BANKA
SLOVENIJE
EVROSISTEM

DELOVNI ZVEZKI BANKE SLOVENIJE/
BANK OF SLOVENIA WORKING PAPERS:

TOWARDS THE DEVELOPMENT OF AN OVERHEATING IDENTICATION FRAMEWORK: THE CASE OF SLOVENIA

Georgios Papadopoulos

4 / 2018

Title/*Naslov*: Towards the development of an overheating identification framework: The case of Slovenia

No./*Številka*: 4/2018

Published by/*Izdajatelj*: BANKA SLOVENIJE
Slovenska 35
1505 Ljubljana
tel.: 01/+386 1 47 19 000
fax: 01/+386 1 25 15 516
http://www.bsi.si

The BANK OF SLOVENIA WORKING PAPERS collection is drawn up and edited by the Bank of Slovenia's Analysis and Research Department (Tel: +386 01 47 19680; Fax: +386 1 47 19 726; Email: arc@bsi.si).
The views and conclusions expressed in the papers in this publication do not necessarily reflect the official position of the Bank of Slovenia or its bodies.

Zbirko DELOVNI ZVEZKI BANKE SLOVENIJE *pripravlja in ureja Analitsko-raziskovalni center Banke Slovenije (telefon: 01/ 47 19 680, fax: 01/ 47 19 726, e-pošta: arc@bsi.si).*
Mnenja in zaključki, objavljeni v prispevkih v tej publikaciji, ne odražajo nujno uradnih stališč Banke Slovenije ali njenih organov.

https://www.bsi.si/publikacije/raziskave-in-analize/delovni-zvezki-banke-slovenije

Katabožni zapis o publikaciji (CIP) pripravili v Narodni in univerzitetni knjižnici v Ljubljani
COBISS.SI-ID=296222464
ISBN 978-961-6960-25-0 (pdf)

Towards the development of an overheating identification framework: The case of Slovenia

Georgios Papadopoulos*

Analysis and Research Centre, Bank of Slovenia, Ljubljana, Slovenia

georgios.papadopoulos@bsi.si

Abstract

The high costs associated with financial crises highlight the need of early identification of vulnerabilities' build up, while the policy makers can effectively implement preventive measures. This study develops a framework that provides such information about the status of the Slovenian economy and in particular issue a warning when it follows an excessively expansionary path. The method used is the univariate signaling approach given its transparency and applicability in data constrained environments. The results from the empirical application indicate that the Slovenian economy is not in a state of overheating for the time being. However, constant monitoring is appropriate because it will give sufficient reaction time to policy makers in case prevailing conditions change.

JEL classification: C52; E37; E44
Keywords: Overheating; Systemic crises; Early warning systems; Signal extraction

1 Introduction

Time and again, in the aftermath of financial crises research in academia and policy making institutions tries to offer insights about the causes of such events and explores new ways to warn policy makers before these materialize. And indeed, policy makers' concerns are well justified. The findings of research on the impact of banking crises (for an early study see Hoggarth et al. (2002)

*The author would like to thank the participants of internal Bank of Slovenia seminars for useful comments and suggestions. This paper should not be reported as representing the views of the Bank of Slovenia (BoS) or the European System of Central Banks (ESCB). The views expressed are those of the author and do not necessarily reflect those of the BoS or the ESCB.

while for a more recent one see Laeven and Valencia (2012)[1]) indicate that – in advanced economies – they can result in a median reduction of GDP by one third in respect to its pre-crisis trend, while in some extreme cases more than halve it. These gloomy prospects for societies' well-being associated with the large costs of financial crises highlight the need for an early warning system which will provide policy makers with timely information about the status of the economy and allow them to consider preemptive actions in case it is necessary.

This study aims to develop a framework which can inform policy makers about the current state of the economy and in particular issue a warning when it follows an excessively expansionary path. The framework is designed to have several desirable properties such as being transparent, data-driven and suitable for data constrained applications. Such properties are especially relevant in a single country context, where the number of observations as well as the number of overheating periods is limited.

This paper is closer in spirit to the work by Hermansen and Röhn (2015) where the authors instead of focusing on predicting a particular type of crisis (currency, banking, etc.) they rather try to predict severe recessions. Similarly, this paper tries to identify patterns in a large set of variables during the run-up to systemic crises and exploits that information to assess the status of specific sectors and ultimately of the whole economy.

This study belongs to the literature strand of early warning system development using the univariate signaling approach. This approach examines the behavior of individual variables around crisis episodes and issues signals defined in terms of specific thresholds. Non-parametric methods are used to extract the trend for each variable while deviations from that trend above a specific threshold signal the occurrence of a crisis within a particular horizon. This approach has the advantage of being transparent, straightforward to apply and is feasible for data constrained applications. After the seminal works of Kaminsky et al. (1998) and Kaminsky and Reinhart (1999) many studies have examined and extended the capabilities of the signaling approach in identifying the build-up phase of crises.

One of the earliest works using this approach is the one of Gourinchas et al. (2001). The main purpose of their study is to identify lending boom episodes in 91 countries over the period 1990 - 1996. Thus, it is indirectly related to financial or banking crisis prediction in so far as lending booms are associated with such events (Corsetti et al., 1999). An early example of applying the signaling approach to predict banking crises is the work by Borio and Lowe (2002a). In that study the authors introduce the concept of "gaps" ie. variables' deviations from their long-term trends and also propose the construction of a composite indicator based on the warning signals issued by the individual ones[2]. A non-exhaustive list of studies applying the signaling approach includes works which aim to predict financial distress (Ito et al., 2014), recessions (Hermansen and

[1]The authors also consider twin and triplet crises (currency - sovereign debt - banking) but they don't distinguish the effects among the different combinations.

[2]In an extension of her previous work Kaminsky (1999) has also examined the performance of four composite indicators.

Röhn, 2015), banking (Borio and Drehmann, 2009; Alessi and Detken, 2009), sovereign debt (Knedlik and Von Schweinitz, 2012) and currency crises (Edison, 2003). For an early survey of the approaches used for predicting currency crises the reader is referred to the work of Abiad (2003). Regarding approaches used to predict banking crises, the work of Kauko (2014) provides excellent reference.

The second major branch in the early warning system development literature employs multivariate methods. These incorporate information from many variables into one single indicator denoting the probability of a crisis occurring in a specific horizon. The most common method used is logit/probit regressions (Demirgüç-Kunt and Detragiache, 1998, 2005; Bussière and Fratzscher, 2006; Caggiano et al., 2016, among others) but other choices range from Markov switching models (Abiad, 2003) to machine learning methods (Holopainen and Sarlin, 2017; Joy et al., 2017). The advantage of these methods is that they provide a unifying framework and combine information from many indicators into a single measure. Nevertheless, this comes at a cost of being considerably more data demanding (in terms of number of observations as well as crisis events) and being computationally complex.

Finally, studies comparing the performance between the signaling and several multivariate approaches suggest that while the latter are "*better suited to a global early warning system (EWS)*", the former "*may be better suited to country-specific EWS*" (Davis and Karim, 2008, p. 117). Thus, taking also into account the limitations imposed by data availability, the signaling approach will be the one used for the development of the overheating identification framework for Slovenia.

The paper is organized as follows. Section 2 describes the data used for the empirical application. It briefly discusses the selection of variables and provides a summarized picture of their evolution prior to the global financial crisis. Section 3 presents the methodology employed for the development of the overheating identification framework. Section 4 presents the results of the empirical application in Slovenia. In addition, it presents the results of the robustness analysis which tests the out-of-sample performance of the framework and the stability of the indicators used. Finally, section 5 concludes.

2 Data

In search for overheating patterns in Slovenia, the study uses a group of variables covering a large part of the economy. The selection of variables is guided by the literature of early warning indicators (Borio and Lowe, 2002b; Borio and Drehmann, 2009; Drehmann et al., 2011; Hermansen and Röhn, 2015) and by the scope of the framework under development which is to provide an overall picture of the economy's state.

As pointed out in the study of Borio and Lowe (2002b), rapid growth in credit or asset prices alone poses little threat to the economy but it is "*rather what **combination** of events in the financial and real sectors exposes the financial system to a materially increased level of risk*" (Borio and Lowe, 2002b, p.

11). In two recent studies Röhn et al. (2015) and Hermansen and Röhn (2015) examine a large set of indicators covering both domestic and international sides for assessing the economic resilience of OECD countries. Their results provide empirical evidence that many indicators, in addition to the commonly used credit and asset market related ones, exhibit good signaling performance and would have been helpful in uncovering a country's vulnerabilities before a severe recession hits the economy.

Therefore, analysing a pool of variables has several benefits. First, it can uncover specific sectors of the economy that might be excessively growing and would otherwise be overlooked. Moreover, it allows for an assessment of the overall economic environment by providing information of the developments in various parts of the economy.

In this study 18 variables are examined capturing developments in asset markets, financial, non-financial and public sectors as well as overall economic conditions. Most data are collected from the Statistical Office of the Republic of Slovenia (SORS) while other sources include internal Bank of Slovenia (BoS) databases, Eurostat and Reuters. Data are in quarterly frequency and span from 2000q1, or as early as possible, until the most recent available which for most variables is 2017q3. Data prior to 2000q1 are not considered because the transition process of the previous years affects the behaviour of the variables.

In Table 1 are reported the variables examined in the development of the framework, their description, time range and source of data.

[Table 1 about here]

In addition to covering a broad part of the economy, the selection of these specific variables serves the purpose of comparing and cross-checking the results with similar tools used internally based on expert judgment.

Table 2 provides a summarized picture of the variables' evolution before the onset of the 2008 crisis. More specifically, it shows the minima, maxima, means and standard deviations for each of the 18 variables of Table 1 in two subperiods; from the beginning of the sample until 2005q3 (Period 1) and during the 3 years preceding 2008q4 (Period 2), when the crisis event occurred. It also reports the results from distributional equality tests, namely Levene's W_0 test for equality of variances (Levene, 1960) and two tests for the equality of distributions; Kolmogorov-Smirnov (KS) and Epps and Singleton (ES (Epps and Singleton, 1986) tests.

[Table 2 about here]

The evolution of the variables presented in Table 2 reveals an interesting pattern. During the second subperiod, 3 years before the emergence of the global financial crisis, the vast majority of variables follows a significantly stronger expansionary path compared to their earlier evolution.

This is reflected in statistically significant differences in distributional properties between the two subsamples. First, several variables exhibit considerable changes in their volatilities during the second subperiod. Levene's W_0 test for equality of variances between the two subperiods results in the rejection of the null of equal variances at 5% significance level for CONSTINV, CINF, HOUSEP, CRDTOT, CRD2NFC, URT and SBI. With the exception of CINF whose variation is halved in the second subperiod, the aforementioned variables have approximately twice as large standard deviations compared to their first subperiod figures. Tests for the equality of distributions uncover statistically significant differences between the two subperiods. Both Kolmogorov-Smirnov and Epps and Singleton (implemented in Stata by Goerg and Kaiser (2009)) tests reject the null that the two subperiods have been drawn from the same population at 5% significance level for every variable apart from TRADEBAL, COMP, COMPXOPQ and SBI. Two marginal cases are CONSTINV and UNCERT where equality of distributions can be rejected at 10% significance level.

These shifts in distributional properties such as spread as well as differences in the distribution itself, indicate that the variables' underlying data generating processes undergo changes while moving towards unstable conditions. This study examines whether these changes can be exploited in order to extract information when the economy is in an excessively expansionary path, thus providing policy makers an early warning while there is sufficient time for preemptive action.

3 Methodology

The methodology adopted for the development of the overheating identification framework is based on the univariate, signaling approach. Under this approach, a signal is issued whenever a variable exceeds a specific threshold above or below its past trend. This approach meets all the desired properties of simplicity, robustness and applicability in data constrained environments such as the case of an individual country with limited number of observations and a single crisis event in its recent history.

