}{\partial \theta_2}\frac{\partial \ln L_2(\theta_1,\theta_2)}{\partial \theta_2'}\right], \quad \Sigma_{21} = E\left[\frac{1}{T}\frac{\partial \ln L_2(\theta_1,\theta_2)}{\partial \theta_2}\frac{\partial \ln L_1(\theta_1)}{\partial \theta_1'}\right], \\
\Sigma_{12} &= E\left[\frac{1}{T}\frac{\partial \ln L_1(\theta_1)}{\partial \theta_1}\frac{\partial \ln L_2(\theta_1,\theta_2)}{\partial \theta_2'}\right], \quad H_{11}^{(1)} = E\left[\frac{1}{T}\frac{\partial^2 \ln L_1(\theta_1)}{\partial \theta_1 \partial \theta_1'}\right], \\
H_{22}^{(2)} &= E\left[\frac{1}{T}\frac{\partial^2 \ln L_2(\theta_1,\theta_2)}{\partial \theta_2 \partial \theta_2'}\right], \quad H_{21}^{(2)} = E\left[\frac{1}{T}\frac{\partial^2 \ln L_2(\theta_1,\theta_2)}{\partial \theta_2 \partial \theta_1'}\right].
\end{aligned}
$$

[8] See, for example, Greene (2012).

The first term in Equation (4) is the robust variance-covariance matrix of the second stage model; the other three terms reflect the correction for generated regressors. The second term captures the direct effect of estimation uncertainty, while the third and the fourth term account for the additional indirect estimation uncertainty through the correlation between the error terms of the two stages. Details on the sample estimator for V_2 as well as formulas for the derivatives of the various likelihood functions are given in Appendix A.

Finally, we acknowledge that the definition of systemic risk provided by the IMF, BIS, and FSB does not explicitly imply a hierarchical testing framework. We opt for a hierarchical testing framework rather than a joint testing framework because it closely approximates a joint testing framework – as we document in Section 5.3 – and, most importantly, given the non-standard distributional assumptions, is easy to implement.

3 Data

3.1 Candidate indicators of systemic risk

We select our set of candidate indicators based on their relevance in the literature (see, for example, Adrian and Shin (2008); Reinhart and Rogoff (2009); Geanakoplos (2010); Claessens, Kose, and Terrones (2012); Gilchrist and Zakrajšek (2012); Schularick and Taylor (2012); Jordà, Schularick, and Taylor (2015)) and on their availability for G7 countries. This set reflects a small sub-sample of all suggested variables, but it allows us to generate first insights on indicators of systemic risk.[9]

Specifically, in our benchmark analysis we include six variables that have been suggested for the monitoring of financial (in)stability. These variables measure the time series (or cyclical) dimension of systemic risk and should therefore quantify the build-up of systemic risk over time: credit growth, growth of prices of the major asset classes (stocks, bonds, and housing), the credit-to-GDP gap, and a composite financial cycle indicator that combines some of these variables.[10] Results for other transformations of the variables and a range of further indicators are presented in Section 5. Time series plots of our candidate indicators are presented in Appendix B.

Credit growth: Not least since the seminal study by Schularick and Taylor (2012), it is widely accepted that the occurrence of future financial crises is correlated with strong growth in credit aggregates, even though causality is far from obvious (see, for example, Gomes, Grotteria, and Wachter (2019)). The idea is that strong growth rates of credit indicate lending booms which go along with an increase in financial leverage and therefore create financial fragility. By construction, disruptions in credit supply – one measure of an impairment of the financial system (Romer and Romer (2017)) – are directly mirrored in the growth of credit. Therefore, we include credit growth as our first indicator of systemic risk.

More precisely, we use deflated total credit to the non-financial private sector, as this is the standard approach standard in the literature. We download the nominal series from the BIS webpage and transform levels to quarterly real growth rates.

[9]For a synopsis of systemic risk indicators proposed in the literature, please see Bisias et al. (2012) and Aikman et al. (2017). These studies can, however, only provide an initial overview, since the set of indicators is steadily increasing.

[10]We use the GDP implicit price deflator from the OECD Main Economic Indicators database to deflate the credit and asset price series.

House price growth: According to Reinhart and Rogoff (2009), almost all severe banking crises in advanced economies since World War II have involved imbalances in the housing market. The most prominent example of this is the development of the US housing market prior to the onset of the global financial crisis.

House prices are arguably an important indicator of financial fragility because a typical household balance sheet is very sensitive to house price movements. First, housing wealth typically makes up the largest fraction of households' total wealth, and, second, it is used as collateral for borrowing (Iacoviello (2005)). A potential mechanism for how this nexus between house prices and financial fragility arises is exemplified by the theory of leverage cycles as discussed by Geanakoplos (2010). A disruption in credit supply affects the affordability of houses and, through this, its prices: strong disruptions in credit supply lead to a fall in house prices. This in turn increases households' balance sheet leverage. Increased leverage implies a further deterioration of financing conditions, causing house prices to drop even further, and so on and so forth.

On this basis, we include house price growth measured through real residential property prices. We take the nominal series from the OECD Main Economic Indicators database, obtained through Haver Analytics. Similarly to credit growth, we transform the data into quarterly real growth rates.

Stock returns: A series of studies shows that stock returns may be associated with cycles in balance sheets as well. Therefore, it can be argued that they measure financial fragility (e.g. Adrian and Shin (2008); Mendoza (2010)). In view of the work by Miranda-Agrippino and Rey (2019) and Breitung and Eickmeier (2014), stock returns may also capture cross-country spillovers of disruptions to credit supply such as those that occurred during the last global financial crisis.

Empirically, Claessens et al. (2012) find that recessions associated with house and equity price busts tend to be longer than other recessions. Similarly, Jordà et al. (2015) also report that most build-ups of systemic risk since World War II have involved both equity and house prices.

We measure real stock returns via the quarterly real growth rates of a country's broad stock market index. The stock market indices are downloaded from the OECD Main Economic Indicators database.

Bond price growth: Gilchrist and Zakrajšek (2012), amongst others, argue that corporate bond spreads are an important indicator of the soundness of the financial system. Gilchrist, Yankov, and Zakrajšek (2009) find that unexpected increases in corporate credit spreads may be associated with large and persistent declines in real economic activity. Corporate spreads are claimed to capture the link between the quality of borrowers' balance sheets and their access to external finance. An impairment of the financial system should thus be reflected in corporate credit spreads.

In this paper, we use yields on bond indexes instead of high-frequency yields on individual bonds. This is because of a lack of available data for the G7 countries. Clearly, index yields suffer from a loss of information relative to individual yields. However, we argue that lending booms, i.e. longer-run developments on credit markets and sudden disruptions to it, should also be reflected in a corporate bond index series.

We download corporate bond yields from Global Financial Data.[11] We transform the corporate bond yields such that they reflect filtered bond price growth, in order to establish comparability

[11]For Canada, we use the long-term corporate bond yields obtained from the Bank of Canada up to 2006Q1 and the Bank of America Merrill Lynch Corporate Effective Yield for Canada from Haver Analytics thereafter. For Germany, we use yields on debt securities outstanding issued by residents obtained from the Bundesbank.

to the composite financial cycle indicator (see below).[12] Specifically, to be in line with the interpretation of house and equity price growth, the yields are transformed via $p_t = 1/(1 + y_t)$, where p and y denote the price and the yield, respectively. As we analyze bond indices, not all required information is available for a more precise transformation. Specifically, we assume that all yields are zero coupon yields and that the indices are derived from a portfolio of bonds with constant maturity.

Credit-to-GDP gap: The Basel III credit-to-GDP gap arguably represents the most prominent proxy for financial cycles. In Basel regulations (Basel Committee on Banking Supervision (2010)) it is set as the leading indicator for triggering the countercyclical capital buffer (CCyB), which is today legislated in 73 countries around the world. This prominent role is justified by its early warning properties, i.e. a capacity to indicate "imbalances" in the financial system ahead of a financial crisis for a broad set of countries.[13] If the credit-to-GDP gap exceeds a certain threshold, authorities are advised to consider activating (or increasing) the CCyB.

The credit-to-GDP gap is constructed from the credit-to-GDP ratio, where credit refers to total credit to the private non-financial sector. A rising ratio indicates that credit is expanding faster than the economy. Long-term trends in the credit-to-GDP ratio may be justifiable by fundamentals, such as financial innovations or demographic change. However, a substantial deviation of the ratio from its long-term trend, measured by the credit-to-GDP gap, is deemed to indicate excessive credit growth or leverage. Similarly to the indicators described above, a disruption to credit supply should lead to a decline in the credit-to-GDP gap.

We download the credit-to-GDP ratio data from the BIS webpage and construct the gap measure following the official procedure, i.e. we detrend the series via a one-sided Hodrick and Prescott (1997) filter with a smoothing parameter of 400,000. This smoothing parameter implies that cycles of durations exceeding about 40 years are identified as the long-term trend.

It is important to stress that due to the use of the HP filter, the credit-to-GDP gap has been criticized for weak real-time properties (Edge and Meisenzahl (2011)) and for an artificial periodicity which is related to the constant and large smoothing parameter (Schüler (2018, 2020)).[14] For this reason, we go on to include another proxy of financial cycles that circumvents this criticism.

Composite financial cycle: The composite financial cycle has been proposed by Schüler et al. (2020) and relates to the idea of leverage cycles as defined by Geanakoplos (2010). Specifically, the authors identify financial cycles via the balance sheet channel. One implication of theories involving credit market frictions is that the state of agents' balance sheets is an important determinant of their ability to borrow and lend. Changes in asset prices alter an agent's net worth, which influences the scale of borrowing and lending. Using this idea, the authors

For France, we take the average yields of first-class private bonds provided by the Banque de France. For Italy, the average corporate bond yield is provided by Banca d'Italia. For Japan, data up to October 2011 are from The Economist. From November 2011 on, we use the yield on Nomura Securities bonds. For the UK, data up October 2011 are taken from an index of corporate bond yields calculated by the Financial Times. From November 2011 on, the EIB 2028 bond is used. For the U.S., we use Moody's Corporate BAA yield, which includes bonds with time to maturity as close as possible to 30 years.

[12] In Section 5, we also present results for corporate credit spreads as potential indicators of systemic risk, which is closer to the idea of Gilchrist and Zakrajšek (2012), but does not fully capture the cyclical dimension of systemic risk.

[13] See, for instance, Detken, Weeken, Alessi, Bonfim, Boucinha, Castro, Frontczak, Giordana, Giese, and Jahn (2014) and references therein.

[14] Spurious periodicities can be problematic from a policy point of view, as they imply that the credit-to-GDP gap is chiefly determined by cycles up to 40 years. Schüler et al. (2020) show that this might be a reasonable assumption for the U.S. However, financial cycles typically differ strongly across countries (see also Hiebert, Jaccard, and Schüler (2018)).

measure financial cycles as the common expansions of credit and asset prices. These common expansions may lead to financial instability (see also Reinhart and Rogoff (2009)), similarly to the theory of leveraged asset price bubbles discussed by Jordà et al. (2015). The composite financial cycle may thus capture impairments of the financial system that suppress credit and asset prices simultaneously.

Schüler et al. (2020) measure credit – similarly to the credit-to-GDP gap – by total real credit to the non-financial private sector. The set of real asset prices includes house prices, equity prices, and corporate bond prices.

Methodologically, the indicator is constructed on the basis of the quarterly standardized growth rates of the above-mentioned components. Standardization is conducted using each variable's empirical cumulative distribution function, in order to align the different means and variances of the underlying indicators before aggregating them into a composite financial cycle. Growth rates have stable real-time properties and do not induce spurious periodicities, which is a known problem of the HP filter and thus of the credit-to-GDP gap.[15] Clearly, a disadvantage of quarterly growth rates relative to the HP filter is that quarterly growth rates are very volatile, thus potentially delivering imprecise signals. The authors address this issue by (i) aggregating up the different standardized variables, which mutes idiosyncratic movements and smooths the time series, and (ii) smoothing the aggregated index via a one-sided moving average.[16] Overall, the resulting composite financial cycle thus represents common fluctuations in credit and asset prices.

3.2 Periods of financial disruptions

We rely on the disruption variables provided by Romer and Romer (2017) for the left-hand side of the logit regressions in Stage 1.[17] Following a narrative approach, these authors identify periods of disruption to credit supply in a large panel of countries and cluster these disruptions by their severity into 15 different categories. Since our analysis requires a binary dummy variable, we pool the 15 different disruption categories into one. We choose this dummy as our benchmark because, due to its granularity, it also reveals many small disruptions which have not led to systemic financial crises, but can still be viewed as periods of distorted credit supply, arguably reflecting changes in systemic risk. The Romer and Romer (2017) data are available from 1973 to 2017, which determines our sample period throughout the paper.

A disadvantage of the Romer and Romer (2017) variable is its limited availability (half-year frequency), which requires us to transform candidate indicators to half-year frequency as well. Therefore, in Section 5, we present additional results using, amongst others, the quarterly crisis dates provided by Laeven and Valencia (2018). For this robustness check we use the same sample period as for the benchmark analysis to keep the results comparable. A more detailed comparison of these two and some other crisis dummy variables is also presented in Appendix B.

Finally, in Stage 2 of our testing framework we use real GDP growth as a measure for real economic activity. We obtain this series from the OECD National Quarterly Accounts.

[15] For the G7 economies, Schüler et al. (2020) find that the composite financial cycle significantly outperforms the credit-to-GDP gap in predicting financial crises and the "vulnerability periods" ahead of them.

[16] Aggregation is carried out using a time-varying linear combination of the standardized growth rates. The linear combinations take into account pairwise time-varying correlations between components, so as to emphasize the subset of variables that positively co-move the most strongly.

[17] The data are available on the webpage of the authors: `https://eml.berkeley.edu/~dromer/#data`

4 Main results

For the sake of clarity and readability, we summarize the majority of our results graphically. Figure 1 depicts our main findings. Each subplot illustrates the test results for one candidate variable at a time. We run our test for the G7 countries and for horizons up to three years ahead. For completeness, we also include results from contemporaneous regressions, i.e. we have $h = 0, \dots, 6$ in our semiannual framework.

The color code is as follows. A white square indicates that the candidate variable is insignificant in Stage 1 of our test, i.e. in the logit regression. A gray square means that the variable passes Stage 1, but not Stage 2 (neither for the mean nor for the quantile regression). If a candidate variable passes both stages (for the mean or the quantile regression or both), then the color of the square is determined by the sign of the sum of the slope coefficients in the first stage (red = positive, blue = negative). This distinction turns out to be crucial for the interpretation of our results, as will become clear below. The dark vs. light shading indicates whether we have $\delta_\tau < \delta$ in Stage 2, as outlined in Section 2.2. A dark shade implies that the candidate variable explains time variation in the tails of real GDP growth beyond movements in the location of the distribution. We highlight this type of asymmetry because it is one way to express the severity of the "negative consequences for the real economy". Finally, all charts that we present in this paper depict significance at the 10% level in the underlying tests.[18]

Figure 1 gives rise to a couple of interesting observations, the most important of which concerns the credit-to-GDP gap. This variable almost always passes Stage 1 of our test, but it does so with varying signs across countries, as indicated by the red and blue colors. This pattern impedes a coherent economic interpretation of the credit-to-GDP as an indicator of systemic risk. In Germany and Japan, the results imply that systemic risk is elevated when the level of the credit-to-GDP gap is low, whereas in France, Italy and the U.S., large systemic risk aligns with high levels of the credit-to-GDP gap. Admittedly, the credit-to-GDP gap passes Stage 2 of our test for up to three years ahead in France, Italy and the U.S., i.e. its predictive power is great, but this apparent success is undermined by the incoherent signs in Stage 1. A likely reason for the inconclusive signs across countries is related to the methodology used to construct the indicator: the one-sided Hodrick and Prescott (1997) filter. Schüler (2018, 2020) shows that the filter specification recommended under Basel III introduces artificial boom-bust cycles in the estimated gap. These spurious cycles likely mask country-specificities that have been pointed out by a series of other studies.[19] Therefore, other transformations of the ratio of credit to GDP, such as simple growth rates, might perform better.

The composite financial cycle results in a coherent interpretation for all countries except Canada. That is, combining credit growth with asset price growth aligns the signs of the coefficients in Stage 1. In five out of seven countries, a low value of the financial cycle indicates elevated systemic risk up to one year ahead, and even longer for the U.S. and Japan. The predominantly blue color of the plot seems surprising at first glance, given the popular narrative of "credit booms gone bust" (Schularick and Taylor (2012)), according to which excessive credit growth is a signal of elevated systemic risk. In contrast to the latter and other studies, we predict the whole period of a financial disruption, and not just its onset or a "vulnerability period" that precedes it. Clearly, elevated levels of credit growth may be observed prior to the onset of financial crises. However, our results are well in line with the impossibility of predicting turning points. Instead, our findings suggest that higher systemic risk goes hand in hand with lower

[18] Significance at the 10% level seems to be a weak requirement at first glance. However, due to our hierarchical testing framework, we already reject the null hypothesis "no systemic risk indicator" less often than in a simple one-stage test. We elaborate on the impact of the hierarchical structure in Section 5.3.

[19] See, for example, Galati, Hindrayanto, Koopman, and Vlekke (2016); Schüler et al. (2015, 2020); Hiebert et al. (2018); Rünstler and Vlekke (2018); Schüler (2018); Mandler and Scharnagl (2019); Strohsal, Proaño, and Wolters (2019).

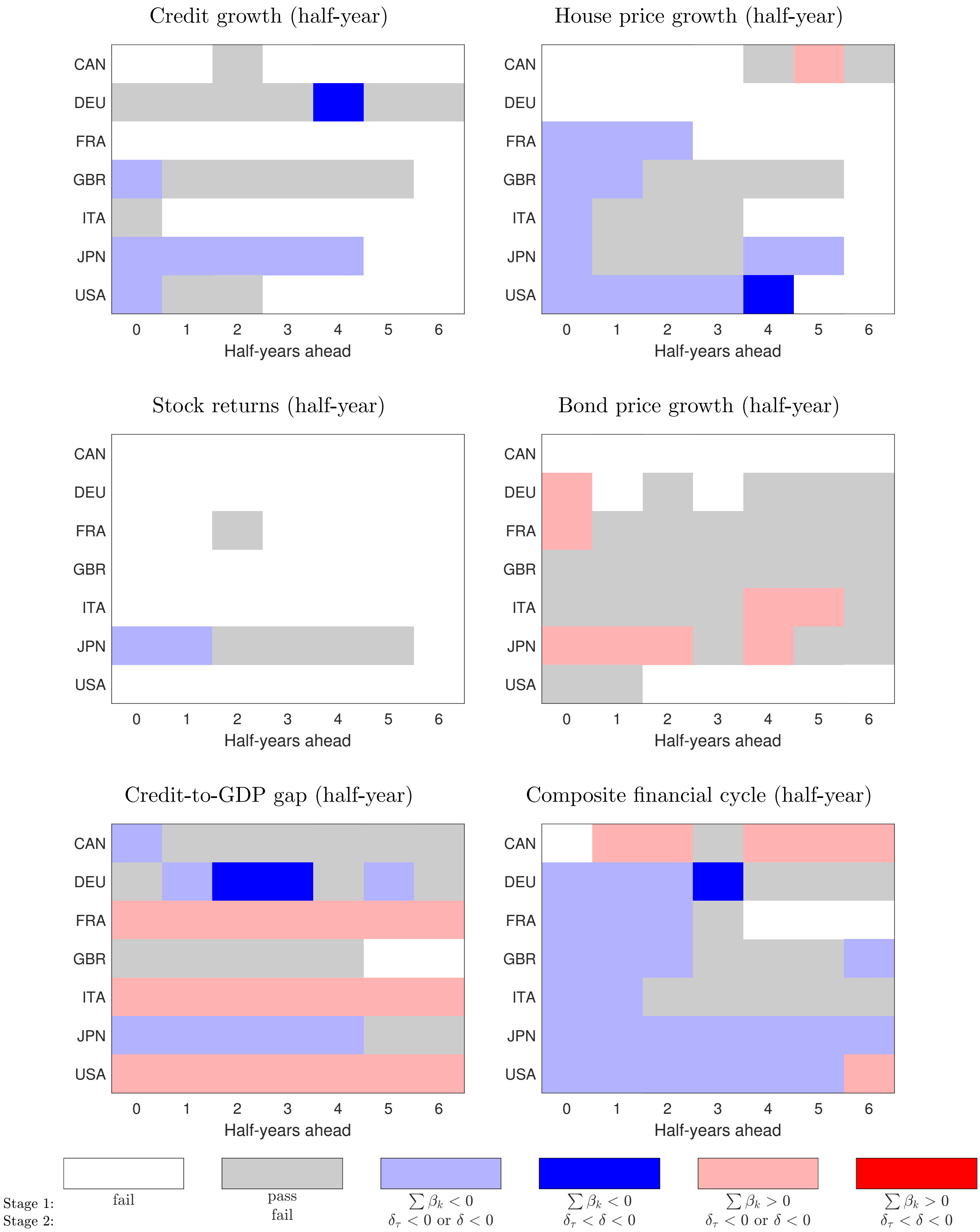

Figure 1: Two-stage regression results

Notes: The figure shows the results from our two-stage regressions in the form of a heatmap. White color indicates that the variable fails in Stage 1 of the test. Gray color indicates that the variable fails in Stage 2 of the test. The different shades of blue and red indicate whether Stage 2 is passed only for OLS or quantile regressions (light color) or for both OLS and quantile regressions (dark color), where blue (red) color means that the sum of the slope coefficients in Stage 1 is negative (positive). The dependent variable in Stage 1 is the half-yearly crisis dummy based on Romer and Romer (2017).

values of the financial cycle which occur after a boom period. Finally, the results for Canada have to be taken with a grain of salt. None of the candidate indicators qualify as an indicator of systemic risk for Canada. A potential explanation is that Canada exhibits the fewest disruption periods in our sample (see Appendix B). In fact, if the alternative Laeven and Valencia (2018) dummies analyzed in Section 5.1.1 are used, Canada does not have a single disruption period after 1973.

