

A Bayesian signals approach for the detection of crises

Panayotis G. Michaelides¹, Mike G. Tsionas² *and* Panos Xidonas³

Second revision, March 2019

Abstract

In this paper, we consider the signals approach as an early-warning-system to detect crises. Crisis detection from a signals approach involves Type I and II errors which are handled through a utility function. We provide a Bayesian model and we test the effectiveness of the signals approach in three data sets: i) Currency and banking crises for 76 currency and 26 banking crises in 15 developing and 5 industrial countries between 1970 and 1995, ii) costly asset price booms using quarterly data ranging from 1970 to 2007, and iii) public debt crises in Europe in 11 countries in the European Monetary Union from the introduction of the Euro until November 2011. The Bayesian model relies on a vector autoregression for indicator variables, and incorporates dynamic factors, time-varying weights in the latent composite indicator and special priors to avoid the proliferation of parameters. The Bayesian vector autoregressions are extended to a semi-parametric context to capture non-linearities. Our evidence reveals that our approach is successful as an early-warning mechanism after allowing for breaks and nonlinearities and, perhaps more importantly, the composite indicator is better represented as a flexible nonlinear function of the underlying indicators.

Key Words: Predicting Crises; Early Warning System; Bayesian Analysis; Leading Indicators.

Acknowledgments: The authors wish to thank four anonymous reviewers and the Co-Editor Sushanta Mallick for many useful comments on earlier versions. The usual disclaimer applies.

¹ National Technical University of Athens, Greece, pmichael@central.ntua.gr.

² Lancaster University Management School, LA1 4YX, UK, m.tsionas@lancaster.ac.uk (*contact author*).

³ ESSCA Grande Ecole, France, panos.xidonas@essca.fr.

1. Introduction

Predicting crises is important as, among other things, the cost of crises is considerable. Caprio and Klingebiel (1996) estimated that average bailout amounts to 10 percent of GDP and cumulative losses in output are, approximately, 5.6 percent of GDP (Hoggarth et al., 2002). Lo Duca et al. (2017) estimate that output losses from financial crises in the European Union amount, on average, to 8% of GDP, and Laeven and Valencia (2012) estimate that during banking crises across a large sample of countries worldwide output losses amounted on average to 23% of GDP. The cost of crises necessitates to have indicators of systemic risk that allow sufficient time for policy makers to engage in countercyclical measures.

The signals approach, as an EWM for crises, goes back to the path-breaking paper of Kaminsky et al. (1998), followed by Kaminsky and Reinhart (1999). Its applications include EWMs for: (i) debt crises (see Knedlik and von Schweinitz, 2012), (ii) asset price bubbles (see Alessi and Detken, 2011), (iii) banking crises (see Borio and Drehmann, 2009), and even (iv) currency crises (see Edison, 2003). Meanwhile, binary choice models predict a binary crisis variable in the spirit of Frankel and Rose (1996); Berg and Pattillo (1999); Kamin et al. (2001); Bussière and Fratzscher (2006). Other authors use the Markov-switching technique in an attempt to provide better forecasting performance compared to binary choice models or the signals approach (Abiad, 2003; Mariano et al., 2004; Kittelmann et al., 2006; Knedlik and Scheufele, 2008). However, the interpretation of the regimes in economic terms is somewhat problematic.⁴ **Another related approach is presented in Gomez-Puig and Sosvilla-Rivero (2016).**

In an important paper, El-Shagi, Knedlik, and von Schweinitz (2013, EKS throughout) test the null hypothesis of no correlation between indicators and crisis probability in three applications of the signals approach to different crisis types. The topic is of considerable interest given the 2008-2009 subprime crisis. Constructing indicators and signals that can provide an Early Warning Mechanism is, subsequently, of considerable importance for policy authorities, systemic banks and other authorities like central banks.

In this paper, we consider the signals approach as an EWM to detect crises. We provide a novel Bayesian model and we test the effectiveness of the signals approach in three data sets. The adopted Bayesian approach has several advantages over the traditional approaches. For instance, it overcomes the so-called over-fitting issue and has an increased flexibility. However, the most important feature of the Bayesian approach is the fact that it mixes various pieces of information, such as sample information, prior information etc., so as to construct a representation which accounts for the stochastic nature of the variables. More precisely, the principal reason for choosing a Bayesian model is that it takes fuller account of the many uncertainties that are associated with the model and the parameters of interest. We use a Bayesian modeling approach organized around the Sequential Monte Carlo (SMC) and Markov Chain Monte Carlo (MCMC) procedures. See, for instance, Wasserman (2004), Robert (2001) and Carlin and Louis (2000). **The reader should keep in mind that the approach in EKS is non-parametric, whereas our analysis is based on flexible functional forms using artificial neural networks. Therefore, our approach is *semi-parametric* and allows for breaks and non-linearities.**

The Bayesian model relies on a Vector Autoregression for indicator variables, it incorporates dynamic factors and special priors to avoid the proliferation of parameters. The Bayesian Vector Autoregression model is extended to a semi-parametric context to capture non-linearities in economic time series. Crisis detection from a signals approach involves Type I and II errors which are handled through a utility function. *Our evidence reveals that non-linear*

⁴ For an extensive review, see EKS and the references therein.