The procedure followed aims to identify patterns in the evolution of the variables studied which could signal that the economy is in a state of potentially unsustainable expansion. In particular, the search for these patterns will focus in the period between 1 and 3 years before a crisis event. The choice of this specific window is standard in the literature (Behn et al., 2013, 2016; Sarlin, 2013) since it gives sufficient time to policy makers to design and implement any needed measures for reducing or even reversing the build-up of vulnerabilities.

The procedure comprises of the following elements:

- Extraction of each variable's trend
- Definition of the "normal" conditions range
- Assessment of the signaling performance of each indicator.

Each of these elements is described in detail in the remainder of Section 3.

3.1 Trend extraction

The trend for each variable is extracted using three methods; simple moving average, exponential moving average and the, frequently used in the early-warning literature, Hodrick-Prescott (HP) filter (Hodrick and Prescott, 1997).

Among all trend extracting methods, the simple -or equally weighted- moving average is the less sophisticated one and therefore it is a natural starting point. It is defined as:

$$SMA_t^{y,n} = \frac{1}{n} \sum_{i=0}^{n-1} y_{t-i} \tag{1}$$

where y is the variable examined and $n-1$ the number of past observations included in the moving average. Three different windows are used in the empirical part, ranging from a quickly adjusting 6 quarter window (SMA_6) to a slowly adjusting 16 quarter window (SMA_{16}). The 8 (SMA_8) and 12 quarter windows (SMA_{12}) are considered as two moderately adjusting schemes.

A slightly more complicated version of the moving average is the exponentially weighted moving average. It is defined as:

$$EMA_t^{y,\alpha} = \begin{cases} y_1, & t = 1 \\ \alpha \cdot y_t + (1-\alpha) \cdot EMA_{t-1}^{y;\alpha}, & t > 1 \end{cases} \tag{2}$$

The weights assigned to each observation decrease exponentially with more recent receiving larger weights than the older ones. The coefficient $0 < \alpha < 1$ determines the speed of moving average's adjustment, that is how quickly the weights for the oldest observations will decrease. Small values of α will discount older observations more slowly, whereas large values will quickly discount them. In the empirical application two adjustment levels are considered; a fast one (EMA_f) with $\alpha = 0.8$ and a more slowly moving one (EMA_s) with $\alpha = 0.3$.

The last method used to extract each variable's trend is the widely used HP filter. Two of the earliest uses of this method for developing early warning indicators include the works of Gourinchas et al. (2001) and Borio and Lowe (2002b). Ever since, it has become the standard approach in the respective literature, in particular the strand following the signaling approach. A certainly non-exhaustive list of studies using the HP filter includes central banks (Alessi and Detken, 2009; Behn et al., 2013; Gerdrup et al., 2013; Ito et al., 2014), international institutions (Borio and Drehmann, 2009; Drehmann et al., 2011; Mitra et al., 2011; Hermansen and Röhn, 2015) as well as private financial intermediaries (Weistroffer and Vallés, 2008; Lanzeni and Weistroffer, 2013).

Using the HP filter, the trend y_t^* of a variable y_t is estimated by minimizing Equation 3 below:

$$\min_{y_t^*} \left(\sum_{t=1}^{T} (y_t - y_t^*)^2 + \lambda \cdot \sum_{t=2}^{T-1} \left[(y_{t+1}^* - y_t^*) - (y_t^* - y_{t-1}^*) \right]^2 \right) \tag{3}$$

The smoothing parameter λ controls the speed of the filter's adjustment or, equivalently, the level of smoothing. Higher values of λ imply slower adjustment, whereas lower values correspond to faster one. Following Hermansen and Röhn (2015), two values for λ are considered; a slowly adjusting trend (HP_s) with $\lambda = 400000$ and a fast adjusting one (HP_f) with $\lambda = 26000$.

It should be mentioned that trends are estimated in a quasi-real time manner so that they reflect, to the extent possible,[3] only the information available at the policy maker at each point in time. For both simple and exponential moving averages this is done by construction. For the case of HP filter the commonly used one-sided, recursive estimation is applied.

The next step in the procedure is closely linked to the previous one and focuses on the definition of "normal" conditions.

3.2 Defining "normal" conditions

The definition of "normal" conditions is based on the estimation of each variable's gap, i.e. its deviation from its trend. Under this approach, large gaps -which can be the result of either rapid growth in a short period of time or of persistent growth above a variable's trend over an extended period (Borio and Lowe, 2002b)- are interpreted as a signal of instability build-up. Thus, overheating is defined as a variable's excessive deviation from its trend.

Following Ito et al. (2014), the deviation of each variable from its trend is defined as:

$$\sigma_t^{y,m} = \sqrt{\frac{1}{N_T - 1} \sum_{t=1}^{N_T} (y_t - y_t^{*,m})^2} \tag{4}$$

with N_T denoting the number of observations up to time T and $y_t^{*,m}$ the trend of variable y_t as estimated by method m, using Equations 1 to 3 respectively. As mentioned, eight trend extracting methods are used in this study, hence $m = \{SMA_6, SMA_8, SMA_{12}, SMA_{16}, EMA_f, EMA_s, HP_f, HP_s\}$. Similar to trends, $\sigma_t^{y,m}$ is estimated recursively incorporating information available up to the point of estimation.

Finally, the "normal" conditions range for each variable is defined as τ times $\sigma_t^{y,m}$ around its trend. Since the threshold above which a deviation can be regarded as excessive is not known, its level is determined by the resulting signaling performance examined over a range of values for τ. In the empirical application four different values for τ are considered; 0.75 $\sigma_t^{y,m}$, 1 $\sigma_t^{y,m}$, 1.25 $\sigma_t^{y,m}$ and 1.5 $\sigma_t^{y,m}$. The threshold yielding the optimal[4] signaling performance for each specific variable - trend extracting method pair is the one which will be used in the analysis.

It should be noted that classifying a threshold breach as a warning signal is linked to the variable examined. Therefore, for variables such as CA, TRADE-

[3] Due to lack of vintage data only the latest information available is used. That is, data revisions are not taken into account.

[4] In the sense of achieving the maximum amount of correct alarms at the lowest possible number of false ones.

BAL, UNCERT and URT, surpassing the lower threshold (i.e. rapid decline) will be an indication of overheating. For the remaining variables exceeding the upper boundary constitutes a warning signal.

Each specific combination of $\{y, m, \tau\}$ will be referred hereafter as **indicator**. Also, each particular combination of $\{m, \tau\}$ for a certain variable y will be mentioned as **subindicator** of y. The signaling performance of the various subindicators is assessed using various criteria from the respective literature as described below.

3.3 Signaling performance assessment

In total, 576 (18 variables, y × 8 trend extracting methods, m × 4 thresholds, τ) indicators are considered in the analysis.

Their performance assessment is based on their ability to provide early warning signals during the period preceding the materialization of a crisis, while at the same time displaying the lowest acceptable number of false alarms. More specifically, the warning period considered in the analysis spans from 4 to 12 quarters before the onset of a crisis. This particular window is commonly used by the literature (Behn et al., 2013, 2016; Sarlin, 2013) since it leaves enough time for the policy makers to investigate further and take appropriate action if needed.

The measures used to assess the signaling performance of the examined indicators are Accuracy, policy maker's loss function L (Sarlin, 2013) and the AUROC (Area Under Receiver Operating Characteristic) curve. All these measures are based on the contingency matrix presented in Table 3.

[Table 3 about here]

Each observation of a given indicator resides in one of the four quadrants of Table 3. If a warning signal issued is followed by the occurrence of a crisis 4 to 12 quarters ahead, then this is considered as a True Positive (TP). However if it is not, then this is considered a False Positive (FP). Conversely, if no warning signal is issued but a crisis occurred $4 - 12$ quarters in the future, then this is classified as a False Negative (FN). Similarly, if no warning signal is issued and there is no event in the given horizon, then this is regarded as a True Negative (TN). Two types of errors can be distinguished from Table 3. Type 1 errors are associated with missing a crisis and are defined as $T_1 = \frac{FN}{TP+FN}$, while Type 2 errors denote false alarms and are defined as $T_2 = \frac{FP}{FP+TN}$.

Accuracy is defined as the ratio of correct signals (TP + TN) over the total number of signals, defined in Equation 5:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

The closer the Accuracy of an indicator is to 1, the better that indicator is.

The second measure for assessing the performance of an indicator, is the policy maker's loss function which is initially introduced by Bussière and Fratzscher

(Bussière and Fratzscher, 2006, 2008) and later extended by the works of Alessi and Detken (2009) and Sarlin (2013). The policy maker's loss function L is defined as follows:

$$L(\mu, \tau) = \mu P_1 T_1(\tau) + (1 - \mu) P_2 T_2(\tau) \tag{6}$$

where $\mu \in [0, 1]$ is the policy maker's preference parameter, $T_1(\tau)$ and $T_2(\tau)$ are Type 1 and Type 2 errors respectively and τ the threshold used. With P_1 is denoted the ratio of the number of periods in which a signal should be issued to the total number of observations, $P_1 = \frac{TP+FN}{TP+TN+FP+FN}$ while P_2 is its complementary, $P_2 = 1 - P_1$.

The preference parameter μ indicates the type of errors that a policy maker is more averse to making. Larger values of μ emphasize aversion towards missing a crisis (Type 1 errors), whereas smaller values emphasize aversion toward false alarms (Type 2 errors). For a perfectly balanced policy maker, μ would assume the value of 0.5.

In the early works using L, the values of μ for horizons similar to those examined in this study are set to below 0.5 (Bussière and Fratzscher, 2006). However, despite the fact that they also regard μ's close to 0.5 be a reasonable choice for a central banker, Alessi and Detken (2009) note that "*the recent financial crisis might have increased the average value [of the preference parameter]*" (Alessi and Detken, 2009, footnote 15). Indeed, various recent studies (Behn et al., 2013, 2016; Ito et al., 2014) use μ's above 0.8, thus putting more emphasis in avoiding Type 1 errors. In this study the preference parameter is set to $\mu = 0.85$.

The final statistical measure used to assess indicators' performance is the AUROC curve. This measure is used in the empirical literature of early warning system development (Drehmann and Juselius, 2014) to evaluate the discriminatory power of a signaling method. The closer the value of the AUROC curve is to 1, the higher the discriminatory power of that method. On the contrary, a completely uninformative method would have an AUROC = 0.5

In order to identify the best performing indicators and, similar to Lo Duca et al. (2017), to narrow down the list of indicators used, three conditions must be met. To be considered in the final result a subindicator of variable y should exhibit Accuracy ≥ 0.65, $L \leq 0.2$ and AUROC ≥ 0.7. It should be noted that for any variable y only the best performing subindicator is retained.

3.4 Addressing the "post-crisis bias"

Before proceeding further with the analysis, the "post-crisis bias" is dealt. The effect of the conditions prevailing during and just after crises in the development of early warning systems is known as the "post-crisis bias" in the literature (Bussière and Fratzscher, 2006). The fact that the behaviour of the variables examined is very different during tranquil times as compared to recovery episodes can lead to an important bias. Especially in the case of this study, trend extracting methods are affected by crisis events, resulting in incorrect classification of recoveries as overheating episodes. Crisis events affect the framework by pushing trends substantially downwards, thus misclassifying subsequent upward movements as overheating periods. This problem can be mitigated by excluding the data around crisis periods as in Demirgüç-Kunt and Detragiache (1998).

As is obvious, dealing with the "post crisis bias" requires information about the starting dates and duration of crisis events. Exact dating of crises is a challenging task and very active field of research. Various approaches show significant overlap but also display differences regarding the specific dates of various crisis episodes (Kaminsky and Reinhart, 1999; Laeven and Valencia, 2008, 2010, 2012; Reinhart and Rogoff, 2011). The dates used in the empirical application are taken from the ECB/ESRB EU crises database (Lo Duca et al., 2017), in particular the ones referring to events classified as "systemic".