Notably, none of the single components of the financial cycle perform particularly well in Figure 1. Credit growth does not even pass Stage 1 of our test in most cases, which contradicts the suggestive conclusions drawn by Schularick and Taylor (2012). Corporate bond price growth does pass Stage 1 of our test, but the predicted disruption probabilities seem largely unrelated to subsequent GDP growth rates. If anything, house price growth works best, passing our test at least comtemporaneously in five out of seven countries. However, it seems somewhat tailored to the U.S., where it has predictive power up to two years ahead. The blue color again indicates that, if at all, elevated systemic risk accompanies low house price growth, such as in the period after a house price boom. In Germany, house price growth does not indicate systemic risk at all.

Finally, we emphasize that in the benchmark setup we do not see pronounced evidence of nonlinearities, indicated by the few dark shaded squares. That is, even though the composite financial cycle performs best, it does not predict particularly severe real economic consequences in the sense of additional movements of the tail of real GDP growth on top of movements of the center. However, this lack of nonlinearity does not automatically mean that the potential losses to real GDP upon a surge of the disruption probability are small, as we show next.

So far, we have relied on a purely graphical representation of our results, and we will continue doing so throughout the rest of the paper. However, to reinforce our main points, we present some of the exact statistical results upon which Figure 1 is based in Tables 1 and 2. Similar tables for the other candidate variables can be found in Appendix B.[20] We report the (sum of the) coefficients of Stage 1, the number of lags used in Stage 1 in brackets, as well as the mean and quantile coefficients from Stage 2. Significance of all slope coefficients is indicated by the familiar one, two, and three-star notation.

First of all, one can see that the mean and quantile coefficients in Stage 2 can in fact be very sizeable, below -30 at times. For instance, for the financial cycle in Germany, a one percentage point increase in the disruption probability lowers the 5% quantile of real GDP growth by an annualized 0.39 percentage point. Besides the impressive size of some coefficients, it can also be seen that the credit-to-GDP gap and the composite financial cycle pass Stage 1 of our test even at the 1% significance level in most cases. In other words, the link between these two variables and the disruption dummy variables of Romer and Romer (2017) is very strong across countries and forecast horizons. However, the issue with the incoherent signs of the slope coefficients in Stage 1 for the credit-to-GDP gap also becomes evident from the table. Notably, the negative slope coefficients for Germany and the positive slope coefficients for the U.S. are even of the same order of magnitude.

The tables also indicate that for both candidate variables the quantile coefficients are in fact much larger than the mean coefficient in many cases, most strikingly for Germany, but also for the U.S. However, the difference is not enough to justify dark shading for the respective square in the figure because the confidence bands around these coefficients (not reported in the table) overlap.

[20] Additional tables for the specifications and robustness checks presented later in the paper are omitted for the sake of brevity, but are available upon request.

Table 1: Credit-to-GDP gap (half-year)

Years ahead	Stage	Regression	CAN		DEU		FRA		GBR		ITA		JPN		USA	
0	1		-0.03***	[3]	-0.48***	[0]	0.18 ***	[0]	-0.07***	[1]	0.03***	[2]	-0.08***	[2]	0.05***	[1]
	2	Mean	-6.75*		-1.27		-6.1 **		-1.25		-3.43*		-3.02*		-4.75**	
		Quantile	-9.57***		-8.33		-1.7		-0.67		-4.22		-1.14		-8.3**	
0.5	1		-0.18***	[3]	-0.55***	[0]	0.2 ***	[0]	-0.04***	[1]	0.05***	[2]	-0.06***	[2]	0.31***	[4]
	2	Mean	0.32		-3.68*		-4.55 **		-0.17		-4.53**		-5.69**		-2.33*	
		Quantile	7.2		-13.61		-1.87		12.34		-9.58		-1.16		-6.95*	
1	1		-0.09**	[1]	-0.52***	[0]	0.23 ***	[0]	-0.03***	[1]	0.08***	[2]	-0.05***	[2]	0.31***	[3]
	2	Mean	-0.92		-4.53*		-6.33 ***		-0.65		-4.12**		-5.7**		-3.4**	
		Quantile	19.71		-17.85***		-11.12		16.99		-4.81*		9.07		2.88	
1.5	1		-0.16**	[1]	-0.46***	[0]	0.26 ***	[0]	-0.01**	[1]	0.11***	[2]	-0.04***	[2]	0.28***	[2]
	2	Mean	-1.02		-3.78		-7.02 ***		-2.8		-4.16**		-5.53**		-4.25***	
		Quantile	14.47		-20.94***		-23.96 *		-1.43		-5.72***		-2.9		-7.44	
2	1		-0.22**	[1]	-0.38***	[0]	0.28 ***	[0]	0*	[1]	0.16***	[3]	-0.02***	[2]	0.28***	[1]
	2	Mean	-3.07		-2.09		-7.3 ***		0.73		-6***		-5.84**		-3.66**	
		Quantile	-28.89		-24.89		-7.8 *		14.67		-11.47		-3.44		-5.19	
2.5	1		-0.23***	[0]	-0.3***	[0]	0.28 ***	[0]	0.02	[1]	0.16***	[2]	-0.01***	[2]	0.31***	[1]
	2	Mean	4.86		-0.09		-6.86 ***		1.2		-5.21***		-3.61		-4.33***	
		Quantile	14.72		-18.23*		-5.06 *		34.42		-3.7		2.82		-4	
3	1		-0.3***	[0]	-0.28***	[0]	0.27 ***	[0]	0.01	[0]	0.2***	[2]	0***	[2]	0.3***	[0]
	2	Mean	0.18		0.84		-6.11 ***		-9.75		-3.27**		-2.74		-4.8**	
		Quantile	-28.19		-23.54		-3.65 *		138.76		-4.47		17.05		-2.67	

Notes: For Stage 1 we report the sum of the slope coefficients of the candidate variable, the significance according to the likelihood ratio test, and the number of lags of the candidate variable (in brackets; as determined by the BIC). For Stage 2 we report (both for mean and quantile regressions) the slope coefficients and the significance according to the adjusted one-sided t-test. One, two, and three stars denote significance at the 10%, 5%, and 1% level respectively.

Table 2: Composite financial cycle (half-year)

Years ahead	Stage	Regression	CAN		DEU		FRA		GBR		ITA		JPN		USA	
0	1		-0.47	[0]	-7.56***	[0]	-6.88***	[0]	-19.98 ***	[0]	-9.52***	[0]	-17.16***	[0]	-6.69***	[0]
	2	Mean	-308.21		-13.42**		-4.1***		-13.29 ***		-4.65**		-4.42***		-10.27***	
		Quantile	-464.19		-38.14**		-7.57		-17.88 ***		-15.46**		-0.68		-14.02***	
0.5	1		6.95**	[2]	-9.74***	[0]	-6.49***	[0]	-19.41 ***	[0]	-8.37***	[0]	-16.72***	[0]	-6.16***	[0]
	2	Mean	-16.57**		-11.27***		-3.57**		-9.3 **		-3.74**		-3.88***		-10.02***	
		Quantile	-23.1*		-15.42*		-11.63		-18.78 ***		-7.79***		-0.65		-19.1***	
1	1		7.44**	[1]	-9.84***	[0]	-5.33***	[0]	-17.02 ***	[0]	-7.35***	[0]	-16.44***	[5]	-5.6***	[0]
	2	Mean	-13.79**		-15.09***		-5.35**		-5.69 **		-2.49		-2.98**		-11.21***	
		Quantile	-38.95*		-34.14***		-12.68		-4.31		-5.09		2.28		-21.52***	
1.5	1		7.82***	[0]	-11.08***	[0]	-3.61*	[0]	-15.02 ***	[0]	-6.84***	[0]	-9.72***	[0]	-3.89***	[1]
	2	Mean	-4.9		-10.48**		-3.99		0.78		0.52		-3.22**		-10.66***	
		Quantile	7.59		-39.34***		8.63		-2.43		2.13		-1.8		-22.83***	
2	1		9.89***	[0]	-11.34***	[0]	-2.19	[0]	-13.01 ***	[0]	-6.23***	[0]	-7.47***	[0]	-1.98***	[1]
	2	Mean	-12.61*		-3.88		1.62		3.72		2.56		-3.94**		-10.99**	
		Quantile	-43.72		-12.66		17.21		18.75		-1.38		-5.13		-20.69**	
2.5	1		9.56***	[0]	-11.57***	[0]	-0.57	[0]	-11.22 ***	[0]	-5.34***	[0]	-5.6***	[0]	-0.46*	[1]
	2	Mean	-19.45**		-0.94		15.11		1.09		1.87		-6.65***		-7.87*	
		Quantile	-28.91		-9.64		149.93		26.28		1.95		-7		-20.06	
3	1		7.69***	[0]	-11.59***	[0]	0.85	[0]	-8.99 ***	[0]	-4.34***	[0]	-4.35***	[0]	4.3***	[4]
	2	Mean	-19.03*		3.66		-19.08		-8.53 *		2.42		-7.44**		-2.85*	
		Quantile	13.78		-21.76		-116.22		-44.94 **		3.11		-2.63		3.47	

Notes: For Stage 1 we report the sum of the slope coefficients of the candidate variable, the significance according to the likelihood ratio test, and the number of lags of the candidate variable (in brackets; as determined by the BIC). For Stage 2 we report (both for mean and quantile regressions) the slope coefficients and the significance according to the adjusted one-sided t-test. One, two, and three stars denote significance at the 10%, 5%, and 1% level respectively.

5 Robustness checks

Throughout the process of designing and applying our testing framework, we have had to make a number of choices. We now present robustness checks for some of these choices.

5.1 Alternative dependent variables for hierarchical test

5.1.1 Stage 1: Alternative dummy variables for periods of financial disruptions

Besides the narrative approach of Romer and Romer (2017), several other authors have produced dummy variables for systemic banking crises or disruptions to the financial system. Most of these databases cover a wide range of countries and long histories of banking crises, but we restrict ourselves to the G7 countries and (if possible) to the sample from 1973 to 2017 for the sake of comparability.

First, we employ the crisis dummies of Laeven and Valencia (2018), which can be downloaded from the webpage of the IMF. The authors define a crisis as an event that meets two conditions: (i) significant signs of financial distress in the banking system, and (ii) significant banking policy intervention measures are implemented in response to the distress. The original time series are monthly. We aggregate the data up to quarterly frequency because most of our candidate variables are quarterly.

Second, Reinhart and Rogoff (2009) collect data on banking crises, currency crises, sovereign default, and inflation crises as well as major stock market crashes for a wide range of countries. We use the data that underlie Figure 16.2 in their book, and from the various dummies we select the dummies for banking crises marked as systemic. The data can be downloaded from the webpage of the authors. The time series are annual, ending in 2014.[21] To break down the annual time series into semiannual time series, we follow Romer and Romer (2017).

Third, we identify periods of financial disruptions via the European Systemic Risk Board (ESRB) dummies that define financial crises by combining a quantitative approach based on a financial stress index with expert judgement from national and European authorities. The ERSB offers data up to 2016 in its European financial crises database.[22] From this database we extract the "systemic crises" and disregard the "residual events". We aggregate the resulting time series, which are monthly, up to a quarterly frequency.

A graphical comparison of the different crisis dummy variables is presented in Appendix B. The dummy variables tend to agree on the strongest financial disruption - for Germany, France, the United Kingdom and the U.S., this is the global financial crisis. For Italy, all but the ESRB dummy mark the global financial crisis. For Japan, the dummies agree on the Asian financial crisis. Still, we observe differences – even for these major events – in terms of the start and/or length of the disruption periods. For instance, while Romer and Romer (2017) and the ESRB dummies indicate a protracted disruption for Germany in the aftermath of the global financial crisis, those of Laeven and Valencia (2018) or Reinhart and Rogoff (2009) measure a disruption for about half of that period. Therefore, we expect differences in the results.

The results for these alternative disruption dummies are depicted in Figures 2, 3, and 4. In all figures, missing data on disruption dummies are indicated by black color for the respective countries. Overall, our main results are robust to the use of alternative dummies. However, there are two notable exceptions: the inconclusive signs of the slope coefficients for the credit-to-GDP gap disappear to some extent, and for the composite financial cycle the signs of some slope coefficients flip for longer forecast horizons.

[21] `www.carmenreinhart.com/this-time-is-different`

[22] `www.esrb.europa.eu/pub/financial-crises` or Lo Duca, Koban, Basten, Bengtsson, Klause, Kusmierczyk, Lang, Detken, and Peltonen (2017).

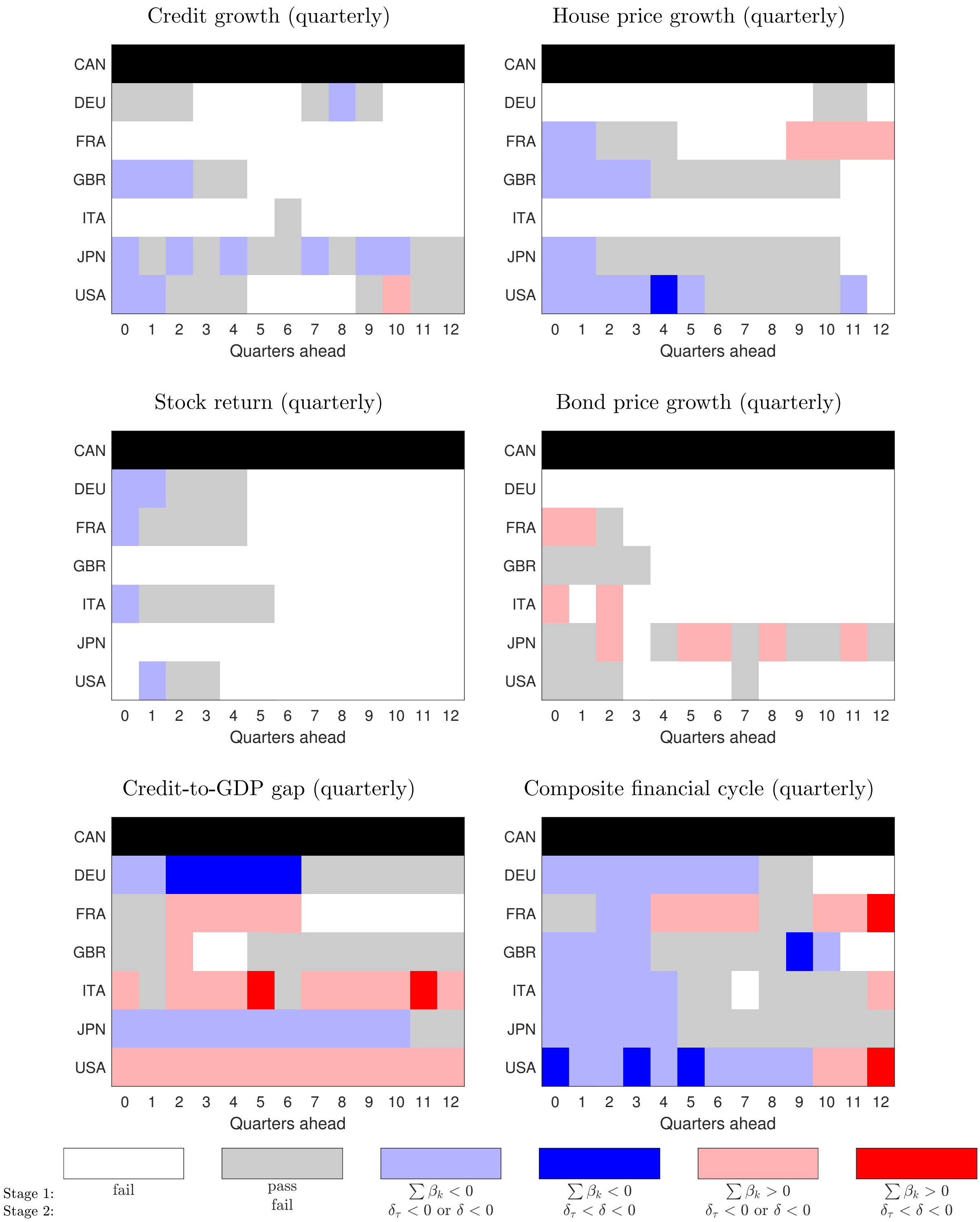

Figure 2: Laeven and Valencia (2018) dummy (quarterly) as a measure of financial disruptions

Notes: The figure shows the results from our two-stage regressions in the form of a heatmap. The color code is the same as in Figure 1. Results for Canada are colored black because the dataset contains no disruption observations for Canada.

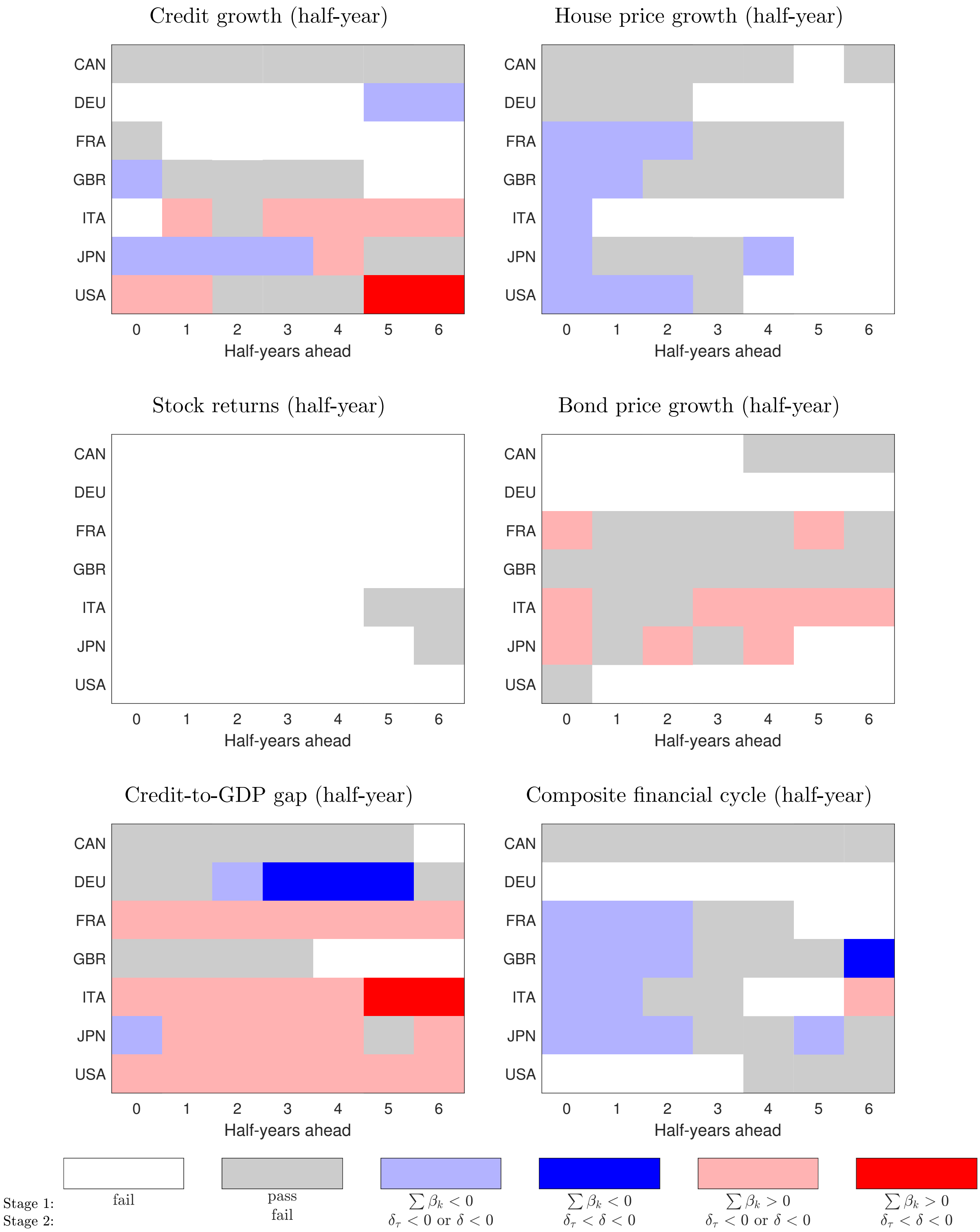

Figure 3: Reinhart and Rogoff (2009) dummy (half-year) as a measure of financial disruptions

Notes: The figure shows the results from our two-stage regressions in the form of a heatmap. The color code is the same as in Figure 1.