models are quite successful as early-warning systems but linear models, even after allowing for breaks, are not so successful. This is important for policy making as well as future work on crises and their timely detection. To understand how the present paper contributes to the literature and can help in terms of future research we need to mention the following. For example, Guerreiro (2014) argues that “[s]ince it is always difficult to analyze a crisis on the spot, the likely roots of interest rate differentials in EMU are still an open question” and finds that persistent balance of payments disequilibria trigger interest rate increases, debt crises, and subsequent implementation of strict programmes in the context of bailouts. Heryán and Tzeremes (2017) document that commercial banks react to monetary policy shocks differently in crisis periods. Specifically, the bank lending channels in old EMU countries have become more sensitive to changes in M2 than in short-term interest rates during the subprime crisis but the bank lending channels in *new* EU countries are more sensitive to short-term market interest rates. This perplexes the issue of contagion of crises and the nature of crises in the EU. The role of nonlinearities has been examined formally by Basak, Das and Rohit (2017) who find that, under nonlinearity, it becomes harder to absorb shocks and hence an economy is more likely to tumble into a crisis. In a related study, Ureche-Rangau and Burietz (2013) find a statistically significant link between the amount of capital injections provided by European governments following the subprime crisis to bail out banks and maintain stability. Moreover, they find that ECB interventions do not produce statistically significant results, while the outcome of guarantees is rather inconclusive, depending on estimation techniques. Drakos and Kouretas (2015) document a nonlinearity in the Taylor rule showing that there is a structural break in the implementation of the monetary policy of ECB during the peak of the financial crisis. Specifically, they show that the ECB monetary policy followed the Taylor rule before the crisis but not so in the post-crisis period. Qin and Luo (2016) consider capital account openness as an EWM for banking crises in the G20 countries from 1989 to 2010. Their results indicate that the capital account openness has significant predictive power for systemic banking crises, and the impact is closely related with economic development (real GDP per capita). The authors demonstrate that the particular EWM has good in-sample and out-of-sample performance, and it is robust for different sampling periods.

Several econometric models have been proposed to develop an EWM. Fuertes and Kalotychou (2006) applied pooled logit models to predict debt crises in emerging economies. Caggiano et al. (2016) used both binomial and multinomial logit econometric models in the context of building EWMs for systemic banking crises. Billio et al. (2016) proposed an entropy-based EWM for systemic risk. Cheng and Zhao (2019) provide an alternative system based on input-output analysis as cross-holding of claims and obligations among financial institutions can be treated in the context of input-output linkages and, therefore, model the financial system and contagion by using the classic Leontief framework. It is instructive that the authors examine the effect of two policies -capital injections and forced mergers of financial institutions. The authors mention that “*capital injections is an effective instrument for mitigating financial contagion [...]. However, forced mergers is not always effective. Although forced mergers help to recapitalize financial institutions, they also lead to restructuring of the whole financial system, and this restructuring mechanism may influence systemic risk.*” This outcome is quite important not only because capital injections are effective albeit costly but mostly because a restructuring of the financial system necessitates a re-examination of systemic risk and EWMs. This is likely to introduce structural breaks and nonlinearities that change in a significant way the operation of EWMs. Thus, EWMs need to account for such breaks and nonlinearities. An operational way to do so is provided in the present paper, using a novel semi-parametric Bayesian Vector Autoregressive Model.

The paper is structured as follows: Section 2 offers some preliminary information on

the model. Section 3 presents the statistical model. Section 4 sets out the empirical analysis and an examination of spurious predictive power is taken up in Section 5. Section 6 offers some concluding remarks.

2. Preliminaries

We follow EKS who test the null hypothesis of no correlation between indicators and crisis probability in three applications of the signals approach to different crisis types using the bootstrap. Suppose time periods are denoted by $t \in \{1, \dots, T\}$, countries by $n \in \{1, \dots, N\}$ and there are indicator variables $X_{tn,m}, m \in \{1, \dots, M\}$. If there has been a crisis in country n and period t we have $C_{tn} = 1$ and zero, otherwise. Define an early warning signal with early warning horizon as:

$$W_{tn} = \begin{cases} 1, & \text{if } \exists k \in \{1, \dots, h\}: C_{t+k,n} = 1, \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

Therefore, W_{tn} is an indicator which is equal to one if, for a given country, it can predict whether there is a crisis at a future date $t + k$ for some $k \in \{1, \dots, h\}$ where h is the time horizon.

For each indicator variable there is a threshold τ_m so that we have a binary signal:

$$S_{X_{tn,m}} = \mathbb{I}(X_{tn,m} \geq \tau_m). \quad (2)$$

Therefore, we have a division of the observations into four subsets A, B, C, and D. The possibilities are shown in Table 1.

PLEASE INSERT TABLE 1 AROUND HERE

The noise-to-signal ratio (NSR) is defined as

$$NSR = \frac{B/(B + D)}{A/(A + C)}. \quad (3)$$

Moreover, a utility function based on Type I and Type II errors can be defined (Bussière and Fratzscher, 2008; Alessi and Detken, 2011):

$$U(\theta) = \min(\theta, 1 - \theta) - \theta \frac{C}{A + C} - (1 - \theta) \frac{B}{B + D}. \quad (4)$$

An indicator with $NSR > 1$ or $U(\theta) < 0$ can be ignored. The parameter $\theta \in (0,1)$ can be set by the decision maker to balance the costs from the Type I and Type II errors. The objective is to find the thresholds $\{\tau_m, m = 1, \dots, M\}$ so that NSR is minimized or utility is maximized. Finally, one can define a composite indicator

$$CI = \sum_{m=1}^M \omega_m S_{X_m}. \quad (5)$$

The key advantage of the composite indicators approach is that it does not rely on a particular model. However, there are some statistical problems. More precisely, El-Shagi, Knedlik, and von Schweinitz (2013, pp. 79, and 81-82) argue that:

- One can fix parameter θ in advance (to 0.4, say) or, instead, provide conditional probabilities of a crisis given a particular range of the indicator variables.
- However, they also notice that spurious correlation may arise particularly when the threshold values are adjusted to maximize in-sample fit.

- Optimally adjusted thresholds *in-sample*, may produce some degree of *out-of-sample* correlation, even if in-sample fit is due entirely to chance.

This is used here to set forth the state of the art and suggest ways for improvement. Besides the moving block bootstrap and the conditional probability bootstrap, El-Shagi, Knedlik and von Schweinitz (2013) use a Panel VAR bootstrap following Bai and Ng (2005). They assume that every indicator has a global and a national component, and both follow an individual VAR process. The objective of all three approaches is to test the null hypothesis that the utility or noise-to-signal ratio for individual indicators could have been achieved by chance and test for the null hypothesis that a composite indicator (based on the noise-to-signal ratio or utility optimization) of the whole set of indicators in one application could have been the result of a random process. They reevaluate the economic findings of Kaminsky and Reinhart (1999), Alessi and Detken (2011) and Knedlik and von Schweinitz (2012) and assess the statistical significance of their results, concentrating on specific findings. Against this background, EKS then present general results on the application of the signals approach, and the results of the different tests for four applications are presented.