4 Results

In the empirical part of the study the aforementioned methodology is applied to the Slovenian economy to provide an assessment of its current state. Using the period preceding the global financial crisis of 2008 as the "calibrating event", the various parameters of the framework are determined based on their signaling performance as described in Section 3. Also, given that the use of a single event to identify the best performing indicators raises concerns about the out-of-sample performance of the framework, its robustness is tested in a similar context using data for Germany and examining its signaling ability prior to 2008.

4.1 The case of Slovenia

The application of the described methodology on the data of Table 1 reveals some interesting patterns regarding the evolution of the Slovenian economy over time. In addition, the analysis provides useful information regarding the signaling performance of the employed indicators.

For the case of Slovenia the early warning period within which the indicators examined are desired to provide a warning signal is between 2005q4 and 2007q4. As can be seen, the crisis is assumed to have started in 2008q4 which is the first occurrence of negative annual GDP growth. According to the ECB/ESRB EU crises database (Lo Duca et al., 2017) the crisis lasted until 2014q4[5] thus, this entire period is excluded from the analysis.

Following the procedure described in Section 3, eight trend extracting methods and four thresholds are used, therefore generating 32 subindicators per variable. The first step in the selection process is to estimate each subindicator's Accuracy and loss function L. The period during which these measures are estimated is between 2000q1 (2002q1 for GDEBT and CINF) and 12 quarters

[5] More precisely, it is noted that the *crisis management* lasted until 2014q4. However, since the economy is not classified to have returned "back to normal" yet, the previous date is used for the rest of this study as the end of the crisis.

before the last available observation. The rationale behind the latter cut-off point is that any signal is tested against reality up to a maximum of 12 quarters in the future. Thus, the predictive horizon of signals issued more recently than the past 12 quarters extends to the future where information about the occurrence of a crisis event is evidently not available. In the empirical application for Slovenia this essentially means that the performance assessment period spans from the beginning of the sample until 2008q3 since the 12 quarter mark before 2017q3 lies within the exclusion window of 2008q4 – 2014q4.

As a representative example, the Accuracy versus L of GDP's subindicators is plotted in Figure 1 showing several interesting findings.

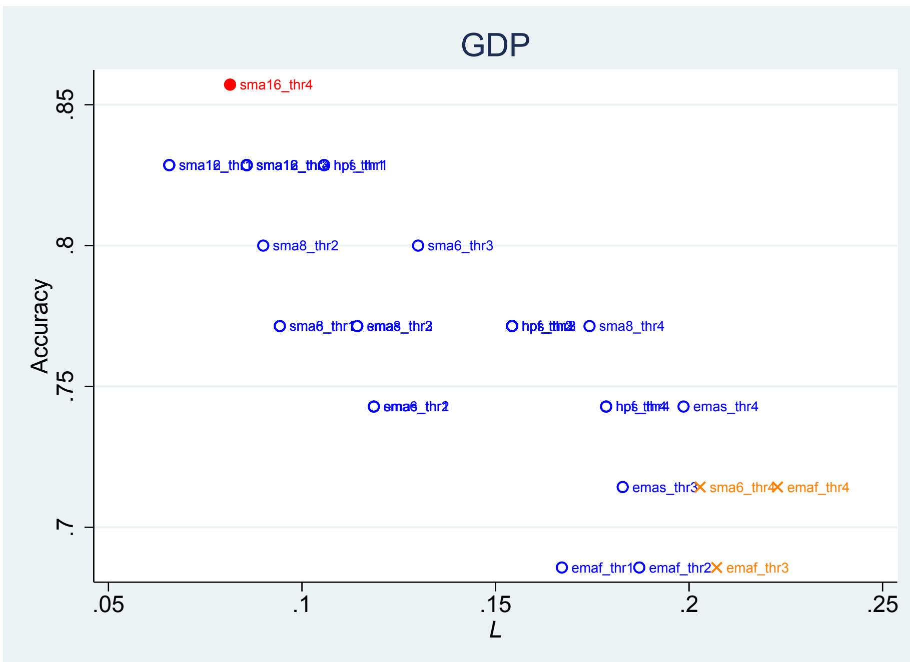

Figure 1: Signaling performance of GDP's subindicators. With red filled circle is marked the preferred one, with blue hollow circles the ones that meet the desired performance conditions and with orange X's those that don't.

A first observation is that many subindicators exhibit Accuracy ≥ 0.65 and $L \leq 0.2$. This is the case for many variables, however for some of them only few subindicators meet the desired criteria for further consideration as can be seen in Appendix A. Moreover, for 8 of them none meets the criteria and therefore these variables are excluded from the subsequent analysis. Another pattern that is also exhibited by three additional variables (GDEBT, HOUSEP and SENTI) is that the subindicator with the highest Accuracy doesn't have also the lowest L. In all cases, the difference between the two choices (maximizing Accuracy or mininizing L) results in the selection of a different threshold, while the best performing trend extracting method remains the same (SMA_{16}). Maximizing Accuracy tends to choose a higher threshold, whereas minimizing L selects lower ones. This is due to the fact that the employed preference parameter $\mu = 0.85$ puts substantial weight on the avoidance of Type 1 errors (missing a crisis), therefore low thresholds which issue more warning signals are preferred over higher ones which tend to produce fewer warnings. Despite this difference, the results are qualitatively similar and this measure ordering will be used in the analysis. A final interesting pattern, common in all variables, is that simple moving average schemes (and in the majority of cases, the slowly moving ones) outperform both exponential moving averages as well as the HP filters. This finding corroborates the results of similar studies (Ito et al., 2014; Hermansen and Röhn, 2015) which have documented a very good performance of simple moving averages in early warning system development. While the HP filter is the predominant approach in the early warning literature, these findings warrant further investigation of the performance of alternative methods for trend extraction and their incorporation in early warning systems.

The final step is to assess the signaling performance of the selected subindicator (marked with a red, filled circle in Figure 1) using the AUROC curve.

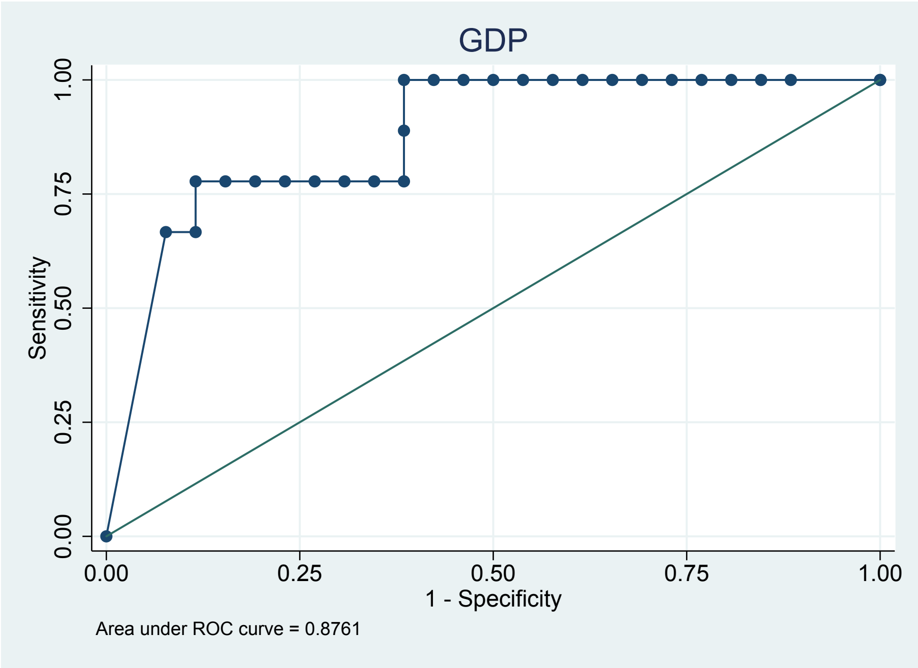

Figure 2: The AUROC curve for the preferred subindicator of GDP.

In Figure 2, the correct warning signal ratio or $Sensitivity = \frac{TP}{TP+FN}$ is plotted against the noise ratio or $1 - Specificity = 1 - \frac{TN}{TN+FP}$ for the selected subindicator for GDP. As shown, this subindicator has substantial discriminatory power with the area under ROC curve being around 0.88. Thus, its signals will be used in the early warning framework.

For illustrative purposes, the evolution of the "normal" conditions range based on the aforementioned subindicator for GDP is presented in Figure 3. The grey area extends $1.5 \cdot \sigma_t$ times around GDP's SMA_{16} trend, while the orange area denotes the period within which a warning signal is desired to be issued. The empty area between 2008q4 and 2014q4 where only the historical values of GDP are plotted is the crisis period which is excluded from the analysis. As is obvious, during that period the behaviour of GDP is completely different compared to the rest of the sample before and after it.

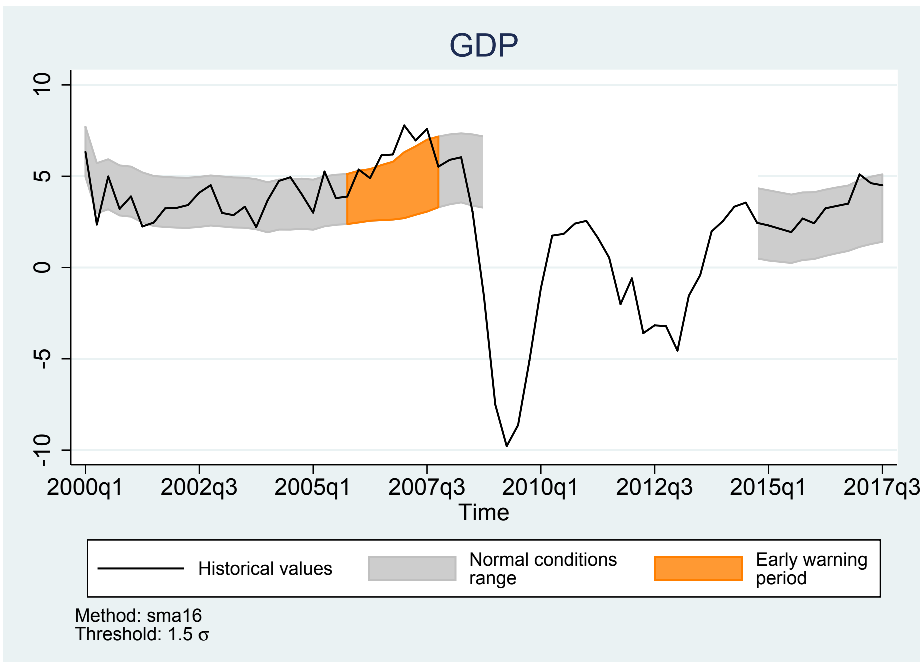

Figure 3: The evolution of GDP and its "normal" conditions. Also noted is the trend extracting method and the threshold used.

One can see that for the most part GDP stays within the "normal" conditions range. However, before 2008q4 there were several occasions where it was very close to the upper boundary and even breached it. Before the warning period this pattern is observed in the second half of 2004 and, after a short correction, in 2005q2 when GDP is either very close to or surpasses the "normal" conditions boundaries. A possible explanation is that the indicator captures the favourable economic environment and buoyant sentiment (a pattern observed in several other indicators as well; see Appendix A) following Slovenia's accession in the EU in May 2004.

During the warning period, GDP consistently exceeds the limits from 2006q1 until 2007q3[6] when it started its declining path. Thus for the most part of the desired period this indicator issued warning signals which were followed by a crisis event.

At the right edge of Figure 3, GDP is identified to have marginally gone beyond the ”normal” conditions range in 2017q1 while in the quarters that followed it returned back within limits and is currently moving slowly towards its long-term trend.

The output from all indicators satisfying the desired performance criteria is presented in a compact and informative way in the form of a heatmap in Figure 4. Warm colors denote an excessively expansionary evolution of the respective variable while cold colors an excessively contractionary one.