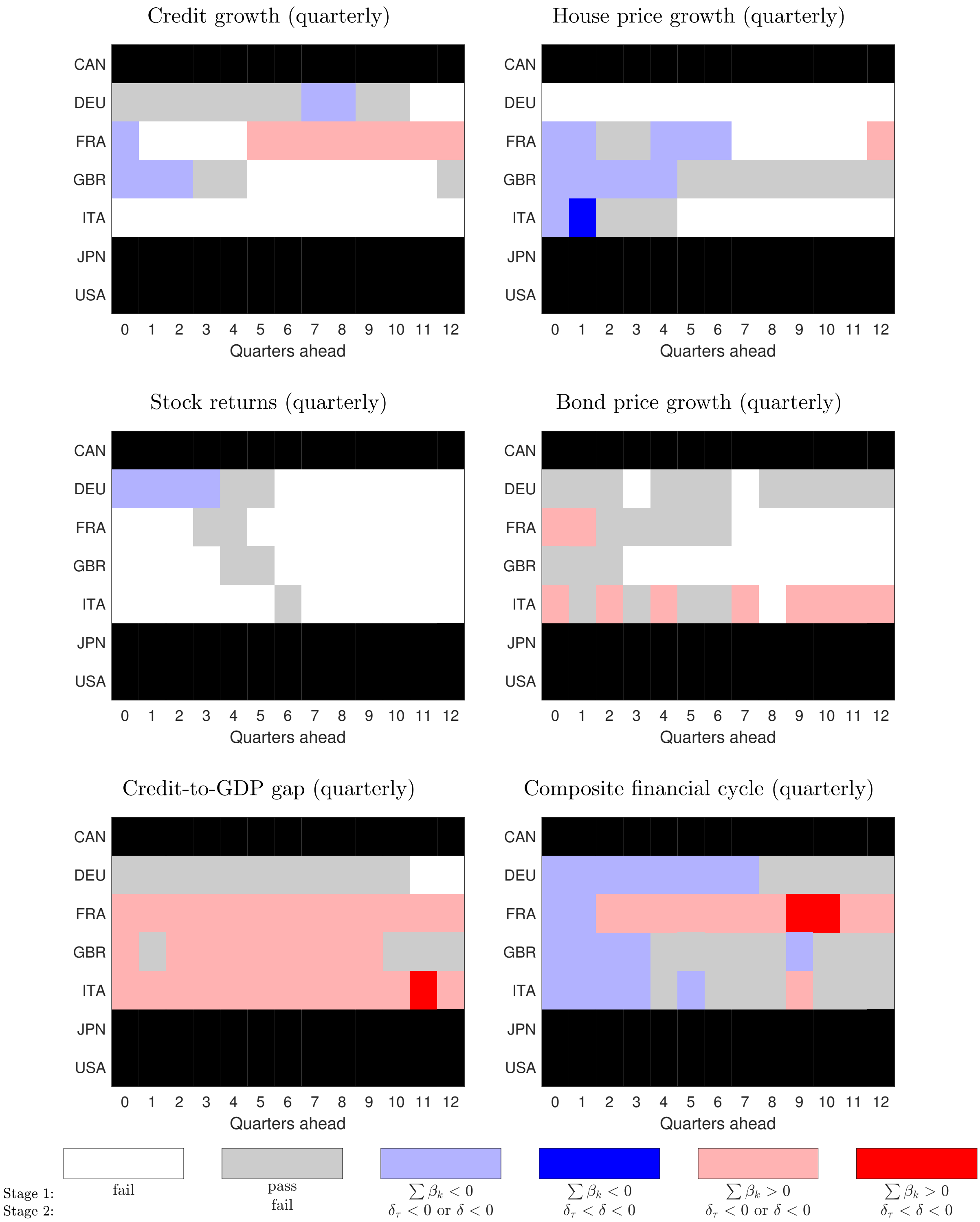

Figure 4: ESRB dummy (quarterly) as a measure of financial disruptions

Notes: The figure shows the results from our two-stage regressions in the form of a heatmap. The color code is the same as in Figure 1. Dummies are available for only four of the G7 countries. Black squares indicate missing data.

More precisely, for Laeven and Valencia (2018) the sign of the coefficient of the composite financial cycle in Stage 1 flips for France and Germany at longer forecast horizons.

With the Reinhart and Rogoff (2009) dummy, the composite financial cycles stops to predict crisis probabilities for Germany and the U.S. Yet, for a forecast horizon of up to 2.5 years ahead, the composite financial cycle still produces consistent signs. This is in contrast to the credit-to-GDP gap. Here, the results become slightly more favorable with the Reinhart and Rogoff (2009) dummies since Japan now becomes a "red country" for most periods. However, for Germany we still see the inconsistently negative sign of the slope coefficient in Stage 1.

With the quarterly ESRB dummy, greater changes appear. For the credit-to-GDP gap, the UK turns from a "gray country" to a "red country". Moreover, the puzzling inconclusive signs of the slope coefficients disappear for the credit-to-GDP gap in Germany. In turn, the composite financial cycle in France produces inconsistent signs, but again only for the later horizons. Importantly, the credit-to-GDP gap no longer indicates systemic risk in Germany. The composite financial cycle indicates systemic risk for all countries with the same sign at the short horizon. Finally, similarly to the main analysis, we do not see pronounced nonlinearities with any of the alternative disruption dummies.

5.1.2 Stage 1: Business cycle dummies

Since Stage 2 of our test involves modeling the conditional distribution of real GDP growth, another interesting exercise might be to use standard business cycle data in Stage 1 of our test as well. Figure 5 presents results for the quarterly business cycle peak-to-trough dummies from the webpage of the Economic Cycle Research Institute (ECRI).[23]

Interestingly, some of our candidate variables are also informative when it comes to the timing and the severity of booms and recessions. Both stock price and corporate bond price growth predict recessions, but house price growth and credit growth do not perform particularly well in this exercise. The composite financial cycle, which combines all four aforementioned variables, inherits some of the properties of its components and thus has explanatory power for business cycles. This points to the well-known tensions between financial cycles and business cycles. The fraction of business cycle peaks and troughs that is explained by the financial cycle in Stage 1 contributes largely to the overall variation in the center and the tails of real GDP growth. On the other hand, the results for the credit-to-GDP gap are inconclusive. This is intuitive. On the one hand, by construction, business cycle variation is masked by the method of constructing the credit-to-GDP gap (Schüler (2020)). On the other hand, credit growth alone also has a weak link to business cycles.

5.1.3 Stage 2: Alternative measures of economic activity

The definition of systemic risk presented above targets "serious negative consequences for the real economy". Obviously, aggregate GDP growth is just one of many variables that quantify economic activity. Therefore, we also run our test with two alternative dependent variables for Stage 2, namely the growth rate of industrial production and the aggregate unemployment rate. The data are again obtained from the OECD National Quarterly Accounts.

The results for both measures are depicted in Figures 6 and 7. The main takeaway is that our results are by and large robust to the use of alternative measures of real economic activity. For the growth rate of industrial production, the second stage delivers fewer significant coefficients as compared to the benchmark case with GDP growth (Figure 1). However, this finding is sensitive to the choice of the dummy variable for Stage 1. Figures 6 and 7 are based on the

[23] `www.businesscycle.com`

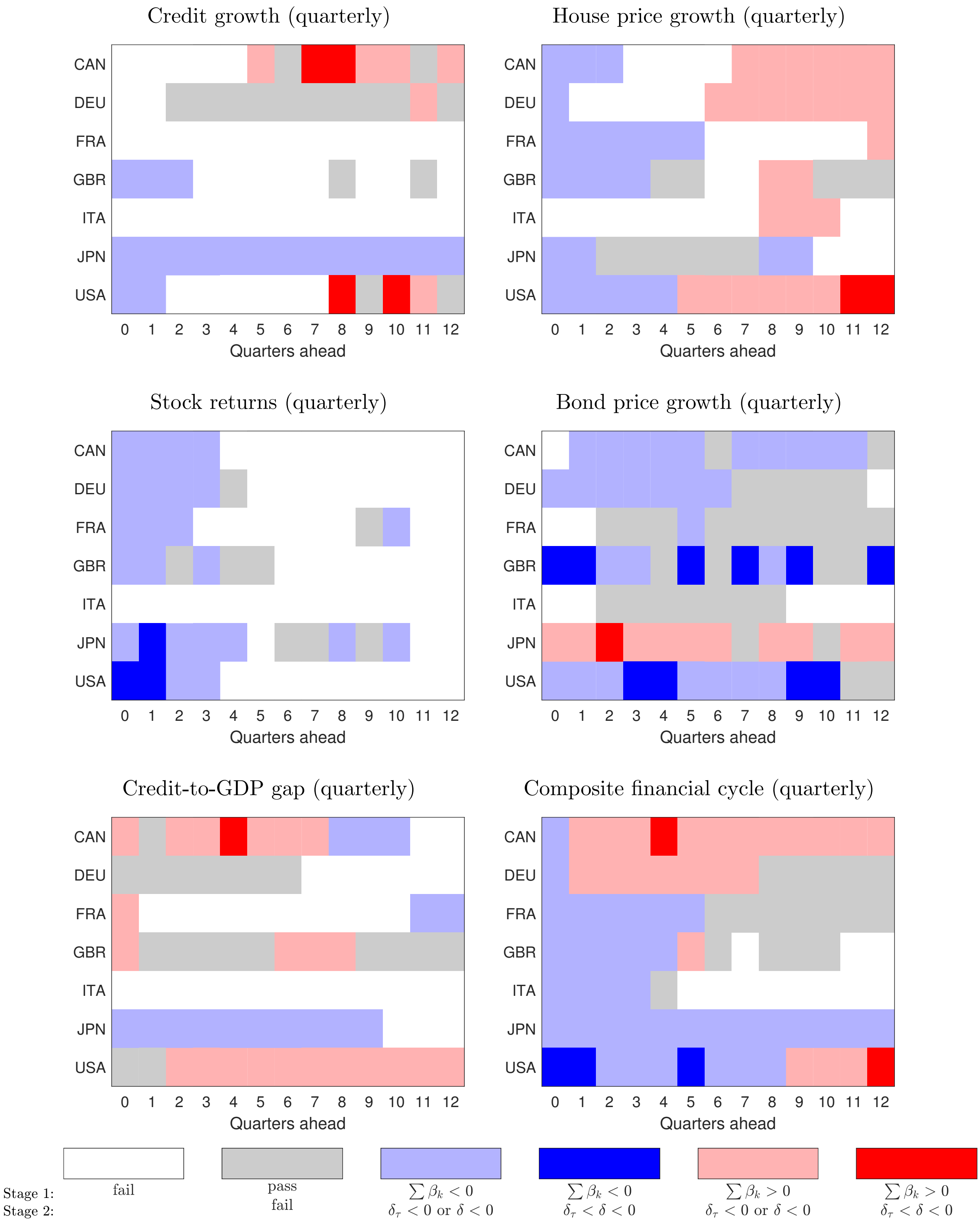

Figure 5: Peak-to-trough dummy (quarterly) as a measure of financial disruptions

Notes: The figure shows the results from our two-stage regressions in the form of a heatmap. The color code is the same as in Figure 1. The dependent variable in Stage 1 is the Economic Cycle Research Institute peak-to-trough recession dummies.

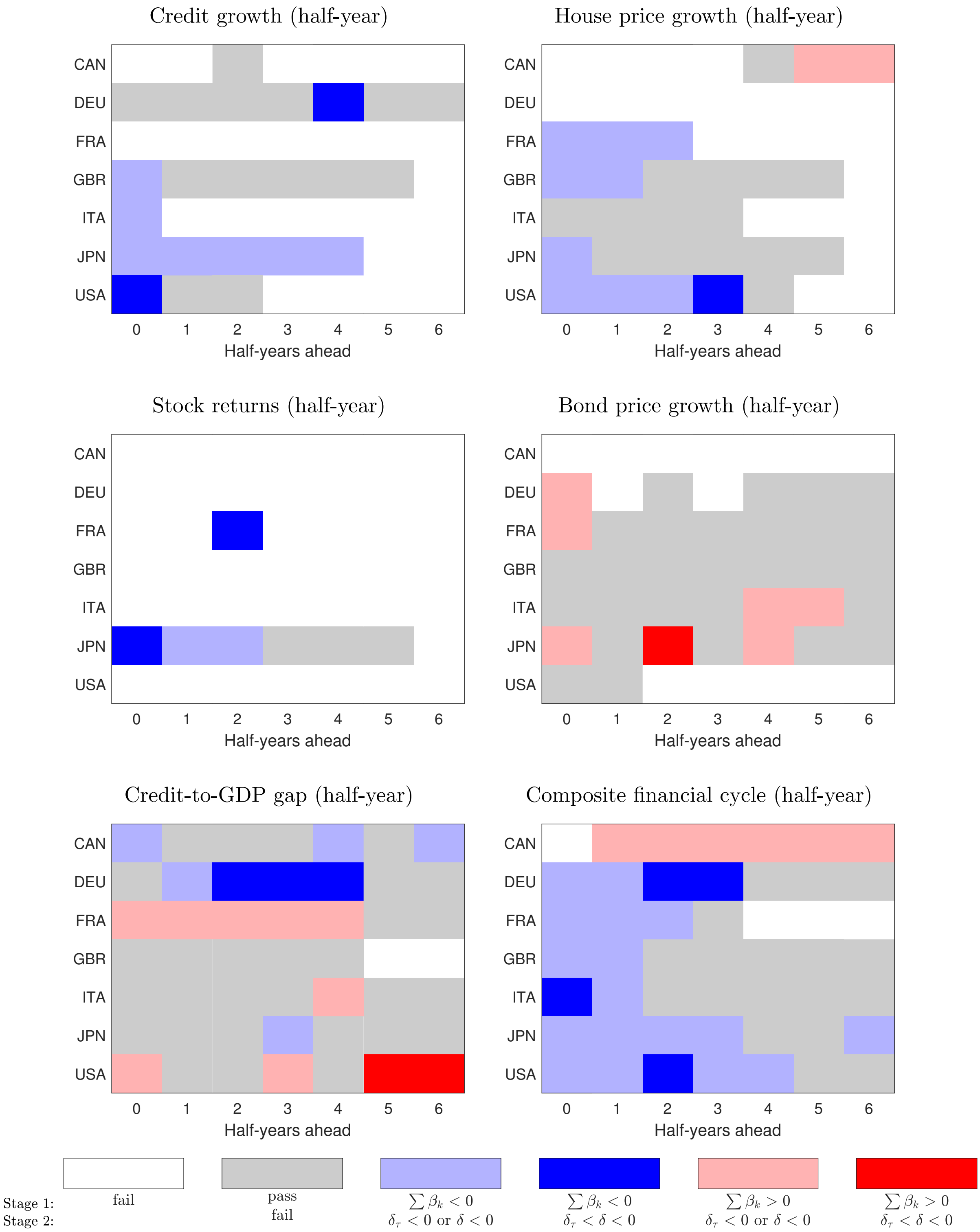

Figure 6: Industrial production (half-year) as a measure of real activity

Notes: The figure shows the results from our two-stage regressions in the form of a heatmap. The color code is the same as in Figure 1. The dependent variable in Stage 1 is the crisis dummies based on Romer and Romer (2017). The dependent variable in Stage 2 is the half-year growth rate of industrial production.

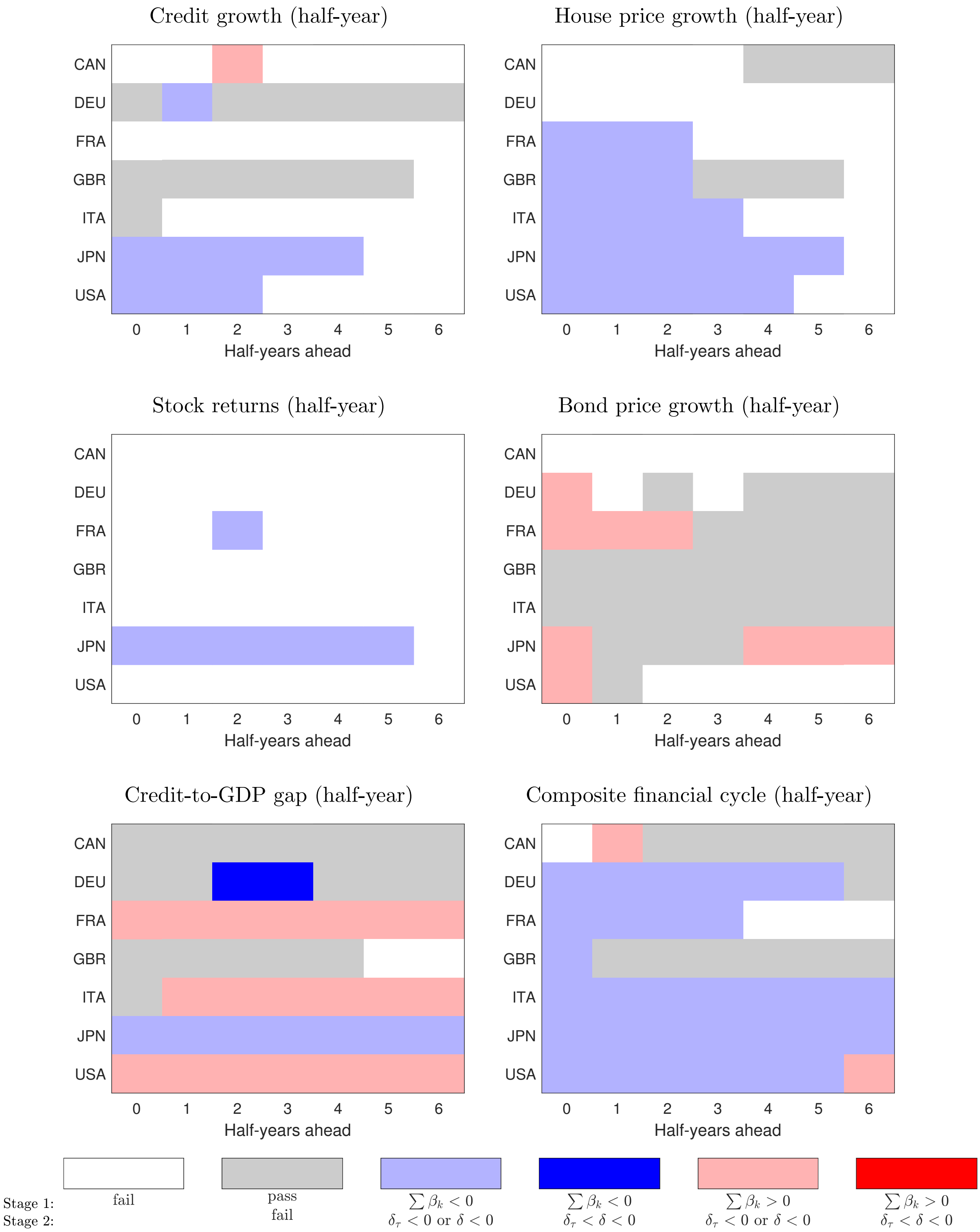

Figure 7: Unemployment rate (half-year) as a measure of real activity

Notes: The figure shows the results from our two-stage regressions in the form of a heatmap. The color code is the same as in Figure 1. The dependent variable in Stage 1 is the crisis dummies based on Romer and Romer (2017). The dependent variable in Stage 2 is the half-year unemployment rate.

Romer and Romer (2017) dummies. With Laeven and Valencia (2012) or Reinhart and Rogoff (2009) dummies (results not shown here in the interest of brevity), we get most of the significant coefficients that we have seen before. Regarding the unemployment rate, we do not observe any major changes to the benchmark results with GDP growth.

5.2 Further candidate indicators of systemic risk

5.2.1 Financial conditions indices

The universe of candidate variables to measure systemic risk naturally comprises more than the six variables we have presented above. In our benchmark setup, we studied measures for the time series dimension of systemic risk which are linked to the notion of financial cycles, as these variables feature prominently in the macroprudential policy debate. An alternative notion of systemic risk focuses on its cross-sectional dimension, capturing contagion and spillover effects.

For instance, on a weekly basis the Federal Reserve Bank of Chicago publishes the National Financial Conditions Index (NFCI) developed by Brave and Butters (2012). Adrian et al. (2019) choose this indicator to demonstrate that U.S. financial conditions affect the lower tail of GDP growth. The NFCI captures financial stress in traditional and newly developed financial markets as gauged by 105 different variables, relying on a dynamic factor model. The 105 variables describe credit risk, volatility, credit growth and leverage. Variables receiving high weights are, for instance, commercial paper spreads, interest rate swap spreads and the TED spread – expressing investors' perceptions of credit risk – and the VIX, which is often viewed as a measure of risk aversion and financial uncertainty. Furthermore, the indicator includes survey-based measures of economic conditions for consumers and businesses. The NFCI thus captures financial instability from various sources or mechanisms. All of these relate to contagion, spillovers, and interlinkages within the financial system.

In a similar fashion, the Country-Level Index of Financial Stress (CLIFS) was developed by Duprey, Klaus, and Peltonen (2017) as a monthly indicator of financial stress for European countries. It monitors three financial market segments: equity, bond, and foreign exchange. Stress in these markets is captured, first, by monthly realized volatilities and, second, by large downward spikes (as measured, for example, through the monthly cumulative loss in an index). The individual stress measures are then aggregated in such a way that time periods of high co-movement of sub-indices are emphasized. Thus, the indicator captures, similarly to the NFCI, spillovers and contagion within and across financial markets. While the NFCI has a much broader underlying dataset, the two indicators overlap in the monitoring of equity and bond markets.

Figure 8 depicts the results for these financial conditions indices (FCIs) with the four disruption dummies outlined above. We find that the FCIs, having been disciplined to forecast events of financial disruption, do not serve as indicators of systemic risk. This is in stark contrast to the direct link between financial conditions and the tails of GDP growth exemplified by Adrian et al. (2019). For the U.S., the indices already fall short in Stage 1 of our test, and for the other countries the results are again very incoherent. We conclude that FCIs have no predictive power for financial disruptions in time series regressions. The possibility to explain movements in tail risk of real GDP growth, which is strongly advocated by Adrian et al. (2019), becomes insignificant in our two-stage hierarchical testing framework derived from the definition provided by the IMF, BIS, and FSB.