Potential problems with this approach, despite its attractiveness and usefulness are the following: (i) The bootstrap has only an asymptotic justification, making it problematic with typical sample sizes in applied econometrics. (ii) The bootstrap or the panel-VAR-bootstrap do not allow for possible structural breaks inside or outside the sample. (iii) The probabilities that results could have been the result of a random process or chance reflects a sampling-theory probability, not a posterior probability given the data at hand. In view of these problems, it is interesting to examine whether a Bayesian approach, along with a new estimation and inference methodology could, potentially, be useful in predicting crises. **This is not merely a new formal approach to EWMs but rather an attempt to re-examine the ability of EWMs to predict crises when richer models are considered using a different methodological approach. Our methodological advance extends to “endogenizing” the composite indicator in that weights are estimated from the data, and, in fact, the composite indicator need not be an affine function of underlying indicators. In practice, we use a flexible specification incorporating artificial neural network components involving the underlying indicators.**

3. Statistical model

The problem, from the perspective of our study, is that the model should be flexible and realistic. “Flexible” means that we should allow for at least some non-parametric components, and “realistic” means that structural breaks and causality must be accounted for. Suppose that for country $n \in \mathbb{N} = \{1, \dots, N\}$, time period $t \in \mathbb{T} = \{1, \dots, T\}$, and $m \in \mathbb{M} = \{1, \dots, M\}$ the relevant indicator variables $\mathbf{X}_t^{(n)} = \{X_{tn,m} \forall n \in \mathbb{N} \forall m \in \mathbb{M}\}$, follow a factor VAR model **or more precisely a Factor-Augmented Panel Vector Autoregression (FA-PVAR)**. Conditional on $1 \leq G < n$ global factors \mathbf{f}_t we have:

$$\mathbf{X}_t^{(n)} = \mathbf{A}_n(L)\mathbf{f}_t + \mathbf{v}_t^{(n)}, \forall n \in \mathbb{N}, \quad (6)$$

where $\mathbf{A}_n(L)$ represents a matrix polynomial in the lag operator, L .⁵ The global $G \times 1$ factor \mathbf{f}_t follows a VAR process:

$$\mathbf{f}_t = \mathbf{B}(L)\mathbf{f}_t + \mathbf{C}(L)\mathbf{X}_t^* + \boldsymbol{\xi}_t, \quad (7)$$

⁵For a variable \mathbf{x}_t , $L^l \mathbf{x}_t := \mathbf{x}_{t-l}$. Moreover, $\mathbf{A}_n(L) := \mathbf{A}_0 \mathbf{f}_t + \mathbf{A}_1 \mathbf{f}_{t-1} + \mathbf{A}_2 \mathbf{f}_{t-2} + \dots + \mathbf{A}_l \mathbf{f}_{t-l}$, where the matrices \mathbf{A}_j have dimension $M \times G$, for $j = 0, 1, \dots, l$.

where X_t^* is an aggregation or *composite indicator function* of $\{\mathbf{X}_t^{(n)}, n \in \mathbb{N}\}$ whose specification will be taken up later. For the error term we assume

$$\boldsymbol{\xi}_t \sim \mathcal{N}(\mathbf{0}, \Sigma_{\boldsymbol{\xi}}) \forall t \in \mathbb{T}. \quad (8)$$

This is a novel feature relative to EKS. For the individual country-specific components we assume:

$$\mathbf{v}_t^{(n)} = \mathbf{D}(L)\mathbf{v}_t^{(n)} + \mathbf{c}_t^{(n)} + \boldsymbol{\varepsilon}_t^{(n)}, \quad (9)$$

where $\mathbf{c}_n, n \in \mathbb{N}$ represent country-specific effects. Relative to EKS we assume that the country-specific effects can be also time-varying. These country-specific effects have the following structure:

$$\mathbf{c}_t^{(n)} = \mathbf{c}_{t-1}^{(n)} + \mathbf{1}_M \cdot \gamma_b^{(n)} D_{b,t}^{(n)} + \boldsymbol{\zeta}_t^{(n)} \forall t \in \mathbb{T}, \forall n \in \mathbb{N} \quad (10)$$

where $D_{b,t}^{(n)} = 1$ if there is a structural break at date t for country n , and zero otherwise. The number of breaks we allow is $b \in \mathbb{B} = \{1, \dots, \bar{B}\}$ where $\bar{B} \ll T$. Moreover, $\mathbf{1}_M$ is an $M \times 1$ vector of ones indicating that the structural break(s) occurs simultaneously for all terms in $\mathbf{c}_t^{(n)}$. Although different breaks at different periods can be accommodated econometrically, this approach is probably excessive for the number of time series and observations, in most applications.

For the error terms we assume:

$$\boldsymbol{\zeta}_t^{(n)} \sim \mathcal{N}(\mathbf{0}, \Sigma_{\boldsymbol{\zeta}}^{(n)}), \boldsymbol{\varepsilon}_t^{(n)} \sim \mathcal{N}(\mathbf{0}, \Sigma_{\boldsymbol{\varepsilon}}^{(n)}) \forall n \in \mathbb{N}. \quad (11)$$

For the aggregation or composite indicator function we assume a *stochastic weighting scheme*:

$$X_t^* = \sum_{n \in \mathbb{N}} \alpha_{t,n} \mathbf{X}_t^{(n)} + u_t := \boldsymbol{\alpha}_t' \mathbf{X}^{(t)} + u_t \forall t \in \mathbb{T}, \quad (12)$$

where $\mathbf{X}^{(t)} := \{\mathbf{X}_t^{(n)}, t \in \mathbb{T}\}$, the weights $\alpha_{t,n} \geq 0, \forall n \in \mathbb{N}$ and $\sum_{n \in \mathbb{N}} \alpha_{t,n} = 1 \forall t \in \mathbb{T}$. The weighting scheme assumes that the representation is not exact as we have the error term u_t . Dynamics in (12) can be accommodated using an AR(R) process for the error term:

$$u_t = \sum_{r=1}^R \rho_r u_{t-r} + \epsilon_t, \epsilon_t \sim \mathcal{N}(0, \sigma_{\epsilon}^2) \forall t \in \mathbb{T}. \quad (13)$$