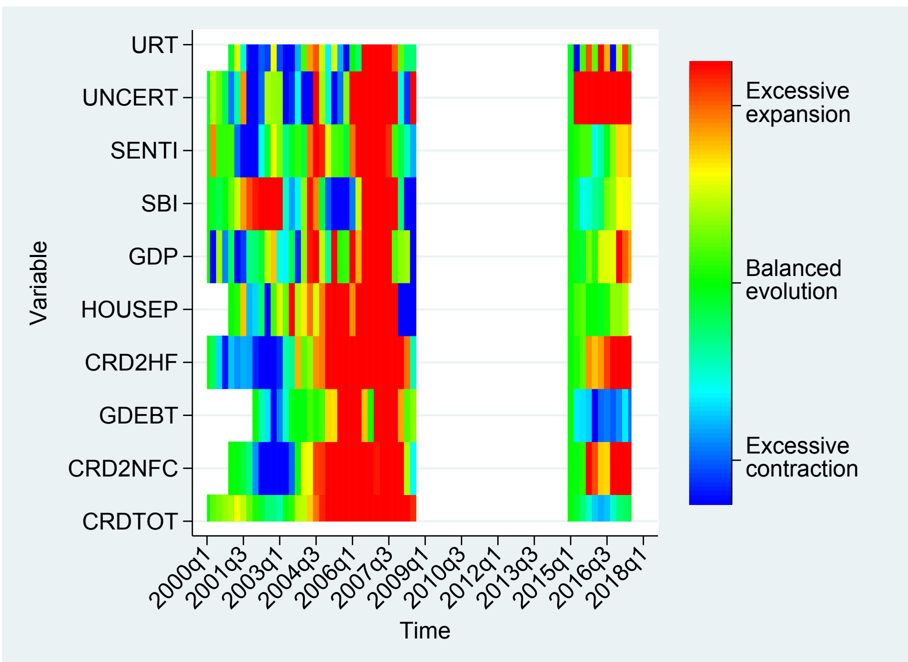

Figure 4: The overheating heatmap for Slovenia.

As can be seen in Figure 4, from a total of 18 variables (576 subindicators), 10 meet the signaling performance conditions. As of 2017q3 most of them lie within the ”normal” conditions range with one of them even be at the lower end of the scale (GDEBT).

Another group of variables is CRDTOT, HOUSEP, SBI, SENTI and URT. These exhibit a balanced evolution which means that their current values are close or slightly exceeding their long term trends. Nonetheless, with the exception of URT which has a quite volatile pattern and CRDTOT which is very

[6]With the exception of 2006q2 when it is very close to the upper boundary.

close to its long term trend, all the remaining variables of this group show mildly upward trends towards their upper boundaries. This can be seen in the trend plots in Appendix A and in their colors in Figure 4 which change from dark to light green shades (HOUSEP and SBI) or from yellowish to orange (SENTI). Thus, while not yet alarming, their developments should be followed in case their growth further accelerates and goes beyond limits.

Apart from the variables which are in the "safe zone", there are some which already display excessive deviations from their trends and warrant closer monitoring. These variables are CRD2NFC, CRD2HF, UNCERT and GDP.

Starting with the variables related to the growth rate of loans to households and NFCs (CRD2HF) and solely to NFCs (CRD2NFC) the indicators suggest that both are surpassing their trends by a large margin. However, a closer inspection of the respective plots (see Appendix A) reveals that both variables have recovered from a significantly negative territory following the end of the crisis in 2014q4 and only recently returned into positive ground. In addition, the indicator related to the total credit to GDP ratio (CRDTOT) is well within limits. Therefore, taking all information into account, it is reasonable to assume that credit growth is not a major concern for the time being.

The signals from the subindicator for uncertainty (UNCERT) indicate a persistent and excessive decline after 2014q4. Observing the evolution of UNCERT's historical values (see Appendix A) it is evident that it hasn't yet returned to its pre-crisis levels. Hence, despite the undeniable fact that it is declining at a very fast pace as captured by the respective indicator, perhaps for this specific variable the window required to overcome the "post-crisis bias" is longer.

Regarding annual real GDP growth (GDP), as discussed above and can be seen in Figure 3, since 2017q1 when it exceeded its "normal" conditions threshold is in a mild path of return within boundaries. Indeed, in the last two observations (2017q2 and 2017q3) its values exhibit a slow movement towards its long term trend. Nevertheless, attention is still needed in order to draw firm conclusions regarding its status.

A final output which summarizes the signals from each individual indicator and provides an overall picture of the status of the economy based on those, is a composite indicator. Several methods have been examined in the literature for developing composite early warning indicators. The approach of Borio and Lowe (2002a) assumes that a composite indicator issues a warning signal when all its constituent indicators exceed their respective thresholds. The method followed in this study is closer to the one of Kaminsky (1999) where the composite indicator is constructed as a weighted average of the individual signals based on their performance[7]. In this study two simple methods of aggregation are considered for the construction of composite indicators; the equally weighted average and the median of the signals from all individual indicators.

[7] In particular, Kaminsky (1999) uses the inverse noise-to-signal ratio of each indicator to weight their signals.

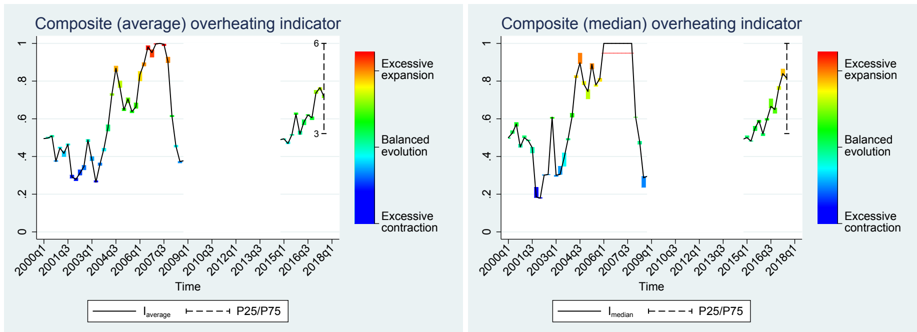

Figure 5: Average (left) and median (right) composite indicators.

In Figure 5 is presented the evolution of the average and median composite indicators. The composite indicators range from 0 (excessive contraction) to 1 (excessive expansion) and follow the same colouring convention as in Figure 4. It should be mentioned that the numbers do not represent the probability of a crisis occurring but rather the average (median) signal from the pool of indicators.

A potential problem with using such aggregation methods for the development of composite indicators is that the final result depends heavily on the number of individual indicators used and their signals' distribution. If too few indicators are used then the final outcome could misleadingly point to either in the direction of overheating or the opposite. As regards with the distribution of signals, it could be the case that some indicators exhibit extreme signals which will dominate in the composite one.

Thus, the results should be interpreted with caution. Of course, when many indicators point to the same direction it will be reflected in the composite one and will strengthen the confidence in the signal issued. Given that the average composite indicator is more vulnerable to extreme values, a visual aid about the distribution of the individual signals is also plotted. The capped vertical dashed line in the last observation shows the first and third quartiles of the signal distribution at that point in time. In addition, the numbers at the edges denote the number of indicators giving a signal below (bottom edge) and above or equal to (top edge) the average. In the case of the median composite indicator such information is not provided since by construction it is expected to be in the middle of the range.

Figure 5 reveals some interesting patterns. As expected, both composite indicators are capturing well the pre-2008 period displaying either clear signals (equal to 1) or at least substantially high values from as early as 2006q1. An interesting feature present in both plots is the elevated figures during the second half of 2004. Despite that the framework is calibrated to a different event, it detects a period of high growth (as opposed to overheating) well before 2008. Indeed, 2004's Annual Report of the Bank of Slovenia (Bank of Slovenia, 2004) confirms these developments. As has been recorded at the time, "*economic* *growth was relatively high*" and primarily influenced by a "*high level of activity by financial intermediaries*" (Bank of Slovenia, 2004, p. 12). In addition, it was observed that "*housing prices rose higher than average*" (Bank of Slovenia, 2004, p. 15). All these facts are captured individually by the respective indicators in Figure 4 and also reflected in both composite indicators.

Comparing the two composite indicators one can see that the median exhibits sharper changes in its evolution as opposed to the average one. Apart from that difference, their patterns are similar and both point to the direction that currently the economy does not suffer from overheating pressures. Looking in particular at the average composite indicator in the left panel of Figure 5, one can observe that the signals from individual indicators are balanced with 5 having higher and 4 lower values than the average[8].

It should be mentioned that the results rest on two fundamental assumptions; indicators' performance being constant over time and recurrence of past patterns yielding similar outcomes in the future. Having that in mind, the graphs in Figure 4 are Figure 5 indicate that the Slovenian economy is growing at a largely balanced pace without any severe signs of overheating for the time being. However, close monitoring of the developments in credit supply and asset prices is warranted so that preemptive action can be taken early on if deemed necessary in case prevailing conditions change.

4.2 Robustness check

The purpose of the robustness check in this study is twofold. First, it aims to test the out-of-sample signaling performance of the framework and second to assess the stability of the identified indicators across time.

For that reason the dates of the systemic events identified in the ECB/ESRB EU crises database (Lo Duca et al., 2017) are used. An obvious requirement for a country to be considered in the robustness check is that it must have experienced more than one crisis events. Moreover, the fact that data availability is limited for many of the variables examined introduces an additional constraint. Therefore, given that some year's data are needed to estimate each variable's trend[9], events that happened before late 1990s can hardly be exploited.

For the majority of EU countries the global financial crisis of 2008 and the subsequent sovereign debt crisis have been the most recent systemic events as identified in the ECB/ESRB EU crises database. However, few of them experienced another systemic crisis around late 1990s or early 2000s. The small list of potential candidate countries includes Cyprus (crisis start date 2000m1), Croatia (crisis start date 1998m4) and Germany (crisis start date 2001m1). From these three countries Croatia is excluded because the systemic event is classified as a "transition" one, thus its very nature is different from what this

[8] Due to the fact that data for HOUSEP are missing for 2017q3, the total number of indicators is 9 instead of 10 at that point in time.

[9] Drehmann and Juselius (2014) use a minimum of 6 years of data to ensure that the estimated trends are sufficiently stable.

framework aims to capture. Also Cyprus cannot be considered in the robustness check due to missing data for many of the variables of interest.

Therefore, Germany is used for testing the predictive power of the framework and the stability of the various indicators. The period of the "calibrating" crisis event spans from 2000q4 until 2003q4 and will be excluded from the analysis as mentioned previously in order to mitigate the "post crisis bias". Consequently, the early warning period ranges from 1997q4 up to 1999q4, i.e. 12 to 4 quarters before the crisis started.

The data used are collected from various sources attempting to use as similar variables as possible -conditional on their availability- to the ones used in the application for Slovenia. A description of the data and their sources is provided in Table 4.

[Table 4 about here]

Overall, 480 (15 variables, y × 8 trend extracting methods, m × 4 thresholds, τ) indicators are examined by applying the same procedure as for the case of Slovenia on the data of Table 4 until the cut-off date of 2006q3. This is one year before the most recent systemic crisis event in Germany as reported in the ECB/ESRB EU crises database. Despite all efforts to use as similar data as possible compared to the ones used for Slovenia, it wasn't possible to include several of them due to data availability. Moreover, for some of them the closest available proxies are used such as gross capital formation (CAPFORM_{DE}) instead of gross fixed investment (FIXEDINV) and the index of real unit labour cost based on persons (RULC_{DE}) in place of nominal unit labour cost (ULC).

It should be noted that according to Lo Duca et al. (2017), the systemic crisis that hit Germany in 2007q3 originated purely from abroad[10]. This could potentially affect the results since domestic economy at the time might have been at a balanced evolution path which was nevertheless disrupted by an exogenous event.

As previously, the results are presented as a heatmap in Figure 6, followed by the average and median composite indicators in Figure 7. A more detailed picture of the individual indicators and their performance can be found in Appendix B.