Although our framework does not allow for causal interpretations, the insignificance does not come as a surprise, since financial conditions are known to be closely linked to the business cycle rather than to longer-term movements of the financial cycle. From the two right-hand plots (with Laeven and Valencia (2018) as well as ESRB dummies), one may tentatively conclude

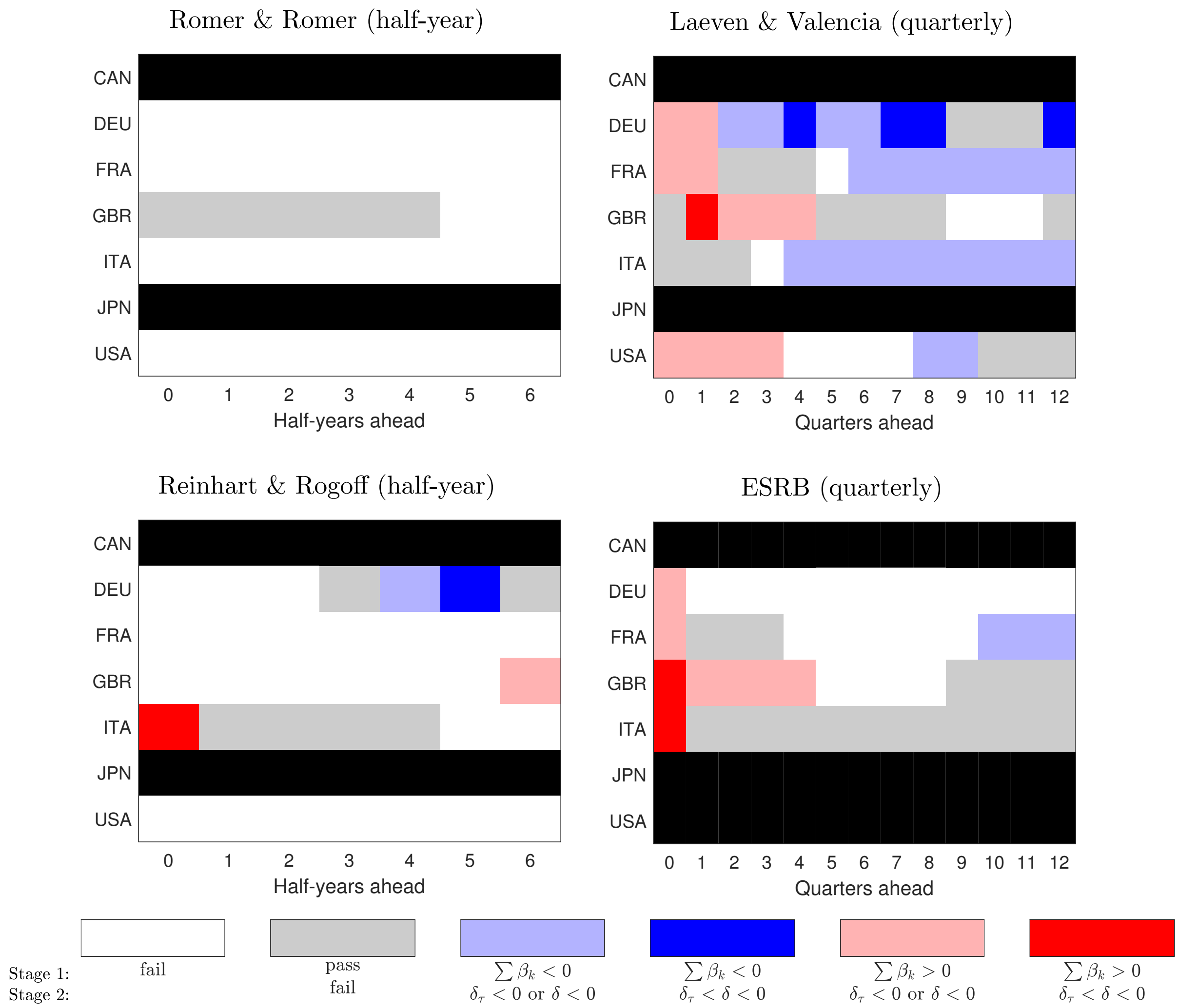

Figure 8: Financial conditions indices (half-year and quarterly) as a candidate indicator of systemic risk

Notes: The figure shows the results from our two-stage regression in the form of a heatmap. The color code is the same as in Figure 1. The independent variables in this figure are various financial conditions indices. The dependent variable in Stage 1 is the crisis dummies based on Romer and Romer (2017) (top left), Laeven and Valencia (2018) (top right), Reinhart and Rogoff (2009) (bottom left) and the ESRB (Lo Duca et al. (2017); bottom right).

that FCIs work in *contemporaneous* regressions, i.e. they may "set off the alarm" once the financial system is disrupted and real GDP growth is already declining. However, they fall short of indicating systemic risk in advance, which is the thematic core of our paper.

5.2.2 Business cycle indicators

Since Stage 2 involves regressions of real GDP growth on some predictor variables, a concern may be that standard business cycle indicators, which are essentially unrelated to systemic risk, may also pass our test. Therefore, as a "placebo test", we run our procedure with the term spread, one of many established business cycle measures.

We calculate the term spread as the difference between long-term and short-term interest rates, which are mainly taken from the OECD Main Economic Indicators database.[24]

[24] More precisely, as short-term rates we use: the 3-month prime corporate rate for Canada; the 3-month

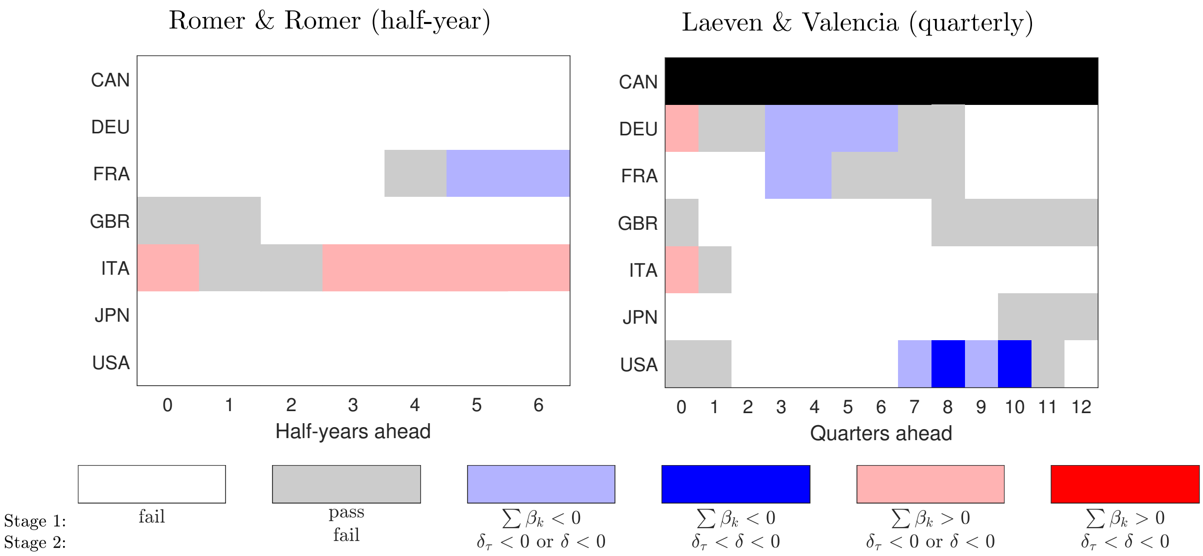

Figure 9: Term spread (half-year and quarterly) as a candidate indicator of systemic risk

Notes: The figure shows the results from our two-stage regressions in the form of a heatmap. The color code is the same as in Figure 1. The independent variable in this figure is the term spread. The dependent variable in Stage 1 is the crisis dummy based on Romer and Romer (2017) (left) and Laeven and Valencia (2018) (right).

Figure 9 depicts the results. Consistent with our theory, the term spread largely fails to predict periods of financial disruptions, as indicated by the many white squares. If it does indeed explain financial disruptions, then the predicted probabilities no longer explain time variation in means or quantiles of GDP growth, as indicated by the remaining gray squares.

5.2.3 Corporate credit spreads

In our benchmark analysis, we apply our hierarchical testing framework to corporate bond price growth. For this exercise, we transform corporate bond index yields to growth rates of prices in order to establish comparability to the composite financial cycle indicator.

Gilchrist and Zakrajšek (2012), amongst others, instead argue that corporate bond credit spreads may be suited to indicating systemic risk. Therefore, as a robustness check, we also apply our framework to credit spreads.

We retain the data that we use to measure bond price growth in Section 3. We define credit spreads as the difference between these corporate bond yields and the long-term Treasury yields used in Section 5.2.2. The results from our test are depicted in Figure 10.

The findings are somewhat mixed. In general, credit spreads perform better in our test than bond price growth, but worse than house price growth (see e.g. Figure 1). Our key results are therefore robust to the use of spreads: a composite measure of the financial cycle is more informative about systemic risk than measures constructed from the individual components.

FIBOR, PIBOR and TIBOR for Germany, France and Japan; the 3-month interbank loan rate for the UK; the average rate on all Treasury bills for Italy; and the 3-month core deposit rate for the U.S. Due to lack of data, we amend these series with 3-month Sterling interbank rates from the Bank of England (up to 1978), discount rates from the Bank of Italy (up to 1977), and discount rates from the Bank of Japan (up to 2002). For long-term rates we use the "long-term rates" data in that database for Canada, Germany, France, UK and U.S.. For Italy we use data from the Bank of Italy and for Japan we take compound bond yields over 10 years from the Japan Securities Dealers Association.

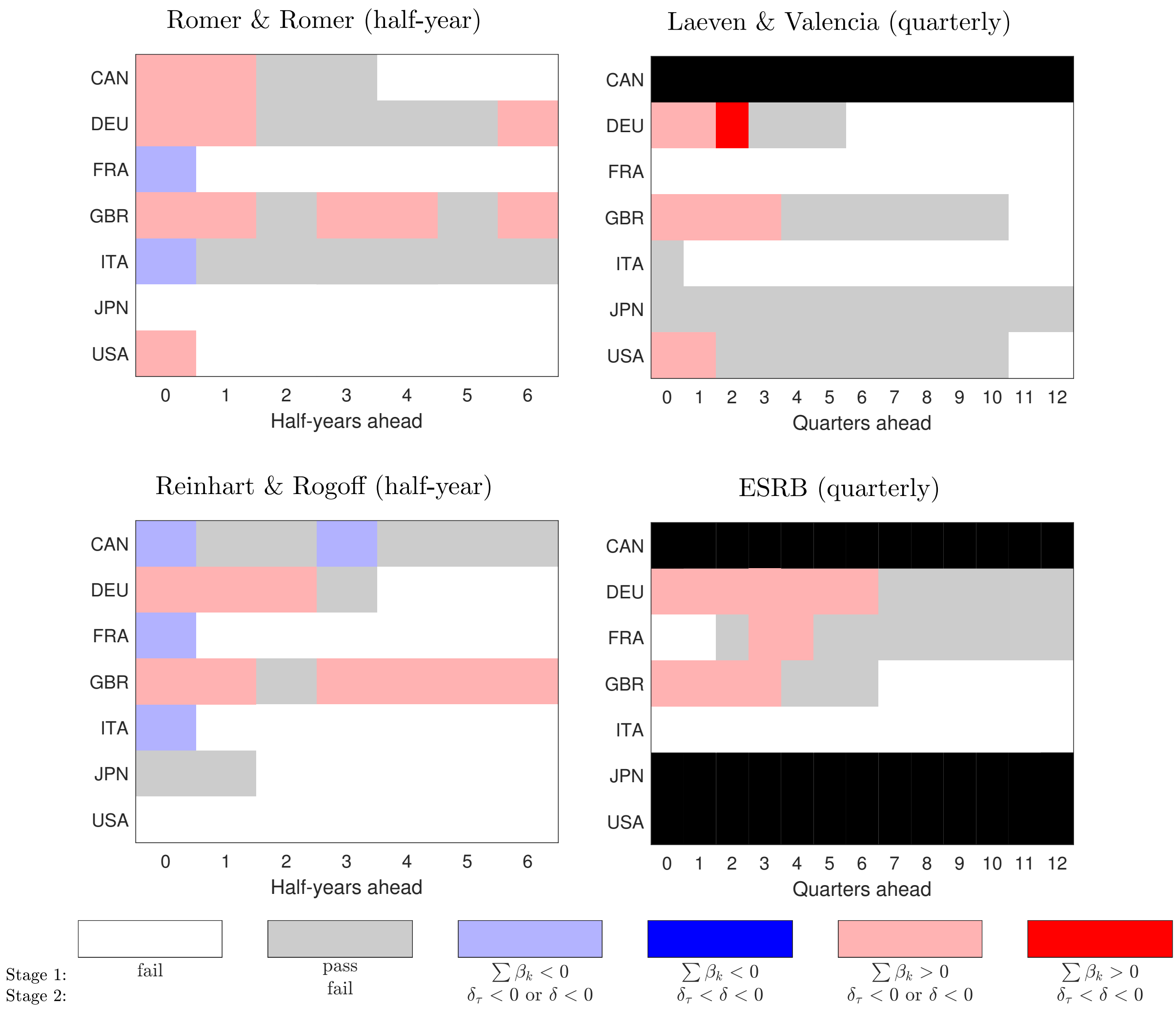

Figure 10: Credit spread (half-year and quarterly) as a candidate indicator of systemic risk

Notes: The figure shows the results from our two-stage regression in the form of a heatmap. The color code is the same as in Figure 1. The independent variables in this figure are various financial conditions indices. The dependent variable in Stage 1 is the crisis dummies based on Romer and Romer (2017) (top left), Laeven and Valencia (2018) (top right), Reinhart and Rogoff (2009) (bottom left) and the ESRB (Lo Duca et al. (2017); bottom right).

Conditional on passing our testing procedure, the predominant color in the figure is red, indicating that high credit spreads signal high systemic risk. This is consistent with the arguments of Gilchrist and Zakrajšek (2012). However, similarly to the credit-to-GDP gap, we also see inconclusive signs of the slope coefficients for credit spreads: the respective squares for France and Italy (and also Canada with the Reinhart and Rogoff (2009) dummies) are blue.

Overall, the credit spread tends to pass the test only contemporaneously or for very short horizons. A notable exception is the United Kingdom, where the credit spread signals systemic risk up to three years ahead. Finally, the performance of the credit spread in the U.S. is very weak. This may support the argument of Gilchrist and Zakrajšek (2012) that their micro-founded GZ spread is better suited for measuring systemic risk than spreads based on broad corporate bond indices, which is what we are using here.

5.2.4 Long-term growth rates of credit, house prices and stock prices

The benchmark setup relies on half-year growth rates of credit and asset prices. We opt for this because it is the natural first choice in our semiannual framework. However, longer-term growth rates of these variables exhibit superior early warning properties (see, for instance, Behn, Detken, Peltonen, and Schudel (2017)). Therefore, as another robustness check, we present results for such longer-term growth rates in Figure 11.

The changes compared to the benchmark results in Figure 1 are minimal, but, particularly with house price growth, we see some of the squares turning red. This indicates that persistent positive growth in house prices may signal elevated systemic risk up to three years ahead, but only for three out of seven countries. For the other candidate variables, none of our major results change.

5.3 Econometric procedure

Our econometric procedure comprises two main elements: (i) the hierarchical structure of the test – the indicator may only advance to Stage 2 if it passes Stage 1 – and (ii) the standard error correction. In the following we analyze the impact of both elements on our main results and shed some light on finite sample issues.

5.3.1 Impact of standard error correction

We start by presenting evidence on the impact of the standard error correction in Table 3. Here we stick to one special case from our benchmark analysis, namely Germany with a forecast horizon of one year ahead and with the Romer and Romer (2017) dummy. We choose this case because it features one of the few nonlinearities in Figure 1.

The table reports for each candidate variable the significance level in Stage 1, the slope coefficients in Stage 2 (both for mean and quantile regressions), the standard errors with and without the correction in Theorem 1, and 80% confidence intervals around the point estimates, based on these standard errors.

The standard error correction has a small but decisive impact on some results. For instance, for the composite financial cycle, the confidence intervals for mean and quantile regression overlap with corrected standard errors, but they do not overlap with non-corrected standard errors. Hence, the respective square in Figure 1 is shaded light blue, whereas the color would be dark blue with non-corrected standard errors.

Moreover, the table also shows that the standard error correction can be extremely large, as is the case for house price growth. Here the logit regression in Stage 1 is heavily misspecified, reinforcing the finding that house price growth has no predictive power for financial disruption periods in Germany.

In rare cases (for instance, for credit growth), the corrected standard errors are smaller than the non-corrected standard errors. This can happen if the term Σ_{21} in formula (4), reflecting the correlation between the errors from Stage 1 and Stage 2, is negative and dominates the other terms. We elaborate on this issue in Section 5.3.3 below.

5.3.2 Impact of hierarchical testing framework

Figures 12 and 13 present the same results as Figure 1, except for one change: we highlight in green all cases where the candidate variable does not pass Stage 1 of the test, but would pass

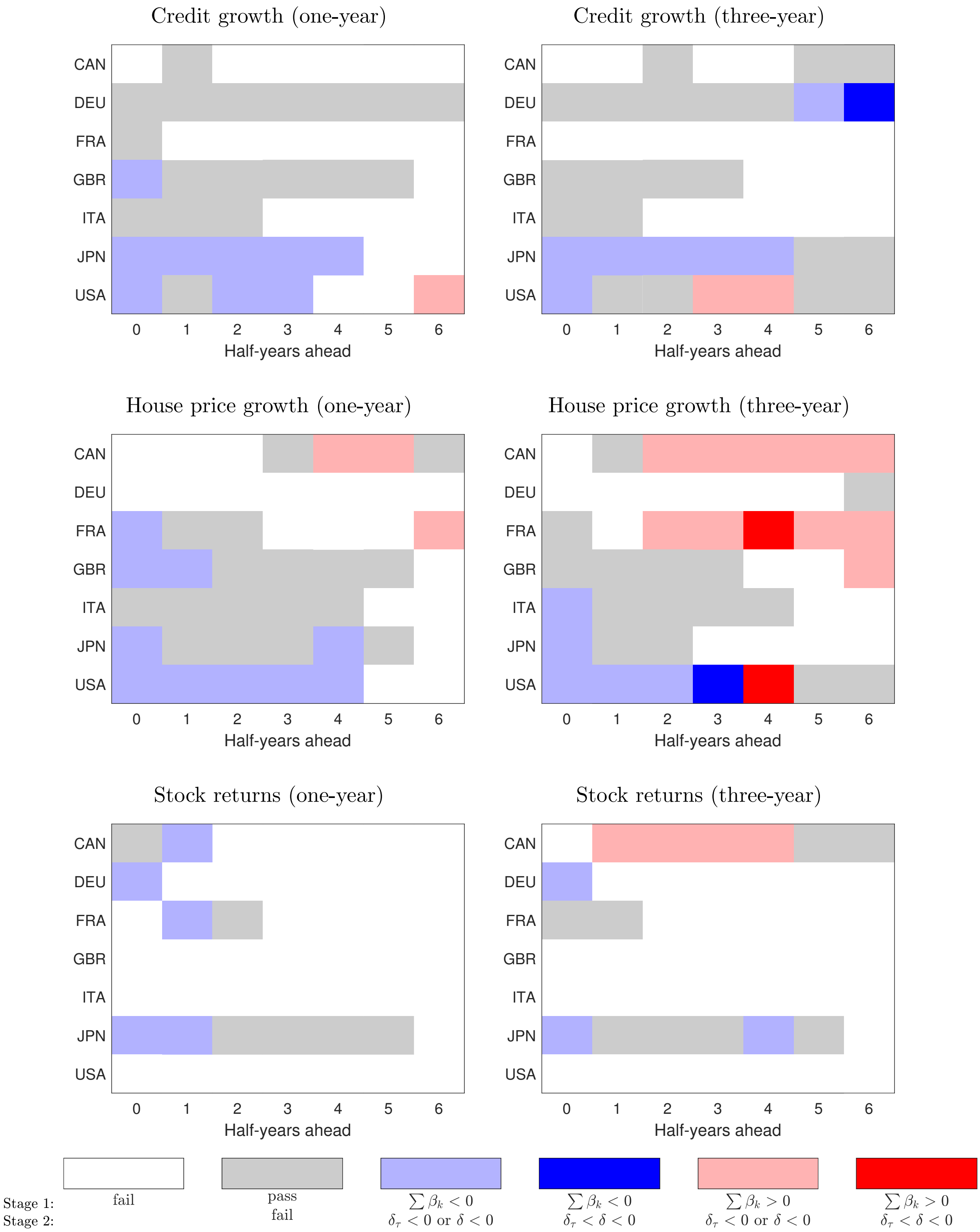

Figure 11: Longer-term growth rates of credit and asset prices as a candidate indicator of systemic risk

Notes: The figure shows the results from our two-stage regression in the form of a heatmap. The color code is the same as in Figure 1. The independent variables in this figure are one-year and three-year of indicated variable. The dependent variable in Stage 1 is the crisis dummy based on Romer and Romer (2017).

Table 3: Example of the impact of our standard error correction

Variable (half-year)	Regression	Correction	Stage 1	Stage 2 $\hat{\delta}_{(\tau)}$	Stage 2 Std.error	80% conf. interval
Credit growth	Mean	no	***	-1.51	2.52	[-4.77 , 1.75]
		yes			2.37	[-4.57 , 1.55]
	Quantile	no		-16.68	18.05	[-40.01, 6.65]
		yes			18.44	[-40.51 , 7.15]
House price growth	Mean	no		14.92	45.39	[-43.75 , 73.59]
		yes			83.19	[-92.6 , 122.44]
	Quantile	no		118.48	81.47	[13.18, 223.78]
		yes			492.71	[-518.35 , 755.31]
Stock return	Mean	no		-21.52	16.89	[-43.35 , 0.31]
		yes			34.54	[-66.16 , 23.12]
	Quantile	no		-43.84	52.51	[-111.71 , 24.03]
		yes			69.14	[-133.2 , 45.52]
Bond price growth	Mean	no	**	-1.12	6.46	[-9.47 , 7.23]
		yes			6.34	[-9.31 , 7.07]
	Quantile	no		1.12	8.55	[-9.93 , 12.17]
		yes			8.6	[-10 , 12.24]
Credit-to-GDP gap	Mean	no	***	-4.53	3.07*	[-8.5 , -0.56]
		yes			3.03*	[-8.45 , -0.61]
	Quantile	no		-17.85	6.26***	[-25.94 , -9.76]
		yes			6.31***	[-26.01 , -9.69]
Financial cycle	Mean	no	***	-15.09	5.02***	[-21.58 , -8.6]
		yes			5.03***	[-21.59 , -8.59]
	Quantile	no		-34.14	9.54***	[-46.47 , -21.81]
		yes			10.93***	[-48.27 , -20.01]
FCI	Mean	no		19.02	16.86	[-2.77 , 40.81]
		yes			30.97	[-21.01 , 59.05]
	Quantile	no		42.23	28.66	[5.19 , 79.27]
		yes			77.24	[-57.6 , 142.06]

Notes: The table displays the impact of the standard error correction on our results. As an example the table reports the results of Germany with a forecast horizon of one year and the Romer and Romer (2017) dummy in Stage 1. For each candidate variable we report the significance in Stage 1 (based on a likelihood test), the slope coefficient in Stage 2 (both for mean and quantile regressions), the standard error without our correction (labeled as "no"), the standard error with the correction outlined in Theorem 1 (labeled as "yes"), and the resulting 80% confidence interval around the point estimate. One, two, and three stars denote significance at the 10%, 5%, and 1% significance level respectively.

Stage 2 in a non-hierarchical test. Figure 12 depicts the results without, Figure 13 with the standard error correction approach of Theorem 1.