As we explain below, we use $R = 2$. The prior for ρ_r s is flat across the real line. For the weights $\boldsymbol{\alpha}_t = (\alpha_{t,n}, n \in \mathbb{N})$ we assume that they are *time-varying*:

$$\boldsymbol{\alpha}_t = 2\boldsymbol{\alpha}_{t-1} - \boldsymbol{\alpha}_{t-2} + \mathbf{v}_{\boldsymbol{\alpha},t}, \quad (14)$$

where

$$\mathbf{v}_{\boldsymbol{\alpha},t} \sim \mathcal{N}_n(\mathbf{0}, \lambda^2 \mathbf{I}), \forall t \in \mathbb{T}. \quad (15)$$

The process in (14) *imposes a (spline) non-parametric structure* on $\boldsymbol{\alpha}_t$, which assumes that they are smooth (but unknown) functions of time. The control of smoothness, as in splines, is controlled by the common parameter λ . This specification has been used by Koop and Poirier (2004) and Koop, Poirier, and Tobias (2005), see for example equation (2.2) in the second study. The prior emerges from the consideration that given ordered data $\{x_t, t = 1, \dots, n\}$ we seek a function $f(x)$ so that the following criterion is minimized: $n^{-1} \sum_{t=1}^n (x_t - f(x_t))^2 +$

⁶We have $\mathbf{B}(L)\mathbf{f}_t = \mathbf{B}_1\mathbf{f}_{t-1} + \mathbf{B}_2\mathbf{f}_{t-2} + \dots$, and similarly for $\mathbf{C}(L)$ and $\mathbf{D}(L)$ below.

$\lambda \int f''(x)^2 dx$. The resulting function is a spline (Wahba, 1990, page xiv).

The parameter is assumed common for simplicity. We do not assume that this parameter is specified in advance. In fact, it has a prior of the form:⁷

$$\frac{\bar{q}_\lambda}{\lambda^2} \sim \chi^2(\bar{n}_\lambda), \quad (16)$$

where \bar{q}_λ and \bar{n}_λ are set to 0.1 and 1 respectively, so that the prior is vague to allow that it can be dominated by the data.

From (11) it is evident that we must impose structure on the many covariance matrices involved. Let us denote generically $\Sigma_\zeta^{(n)}$ and $\Sigma_\varepsilon^{(n)}$, $n \in \mathbb{N}$ by $\Sigma^{(n)}$. Consider the Cholesky decomposition $\Sigma^{(n)} = \mathbf{H}_{(n)}' \mathbf{H}_{(n)}$ and the lower diagonal elements of $\mathbf{H}_{(n)}$ by the vector $\mathbf{h}^{(n)}$. If the matrices have dimension, say $d \times d$ the vector has dimension $d_h = \frac{d(d+1)}{2}$. To address the problem of proliferation of parameters in large systems, we have to impose restrictions that these matrices are not too ‘‘dissimilar’’ across countries but, of course, not the same. This can be accomplished through the following prior:

$$\mathbf{d}_{n,n'} := \mathbf{h}^{(n)} - \mathbf{h}^{(n')} \sim \mathcal{N}_{d_h}(\mathbf{0}, \omega^2 \mathbf{I}) \forall n \neq n' \in \mathbb{N}. \quad (17)$$

Parameter ω controls the degree of similarity and, like before, we adopt a prior of the form:

$$\frac{\bar{q}_\omega}{\omega^2} \sim \chi^2(\bar{n}_\omega), \quad (18)$$

where \bar{q}_ω and \bar{n}_ω are set to 0.1 and 1 respectively. All univariate scale parameters (like σ_ε) have the same prior. In practice, we also impose the constraint⁸

$$\|\mathbf{d}_{n,n'}\| \leq M^*, \quad (19)$$

for some constant large enough, such as $M^* = 10^4$ which has been found to be satisfied always in the applications we examine.

For the VAR model in (6) suppose for simplicity

$$\mathbf{X}_t^{(n)} = \mathbf{A}_n(L) \mathbf{f}_t + \mathbf{u}_t^{(n)} = \mathbf{A}_0^{(n)} \mathbf{f}_t + \mathbf{A}_1^{(n)} \mathbf{f}_{t-1} + \mathbf{v}_t^{(n)} \forall n \in \mathbb{N}, \quad (20)$$

so that we have a single lag but the current-period factors appear in the VAR. This can be a very large system even when the number of indicators, the number of countries and the number of factors is quite moderate. This necessitates to deal with the proliferation of these parameters in a formal way. EKS deal with the problem along the way of Bai and Ng (2005) who extract principal components and then estimate each equation in (20) using OLS. Of course, nothing is wrong, in principle, with this way to proceed as it simplifies estimation considerably. The problem is that the estimates may have a large variance, thus invalidating testing their null hypothesis in a precise way.

A Bayesian way to deal with the proliferation of parameters is to use a so-called LASSO prior (Hans, 2009, Park and Casella, 2008, Tibshirani, 1996) on the parameters of $\boldsymbol{\beta}_A = \text{vec}[\mathbf{A}_0^{(n)}, \mathbf{A}_1^{(n)}]$. LASSO priors are well-known to result in parsimonious representations and better forecasting results. For (9) if $\mathbf{D} = [\mathbf{D}_1, \dots, \mathbf{D}_{l^*}]$, $\boldsymbol{\beta}_D = \text{vec}(\mathbf{D})$ and similarly for \mathbf{C} and \mathbf{B} we can, again, use a LASSO prior to deal with the excessive number of parameters in (9). The model in (10) is already comprehensive as it allows for random-walk behavior and a number of breaks. We impose an upper limit on the number of breaks, $\bar{B} = 5$. For (13) we set after some preliminary investigation that $R = 2$ is adequate. Our prior for σ_ε^2 is similar to (16)

⁷This is, in fact, a *gamma* prior. Its interpretation is that from a sample of size \bar{n}_λ from $\mathcal{N}(0, \lambda^2)$ we obtain a sum of squares equal to \bar{q}_λ .