[10] On the contrary, the origin of the recent crisis in Slovenia is classified to be both domestic as well as external (Lo Duca et al., 2017).

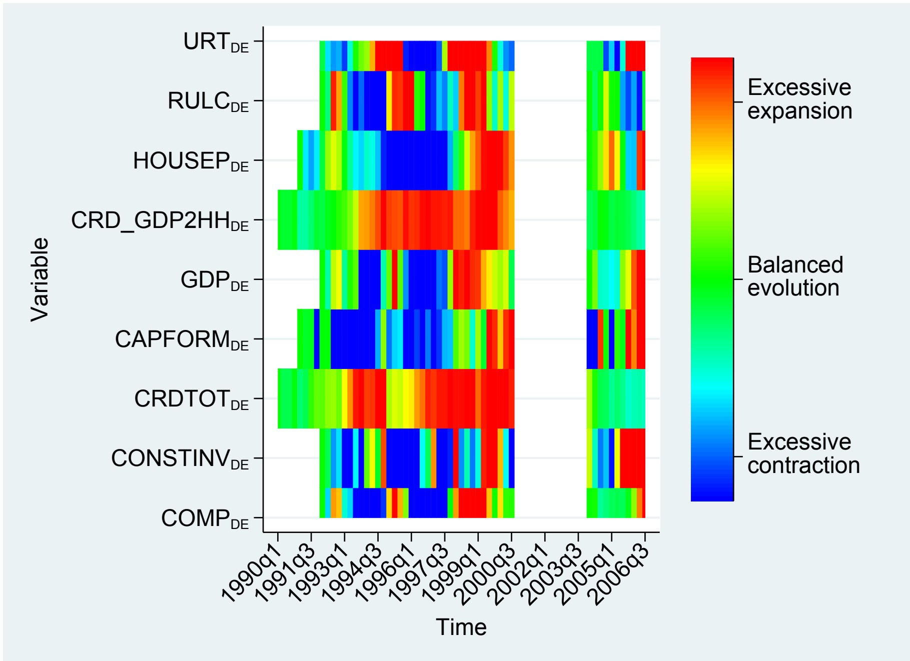

Figure 6: The overheating heatmap for Germany up to 2006q3.

In Figure 6 one can see that the subindicators from 9 variables meet the desired performance conditions. As expected, they all issue warning signals in the period from 1997q4 until 1999q4 while some of them are warm coloured also before, especially around the first half of 1995. At the right edge of the heatmap, one year before the systemic crisis of 2007q3, 6 indicators exceed the "normal" conditions boundaries whereas only 3 stay within limits. Moreover, 3 indicators in particular (URT_{DE}, $\mathrm{CAPFORM}_{DE}$ and $\mathrm{CONSTINV}_{DE}$) have been issuing warning signals from as early as 2005q4. Therefore, despite the non-domestic nature of the subsequent systemic crisis, the framework would have managed to capture some developments which could have been potentially increasing the vulnerability of the economy and would have issued an early warning. Another interesting finding is that the variables which are found to have an adequate signaling performance in both applications are GDP, unemployment, house price growth and the ratio of credit to GDP. In a less strict comparison one would also include in that list credit to households since it is used in both cases; in the case of Slovenia as the growth of credit to households and NFCs; and in the case of Germany as the ratio of credit to households over GDP. Apart from those, the rest are unique to each case. This corroborates the findings of the literature (for an earlier survey on currency crises see Abiad (2003) and for a more recent one on banking crises see Kauko (2014) and references therein) which indicate that although there are some common variables exhibiting similar patterns before a crisis, there is however a large degree of heterogeneity through time and across countries. This could affect methods which pool together and analyze many individual countries, since country-specific variables would have possibly turned out to be statistically insignificant. On the contrary, the proposed approach can exploit a much larger information set and uncover country-specific vulnerabilities which would otherwise be overlooked.

Finally, the composite indicators' evolution is presented in Figure 7.

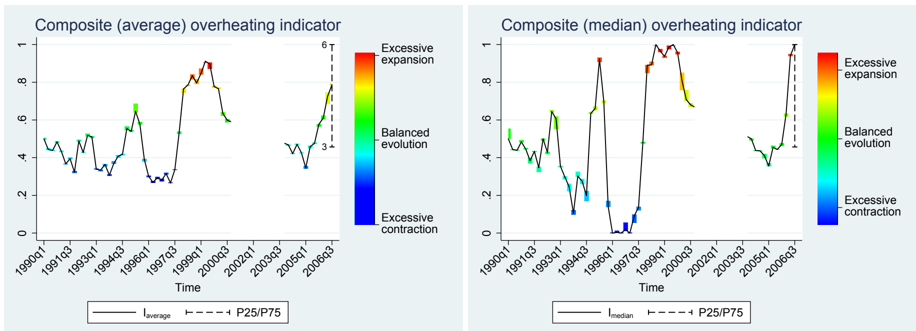

Figure 7: Average (left) and median (right) composite indicators for Germany up to 2006q3.

A similar pattern to the one depicted in Figure 6 is displayed by the composite indicators in Figure 7. Both exhibit high values from 1998q1 until 1999q4 when they begin to decline sharply. This pattern[11] is consistent with the findings of other studies (Behn et al., 2013; Babecký et al., 2014; Drehmann and Juselius, 2014) which document a peak of several indicators during the build-up phase and their fall in the period immediately before the onset of a crisis. Another interesting feature, especially pronounced in the right panel in Figure 7, is the peak of the median composite indicator in 1995q2 and a subsequent trough. This behaviour can be attributed to the "post-crisis bias" of a preceding event. The particular event is the recession in Germany which lasted from 1992q2 until 1994q3 according to Lo Duca et al. (2017). Classified as residual event in the ECB/ESRB EU crises database, its inclusion in the data results in pushing the trends downwards. Hence, during the following recovery, many indicators surpass the overly low thresholds which leads to the markedly high values of the composite ones. However, excluding that period would have limited substantially the available data to the point which this application would have been rendered unfeasible. Finally, both composite indicators point to the direction that in 2006q3 the economy is not following a balanced evolution path. Out of 9 individual indicators, 6 take the value of 1 in 2006q3 which is reflected on both composite ones. The fact that three indicators are slightly below their long term trends affects the average composite indicator (Figure 7, left panel). However, it still shows a high figure of close to 0.8, while the median (Figure 7,

[11]The same behaviour is shown in the evolution of the composite indicators for Slovenia in Figure 5.

right panel) issues a clear warning signal.

Despite the exogenous nature of the systemic event and the fact that data from a recessionary period were included, overall the results indicate that the framework would have been able to detect some potential vulnerabilities and issue an early warning at least one year in advance of the most recent systemic crisis in Germany.

5 Conclusions

In this paper a simple and transparent methodology is applied to develop an early warning system for overheating identification for the Slovenian economy. The low data demands of the employed signaling approach make it particularly suitable for individual country applications. The results from the empirical application indicate that the Slovenian economy is not in a state of overheating for the time being and that most variables studied grow at a moderate pace. However, constant monitoring of their developments is appropriate because it will give sufficient reaction time to policy makers in case prevailing conditions change.

The robustness check, using historical data from the German economy, confirmed the adequate predictive ability of the framework and uncovered some interesting findings regarding the stability of the indicators examined. First, despite being calibrated to a single crisis event and the fact that the next one was purely exogenous to the domestic economy, the framework would have been able to reveal several potential sources of instability and would have issued warnings sufficiently early before the subsequent systemic crisis had begun. Second, the applied methodology identified some common indicators between countries and crises but also revealed the existence of a significant degree of heterogeneity between them. Thus, the proposed approach not only has the advantage of being applicable in data constrained environments but also allows exploiting greater amount of country-specific information compared to similar frameworks which pool many countries together.

A feature present in both country cases examined is that simple trend extracting methods such as the simple moving average outperformed more complicated ones in terms of in-sample signaling ability. Given that similar findings are also documented in other studies, less sophisticated methods should be considered when developing early warning systems in addition to the HP filter which currently dominates the respective literature.

As is obvious, like any other statistical model based on historical data, the predictive power of the framework depends on future stress events following patterns observed in previous ones and therefore the results should be interpreted with caution. Having this caveat in mind, this framework can serve as an additional tool in the policy makers' toolbox by drawing attention to certain sectors in the economy that might be growing excessively and therefore require closer monitoring using other, specialized tools.

References

Abiad, A. (2003). Early warning systems: A survey and a regime-switching approach. *IMF Working Paper No. 03/32*, 03(32).

Alessi, L. and Detken, C. (2009). 'Real time' early warning indicators for costly asset price boom/bust cycles: A role for global liquidity. *ECB Working Paper, No. 1039*.

Babecký, J., Havránek, T., Mateju, J., Rusnák, M., Šmídková, K., and Vašíček, B. (2014). Banking, debt, and currency crises in developed countries: Stylized facts and early warning indicators. *Journal of Financial Stability*, 15:1–17.

Bank of Slovenia (2004). Annual report. Technical report, Bank of Slovenia.

Behn, M., Detken, C., Peltonen, T. A., and Schudel, W. (2013). Setting countercyclical capital buffers based on early warning models: Would it work? *ECB Working Paper, No. 1604*.

Behn, M., Gross, M., and Peltonen, T. (2016). Assessing the costs and benefits of capital-based macroprudential policy. *ECB Working Paper, No. 1935*.

Borio, C. and Drehmann, M. (2009). Assessing the risk of banking crises–revisited. *BIS Quarterly Review*, page 29.

Borio, C. and Lowe, P. (2002a). Assessing the risk of banking crises. *BIS Quarterly Review*, page 43.

Borio, C. and Lowe, P. (2002b). Asset prices, financial and monetary stability: Exploring the nexus. *BIS Working Papers No. 114*.

Bussière, M. and Fratzscher, M. (2006). Towards a new early warning system of financial crises. *Journal of International Money and Finance*, 25(6):953–973.

Bussière, M. and Fratzscher, M. (2008). Low probability, high impact: Policy making and extreme events. *Journal of Policy Modeling*, 30(1):111–121.

Caggiano, G., Calice, P., Leonida, L., and Kapetanios, G. (2016). Comparing logit-based early warning systems: Does the duration of systemic banking crises matter? *Journal of Empirical Finance*, 37:104–116.

Corsetti, G., Pesenti, P., and Roubini, N. (1999). Paper tigers?: A model of the asian crisis. *European Economic Review*, 43(7):1211–1236.

Davis, E. P. and Karim, D. (2008). Comparing early warning systems for banking crises. *Journal of Financial stability*, 4(2):89–120.

Demirgüç-Kunt, A. and Detragiache, E. (1998). The determinants of banking crises in developing and developed countries. *IMF Staff Papers*, 45(1):81–109.

Demirgüç-Kunt, A. and Detragiache, E. (2005). Cross-country empirical studies of systemic bank distress: A survey. *National Institute Economic Review*, 192(1):68–83.

Drehmann, M., Borio, C., and Tsatsaronis, K. (2011). Anchoring countercyclical capital buffers: The role of credit aggregates. *International Journal of Central Banking*, 7(4):189–240.

Drehmann, M. and Juselius, M. (2014). Evaluating early warning indicators of banking crises: Satisfying policy requirements. *International Journal of Forecasting*, 30(3):759–780.

Edison, H. J. (2003). Do indicators of financial crises work? An evaluation of an early warning system. *International Journal of Finance & Economics*, 8(1):11–53.

Epps, T. and Singleton, K. J. (1986). An omnibus test for the two-sample problem using the empirical characteristic function. *Journal of Statistical Computation and Simulation*, 26(3-4):177–203.

Gerdrup, K., Kvinlog, A. B., and Schaanning, E. (2013). Key indicators for a countercyclical capital buffer in Norway - Trends and uncertainty. *Bank of Norway Staff Memo No. 13/2013*.