The plots allow us to assess the impact of the hierarchical structure of our test on our key results. If the standard error correction alone suffices to eliminate "false positives" after Stage 1, then all green squares in Figure 12 should turn white in Figure 13. Apparently, this is not the case, i.e. we still have to rule out the remaining false positives with our hierarchical structure.

This result is important, as one might (erroneously) interpret our hierarchical test using the familiar notion of "necessary" and "sufficient" conditions. Passing Stage 1 would then be a necessary condition, in the sense that we do not allow a candidate variable to pass the whole test if it already fails in Stage 1. Given the standard error correction, the test in Stage 2 could then be thought sufficient to determine whether a candidate variable serves as an indicator of systemic risk. However, the graphical results show that this analogy cannot be drawn here.

Together with the results from Table 3, we therefore conclude that, in order to bear full fruit, our test requires the combination of the two key features (hierarchical structure and standard error correction). However, we emphasize that this combination is not only justified econometrically, but is also in line with our interpretation of the definition of systemic risk outlined in Section 2.1.

5.3.3 Finite sample issues

In the benchmark setup, we do not impose any assumption on the dependence structure of the error terms from Stage 1 and Stage 2, i.e. they can in principle be negatively correlated. In this case, the negative term $\hat{\Sigma}_{21}$ in Equation (4) would reduce the corrected standard errors in Stage 2. The standard errors for credit growth reported in Table 3 are an example of this.

For quantile regressions in small samples, the negative correlation term may occasionally lead to problems because the covariance matrix $\hat{V}_2$ is no longer positive definite. In extensive robustness checks (not reported here) where we switch this correlation term off, we have verified that the described effect occurs very rarely and does not alter any of our major results. Figure 14 in Appendix B depicts one example of such a robustness check, based on the benchmark setup in Section 4. The changes are very small. For instance, for Germany two years ahead, the original gray square for the credit-to-GDP gap is most likely the result of such a numerical instability. On the other hand, for Canada and the composite financial cycle, we see fewer significant results and fewer red squares when neglecting the correlation term, which strengthens our original results.

6 Conclusion

Starting from the definition of systemic risk given by the IMF, BIS, and FSB, we deduce testable hypotheses that allow us to assess whether a candidate variable can serve as an indicator of systemic risk. We then derive a two-stage hierarchical testing procedure that we apply to a set of candidate indicators and data from the G7 countries. The test framework proposed in this paper can be readily applied to further candidate variables or other countries.

Our results provide guidance on which indicators might be preferable over others for a thorough calibration of countercylical macroprudential policy tools. Most importantly, we find that the Basel III credit-to-GDP gap does not indicate systemic risk coherently across G7 countries. It does pass our test for various horizons and countries, but the signs of the coefficients give inconclusive guidance on whether a high or low level of the credit-to-GDP gap implies high systemic risk. A composite financial cycle measure, computed along the lines of Schüler et al. (2020), that combines information from credit growth with the price growth of the major asset

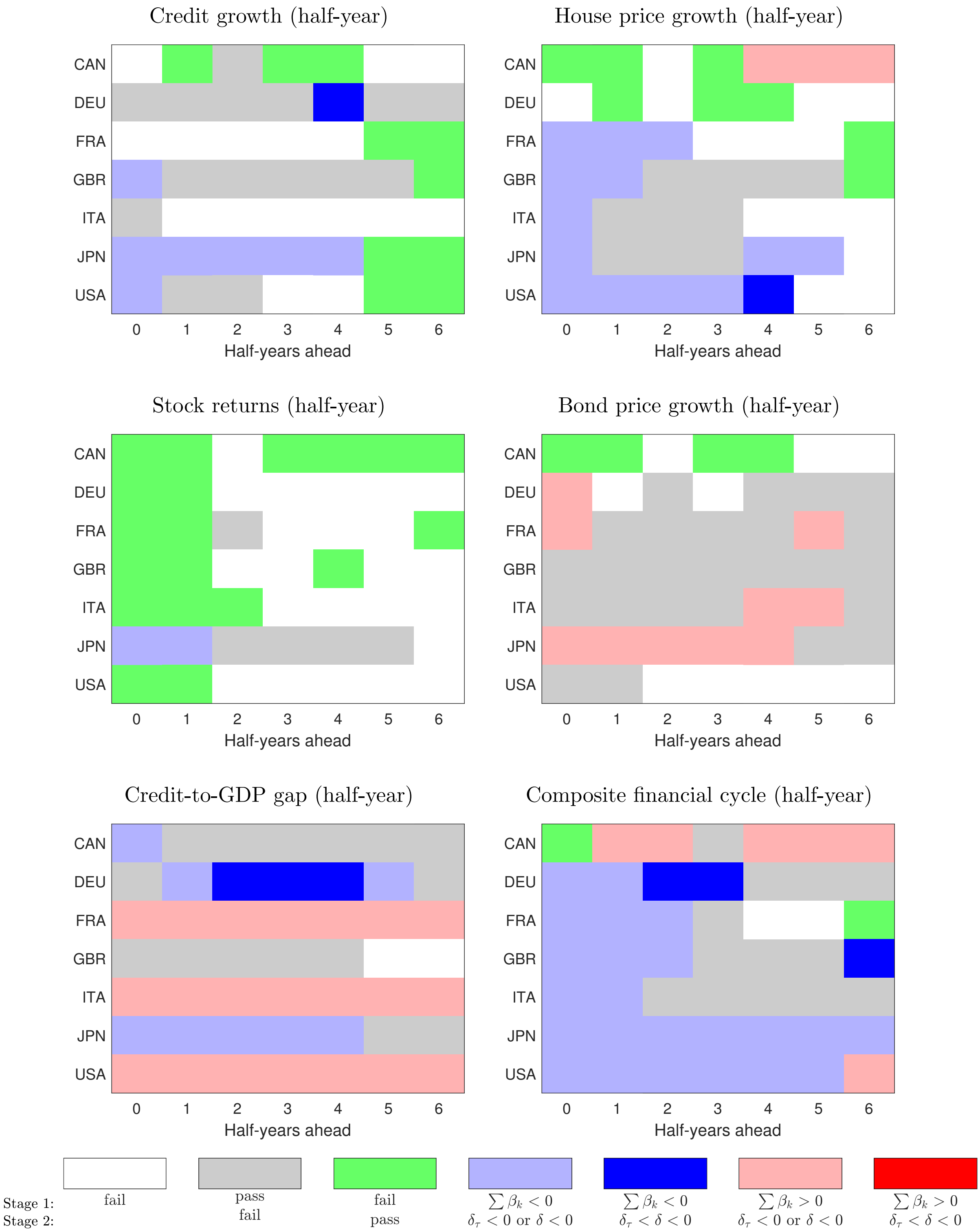

Figure 12: Results without the requirement to pass Stage 1 and with uncorrected standard errors

Notes: The figure shows the results from our two-stage regressions in the form of a heatmap. The color code is the same as in Figure 1. On top of that, green squares indicate cases where the candidate variable passes Stage 2 of our test, but does not pass Stage 1. The regressions in Stage 2 are run without the standard error correction of Theorem 1. The dependent variable in Stage 1 is the crisis dummy based on Romer and Romer (2017).

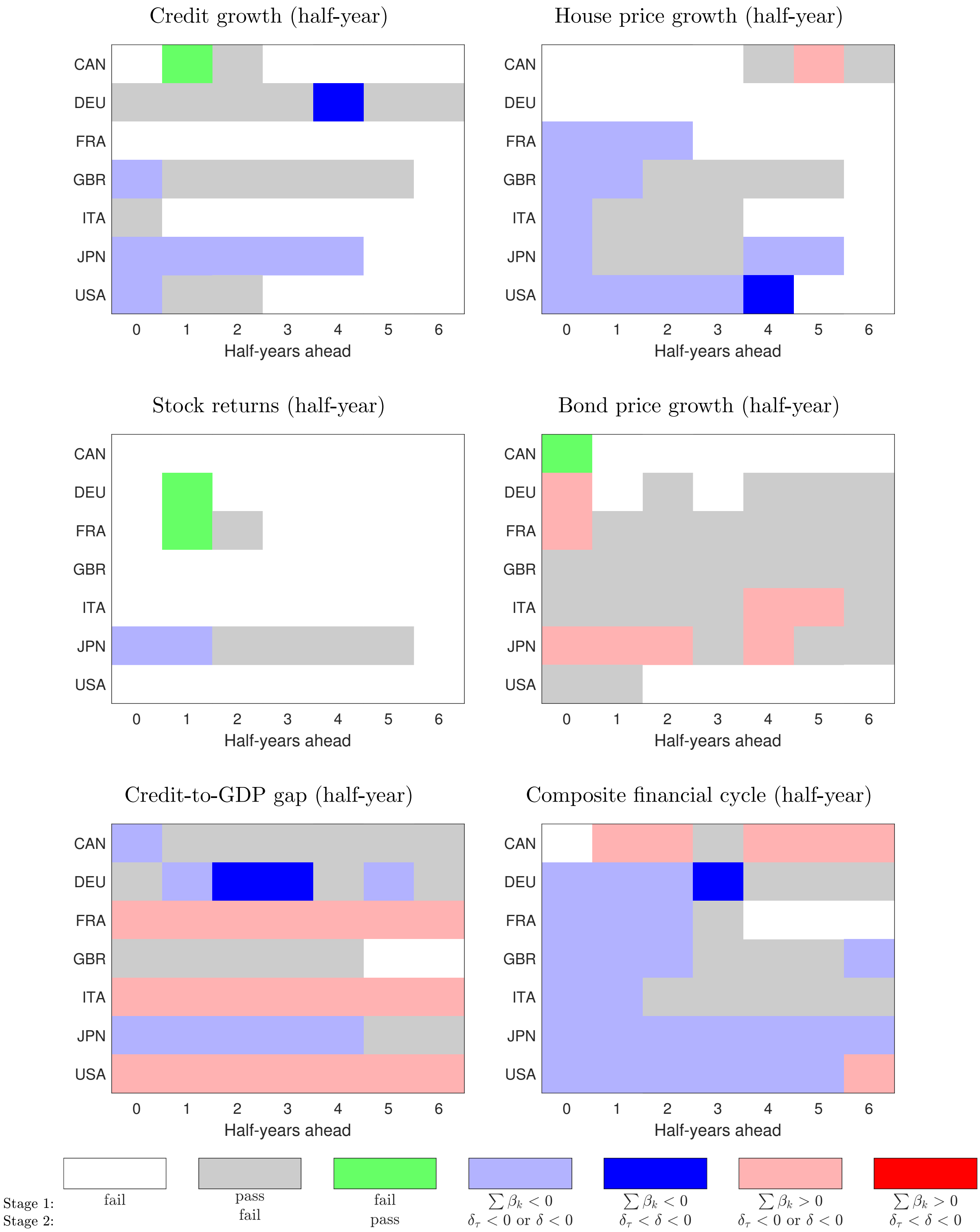

Figure 13: Results without the requirement to pass Stage 1 and with corrected standard errors

Notes: The figure shows the results from our two-stage regressions in the form of a heatmap. The color code is the same as in Figure 1. On top of that, green squares indicate cases where the candidate variable passes Stage 2 of our test, but does not pass Stage 1. The dependent variable in Stage 1 is the crisis dummy based on Romer and Romer (2017).

classes (housing, equity and bonds) provides a much more accurate measurement of systemic risk. Here, a fall in the composite indicator signals a rise in systemic risk. However, the individual components of the composite financial cycle measure do not pass our test on a stand-alone basis, either. Finally, we also document that financial conditions indices capturing contagion and spillover effects do not indicate systemic risk ahead of crises. All these results survive a battery of robustness checks concerning both the structure of our test and the choice of left-hand side and right-hand side variables.

More broadly, the results from our tests also shed light on our theoretical understanding of systemic risk. In particular, it seems that systemic risk can be consistently measured only once the turning points of indicators have been observed. Given the impossibility of predicting turning points, this supports explanations of systemic risk based on leverage cycles in the spirit of Geanakoplos (2010) or Brunnermeier and Sannikov (2014).

This interpretation is also important for policy. Our results support the argument that pre-emptive countercyclical macroprudential policy may not directly reduce systemic risk. Instead, it may have the potential to smooth the financial cycle and to address vulnerability episodes in boom phases, which then indirectly mitigates the amount of systemic risk that can build up in the future.

We see several possible paths for further research. First, our analysis could be extended by exploring real-time and pseudo real-time data, in order to provide more profound practical guidance for policy makers. Second, structural models are clearly needed to rationalize why the joint movement of credit and asset prices is much more informative about systemic risk than any of the individual components. Third, we deliberately choose to run our test separately for each country, neglecting the panel dimension of the data. This allows us to obtain deeper insights on the performance of every candidate variable in every single country. In parallel, it keeps our econometric framework as simple as possible. Notwithstanding, a combination of panel approaches to GDP-at-risk (see, for example, Beutel (2019)) with early-warning models may be a promising line of research to pursue in the future.

References

Adrian, T., N. Boyarchenko, and D. Giannone (2019). Vulnerable growth. *American Economic Review 109*, 1263–1289.

Adrian, T. and H. Shin (2008). Financial intermediary leverage and value-at-risk. *Staff Report 338. Federal Reserve Bank of New York*.

Aikman, D., M. Kiley, S. Lee, M. Palumbo, and M. Warusawitharana (2017). Mapping heat in the U.S. financial system. *Journal of Banking and Finance 81*, 36–64.

Anundsen, A., K. Gerdrup, F. Hansen, and K. Kragh-Sørensen (2016). Bubbles and crises: The role of house prices and credit. *Journal of Applied Econometrics 31*, 1291–1311.

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

Behn, M., C. Detken, T. Peltonen, and W. Schudel (2017). Predicting vulnerabilities in the eu banking sector: The role of global and domestic factors. *International Journal of Central Banking 13*, 147–189.

Beutel, J. (2019). Forecasting growth at risk. *Deutsche Bundesbank mimeo*.

Bisias, D., M. Flood, A. Lo, and S. Valavanis (2012). A survey of systemic risk analytics. *Annual Review of Financial Economics 4*, 255–296.

Brave, S. and R. A. Butters (2012). Diagnosing the financial system: Financial conditions and financial stress. *International Journal of Central Banking 8*, 191–239.

Breitung, J. and S. Eickmeier (2014). Analyzing business and financial cycles using multi-level factor models. *Deutsche Bundesbank Discussion Paper 11/2014*.

Brunnermeier, M. K. and Y. Sannikov (2014). A macroeconomic model with a financial sector. *American Economic Review 104*, 379–421.

Chen, L., A. F. Galvao, and S. Song (2018). Quantile regression with generated regressors. *SSRN mimeo*.

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

Demirgüç-Kunt, A. and E. Detragiache (1998). The determinants of banking crises in developing and developed countries. *IMF Staff Papers 45*, 81–109.

Detken, C., O. Weeken, L. Alessi, D. Bonfim, M. M. Boucinha, C. Castro, S. Frontczak, G. Giordana, J. Giese, and N. Jahn (2014). Operationalising the countercyclical capital buffer: Indicator selection, threshold identification and calibration options. *ESRB Occasional Paper Series 5*.

Duprey, T., B. Klaus, and T. Peltonen (2017). Dating systemic financial stress episodes in the EU countries. *Journal of Financial Stability 32*, 30–56.

Edge, R. M. and R. 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*, 261–298.

Galati, G., I. Hindrayanto, S. Koopman, and M. Vlekke (2016). Measuring financial cycles in a model-based analysis: Empirical evidence for the United States and the euro area. *Economics Letters 145*, 83–87.

Geanakoplos, J. (2010). *NBER Macroeconomics Annual 2009*, Volume 24, Chapter 1: The Leverage Cycle, pp. 1–65. Chicago: University of Chicago Press.

Gilchrist, S., V. Yankov, and E. Zakrajšek (2009). Credit market shocks and economic fluctuations: Evidence from corporate bond and stock markets. *Journal of Monetary Economics 56*, 471–493.

Gilchrist, S. and E. Zakrajšek (2012). Credit spreads and business cycle fluctuations. *American Economic Review 102*, 1692–1720.

Gomes, J., M. Grotteria, and J. Wachter (2019). Foreseen risks. *NBER Working Paper 25277*.

Greene, W. H. (2012). *Econometric Analysis* (7th ed.). New York: Pearson.

Hall, P. and S. J. Sheather (1988). On the distribution of a Studentized quantile. *Journal of the Royal Statistical Society: Series B 50*, 381–391.

Hiebert, P., I. Jaccard, and Y. Schüler (2018). Contrasting financial and business cycles: Stylized facts and candidate explanations. *Journal of Financial Stability 38*, 72–80.

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

Iacoviello, M. (2005). House prices, borrowing constraints, and monetary policy in the business cycle. *American Economic Review 95*, 739–764.

IMF, BIS, and FSB (2009). Guidance to assess the systemic importance of financial institutions, markets and instruments: Initial considerations. *Report to G20 Finance Ministers and Governors*.

Jordà, O., M. Schularick, and A. M. Taylor (2015). Leveraged bubbles. *Journal of Monetary Economics 76*, S1–S20.

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

Koenker, R. and G. Bassett (1978). Regression quantiles. *Econometrica 46*, 33–50.

Komunjer, I. (2005). Quasi-maximum likelihood estimation for conditional quantiles. *Journal of Econometrics 128*, 137–164.

Laeven, L. and F. Valencia (2012). Systemic banking crises database: An update. *IMF Working Paper WP/12/163*.

Laeven, L. and F. Valencia (2018). Systemic banking crisis revisted. *IMF Working Paper WP/18/206*.

Lo Duca, M., A. Koban, M. Basten, E. Bengtsson, B. Klause, P. Kusmierczyk, J. H. Lang, C. Detken, and T. A. Peltonen (2017). A new database for financial crises in European countries – ECB/ESRB EU crises database. *Occasional Paper Series No 194*.

Mandler, M. and M. Scharnagl (2019). Financial cycles across G7 economies: A view from wavelet analysis. *Deutsche Bundesbank Discussion Paper 22/2019*.

Mendoza, E. (2010). Sudden stops, financial crises, and leverage. *American Economic Review 100*, 1941–1966.

Miranda-Agrippino, S. and H. Rey (2019). World asset markets and the global financial cycle. *NBER Working Paper 21722*.

Murphy, K. M. and R. H. Topel (1985). Estimation and inference in two-step econometric models. *Journal of Business and Economic Statistics 3*, 370–379.

Powell, J. (1991). Nonparametric and semiparametric methods in econometrics and statistics. In W. Barnett, J. Powell, and G. Tauchen (Eds.), *Estimation of monotonic regression models under quantile restrictions*, pp. 357–384. New York: Cambridge University Press.

Reinhart, C. M. and K. S. Rogoff (2009). *This time is different: Eight centuries of financial folly.* Princeton University Press.

Romer, C. D. and D. H. Romer (2017). New evidence on the aftermath of financial crises in advanced countries. *American Economic Review 107*, 3072–3118.

Rünstler, G. and M. Vlekke (2018). Business, housing, and credit cycles. *Journal of Applied Econometrics 33*, 212–226.

Schüler, Y., P. Hiebert, and T. Peltonen (2015). Characterising the financial cycle: A multivariate and time-varying approach. *ECB Working Paper 1846*.

Schüler, Y., P. Hiebert, and T. Peltonen (2020). Financial cycles: Characterisation and real-time measurement. *Journal of International Money and Finance, forthcoming*.

Schüler, Y. S. (2018). Detrending and financial cycle facts across G7 countries: Mind a spurious medium term! *ECB Working Paper 2138*.

Schüler, Y. S. (2020). The credit-to-gdp gap und spurious medium-term cycles. *Deutsche Bundesbank mimeo*.

Schularick, M. and A. M. Taylor (2012). Credit booms gone bust: Monetary policy, leverage cycles, and financial crises, 1870-2008. *American Economic Review 102*, 1029–1061.

Strohsal, T., C. Proaño, and J. Wolters (2019). Characterizing the financial cycle: Evidence from a frequency domain analysis. *Journal of Banking and Finance 106*, 568–591.

Appendix A Standard errors for generated regressors

The asymptotics of the standard errors have been described in Theorem 1. The sample estimate of the corrected standard error V_2 in equation (4), denoted by $\hat{V}_2$, is given by

$$\hat{V}_2 = (-\hat{H}_{22}^{(2)})^{-1}[\hat{\Sigma}_{22} + \hat{H}_{21}^{(2)}(-\hat{H}_{11}^{(1)})^{-1}\hat{H}_{21}^{(2)\prime} + \hat{\Sigma}_{21}(-\hat{H}_{11}^{(1)})^{-1}\hat{H}_{21}^{(2)\prime} + \hat{H}_{21}^{(2)}(-\hat{H}_{11}^{(1)})^{-1}\hat{\Sigma}_{12}](-\hat{H}_{22}^{(2)})^{-1}$$

where $\hat{\Sigma}_{22}, \hat{\Sigma}_{21}$ and $\hat{\Sigma}_{12}$ are the typical BHHH estimators

$$\hat{\Sigma}_{22} = \sum_{t=1}^{T} \frac{\partial \ln f_{2t}}{\partial \hat{\theta}_2} \frac{\partial \ln f_{2t}}{\partial \hat{\theta}_2'}, \quad \hat{\Sigma}_{21} = \sum_{t=1}^{T} \frac{\partial \ln f_{2t}}{\partial \hat{\theta}_2} \frac{\partial \ln f_{1t}}{\partial \hat{\theta}_1'}, \quad \hat{\Sigma}_{12} = \sum_{t=1}^{T} \frac{\partial \ln f_{1t}}{\partial \hat{\theta}_1} \frac{\partial \ln f_{2t}}{\partial \hat{\theta}_2'}$$

and the $\hat{H}_{11}$, $\hat{H}_{22}$ and $\hat{H}_{21}$ may be computed as expected Hessians

$$\hat{H}_{11}^{(1)} = \sum_{t=1}^{T} E\left[\frac{\partial \ln^2 f_{1t}}{\partial \hat{\theta}_1 \partial \hat{\theta}_1'}\right], \quad \hat{H}_{22}^{(2)} = \sum_{t=1}^{T} E\left[\frac{\partial \ln^2 f_{2t}}{\partial \hat{\theta}_2 \partial \hat{\theta}_2'}\right], \quad \hat{H}_{21}^{(2)} = \sum_{t=1}^{T} E\left[\frac{\partial \ln^2 f_{2t}}{\partial \hat{\theta}_2 \partial \hat{\theta}_1'}\right].$$

In the following, we present all formulas for the special cases of logit and mean regression as well as logit and quantile regression.