⁸Here, we use the L_∞ -norm in \Re^{d_h} .

and (18) with the same hyper-parameters.

Under these conditions the model in (6) along with (7)-(14) can be estimated jointly as a system to account for all our assumptions. For example, $\boldsymbol{\beta} = [\boldsymbol{\beta}'_A, \boldsymbol{\beta}'_B, \boldsymbol{\beta}'_C, \boldsymbol{\beta}'_D]'$ can be given a single LASSO prior to account for model selection in both VARs in (6) and (9). Lag selection for the polynomials of the lag operators $\mathbf{A}(L), \mathbf{B}(L), \mathbf{C}(L)$ and $\mathbf{D}(L)$ is performed implicitly through the vector $\boldsymbol{\beta}$ which tends to avoid (and thus putting zero weight on) higher lags. We use a maximum of four lags in all cases. The LASSO has been quite effective in restricting the number of lags to no more than two, making it unnecessary to embark on further model comparison using the unknown combination of number of lags for the four matrix polynomials of the lag operator. The LASSO prior has a particularly simple form:

$$p(\boldsymbol{\beta}) \propto \prod_{j=1}^{\dim(\boldsymbol{\beta})} \exp(-\lambda_{\boldsymbol{\beta}} |\beta_j|). \quad (21)$$

This is a Laplace distribution for vector $\boldsymbol{\beta}$ depending on the single ‘‘smoothing parameter’’ $\lambda_{\boldsymbol{\beta}} > 0$. In log form we have a penalty function for the parameters, of the form $\mathcal{P}(\boldsymbol{\beta}) = -\lambda_{\boldsymbol{\beta}} \sum_{j=1}^{\dim(\boldsymbol{\beta})} |\beta_j|$ so the larger the parameter the more we want to penalize non-zero β_j s. Our prior for this parameter is exponential:

$$p(\lambda_{\boldsymbol{\beta}}) = \bar{\vartheta} \exp(-\bar{\vartheta} \lambda_{\boldsymbol{\beta}}). \quad (22)$$

Setting $\bar{\vartheta} = 0.1T^{-1}$ the prior expectation of $\lambda_{\boldsymbol{\beta}}$ is $10T$. Alternatively, we may keep the parameters $\lambda_{\boldsymbol{\beta}}, \lambda^2$ and ω^2 fixed, and choose them to optimize the out-of-sample utility or NSR.

As we are particularly concerned with the linear form of (12) we extend it to allow for a **semi-parametric approach** as follows:

$$\begin{aligned} X_t^* &= \sum_{n \in \mathbb{N}} \alpha_{t,n} X_t^{(n)} + g(\{X_t^{(n)}, n \in \mathbb{N}\}) + u_t \\ &\equiv \boldsymbol{\alpha}_t' \mathbf{X}^{(t)} + g(\{X_t^{(n)}, n \in \mathbb{N}\}) + u_t \quad \forall t \in \mathbb{T}, \end{aligned} \quad (23)$$

where $g(\cdot)$ is an unknown non-linear function of the underlying indicators. A reasonable approach is to use an artificial neural network (ANN) function, due to its good global approximation properties (Hornik, Stinchcombe, and White, 1989):

$$g(\{X_t^{(n)}, n \in \mathbb{N}\}) = \sum_{i=1}^{I^*} \varphi(\mathbf{X}^{(t)'} \boldsymbol{\beta}_{i,1}^{ANN}) \beta_{i,2}^{ANN}, \quad (24)$$

where, again, $\mathbf{X}_t := \{X_t^{(n)}, n \in \mathbb{N}\}$, and $\boldsymbol{\beta}^{ANN} := \{\boldsymbol{\beta}_{i,1}^{ANN}, \beta_{i,2}^{ANN}, i = 1, \dots, I^*\}$ are unknown parameters, and I^* is the order of the ANN. Parameters $\boldsymbol{\beta}_i^{ANN}$ are embedded in the previous $\boldsymbol{\beta}$ vector so that a LASSO prior can be used to impose parsimony restrictions in (24). The activation function is $\varphi(z) = \tanh(z), z \in \mathfrak{R}$ which is commonly used in the literature (Cheng and Titterington, 1994).

Although (6) is standard in the literature, the assumption of linearity in (7) can be put into question and affect seriously the results. Suppose

$$[\mathbf{B}(L) : \mathbf{C}(L)] \begin{bmatrix} \mathbf{f}_t \\ X_t^* \end{bmatrix} = \mathbf{z}_t \quad \forall t \in \mathbb{T}. \quad (25)$$

Then we can modify semi-parametrically (7) as follows⁹:

$$\mathbf{f}_t = \Gamma \mathbf{z}_t + \sum_{j=1}^{J^*} \Phi(\mathbf{z}_t' \boldsymbol{\beta}_{jf,1}) \beta_{jf,2} + \xi_t, \quad t \in \mathbb{T}, \quad (26)$$

where $\Phi(\cdot)$ is a vector function consisting of tanh functions of its elements, as in (24). The linear terms, viz. $\Gamma \mathbf{z}_t$ correspond exactly to (7), i.e. $\Gamma \mathbf{z}_t \equiv \mathbf{B}(L) \mathbf{f}_t + \mathbf{C}(L) X_t^*$. The additional terms allow for a semi-parametric specification along the lines of an artificial neural network (ANN). For simplicity we have the same ANN order J^* and, again, parameters $\boldsymbol{\beta}_{jf,1} \in \mathfrak{R}^{\dim(\mathbf{z}_t)}$ and $\beta_{jf,2}$ ($\forall j = 1, \dots, J^*$) can be made part of the overall, grand vector $\boldsymbol{\beta}$ with an overall LASSO prior as in (21). The functional form of the likelihood / posterior for implementing SMC is described in Appendix A.