Goerg, S. J. and Kaiser, J. (2009). Nonparametric testing of distributions–The Epps-Singleton two-sample test using the empirical characteristic function. *Stata Journal*, 9(3):454.

Gourinchas, P.-O., Valdes, R. O., and Landerretche, O. (2001). Lending booms: Latin america and the world. *Economía*, 1(2):47–99.

Hermansen, M. and Röhn, O. (2015). Economic resilience: The usefulness of early warning indicators in OECD countries. *OECD Economic Department Working Papers*, (1250):1–31.

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

Hoggarth, G., Reis, R., and Saporta, V. (2002). Costs of banking system instability: some empirical evidence. *Journal of Banking & Finance*, 26(5):825–855.

Holopainen, M. and Sarlin, P. (2017). Toward robust early-warning models: A horse race, ensembles and model uncertainty. *Quantitative Finance*, 17(12):1933–1963.

Ito, Y., Kitamura, T., Nakamura, K., and Nakazawa, T. (2014). New financial activity indexes: Early warning system for financial imbalances in Japan. *Bank of Japan Working Paper No. 14-E-7*.

Joy, M., Rusnák, M., Šmídková, K., and Vašíček, B. (2017). Banking and currency crises: Differential diagnostics for developed countries. *International Journal of Finance & Economics*, 22(1):44–67.

Kaminsky, G. (1999). Currency and banking crises; The early warnings of distress. *IMF Working Paper No. 99/178*, 99(178).

Kaminsky, G., Lizondo, S., and Reinhart, C. M. (1998). Leading indicators of currency crises. *Staff Papers*, 45(1):1–48.

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

Kauko, K. (2014). How to foresee banking crises? A survey of the empirical literature. *Economic Systems*, 38(3):289–308.

Knedlik, T. and Von Schweinitz, G. (2012). Macroeconomic imbalances as indicators for debt crises in Europe. *JCMS: Journal of Common Market Studies*, 50(5):726–745.

Laeven, L. and Valencia, F. (2008). Systemic banking crises: A new database. *IMF Working Paper No. 08/224*, 8(224).

Laeven, L. and Valencia, F. (2010). Resolution of banking crises: The good, the bad, and the ugly. *IMF Working Paper No. 10/146*, 10(146).

Laeven, L. and Valencia, F. (2012). Systemic banking crises database: An update. *IMF Working Paper No. 12/163*, 12(163).

Lanzeni, M. L. and Weistroffer, C. (2013). Emerging markets: Who is vulnerable to overheating? Technical report, Deutsche Bank Research Briefing.

Levene, H. (1960). Robust tests for equality of variances. *Contributions to probability and statistics. Essays in honor of Harold Hotelling*, pages 278–292.

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

Mitra, S., Beneš, J., Iorgova, S., Lund-Jensen, K., Schmieder, C., and Severo, T. (2011). Toward operationalizing macroprudential policies: When to act. *Grappling with Crisis Legacies, IMF Global Financial Stability Report*, pages 103–149.

Reinhart, C. M. and Rogoff, K. S. (2011). From financial crash to debt crisis. *American Economic Review*, 101(5):1676–1706.

Röhn, O., Sánchez, A. C., Hermansen, M., and Rasmussen, M. (2015). Economic resilience: A new set of vulnerability indicators for OECD countries. *OECD Economic Department Working Papers*, (1249):1–41.

Sarlin, P. (2013). On policymakers loss functions and the evaluation of early warning systems. *Economics Letters*, 119(1):1–7.

Weistroffer, C. and Vallés, V. (2008). Monitoring banking sector risks: An applied approach. Technical report, Deutsche Bank Research Research Notes 29.

Table 1: Variables examined for overheating signaling performance in Slovenia.

Variable	Description	Transformation	Range	Source
GDP	Real GDP	y-o-y in %	2000q1 - 2017q3	SORS
CA	Current account position	% of nominal GDP	2000q1 - 2017q3	SORS
TRADEBAL	Trade balance	% of nominal GDP	2000q1 - 2017q3	SORS
GDEBT	Gross foreign debt	y-o-y difference of share in GDP, pp.	2002q1 - 2017q3	SORS
FIXEDINV	Total gross fixed investment	% of nominal GDP	2000q1 - 2017q3	SORS
CONSTINV	Construction investment	y-o-y in %	2000q1 - 2017q3	SORS
CINF	Core inflation (excl. energy, food, alcohol and tobacco)	difference between y-o-y growth in SI and EA, pp.	2002q1 - 2017q3	Eurostat
HOUSEP	Prices of used dwellings	y-o-y in %	2001q1 - 2017q2	SORS
CRDTOT	Credit to private sector	% of nominal GDP	2000q1 - 2017q3	BoS
CRD2HF	Loans to NFCs and households	y-o-y in %	2000q1 - 2017q3	BoS
CRD2NFC	Loans to NFCs	y-o-y in %	2001q1 - 2017q3	BoS
COMP	Compensation of employees	Difference between y-o-y growth in wage bill and nominal GDP, pp.	2000q1 - 2016q4	SORS
COMPXOPQ	Compensation of employees excl. OPQ	Difference between y-o-y growth in wage bill (excl. OPQ) and nominal GDP, pp.	2000q1 - 2017q3	SORS
ULC	Nominal unit labour cost	Difference between y-o-y growth in SI and EA, pp.	2000q1 - 2017q2	Eurostat
URT	Unemployment rate	y-o-y in %	2000q1 - 2017q3	SORS
UNCERT	Uncertain economic conditions indicator	Value of seasonally adjusted indicator	2000q1 - 2017q3	SORS
SENTI	Economic sentiment indicator - total economy	Value of seasonally adjusted indicator	2000q1 - 2017q3	Eurostat
SBI	SBI TOP stock market index	y-o-y in %	2000q1 - 2017q3	Reuters

Table 2: Selected descriptive statistics of analyzed variables and distributional similarity tests.

Variable	Period	Minimum	Maximum	Mean	Standard deviation	W_0 p-value	KS p-value	ES p-value
GDP	1	2.213	6.335	3.690	1.062	0.060	0.000	0.000
	2	-1.569	7.784	5.214	2.431			
CA	1	-4.676	0.971	-1.276	1.516	0.500	0.020	0.010
	2	-5.479	-1.216	-3.190	1.604			
TRADEBAL	1	-4.789	1.659	-0.939	1.680	0.420	0.440	0.600
	2	-3.488	1.143	-1.057	1.278			
GDEBT	1	-1.127	9.471	4.269	2.942	0.120	0.000	0.000
	2	6.428	24.547	14.385	4.980			
FIXEDINV	1	24.467	27.931	25.959	0.961	0.170	0.000	0.000
	2	26.766	30.628	28.520	1.313			
CONSTINV	1	-9.690	12.714	2.021	5.455	0.010	0.060	0.140
	2	-4.693	27.519	9.595	9.573			
CINF	1	-0.525	6.051	3.100	1.993	0.000	0.000	0.000
	2	-0.318	2.373	0.832	0.925			
HOUSEP	1	-0.834	16.343	8.270	4.461	0.000	0.020	0.010
	2	-2.874	47.101	21.013	17.176			
CRDTOT	1	0.361	0.518	0.410	0.041	0.000	0.000	0.000
	2	0.541	0.837	0.697	0.101			
CRD2HF	1	12.353	24.481	18.942	3.182	0.160	0.000	0.000
	2	19.622	34.405	27.013	4.343			
CRD2NFC	1	5.703	27.245	17.377	6.485	0.040	0.000	0.000
	2	19.300	32.491	25.938	3.798			
COMP	1	-2.625	4.317	0.270	1.819	0.530	0.440	0.590
	2	-2.379	6.407	0.042	2.397			
COMPXOPQ	1	-2.857	4.925	-0.034	1.897	0.490	0.350	0.730
	2	-2.306	2.945	0.545	1.505			
ULC	1	-0.520	9.429	3.889	2.652	0.090	0.000	0.030
	2	-1.132	4.222	1.227	1.627			
URT	1	5.800	7.500	6.522	0.475	0.600	0.170	0.170
	2	4.100	7.200	5.262	1.001			
UNCERT	1	6.000	21.000	11.609	3.858	0.960	0.020	0.070
	2	2.000	19.000	7.846	4.375			
SENTI	1	96.000	111.733	104.639	3.984	0.090	0.000	0.000
	2	86.633	117.933	109.874	8.218			
SBI	1	-8.777	62.594	21.979	23.610	0.010	0.410	0.170
	2	-57.738	101.594	30.173	47.109			

Period 1 starts at the beginning of the sample until 2005q3. Period 2 ranges from 2005q4 until 2008q4.
Levene's W_0 test for equality of variances. H0: variances are equal.
Kolmogorov-Smirnov (KS) test for equality of distributions. H0: samples have been drawn from the same population.
Epps and Singleton (ES) test for equality of distributions. H0: samples have been drawn from the same population.

Table 3: Contingency matrix.

	Event (at t = T)	No event (at t = T)
Signal (at t = [T-12, T-4])	True Positive (TP)	False Positive (FP)
No signal (at t = [T-12, T-4])	False Negative (FN)	True Negative (TN)

Table 4: Variables examined in the robustness check of the overheating identification framework for Germany.

Variable	Description	Transformation	Range	Source
GDP_{DE}	GDP at market prices, Chain linked volumes (2010), Seasonally and calendar adjusted data, ESA2010	y-o-y in %	1992q1 - 2006q3	Eurostat
$CAPFORM_{DE}$	Gross capital formation, Seasonally and calendar adjusted data, ESA2010	% of GDP	1991q1 - 2006q3	Eurostat
$CONSTINV_{DE}$	Construction investment	y-o-y in %	1992q1 - 2006q3	Eurostat
$HOUSEP_{DE}$	Residential property prices (Index, 1995 = 100)	y-o-y in %	1991q1 - 2006q3	BIS
$CRDTOT_{DE}$	Credit to Private non-financial sector from all sectors at market value	% of GDP	1990q1 - 2006q3	BIS
$CRD2HH_{DE}$	Credit to Households and NPISHs from all sectors at market value - Adjusted for breaks	y-o-y in %	1991q1 - 2006q3	BIS
$CRD\text{-}GDP2HH_{DE}$	Credit to Households and NPISHs from all sectors at market value - Adjusted for breaks	% of GDP	1990q1 - 2006q3	BIS
$CRD2NFC_{DE}$	Credit to Non-financial corporations from all sectors at market value - Adjusted for breaks	y-o-y in %	1991q1 - 2006q3	BIS
$CRD\text{-}GDP2NFC_{DE}$	Credit to Non-financial corporations from all sectors at market value - Adjusted for breaks	% of GDP	1990q1 - 2006q3	BIS
$COMP_{DE}$	Compensation of employees	y-o-y in %	1992q1 - 2006q3	Eurostat
$RULC_{DE}$	Real unit labour cost based on persons (Index, 2005 = 100)	y-o-y in %	1992q1 - 2006q3	Eurostat
URT_{DE}	Unemployment rate, percentage of active population, Seasonally adjusted data, not calendar adjusted data	y-o-y in %	1992q1 - 2006q3	Eurostat
$SENTI_{DE}$	Economic sentiment indicator - total economy	Value of seasonally adjusted indicator	1990q1 - 2006q3	Eurostat
FAZ_{DE}	F.A.Z. stock market index	y-o-y in %	1991q1 - 2006q3	Reuters

Source for residential property prices: National sources, BIS Residential Property Price database (http://www.bis.org/statistics/pp.htm)
Source for credit statistics: BIS total credit statistics (https://www.bis.org/statistics/totcredit/totcredit.xlsx)

A Evolution and signaling performance of individual early warning indicators for Slovenia

Below is presented the evolution of the early warning indicators for Slovenia determined by the multi-step statistical procedure described in the main text. Their "normal" conditions range (grey shaded areas), optimal trend extracting method and threshold used are also displayed. With orange is marked the early warning period, from 2005q4 until 2007q4.