A.1 Logit and mean regression

A.1.1 Logit model

We have

$$P(y_{1t} = 1) = \Lambda(x_{1t}\theta_1)$$

where $\Lambda(x_t\theta) = \frac{\exp(x_t\theta)}{1+\exp(x_t\theta)}$ is the link function. The log-likelihood function of Model 1 is

$$\begin{aligned} \ln L_1(\theta_1) &= \sum_{t=1}^{T} \ln f_1(y_{1t}|x_{1t}, \theta_1) \\ &= \sum_{t=1}^{T} [(1 - y_{1t}) \ln[(1 - \Lambda(x_{1t}\theta_1))] + y_{1t} \ln[\Lambda(x_{1t}\theta_1)]] . \end{aligned}$$

The derivative vector of the log-likelihood of Model 1 is

$$g_{1t}^{(1)} = \frac{\partial \ln f_{1t}}{\partial \theta_1} = x_{1t}'(y_{1t} - \Lambda(x_{1t}\theta_1)) = x_{1t}'u_{1t}$$

where $u_{1t} = y_{1t} - \Lambda(x_{1t}\theta_1)$. The second derivative is

$$g_{11t}^{(1)} = \frac{\partial^2 \ln f_{1t}(\theta_1)}{\partial \theta_1 \partial \theta_1'} = -x_{1t}'x_{1t}\Lambda(x_{1t}\theta_1)(1 - \Lambda(x_{1t}\theta_1))$$

A.1.2 Mean regression model

We have

$$E(y_{2t}|x_{1t}, x_{2t}, \theta_1, \theta_2) = x_{2t}\beta + \sum_{k=0}^{p} \Lambda(x_{1t-k}\theta_1)\gamma_k = z_t\theta_2.$$

The log-likelihood function of Model 2 is

$$\begin{aligned}\ln L_2(\theta_1,\theta_2) &= \sum_{t=1}^{T} \ln f_2(y_{2t}|x_{1t},x_{2t},\theta_1,\theta_2) \\ &= -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln(\sigma^2) - \sum_{t=1}^{T}\frac{1}{2\sigma^2}u_{2t}^2\end{aligned}$$

where $u_{2t} = y_{2t} - z_t\theta_2$. The derivative vectors of the log-likelihood of Model 2 are

$$\begin{aligned}g_{1t}^{(2)} &= \frac{\partial \ln f_{2t}}{\partial \theta_1} = \frac{1}{\sigma^2}n_t' u_{2t} \\ g_{2t}^{(2)} &= \frac{\partial \ln f_{2t}}{\partial \theta_2} = \frac{1}{\sigma^2}z_t' u_{2t}\end{aligned}$$

where

$$\begin{aligned}n_t = \frac{\partial \sum_{j=1}^{k_2} z_{tj}\theta_{2j}}{\partial \theta_1'} &= \sum_{k=0}^{p} x_{1t-k}\Lambda(x_{1t-k}\theta_1)(1-\Lambda(x_{1t-k}\theta_1))\gamma_k \\ &= a_t \odot ((b_t \odot (1_{1\times p+1} - b_t)) \otimes 1_{1\times k_1})\end{aligned}$$

and the second derivative is

$$\begin{aligned}g_{21t}^{(2)} = \frac{\partial^2 \ln f_{2t}}{\partial \theta_2 \partial \theta_1'} &= -\frac{1}{\sigma^2}z_t' n_t + \frac{1}{\sigma^2}m_t' u_{2t} \\ g_{22t}^{(2)} = \frac{\partial^2 \ln f_{2t}}{\partial \theta_2 \partial \theta_2'} &= -\frac{1}{\sigma^2}z_t' z_t\end{aligned}$$

with

$$\begin{aligned}a_t &= [x_{1t}, x_{1t-1}, ..., x_{1t-p}] \\ b_t &= [\Lambda(x_{1t}\theta_1), \Lambda(x_{1t-1}\theta_1), ..., \Lambda(x_{1t-p}\theta_1)] \\ c_t &= [x_{1t}', x_{1t-1}', ..., x_{1t-p}']'\end{aligned}$$

and

$$\begin{aligned}m_t' = \frac{\partial z_t'}{\partial \theta_1'} &= \begin{bmatrix} 0_{(k_2-p+1)\times k_1} \\ x_{1t} \\ x_{1t-1} \\ \vdots \\ x_{1t-p}\end{bmatrix} \odot \left(\begin{bmatrix} 0_{(k_2-p+1)\times k_1} \\ \Lambda(x_{1t}\theta_1)(1-\Lambda(x_{1t}\theta_1)) \\ \Lambda(x_{1t-1}\theta_1)(1-\Lambda(x_{1t-1}\theta_1)) \\ \vdots \\ \Lambda(x_{1t-p}\theta_1)(1-\Lambda(x_{1t-p}\theta_1))\end{bmatrix} \otimes 1_{1\times k_1}\right) \\ &= \begin{bmatrix} 0_{(k_2-p+1)\times k_1} \\ c_t \end{bmatrix} \odot \left(\begin{bmatrix} 0_{(k_2-p+1)\times k_1} \\ b_t' \odot (1_{p+1\times 1} - b_t') \end{bmatrix} \otimes 1_{1\times k_1}\right).\end{aligned}$$

A.1.3 Asymptotic covariance and estimation

Next, we use these conditions to derive the inputs for the corrected asymptotic covariance matrix:

$$\begin{aligned}
\Sigma_{22} &= E\left(\frac{1}{T}\sum_{t=1}^{T} g_{2t}^{(2)} g_{2t}^{(2)\prime}\right) = E\left(\frac{1}{T}\left(\frac{1}{\sigma^2}\right)^2 \sum_{t=1}^{T} u_{2t}^2 z_t' z_t\right) \\
\Sigma_{21} &= E\left(\frac{1}{T}\sum_{t=1}^{T} g_{2t}^{(2)} g_{1t}^{(1)\prime}\right) = E\left(\frac{1}{T}\frac{1}{\sigma^2}\sum_{t=1}^{T} u_{1t}u_{2t} z_t' x_{1t}\right) \\
\Sigma_{12} &= E\left(\frac{1}{T}\sum_{t=1}^{T} g_{1t}^{(1)} g_{2t}^{(2)\prime}\right) = E\left(\frac{1}{T}\frac{1}{\sigma^2}\sum_{t=1}^{T} u_{1t}u_{2t} x_{1t}' z_t\right) \\
H_{11}^{(1)} &= E\left(\frac{1}{T}\sum_{t=1}^{T} g_{11t}^{(1)}\right) = E\left(-\frac{1}{T}\sum_{t=1}^{T} x_{1t}' x_{1t}\Lambda(x_{1t}\theta_1)(1-\Lambda(x_{1t}\theta_1))\right) \\
H_{21}^{(2)} &= E\left(\frac{1}{T}\sum_{t=1}^{T} g_{21t}^{(2)}\right) = E\left(-\frac{1}{T}\frac{1}{\sigma^2}\sum_{t=1}^{T} z_t' n_t\right) \\
H_{22}^{(2)} &= E\left(\frac{1}{T}\sum_{t=1}^{T} g_{22t}^{(2)}\right) = E\left(-\frac{1}{T}\frac{1}{\sigma^2}\sum_{t=1}^{T} z_t' z_t\right)
\end{aligned}$$

Notice that, when the information matrix equality holds, then $\Sigma_{22} = -H_{22}^{(2)}$ and formula (4) reduces to the one presented in Murphy and Topel (1985). Then the quantities for the asymptotic variance matrix may be computed by evaluating the aforementioned expressions at their maximum likelihood estimate. The empirical gradients for the BHHH-Type estimators are

$$\frac{\partial \ln f_1}{\partial \hat{\theta}_1} = x_{1t}'\hat{u}_{1t}, \quad \frac{\partial \ln f_2}{\partial \hat{\theta}_2} = \frac{1}{\hat{\sigma}^2}\hat{z}_t'\hat{u}_{2t}$$

and the expressions for the expected Hessian are

$$E\left[\frac{\partial^2 \ln f_1}{\partial \hat{\theta}_1 \partial \hat{\theta}_1'}\right] = -x_{1t}' x_{1t}\Lambda(x_{1t}\hat{\theta}_1)(1-\Lambda(x_{1t}\hat{\theta}_1), \quad E\left[\frac{\partial^2 \ln f_2}{\partial \hat{\theta}_2 \partial \hat{\theta}_1'}\right] = -\frac{1}{\hat{\sigma}^2}\hat{z}_t'\hat{n}_t, \quad E\left[\frac{\partial^2 \ln f_2}{\partial \hat{\theta}_2 \partial \hat{\theta}_2'}\right] = -\frac{1}{\hat{\sigma}^2}\hat{z}_t'\hat{z}_t.$$

A.2 Logit and quantile regression

The derivations for the logit model are the same as in the previous section.

A.2.1 Quantile regression

We have

$$Q_\tau(y_{2t}|x_{1t}, x_{2t}, \theta_1, \theta_2^\tau) = x_{2t}\beta^\tau + \sum_{k=0}^{p}\Lambda(x_{1t-k}\theta_1)\gamma_k^\tau = z_t\theta_2^\tau.$$

Komunjer (2005) provides the machinery and conditions when the (quasi) maximum likelihood estimation of a possibly non-linear quantile regression yields consistent and asymptotically normal parameter estimates. For the following analysis, we assume that our conditional quantile model is correctly specified such that the assumptions for Corollary 5 (p. 149) are satisfied. The author introduces the family of tick-exponential density functions to study an entire class of quantile regression models, including the linear quantile regression of Koenker and Bassett (1978). The log-likelihood function of Model 2 is

$$\ln L_2(\theta_1, \theta_2^\tau) = \sum_{t=1}^{T} \left[-(1-\tau) \left(\frac{1}{\tau(1-\tau)} (z_t \theta_2^\tau - y_{2t}) 1_{\{y_{2t} \leq z_t \theta_2^\tau\}} \right) + \tau \left(\frac{1}{\tau(1-\tau)} (z_t \theta_2^\tau - y_{2t}) 1_{\{y_{2t} > z_t \theta_2^\tau\}} \right) \right].$$

Assume that the quantile model is continuously differentiable on the parameter space. Then the vector of first order derivatives exists and is continuous with probability one. In particular, it takes the form

$$\begin{aligned} g_{1t}^{(2)} &= \frac{\partial \ln f_2}{\partial \theta_1} = \frac{1}{\tau(1-\tau)} n_t' \left(\tau - 1_{\{y_{2t} \leq z_t \theta_2^\tau\}} \right) \\ g_{2t}^{(2)} &= \frac{\partial \ln f_2}{\partial \theta_2^\tau} = \frac{1}{\tau(1-\tau)} z_t' \left(\tau - 1_{\{y_{2t} \leq z_t \theta_2^\tau\}} \right). \end{aligned}$$

To obtain the second order derivatives, we follow the approach of Komunjer (2005) and first determine the expected value of the first derivatives:

$$\begin{aligned} E\left[g_{1t}^{(2)} \right] &= \frac{1}{\tau(1-\tau)} E\left[n_t' \left(\tau - F_{y_{2t}|z_t\theta_2^\tau}(z_t\theta_2^\tau) \right) \right] \\ E\left[g_{2t}^{(2)} \right] &= \frac{1}{\tau(1-\tau)} E\left[z_t' \left(\tau - F_{y_{2t}|z_t\theta_2^\tau}(z_t\theta_2^\tau) \right) \right]. \end{aligned}$$

Then, second, we differentiate these expressions by making use of the interchangeability of integration and differentiation and of the assumption that the first order conditions are a martingale difference sequence:

$$\begin{aligned} E\left[g_{21t}^{(2)} \right] &= -\frac{1}{\tau(1-\tau)} E\left[z_t' n_t f_{y_{2t}|z_t\theta_2^\tau}(z_t\theta_2^\tau) \right] \\ E\left[g_{22t}^{(2)} \right] &= -\frac{1}{\tau(1-\tau)} E\left[z_t' z_t f_{y_{2t}|z_t\theta_2^\tau}(z_t\theta_2^\tau) \right]. \end{aligned}$$

A.2.2 Asymptotic covariance and estimation

Next, we use these conditions to derive the inputs for the corrected asymptotic covariance matrix:

$$\begin{aligned} \Sigma_{22} &= E\left(\frac{1}{T} \sum_{t=1}^{T} g_{2t}^{(2)} g_{2t}^{(2)\prime} \right) = \frac{1}{\tau(1-\tau)} E\left[\frac{1}{T} \sum_{t=1}^{T} z_t' z_t \right] \\ \Sigma_{21} &= E\left(\frac{1}{T} \sum_{t=1}^{T} g_{2t}^{(2)} g_{1t}^{(1)\prime} \right) = \frac{1}{\tau(1-\tau)} E\left(\frac{1}{T} \sum_{t=1}^{T} u_{1t} (\tau - 1_{\{y_{2t} \leq z_t \theta_2^\tau\}}) z_t' x_{1t} \right) \\ \Sigma_{12} &= E\left(\frac{1}{T} \sum_{t=1}^{T} g_{1t}^{(1)} g_{2t}^{(2)\prime} \right) = \frac{1}{\tau(1-\tau)} E\left(\frac{1}{T} \sum_{t=1}^{T} u_{1t} (\tau - 1_{\{y_{2t} \leq z_t \theta_2^\tau\}}) x_{1t}' z_t \right) \end{aligned}$$

$$\begin{aligned}
H_{11}^{(1)} &= E\left(\frac{1}{T}\sum_{t=1}^{T} g_{11t}^{(1)}\right) = -E\left(\frac{1}{T}\sum_{t=1}^{T} x_{1t}' x_{1t}\Lambda(x_{1t}\theta_1)(1-\Lambda(x_{1t}\theta_1))\right) \\
H_{21}^{(2)} &= E\left(\frac{1}{T}\sum_{t=1}^{T} g_{21t}^{(2)}\right) = -\frac{1}{\tau(1-\tau)}E\left(\frac{1}{T}\sum_{t=1}^{T} z_t' n_t f_{y_{2t}|z_t\theta_2^\tau}(z_t\theta_2^\tau)\right) \\
H_{22}^{(2)} &= E\left(\frac{1}{T}\sum_{t=1}^{T} g_{22t}^{(2)}\right) = -\frac{1}{\tau(1-\tau)}E\left(\frac{1}{T}\sum_{t=1}^{T} z_t' z_t f_{y_{2t}|z_t\theta_2^\tau}(z_t\theta_2^\tau)\right)
\end{aligned}$$

Notice that our formula for the corrected covariance matrix for quantile regressions appears to be similar to the one presented in Theorem 3.2 of the recent discussion paper by Chen, Galvao, and Song (2018). Moreover, if the error term of the quantile regression is identically distributed, i.e., the density of the error terms is independent of the regressors such that $f_{y_{2t}|z_t\theta_2^\tau}(\cdot) = f(\cdot)$ for all t, then the first term, $(-H_{22}^{(2)})^{-1}\Sigma_{22}(-H_{22}^{(2)})^{-1}$, in the formula for the correction reduces to $\tau(1-\tau)\left(E[f(z_t\theta_2^\tau)^2]E\left[T\sum_{t=1}^{T} z_t' z_t\right]\right)^{-1}$, and this is the original variance for quantile regressions in Koenker and Bassett (1978).

The quantities for the asymptotic variance matrix may be computed by evaluating the aforementioned expressions at their maximum likelihood estimate. The empirical gradients for the BHHH-type estimators are

$$\frac{\partial \ln f_1}{\partial \hat\theta_1} = x_{1t}'\hat u_{1t}, \quad \frac{\partial \ln f_2}{\partial \hat\theta_2} = \hat z_t'(\tau - 1_{\{y_{2t}\le \hat z_t\hat\theta_2^\tau\}})$$

and the expressions for the expected Hessian are

$$\begin{aligned}
E\left[\frac{\partial^2 \ln f_1}{\partial\hat\theta_1\partial\hat\theta_1'}\right] &= -x_{1t}' x_{1t}\Lambda(x_{1t}\hat\theta_1)(1-\Lambda(x_{1t}\hat\theta_1), \\
E\left[\frac{\partial^2 \ln f_2}{\partial\hat\theta_2\partial\hat\theta_1'}\right] &= -\frac{1}{\tau(1-\tau)}\hat z_t'\hat n_t \hat f_{y_{2t}|\hat z_t\hat\theta_2^\tau}(\hat z_t\hat\theta_2^\tau), \\
E\left[\frac{\partial^2 \ln f_2}{\partial\hat\theta_2\partial\hat\theta_2'}\right] &= -\frac{1}{\tau(1-\tau)}\hat z_t'\hat z_t \hat f_{y_{2t}|\hat z_t\hat\theta_2^\tau}(\hat z_t\hat\theta_2^\tau).
\end{aligned}$$

We estimate the density of the errors using the kernel method of Powell (1991):

$$\hat f_{y_{2t}|\hat z_t\hat\theta_2^\tau}(\hat z_t\hat\theta_2^\tau) = \frac{1}{2c_T}1(|\hat u_{2t}| < c_T)$$

where $c_T \to 0$ and $\sqrt{T}c_T \to \infty$. For our empirical analysis, we use the default bandwidth

$$c_T = \kappa(\Phi^{-1}(\tau + h_T) - \Phi^{-1}(\tau - h_T))$$

where κ is a robust estimate of scale. The bandwidth h_T is chosen according to Hall and Sheather (1988) and is based on Edgeworth expansions for studentized quantiles:

$$h_T = T^{-\frac{1}{3}} z_\alpha^{\frac{2}{3}}\left[1.5\frac{\phi(\Phi^{-1}(\tau)^2)}{(2(\Phi^{-1}(\tau))^2+1}\right]^{\frac{1}{3}}$$

where $\phi(\cdot)$ and $\Phi^{-1}(\cdot)$ denote the probability density function and the inverse of the cumulative distribution function of the standard normal distribution, and z_α satisfies $\Phi(z_\alpha) = 1 - \alpha/2$.

Appendix B Supplementary figures and tables

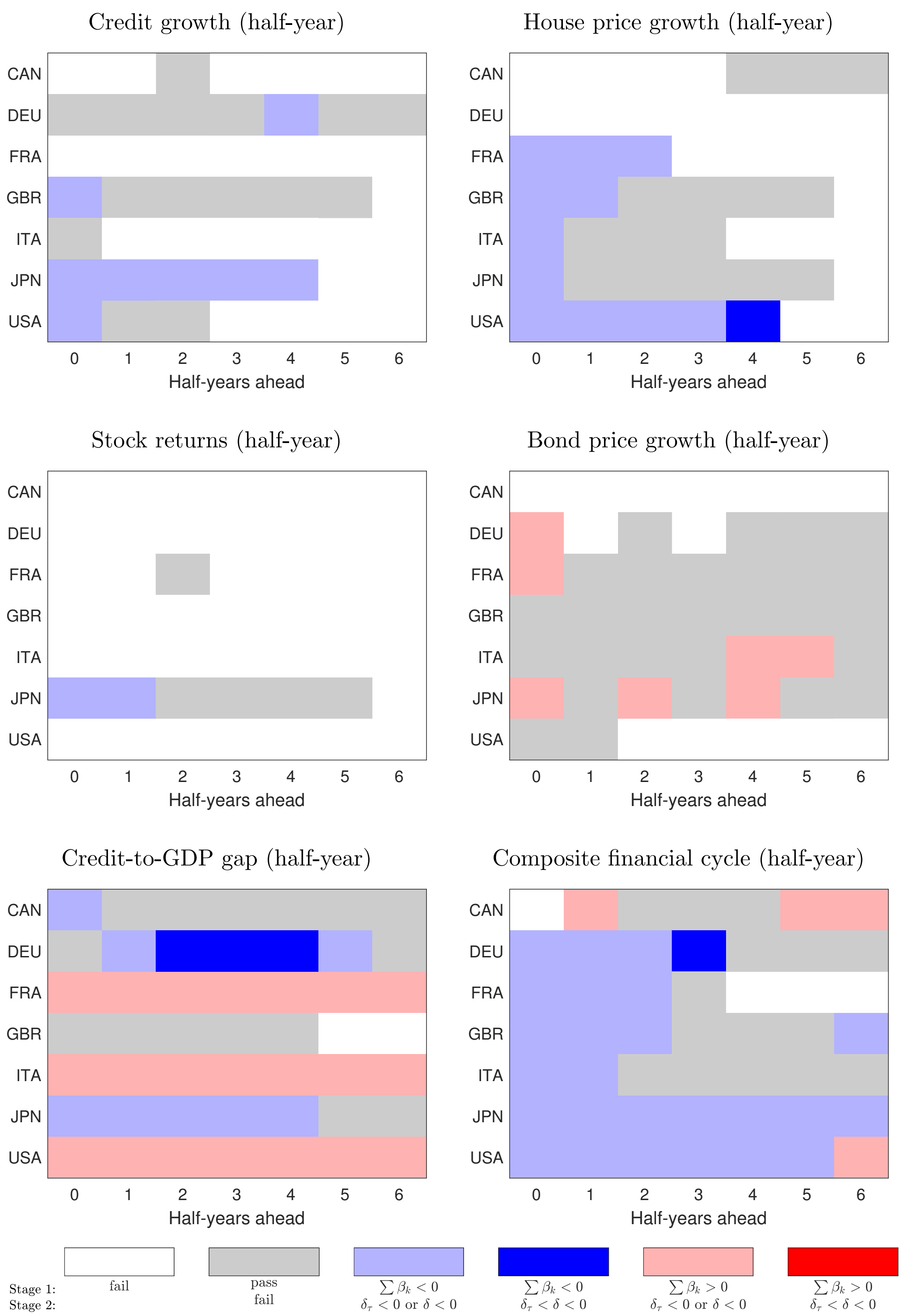

Figure 14: Results assuming independence of error terms in Stages 1 and 2

Notes: The figure shows the results from our two-stage regressions in the form of a heatmap. The color code is the same as in Figure 1. The dependent variable in Stage 1 is the crisis dummy by Romer and Romer (2017).