4. Empirical analysis

4.1. Data

Following EKS we examine three crises:

i) Currency and banking crises (Kaminsky and Reinhart, 1999). The signals approach is first proposed and applied to a total of 76 currency and 26 banking crises in 15 developing and 5 industrial countries between 1970 and 1995. 16 different monthly indicators capturing monetary stress, vulnerability of the banking sector, prices and competitiveness are employed as indicator variables.

ii) Costly asset price booms (Alessi and Detken, 2011). The study applies the signals approach to costly asset price booms in 18 OECD countries. Asset price booms are identified by high growth of asset prices in at least four consecutive quarters. High- and low-cost booms are distinguished by using deviations from potential growth in the years following the boom. Using quarterly data ranging from 1970 to 2007, the authors identify 29 high-cost and 16 low-cost booms until 2002.

“89 indicators constructed from different transformations of 18 underlying quarterly indicators are used. Because of the confidentiality of equity, housing and aggregate asset prices provided by the BIS to Alessi and Detken, we are only able to use 15 of these 18 variables. Since some transformations are based on multivariate systems, we are limited to a total of 50 instead of 89 different indicators. Most of the 15 underlying indicators are available from the OECD Economic Outlook and Main Economic Indicators; domestic credit can be found in the IMF’s International Financial Statistics (IFS); private credit, corrected for structural breaks, was provided by the authors” (El-Shagi, Knedlik, and von Schweinitz, 2013, p. 81).

iii) Public debt crises in Europe (Knedlik and von Schweinitz, 2012) is on the European debt crisis. It analyzes 11 countries in the European Monetary Union (EMU), from the introduction of the Euro until November 2011. A public debt crisis is defined using the spreads of government bond yields over the yield of the average AAA-rated country in the EMU. “A total of 20 monthly indicators out of five different categories (fiscal indicators, competitiveness and domestic demand, asset prices, labor, private and foreign debt) are used. All of them are publicly available from EuroStat, ECB, the OECD Main Economic Indicators, the IMF’s IFS and Morgan Stanley Capital International (MSCI). Because [of] data availability requirements [...] we can only use 17 of the originally proposed indicators” (El-Shagi, Knedlik, and von

⁹As we have Gaussian hidden units, global approximation results are provided by Hartman, Keeler and Kowalski (1990).

Schweinitz, 2013, p.81).

4.2. Statistical inference and results

The model in the previous section (6)-(13) along with the prior distributions is estimated using Markov Chain Monte Carlo (MCMC) organized around Sequential Monte Carlo (or Particle Filtering) techniques (Doucet, Godsill, and Andrieu, 2000, Fearnhead and Clifford, 2003, and Lopes and Tsay, 2011 for a review). These techniques deal effectively with the underlying dynamics of unobserved factors, especially in the case of the non-linear models in (24) and (26). Such models can be estimated using standard techniques such as the Gibbs sampler (e.g. Tierney, 1994) but such techniques are known to converge very slowly for our purposes.¹⁰

Our Sequential Monte Carlo method is described in Technical Appendices A and B. We use 15,000 MCMC iterations omitting the first 5,000 to mitigate start up effects and 10^7 particles. **Convergence of MCMC was tested successfully by visual means as well as using Geweke's (1992) diagnostic.**

When the indicators are on a common scale (say logs in standard units -with mean subtracted and divided by the standard deviation) the assumption of a common threshold τ is, probably, more appropriate. Thus, we have one parameter (τ) that can be used in optimizing out-of-sample utility or NSR. The other parameters can be λ, ω and λ_{β} although it is more reasonable to try vague priors as in (16), (18) and (22). Moreover, we can have linear or semi-parametric models as in (24) and (23). In this respect it is more reasonable to start from (24) and (23) and let the data decide whether the parsimony of linear models is justified in terms of out-of-sample utility or NSR. The procedure can be made fully automatic if we “endogenize” τ so that all models and parameters can be estimated at once. The prior for τ is

$$p(\tau) \propto \mathbb{I}_{[a,b]}(\tau), \quad (27)$$

where a and b are set to **10th and 90th sample percentiles** of the variables involved when they are standardized so that their minimum is 0 and their maximum is 1. As the prior is flat imposition of prior information is minimal. The utilities are calculated for $\theta = 0.5$. For the results of Alessi and Detken (2011), we choose $\theta = 0.4$ in line with these authors. This is a reasonable value to balance the costs of Type I and Type II errors.

The mapping from the parameter vector of utility or NSR is highly non-linear but, in the context of SMC and MCMC, for each parameter draw we can compute easily utility or NSR. These measures are, in turn, averaged over the posterior draws to produce draws for utility or NSR themselves. The problem is (i) what is the posterior probability that a particular indicator is included, and (ii) whether it is included in a linear or non-linear way.

The results for currency and banking crises are reported in Table 2. For the costly asset price boom are reported in Table 3. In Table 4 we report results for the debt crisis. A formal evaluation of spurious predictive power is undertaken in section 6.

PLEASE INSERT TABLES 2, 3 AND 4 AROUND HERE

For the most part, the results of EKS are confirmed using our approach although our models end up being quite parsimonious. Non-linearities through the semi-parametric

¹⁰ All computations are performed in Fortran 77 making extensive use of netlib and IMSL software libraries. The platform is an Intel® Core™ i9 – 7900X CPU @ 3.30 GHz, RAM 32 GB **running Windows 10 and the gnu compiler for Fortran 77**. With 15,000 MCMC iterations omitting the first 5,000 to mitigate start up effects and 10^7 particles computations take, on the average, 38.77 minutes of CPU time.

specifications in (24) and (23) prevail for certain variables although this is clearly not universally true. Posterior distributions of out-of-sample utility and NSR are presented in Figures 1 and 2. These figures are constructed using the out-of-sample draws for utility and NSR of the Composite Indicator.

PLEASE INSERT FIGURES 1 AND 2 AROUND HERE

The posterior distributions of out-of-sample utility and NSR show that the indicator variables and the Composite Indicator are quite useful from a Bayesian perspective, coupled with the semi-parametric specifications and the priors that we have used. *Compared to EKS the reduction in NSR is considerable owing to the model-selection priors.* Posterior statistics for important parameters are reported in Table 5.