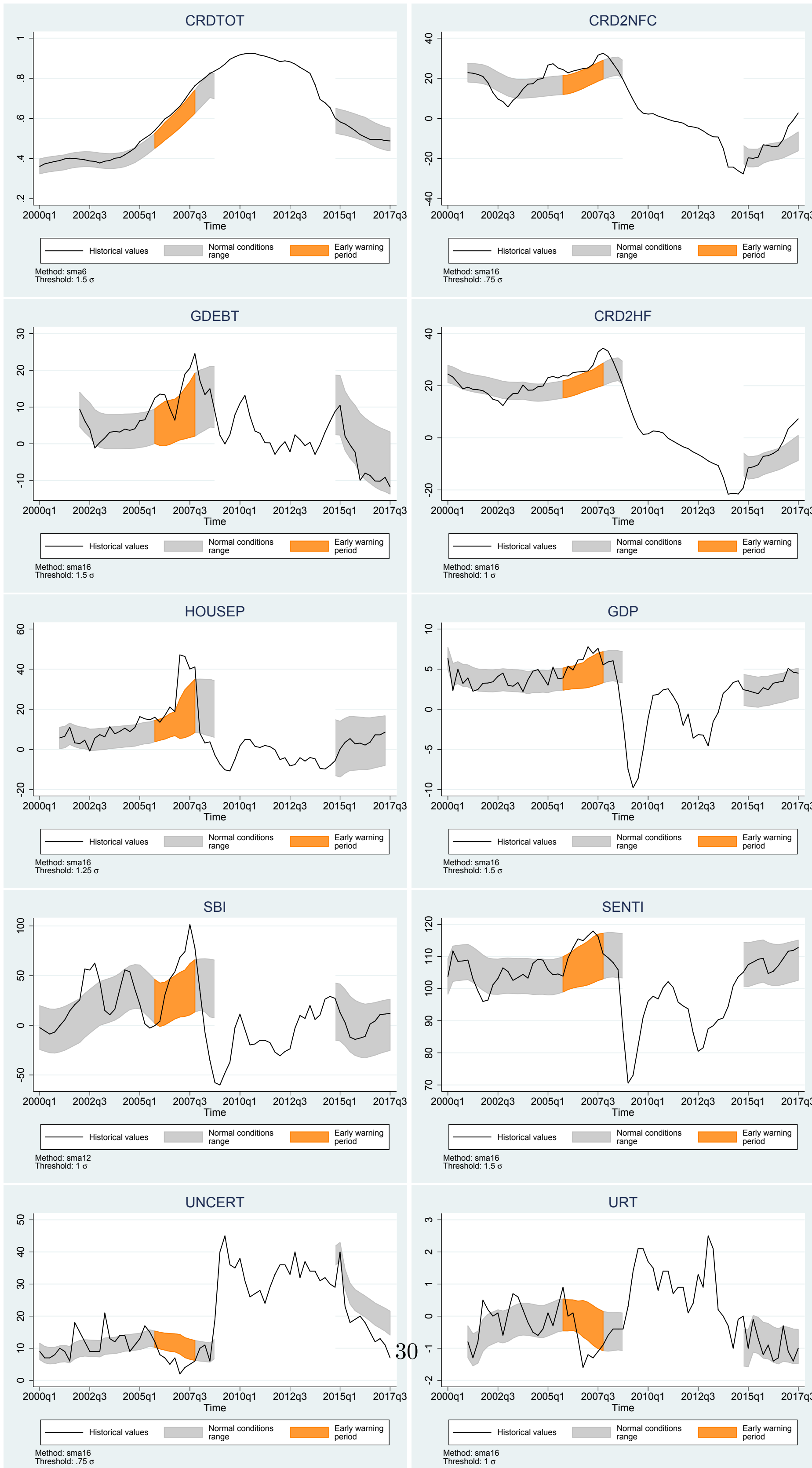

Figure 8: The evolution of individual early warning indicators for Slovenia.

Figures 9 and 10 display the performance measures of the examined subindicators. In the Accuracy - L plots (Figure 9) the red filled circle marks the preferred subindicator, with blue hollow circles the ones that meet the desired performance conditions and with orange X's those that don't.

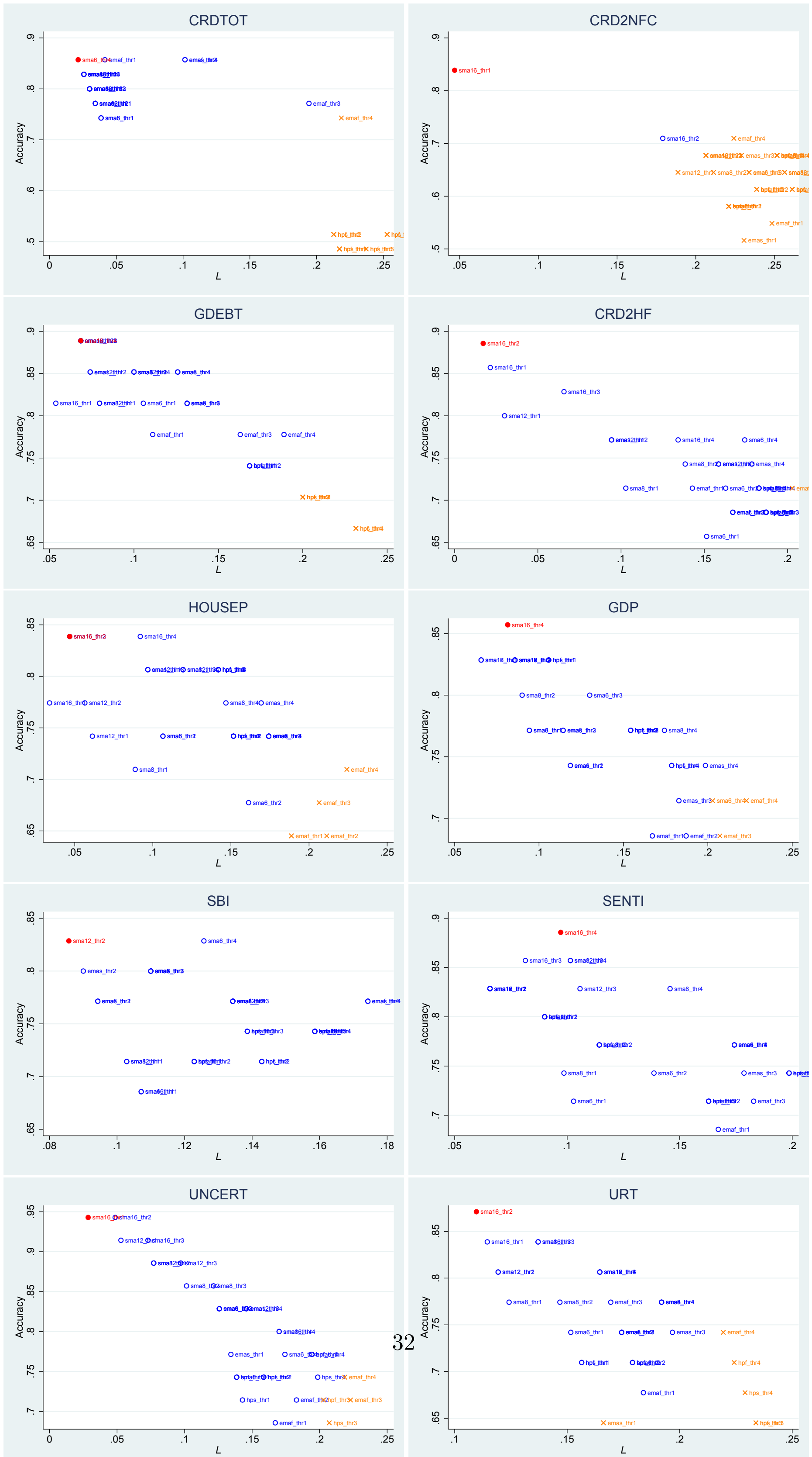

Figure 9: Signaling performance of each variables' subindicators for Slovenia.

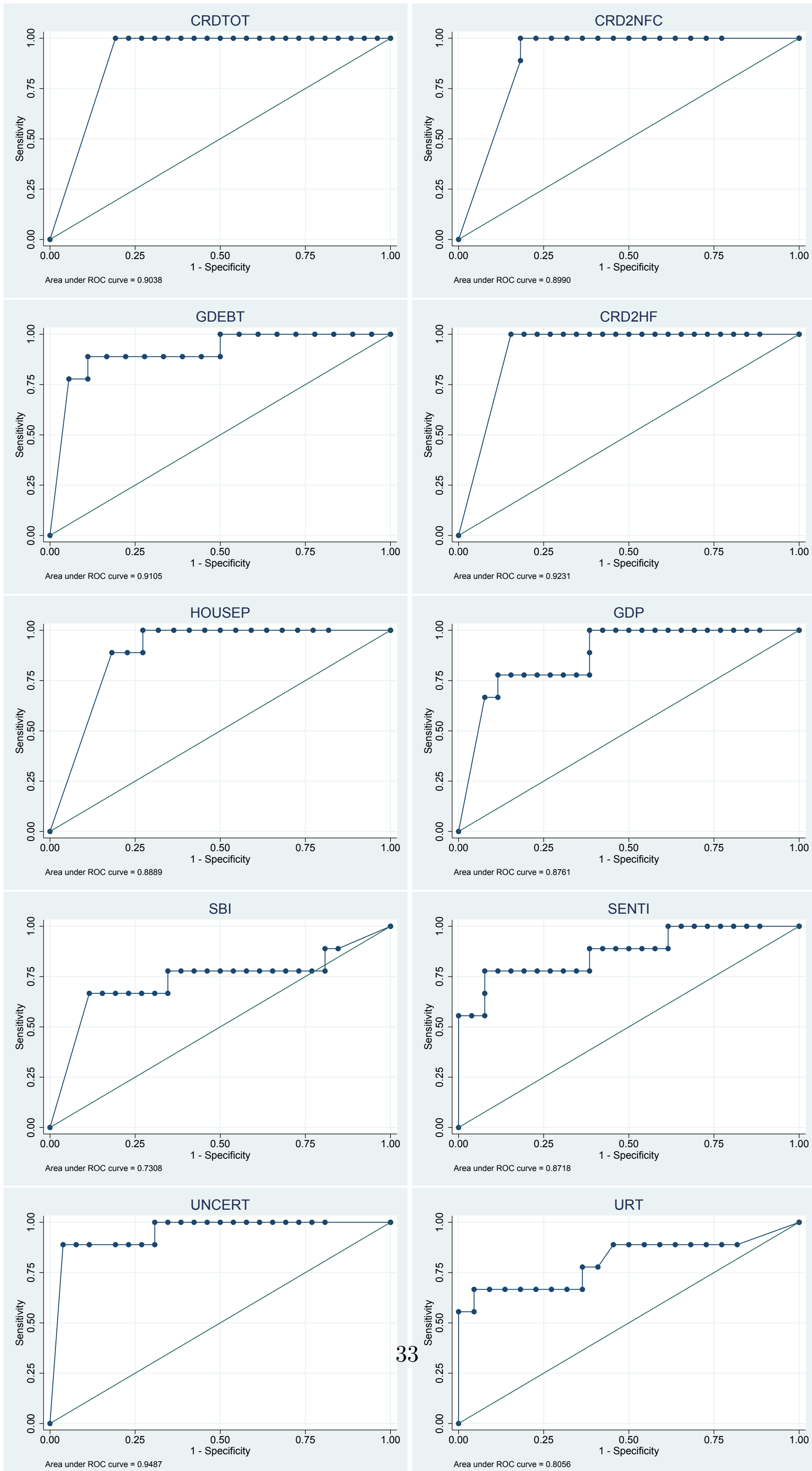

Figure 10: The AUROC curve for the preferred subindicator from each variable for Slovenia.