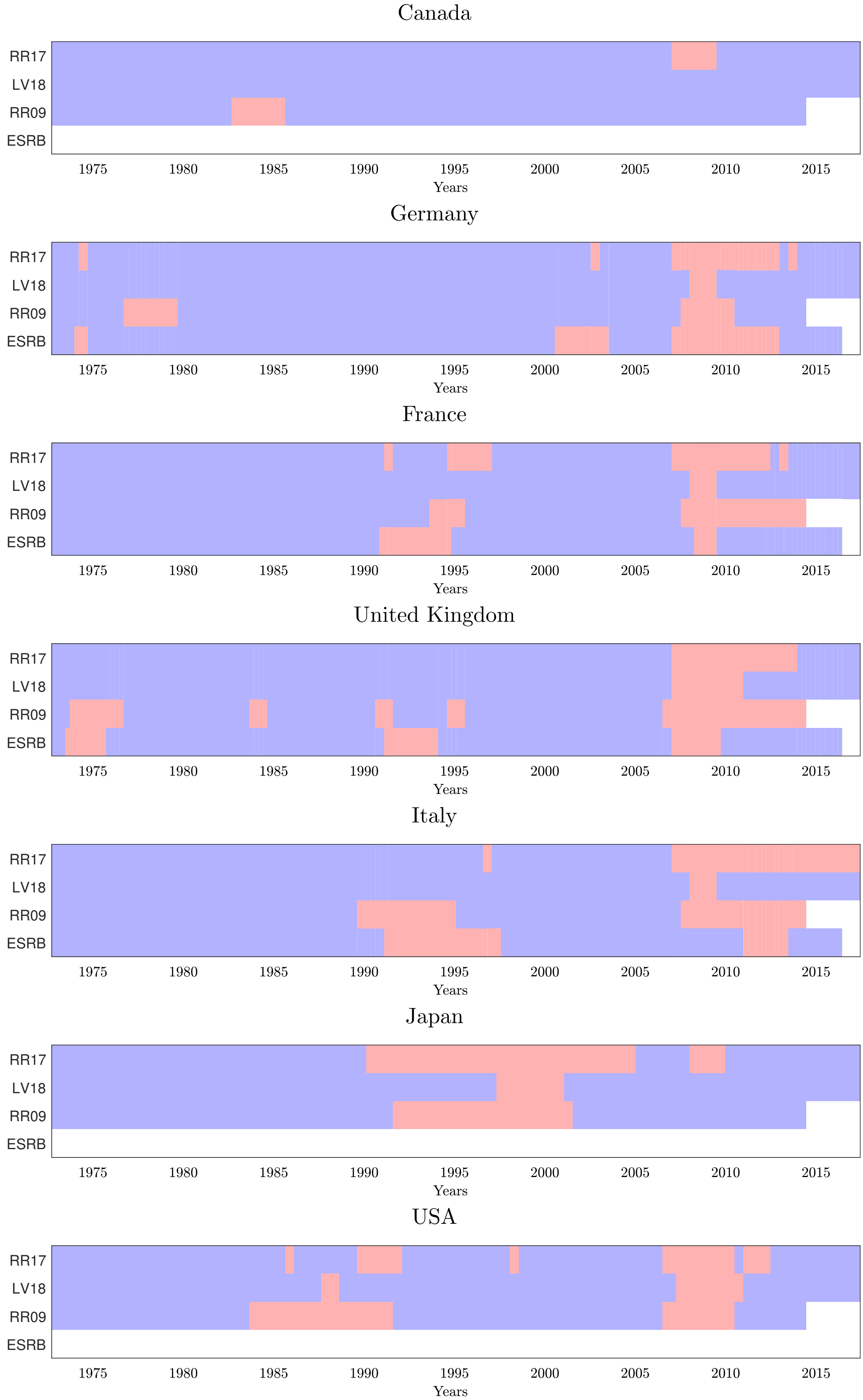

Figure 15: Chronologies of financial disruptions for various crisis indicators

Notes: Red color indicates periods of financial disruption according to the dummy variables of – from top to bottom – Romer and Romer (2017), Laeven and Valencia (2018), Reinhart and Rogoff (2009), and the ESRB (Lo Duca et al. (2017)). Blue color indicates no disruption. White color indicates missing data.

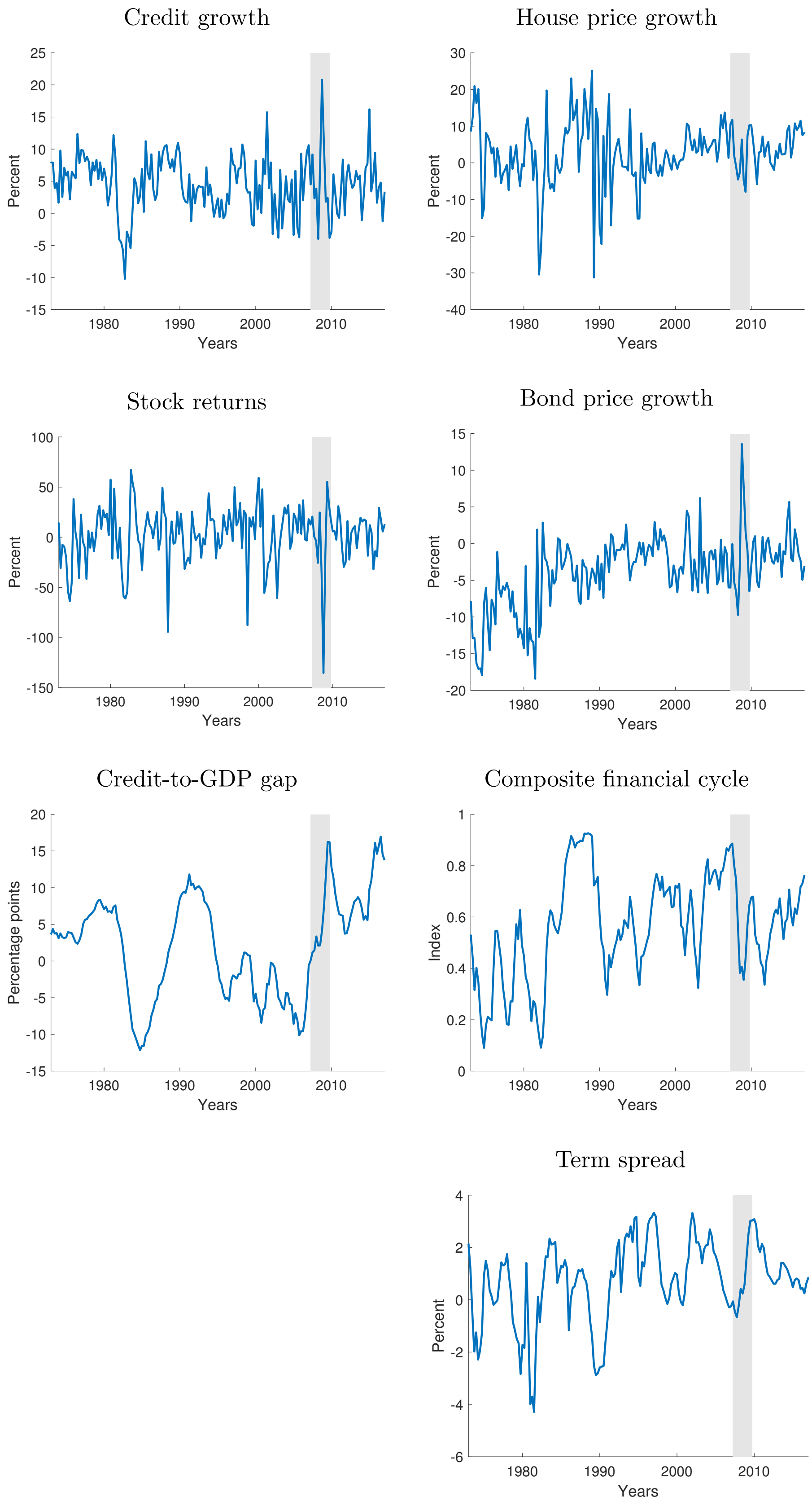

Figure 16: Candidate indicators of systemic risk for Canada

Notes: The gray-shaded area marks periods when the crisis variable of Romer and Romer (2017) is above zero.

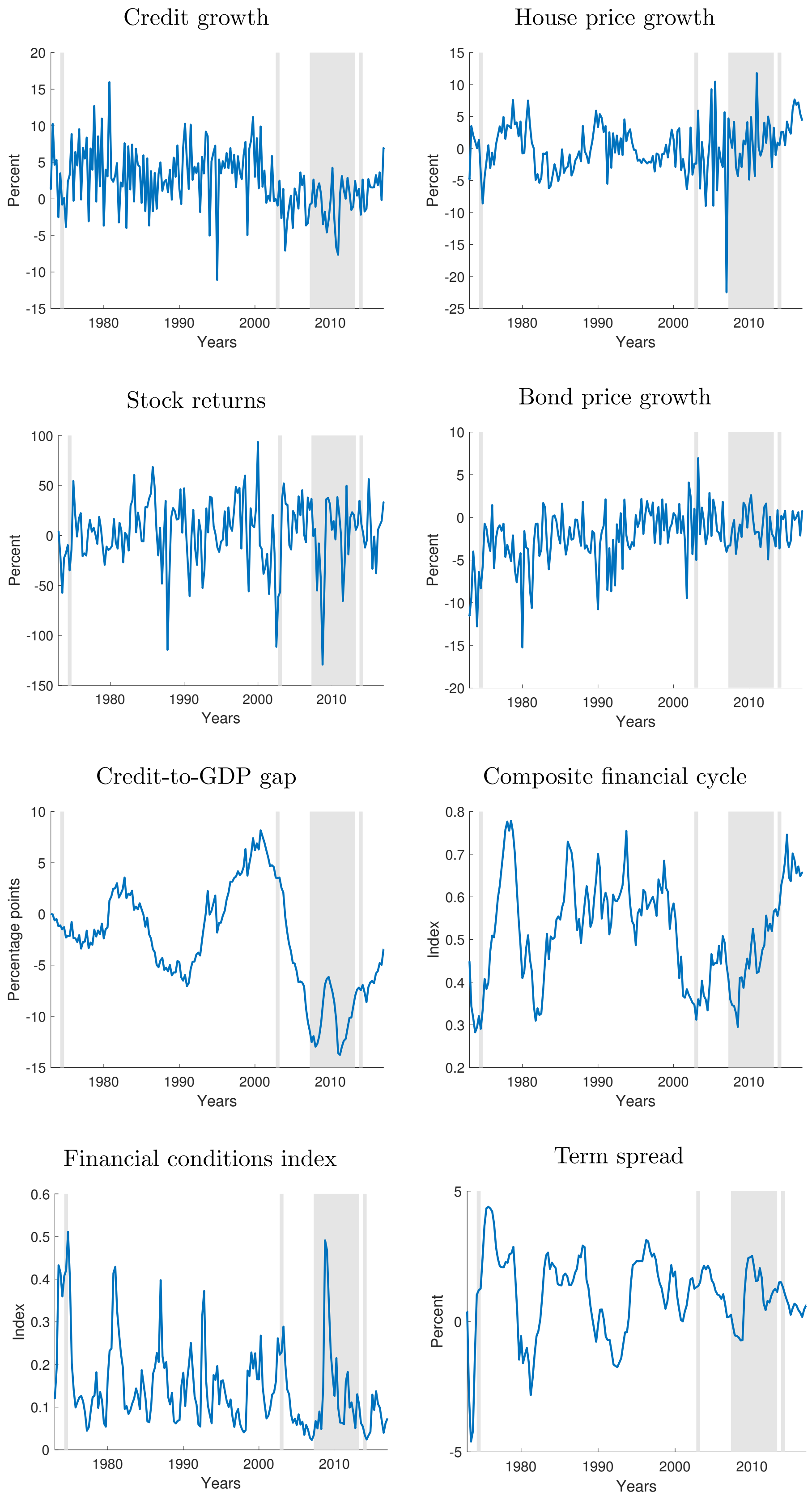

Figure 17: Candidate indicators of systemic risk for Germany

Notes: The gray-shaded area marks periods when the crisis variable of Romer and Romer (2017) is above zero.

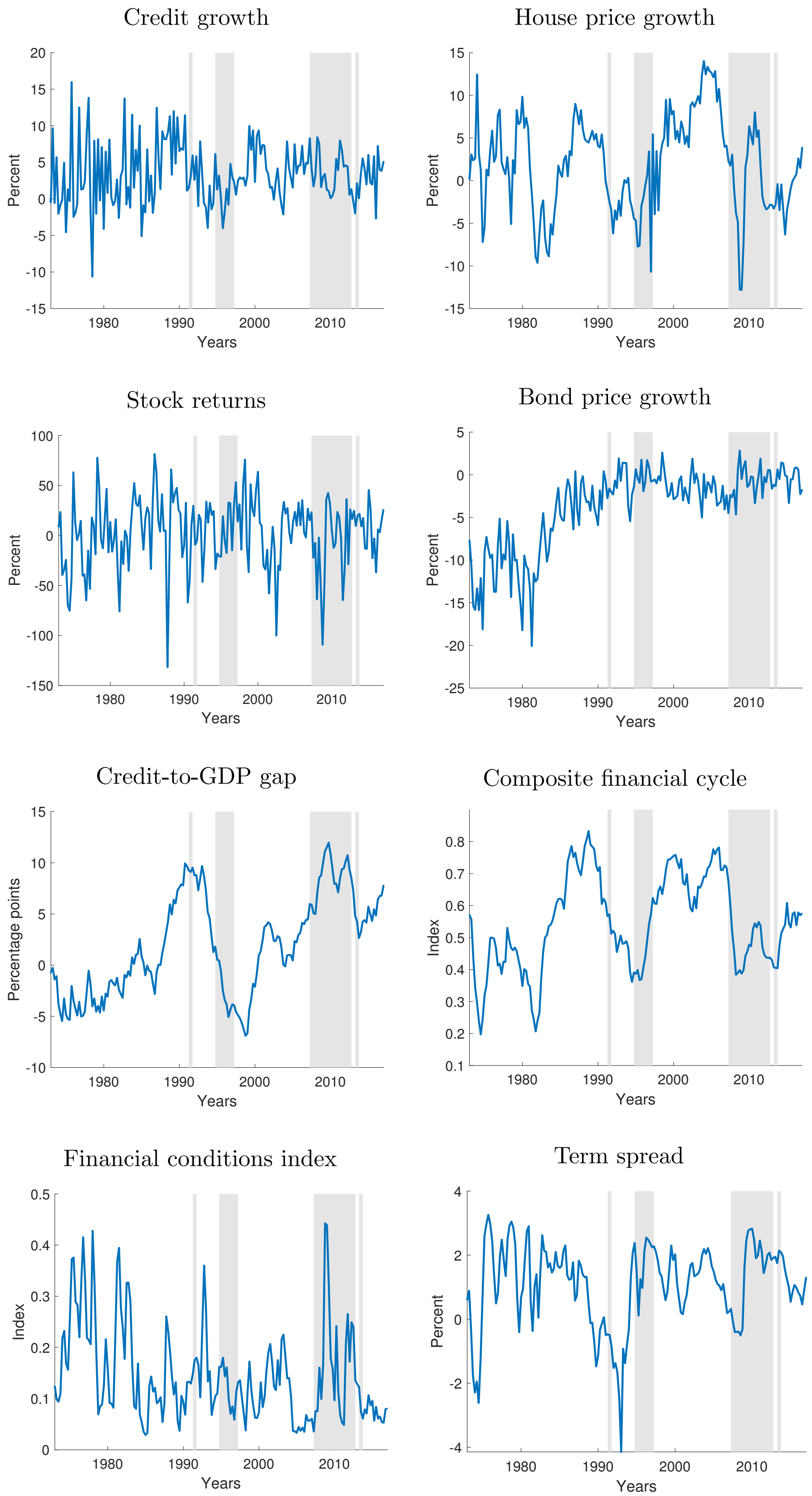

Figure 18: Candidate indicators of systemic risk for France

Notes: The gray-shaded area marks periods when the crisis variable of Romer and Romer (2017) is above zero.

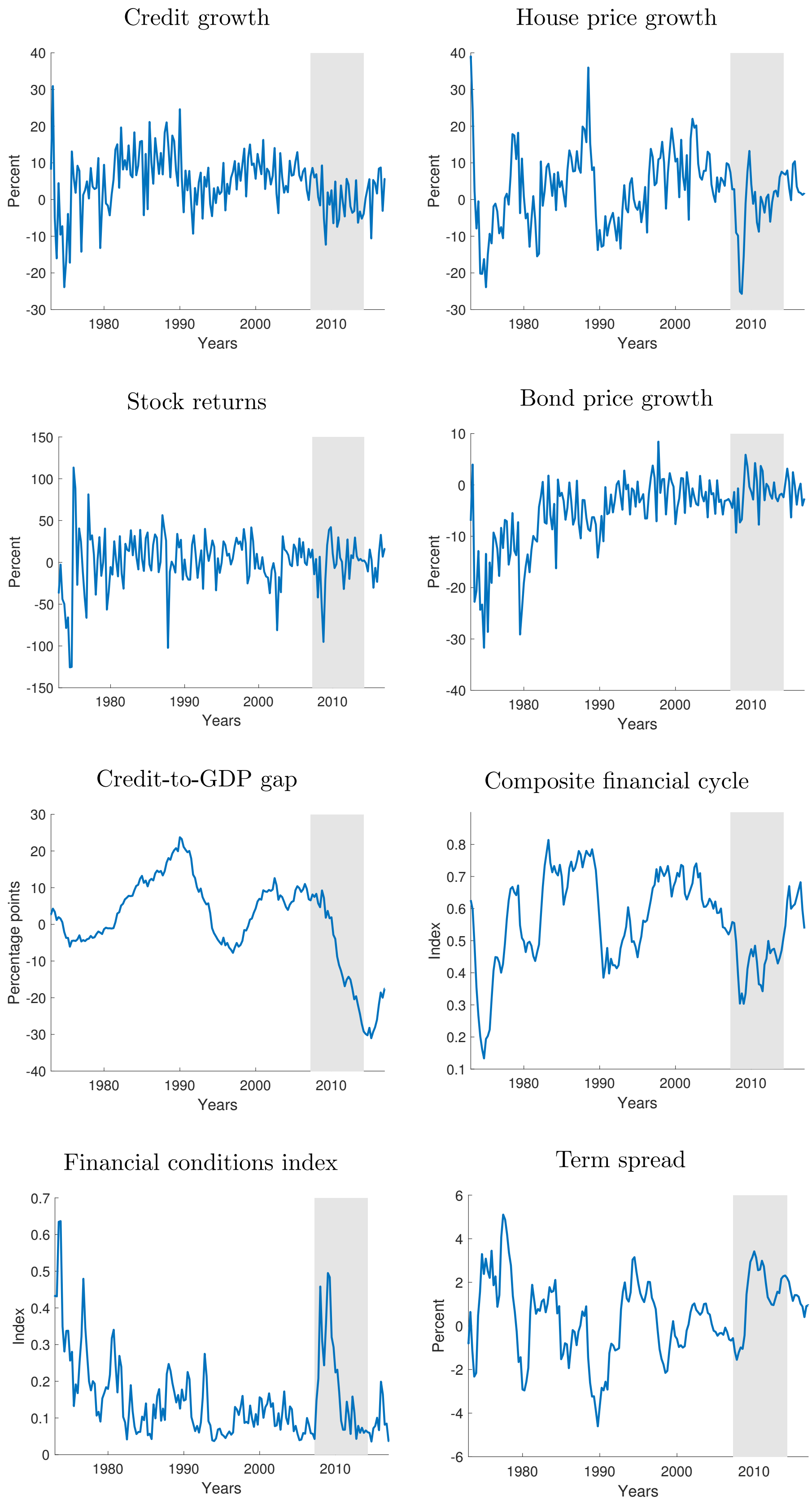

Figure 19: Candidate indicators of systemic risk for the United Kingdom

Notes: The gray-shaded area marks periods when the crisis variable of Romer and Romer (2017) is above zero.