PLEASE INSERT TABLE 5 AROUND HERE

4.3 Sensitivity analysis

The standard argument against any Bayesian analysis is, of course, the specific use of priors. In this sub-section we address this concern by varying the fundamental hyperparameters of our model. Hyperparameters \bar{n}_λ and \bar{n}_ω in (16) and (18) are varied uniformly in the interval (0.1, 10). For the same priors, hyperparameters \bar{q}_λ and \bar{q}_ω are varied uniformly in the interval (0.01, 10). Hyperparameter $\bar{\vartheta}$ in (22) is varied uniformly between $0.1T^{-1}$ and $2T^{-1}$. We consider 1,000 different configurations¹¹ and we repeat the SMC procedure taking 10,000 iterations away from the converged results using the baseline prior. To minimize computational effort, we use 10^5 particles. For selected configurations the results were virtually identical with 10^7 particles.

In Tables 6a and 6b we report absolute percentage changes in key summaries of the posterior distributions of utility and NSR, respectively, for the different data sets. The absolute percentage changes are relative to posterior means using the baseline prior. These percentage changes refer to all 1,000 alternative prior configurations.

Evidently the changes are minor, if not trivial, testifying that the data dominate our baseline priors: Only trivial changes in utility or NSR are observed if we change the prior configuration.

PLEASE INSERT TABLES 6a AND 6b AROUND HERE

5. Spurious predictive power?

As in Kaminsky (2006) and EKS, we use banking crises as an additional indicator for currency crises along with currency crises as an indicator for banking crises. The test of in-sample performance is based on the entire Kaminsky (2006) dataset. As in EKS, to assess out-of-sample performance we produce recursive crisis forecasts for the period January 1996 to June 2003. In Alessi and Detken (2011), the authors used quarterly data from 1970 to 2007, and identified 29 high-cost and 16 low-cost booms until 2002. As in EKS, we use the period 2002.Q2 – 2007.Q4 for out-of-sample prediction.

¹¹ The number of alternative configurations is set to 1,000 to minimize computational effort without sacrificing thorough sensitivity analysis.

Knedlik and von Schweinitz (2012), focus on the European debt crisis and analyze 11 countries in the European Monetary Union, from the introduction of the Euro until November 2011. Since the crises are near the end of the sample, we follow EKS in performing out-of-sample analysis along the cross-sectional dimension.

To study spurious predictive power, possibly due to idiosyncrasies caused by in-sample fit, we re-estimate the models under leave-one-observation-out scenarios, where observations correspond to crises (Kaminsky, 2006 and Alessi and Detken, 2011) or countries (Knedlik and von Schweinitz, 2012).

Instead of using the bootstrap, we rely on the predictive distribution. Given data $\mathbf{X}_{1:T}$ observed in the period 1 to T, latent variables $\mathbf{A}_{1:T}$ and other time-invariant parameters $\boldsymbol{\theta}$, suppose we need to predict an observation \mathbf{X}_o which has not been used in estimation. The predictive distribution is:

$$p(\mathbf{X}_o|\mathbf{X}_{1:T}) = \int p(\mathbf{X}_o, \mathbf{A}_{1:T}, \boldsymbol{\theta}|\mathbf{X}_{1:T})d\mathbf{A}_{1:T}d\boldsymbol{\theta}, \quad (28)$$

which we can re-write as:

$$p(\mathbf{X}_o|\mathbf{X}_{1:T}) = \int p(\mathbf{X}_o|\mathbf{A}_{1:T}, \boldsymbol{\theta}, \mathbf{X}_{1:T})p(\mathbf{A}_{1:T}, \boldsymbol{\theta}|\mathbf{X}_{1:T})d\mathbf{A}_{1:T}d\boldsymbol{\theta}. \quad (29)$$

Provided we have access to the posterior and thus we have draws $\{\mathbf{A}_{1:T}^{(s)}, \boldsymbol{\theta}^{(s)}, s = 1, \dots, S\}$ from the distribution whose density is $p(\mathbf{A}_{1:T}, \boldsymbol{\theta}|\mathbf{X}_{1:T})$, an accurate approximation to the posterior predictive density is:

$$p(\mathbf{X}_o|\mathbf{X}_{1:T}) \cong S^{-1} \sum_{s=1}^S p(\mathbf{X}_o|\mathbf{A}_{1:T}^{(s)}, \boldsymbol{\theta}^{(s)}, \mathbf{X}_{1:T}). \quad (30)$$

From the decomposition $p(\mathbf{A}_{1:T}, \boldsymbol{\theta}|\mathbf{X}_{1:T}) \propto p(\mathbf{A}_{1:T}|\boldsymbol{\theta}, \mathbf{X}_{1:T})p(\boldsymbol{\theta}|\mathbf{X}_{1:T})$, the latent variables can be drawn conditionally on the parameters, and the parameters can be drawn using the Metropolis-Hastings algorithm described in Technical Appendix B. For this approximation, we do not even need to apply MCMC again as the existing samples can be re-weighted to approximate the new posterior using sampling-importance-resampling (SIR). The length of the re-sample is set to 3,000. Moreover, $p(\mathbf{X}_o|\mathbf{A}_{1:T}, \boldsymbol{\theta}, \mathbf{X}_{1:T})$, is available in closed form, due to our normality assumptions.