B Evolution and signaling performance of individual early warning indicators for Germany

Similarly to the case of Slovenia, the evolution of the early warning indicators for Germany is presented in Figure 11. Again, the grey shaded area marks the "normal" conditions range and the orange one the early warning period from 1997q4 until 1999q4.

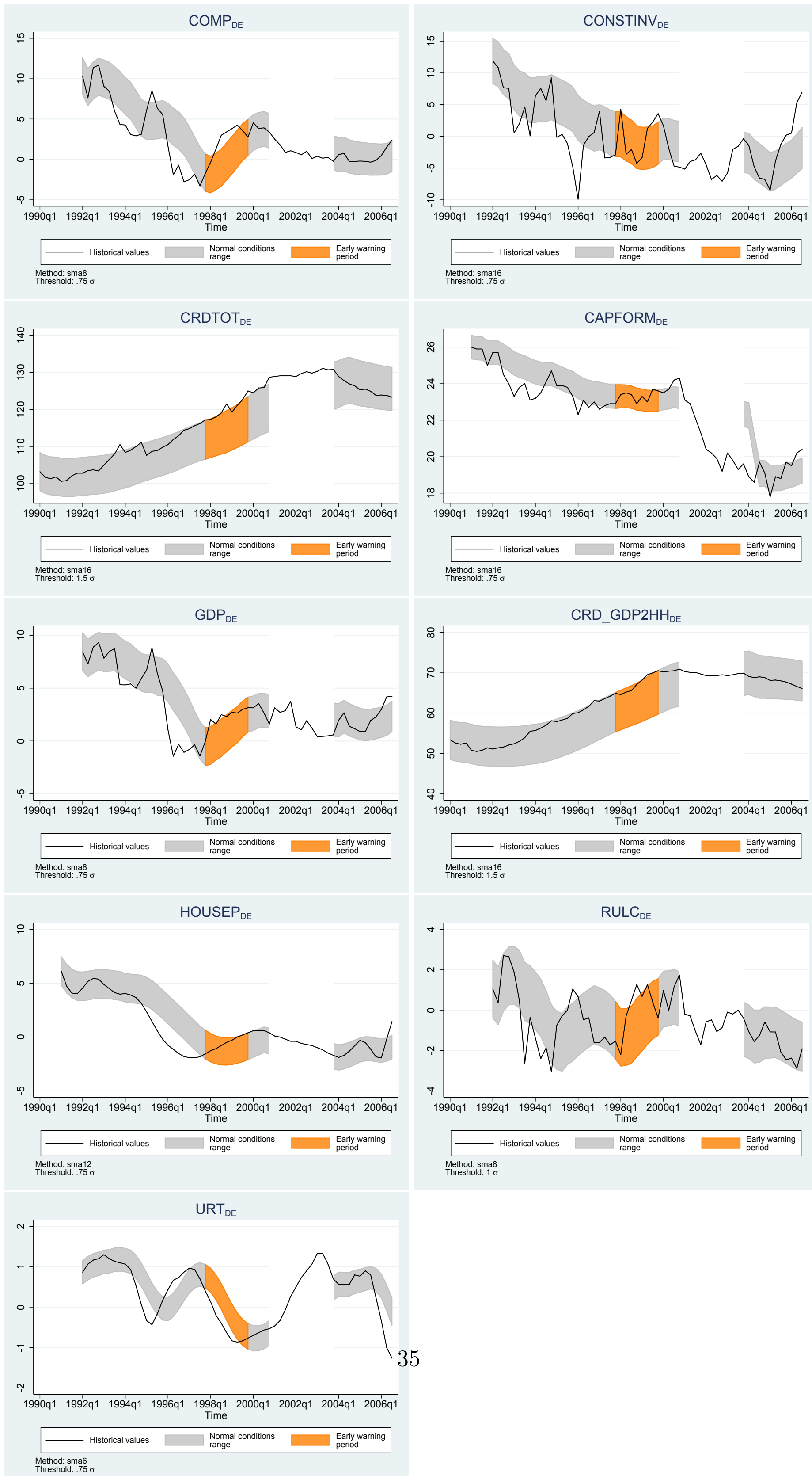

Figure 11: The evolution of individual early warning indicators for Germany.

Finally, Figure 12 shows the signaling performance of the analyzed subindicators for Germany and Figure 13 their AUROC curve measure.

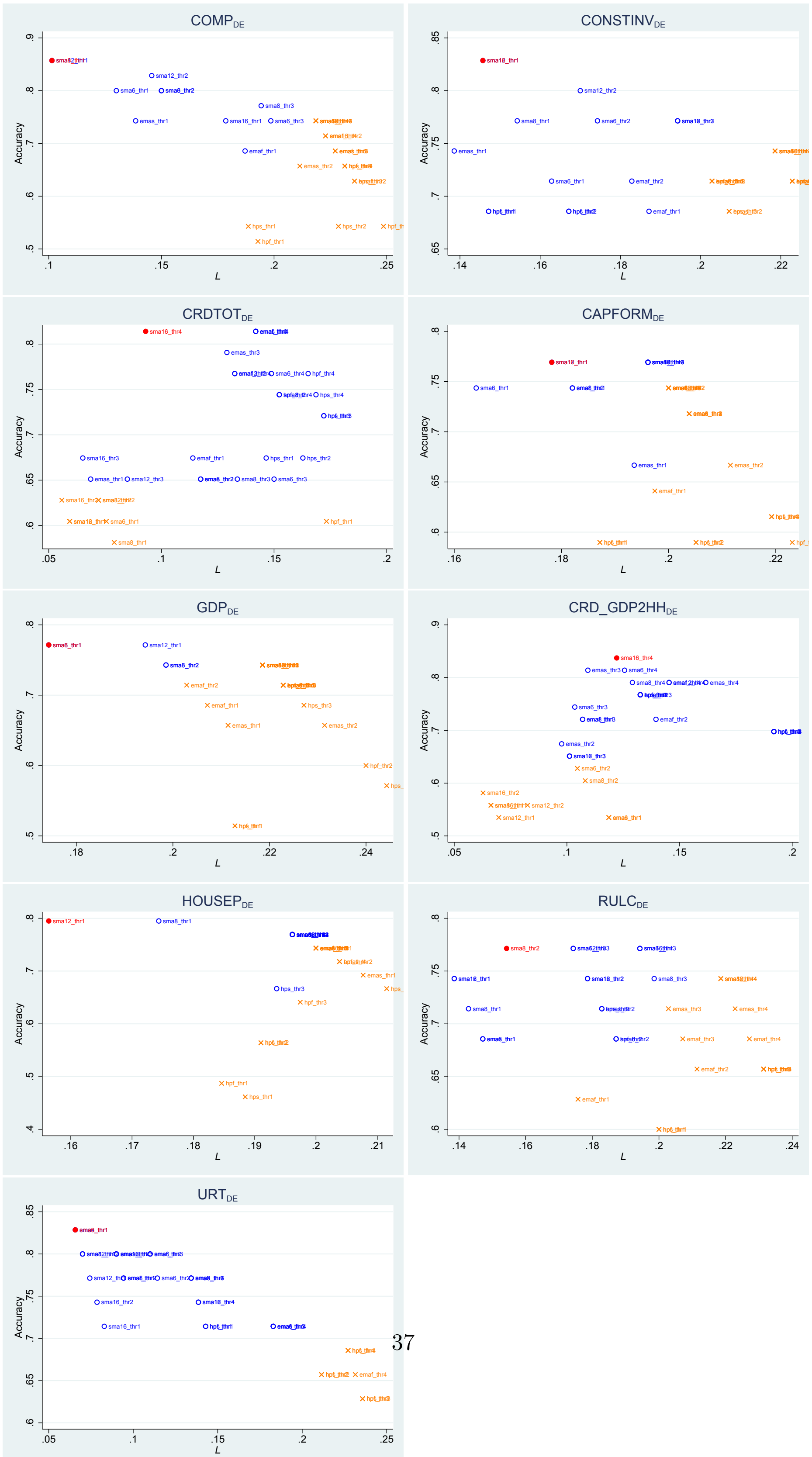

Figure 12: Signaling performance of each variables' subindicators for Germany.

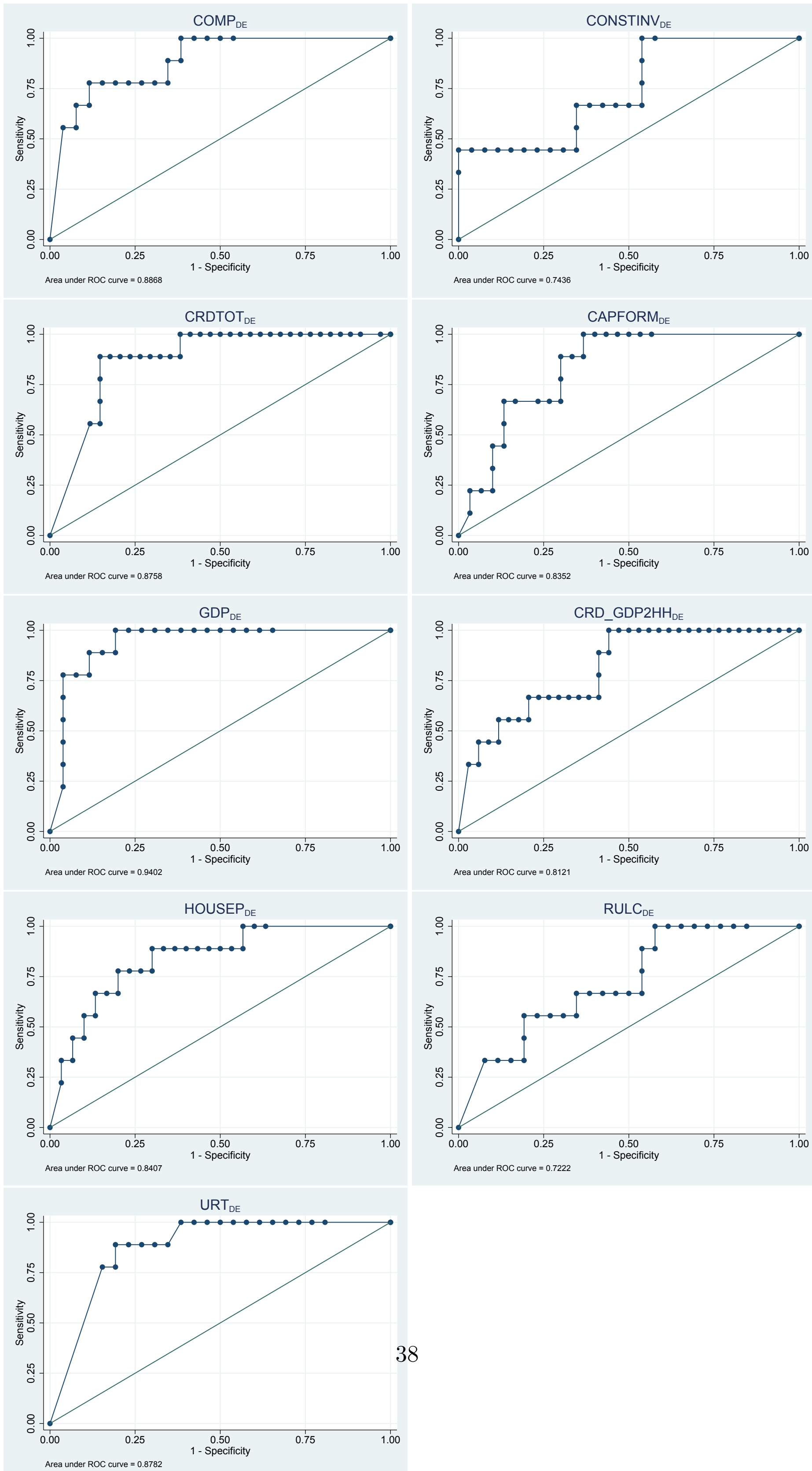

Figure 13: The AUROC curve for the preferred subindicator from each variable for Germany.