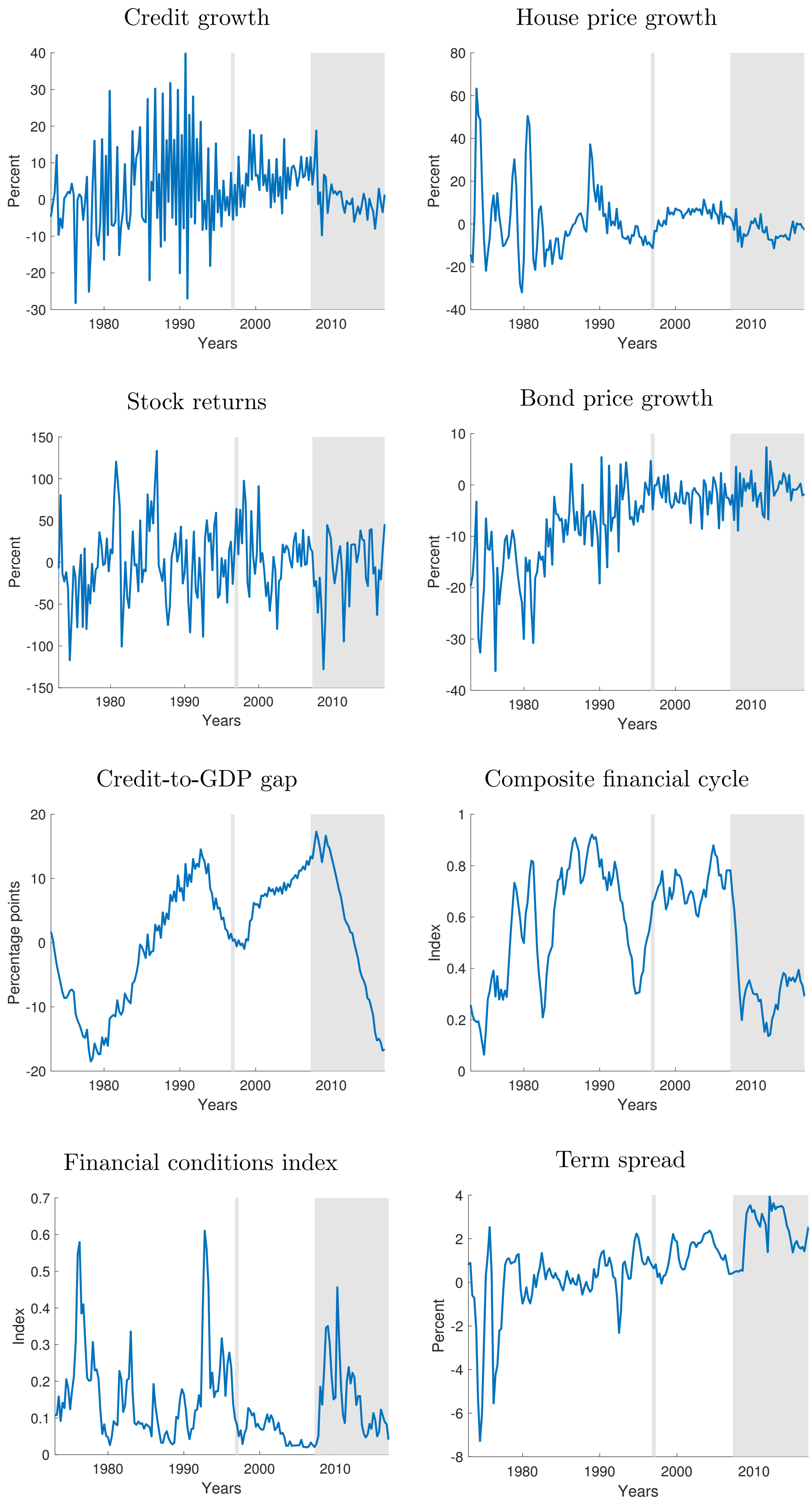

Figure 20: Candidate indicators of systemic risk for Italy

Notes: The gray-shaded area marks periods when the crisis variable of Romer and Romer (2017) is above zero.

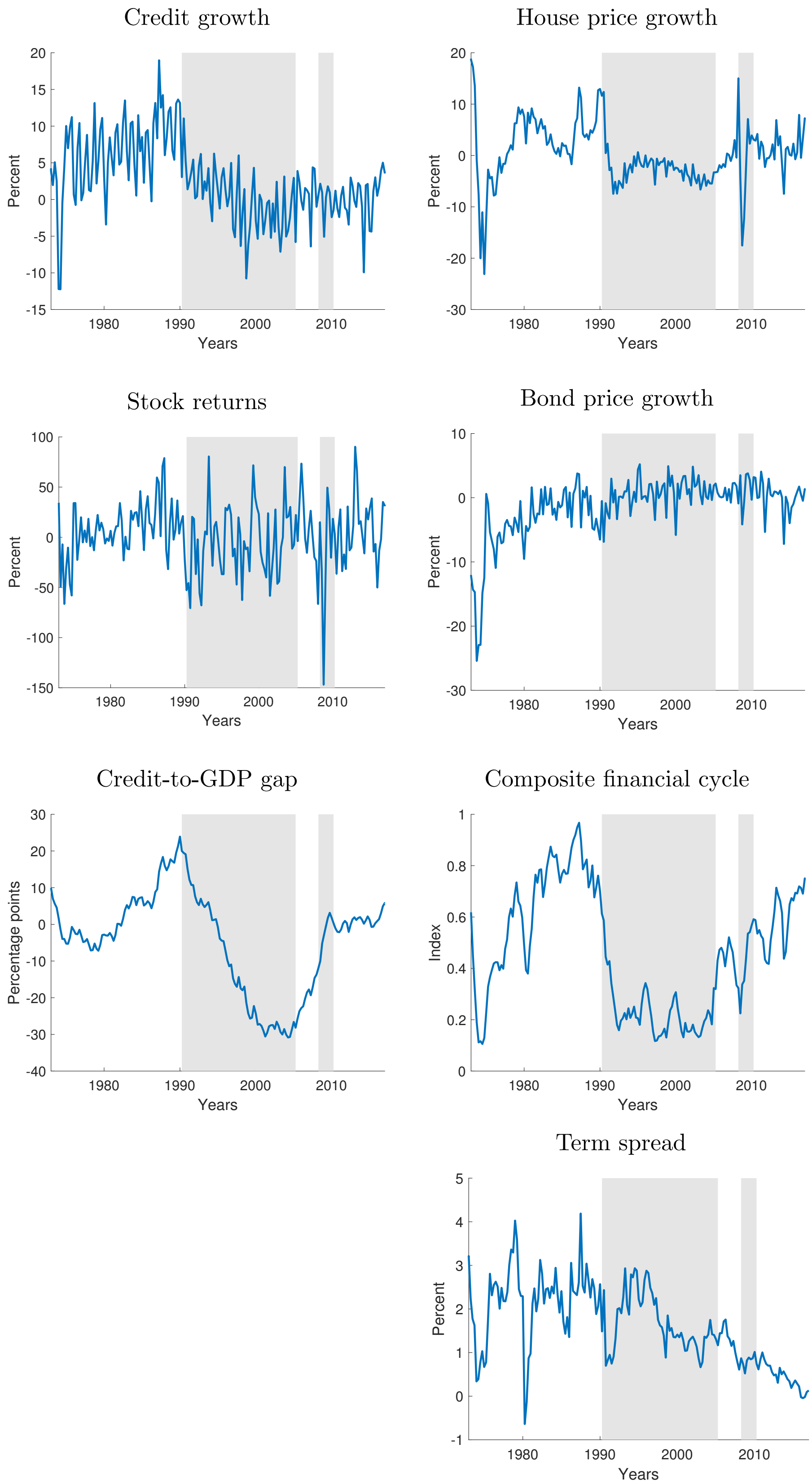

Figure 21: Candidate indicators of systemic risk for Japan

Notes: The gray-shaded area marks periods when the crisis variable of Romer and Romer (2017) is above zero.

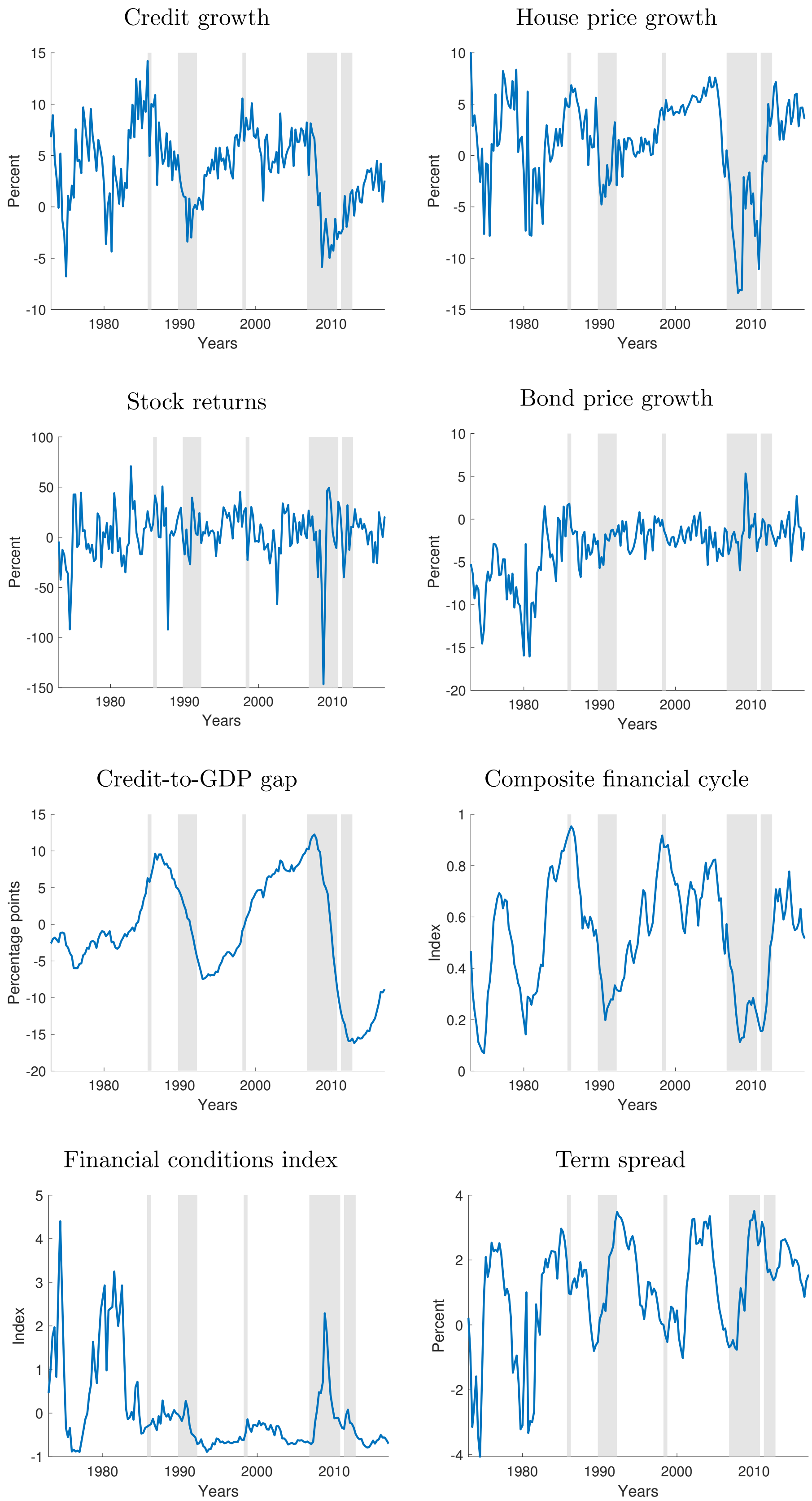

Figure 22: Candidate indicators of systemic risk for the U.S.

Notes: The gray-shaded area marks periods when the crisis variable of Romer and Romer (2017) is above zero.

Table 4: Credit growth (half-year) as a candidate indicator of systemic risk

Years ahead	Stage	Regression	CAN		DEU		FRA		GBR		ITA		JPN		USA	
0	1		0.03	[0]	-0.37***	[1]	-0.06	[0]	-0.22***	[1]	-0.04**	[0]	-0.18***	[1]	-0.2***	[0]
	2	Mean	-6.47		-1.48		-2.56		-6.4**		0.22		-11.86***		-10.59***	
		Quantile	-125.35		1.47		1.77		-10.59*		-13.96		-16.03***		-18.67***	
0.5	1		0.1	[0]	-0.39***	[1]	-0.06	[0]	-0.11***	[0]	-0.03	[0]	-0.14***	[6]	-0.13***	[0]
	2	Mean	-40.58**		-0.36		4.36		-2.02		8.81		-5.61***		-3.02	
		Quantile	-87.83		1.37		18.47		0.62		24.05		-9.02***		18.33	
1	1		0.12*	[0]	-0.45***	[2]	-0.02	[0]	-0.09***	[0]	-0.03	[0]	-0.07***	[0]	-0.09**	[0]
	2	Mean	-7.48		-1.51		18.07		-2.01		-0.08		-12.7**		-3.61	
		Quantile	-14.84		-16.68		42.25		4.33		23.8		-28.1**		16.83	
1.5	1		0.04	[0]	-0.39***	[1]	-0.03	[0]	-0.08***	[0]	-0.02	[0]	-0.06**	[0]	-0.05	[0]
	2	Mean	-43.62		-1.93		-2.44		-1.14		22.76		-11.82**		-7.04	
		Quantile	-205.92		-11.86		43.33		-2.13		68.92		-4.26		26.7	
2	1		0.01	[0]	-0.32***	[1]	-0.01	[0]	-0.06**	[0]	-0.02	[0]	-0.04*	[0]	-0.02	[0]
	2	Mean	-108.47		-1.52		82.02		0.78		14.26		-21.89**		17.26	
		Quantile	-550.81		-28.72**		21.79		16.43		70.12		-33.92*		94.51	
2.5	1		-0.01	[0]	-0.37***	[2]	0	[0]	-0.05*	[0]	-0.01	[0]	-0.03	[0]	0.01	[0]
	2	Mean	189.82		0.03		-953.22		-0.57		74.07		-29.53		-46.42	
		Quantile	371.42		-13.02		-1556		36.72		247.15		-22.57		-142.57	
3	1		-0.04	[0]	-0.43***	[2]	0.02	[0]	-0.04	[0]	-0.01	[0]	-0.02	[0]	0.02	[0]
	2	Mean	19.8		-1.42		-46.92		-7.67		-1.87		-38.06		-24.55	
		Quantile	100.89		-11.76		-107.58		-25.39		17.34		-42.77		-131.61	

Notes: For Stage 1 we report the sum of the slope coefficients of the candidate variable, the significance according to the likelihood ratio test, and the number of lags of the candidate variable (in brackets; as determined by the BIC). For Stage 2 we report (both for mean and quantile regressions) the slope coefficients and the significance according to the adjusted one-sided t-test. One, two, and three stars denote significance at the 10%, 5%, and 1% level respectively.

Table 5: House price growth (half-year) as a candidate indicator of systemic risk

Years ahead	Stage	Regression	CAN		DEU		FRA		GBR		ITA		JPN		USA	
0	1		-0.01	[0]	0.07	[0]	-0.08***	[0]	-0.04**	[0]	-0.03**	[0]	-0.21***	[0]	-0.23***	[2]
	2	Mean	-135.93		11.82		-7.79***		-29.24***		2.1		-8.45***		-5.71**	
		Quantile	-234.34		36.82		-11.28**		-35.26**		-19.85*		-13.83*		-7.99*	
0.5	1		-0.01	[0]	-0.02	[0]	-0.07***	[0]	-0.04**	[0]	-0.03**	[0]	-0.14***	[0]	-0.21***	[1]
	2	Mean	-76.91		-75.49		-4.32*		-18.36**		-1.03		-3.87		-4.39*	
		Quantile	-125.92		-120.79		-3.17		-30.5***		-9.68		14.23		-8.19**	
1	1		0.03	[0]	-0.01	[0]	-0.05**	[0]	-0.04**	[0]	-0.03*	[0]	-0.11***	[0]	-0.2***	[0]
	2	Mean	-14.62		14.92		-6.5*		-4.52		5.35		-0.2		-5.93**	
		Quantile	-23.26		118.48		-8.96		-11.31		11.91		13.78		-14.34***	
1.5	1		0.06	[0]	-0.02	[0]	-0.04	[0]	-0.04**	[0]	-0.02*	[0]	-0.08***	[0]	-0.17***	[0]
	2	Mean	-20.92		-25.32		-4.5		10.33		4.62		-1.43		-3.99**	
		Quantile	-39.51		-92.88		8.19		12.23		11.37		-5.03		-13.65	
2	1		0.08*	[0]	-0.01	[0]	-0.02	[0]	-0.03*	[0]	-0.02	[0]	-0.05*	[0]	-0.1***	[0]
	2	Mean	-14.1		-22.89		6.76		14.66		20.95		-9.18*		-0.24	
		Quantile	-26.12		-146.19		26.91		40.65		44.62		-15.28		-35.72**	
2.5	1		0.08*	[0]	-0.02	[0]	-0.01	[0]	-0.03*	[0]	-0.02	[0]	-0.05*	[0]	-0.04	[0]
	2	Mean	-25.86*		65.75		63.32		12.21		15.76		-10.6*		9.44	
		Quantile	-75.14		114.93		202.61		30.62		11.99		-17.64		44.1	
3	1		0.07*	[0]	0	[0]	0	[0]	-0.02	[0]	-0.02	[0]	-0.04	[0]	-0.02	[0]
	2	Mean	-17.66		1323.76		-6180.56		-15.48		1.03		4.66		26.1	
		Quantile	-59.53		262.42		-18332.42		-59.83		3.4		12.73		50.09	

Notes: For Stage 1 we report the sum of the slope coefficients of the candidate variable, the significance according to the likelihood ratio test, and the number of lags of the candidate variable (in brackets; as determined by the BIC). For Stage 2 we report (both for mean and quantile regressions) the slope coefficients and the significance according to the adjusted one-sided t-test. One, two, and three stars denote significance at the 10%, 5%, and 1% level respectively.

Table 6: Stock returns (half-year) as a candidate indicator of systemic risk

Years ahead	Stage	Regression	CAN		DEU		FRA		GBR		ITA		JPN		USA	
0	1		-0.01	[0]	-0.01	[0]	0	[0]	-0.01	[0]	0	[0]	-0.07***	[4]	-0.01	[0]
	2	Mean	-41.34		-29.61		-21.1		-28.97		-25.54		-5.38***		-33.12	
		Quantile	-76.93		-39.29		-31.6		-66.43		-54.39		-3.95		-45.7	
0.5	1		-0.01	[0]	-0.01	[0]	-0.01	[0]	-0.01	[0]	0	[0]	-0.06***	[3]	-0.01	[0]
	2	Mean	-65.09		-31.45*		-14.8*		-61.13		-17.16		-4.21**		-29.4	
		Quantile	-105.16		-42.28*		-17.35		-91.91		-25.02		-2.78		-118.62	
1	1		-0.01	[0]	0	[0]	-0.01*	[0]	0	[0]	0	[0]	-0.04***	[2]	0	[0]
	2	Mean	-0.48		-21.52		-0.97		-7.8		-8.66		-2.1		-28.75	
		Quantile	0.28		-43.84		-19.14		5.98		-89.22		-8.72		-109.06	
1.5	1		0.01	[0]	-0.01	[0]	0	[0]	-0.01	[0]	0	[0]	-0.02***	[1]	-0.01	[0]
	2	Mean	-51.23		-1.42		0.31		-9.32		-0.18		-0.76		5.14	
		Quantile	-252.45		8.5		8.06		-18.04		-30.22		7.24		52.53	
2	1		0.01	[0]	0	[0]	0	[0]	0	[0]	0	[0]	-0.01***	[0]	0	[0]
	2	Mean	-12.32		26.7		4.48		-30.33		24.64		0.27		7.74	
		Quantile	-116.69		119.99		12.39		-41.07		8.52		4.97		62.33	
2.5	1		0.02	[0]	0	[0]	0	[0]	0	[0]	0	[0]	-0.01**	[0]	0	[0]
	2	Mean	2.94		-0.98		-1.36		30.21		-7.56		-6.96		6.61	
		Quantile	-84.68		35.61		-39.7		75.3		31.12		-1.03		-1.69	
3	1		0.01	[0]	0	[0]	0	[0]	0	[0]	0	[0]	-0.01	[0]	0	[0]
	2	Mean	-31.76		-20.08		-0.62		4.67		-15.89		0.89		71.84	
		Quantile	-83.35		-46.33		-271.53		83.24		-24.13		-6.95		68.89	

Notes: For 1 we report the sum of the slope coefficients of the candidate variable, the significance according to the likelihood ratio test, and the number of lags of the candidate variable (in brackets; as determined by the BIC). For 2 we report (both for mean and quantile regressions) the slope coefficients and the significance according to the adjusted one-sided t-test. One, two, and three stars denote significance at the 10%, 5%, and 1% level respectively.

Table 7: Corporate bond price growth (half-year) as a candidate indicator of systemic risk

Years ahead	Stage	Regression	CAN	DEU	FRA	GBR	ITA	JPN	USA
0	1		0.11 [0]	0.19** [0]	0.28*** [0]	0.12*** [0]	0.55*** [2]	0.25*** [0]	0.14** [0]
	2	Mean	-39.64*	-8.67*	-5.76***	7.61	-1.93	-7.85**	0.79
		Quantile	-52.32	-4.04	-2.62	24.14	3.66	-3.62	28.57
0.5	1		0.08 [0]	0.09 [0]	0.19*** [0]	0.12*** [0]	0.38*** [1]	0.19*** [0]	0.13** [0]
	2	Mean	3.3	14.07	-0.07	7.93	-1.01	-5.68*	13.43
		Quantile	-53.39	-1.02	3.52	27.35	5.6	-2.37	27.94
1	1		0.03 [0]	0.18** [0]	0.13*** [0]	0.11*** [0]	0.32*** [1]	0.15*** [0]	0.08 [0]
	2	Mean	124.55	-1.12	4.23	5.44	-0.37	-6.89**	16.9
		Quantile	298.21	1.12	24.9	20.73	1.54	-19.75	36.66
1.5	1		-0.03 [0]	0.04 [0]	0.11** [0]	0.1** [0]	0.19*** [0]	0.12** [0]	0.04 [0]
	2	Mean	-76.4	48.49	4.43	4.16	-2.16	-5.93	23.68
		Quantile	-180.69	60.38	3.49	33.53	-5.68	-14.97	108.87
2	1		-0.01 [0]	0.14* [0]	0.1** [0]	0.1** [0]	0.39*** [3]	0.12** [0]	0.01 [0]
	2	Mean	-215.82	-6.88	0.17	9.94	-5.34***	-8.85**	52.66
		Quantile	-1228.21	13.5	-5.42	29.01	-7.98**	-5.2	221.42
2.5	1		0 [0]	0.15* [0]	0.12*** [0]	0.09** [0]	0.36*** [2]	0.11** [0]	0.05 [0]
	2	Mean	341.25	-5.38	-0.15	5.02	-3.64**	-2.72	3.99
		Quantile	1621.06	-1.42	-6.78*	45.16	-17.89**	-20.41	49.55
3	1		0.05 [0]	0.16** [0]	0.11** [0]	0.09** [0]	0.28*** [1]	0.08* [0]	0.05 [0]
	2	Mean	20.53	1.02	0.05	4.16	-2.44	-6.13	3.18
		Quantile	143.9	-7.03	8.26	33.4	-1.55	-13.24	70.54

Notes: For Stage 1 we report the sum of the slope coefficients of the candidate variable, the significance according to the likelihood ratio test, and the number of lags of the candidate variable (in brackets; as determined by the BIC). For Stage 2 we report (both for mean and quantile regressions) the slope coefficients and the significance according to the adjusted one-sided t-test. One, two, and three stars denote significance at the 10%, 5%, and 1% level respectively.