For normal data, $LPS_o = \log p(\mathbf{X}_o|\mathbf{X}_{1:T})$ behaves like minus a mean squared error and is known as log predictive score (LPS, Geweke and Amisano, 2010, 2011). Suppose \mathbf{X}_* is an observation whose LPS is maximum relative to all other observations in the full sample, and let $LPS_* = \log p(\mathbf{X}_*|\mathbf{X}_{1:T})$. We define the relative log predictive score as:

$$RLPS = LPS_o - LPS_*. \quad (31)$$

If an observation \mathbf{X}_o can be predicted with relative accuracy, then RLPS should not be too different from zero and, in fact, the relative predictive score:

$$RPS = \exp(LPS_o - LPS_*) \quad (32)$$

provides **evidence** in favor of the hypothesis that an observation \mathbf{X}_o can be predicted equally well compared to the most-well-predicted observation in the full sample. Since RPS is a ratio $RPS < 1$ is evidence against predictability whereas $RPS > 1$ supports predictability. Choosing the observation whose LPS is maximum relative to all other observations in the full

sample is, perhaps, too restrictive. An alternative is to use:

$$LPS_a = T^{-1} \sum_{t=1}^T \log p(\mathbf{X}_t | \mathbf{X}_{s=1, \dots, T, s \neq t}), \quad (33)$$

which is the average log predictive score in the full sample (so it is designed to provide an in-sample measure of fit to compare with out-of-sample accuracy). In turn, we can define a modified RPS as $RPS = \exp(LPS_o - LPS_a)$. Observations with $RPS < 1$ cannot be predicted as well as in-sample observations whereas observations with $RPS \geq 1$ can be predicted at least as better than observations in the estimation sample.

We compare the predictive performance of the latent summary indicator in (12) and (26) using this out-of-sample construction. If the summary indicator performs well by chance or because of spurious dependencies, then RPS should be much lower than 0.32 or 0.37. Whenever RPS exceeds, roughly, 2.71 or 3.16, this¹² would provide overwhelming evidence in favor of out-of-sample performance despite relatively poor in-sample performance. Although rare, this finding cannot be precluded on *a priori* grounds.

In Table 7, we report the range and median value of RPS for each of the three different datasets for the linear and nonlinear models. Evidently, for the linear models, RPS is very close to zero for all data sets, indicating that such models fall short of expectations regarding predictive power. The nonlinear models perform better as median RPS are 0.73 for currency crises, 0.65 for banking crises and 0.87 for asset price booms. These values are well within the bounds we discussed before. The minimum values of RPS are less than the bounds showing that certain crises cannot be predicted. This is not unexpected as a model cannot do a perfect job in terms of prediction. Therefore, we compute also the 5% quantiles for RPS which are 0.39, 0.41 and 0.44 respectively for nonlinear models. For linear models the 5% quantiles are still quite small. In this sense, the nonlinear models perform out-of-sample as well as within-sample and, therefore, predictive power is not spurious. This provides, in turn, direct evidence that the results reported in section 4 are not spurious.

PLEASE INSERT TABLE 7 AROUND HERE

A more stringent test of predictive power due to possibly spurious correlations is provided by comparing directly the in-sample and out-of-sample posterior distributions of utility and NSR. If the in-sample and predictive distributions are quite different, then we have evidence that the in-sample fit does not assure good predictive performance. In econometrics, this is quite common but the question here is whether we can trust the in-sample results from the predictive perspective. Additionally, we provide posterior predictive densities of utility and NSR which are of fundamental interest in predicting crises. Our results are reported in Figures 3 and 4 for utility and NSR respectively. A direct comparison with Figures 1 and 2 shows that in terms of location and spread, the in-sample and predictive posteriors are not very different although, of course, they differ somewhat in terms of shape. This suggests that, roughly, the posteriors are not very different and, therefore, *we cannot attribute predictive ability to pure chance or spurious in-sample correlations*. Under spurious in-sample correlations the in-sample and predictive posterior densities of utility and NSR should have been quite different.

PLEASE INSERT FIGURES 3 AND 4 AROUND HERE

¹² According to Kass and Raftery (1994) these bounds should be $1/e \cong 0.37$ and $e \cong 2.71$. However, these bounds apply to Bayes factors. Here, they are used as rough bounds to determine the “significance” of RPS.

6. Conclusion

In this paper we use a Bayesian approach to early warning mechanisms. We allow for dynamics, variable selection through LASSO, and time-varying coefficients in the crisis indicator. We provide a Bayesian model and we test the effectiveness of the signals approach in three data sets: i) **Currency and banking crises for 76 currency and 26 banking crises in 15 developing and 5 industrial countries between 1970 and 1995**, ii) **costly asset price booms using quarterly data (1970 - 2007)**, and iii) **public debt crises in 11 countries in the European Monetary Union from the introduction of the Euro until November 2011**. The Bayesian model relies on a vector autoregression for indicator variables, and incorporates dynamic factors, time-varying weights in the latent composite indicator and special priors to avoid the proliferation of parameters. **Our EWM is, essentially, a dynamic latent variable that summarizes all relevant information from the model to predict a crisis. This is a novel approach and works quite well in practice.** We find that non-linear models perform better in terms of predictive ability using formal Bayes techniques organized around posterior predictive distributions and log predictive scoring for measures associated with crisis forecasting (utility and signal-to-noise ratio).

The so-called signals approach has certain characteristics compared to traditional early warning systems. In this work, we augmented previous applications of signals and early warning mechanisms with a Bayesian approach, extended to a semi-parametric context to capture non-linearities in economic time series. For all applications considered, we found that the major previous findings are indeed valid but with a more parsimonious approach, whereas the applications showed that *non-linear models are quite successful as early-warning systems, when linear models - even after allowing for breaks - are not so successful*. According to our findings, the proposed model is fully capable of capturing the non-linear crises episodes in a timely manner. Therefore, our approach could be thought of as EWM for crisis detection, which in turn could have important policy implications. In brief, the early identification of bubbles is of utmost importance for policy makers and central bankers as it could prevent: (i) the potential collapse of an economy or an economic sector through regulatory interventions in the relevant markets; (ii) the catastrophic spillover effects in the national or global economy through interest / exchange rate setting.

Constructing indicators and signals that can provide an EWM is, subsequently, of considerable importance for policy authorities, systemic banks and other authorities like central banks and could be a fruitful topic for further investigation. **We find it particularly interesting that a semi-parametric Bayesian approach that allows for breaks, potential non-linearities, and a latent variable approach to crisis detection is successful in out-of-sample detection of crises.** Natural by-products of our analysis include posterior distributions of utility and noise-to-signal ratio which summarize all statistical uncertainty about these quantities. As a Bayesian extension of the signals approach in EKS, the new model is more flexible and more easily adaptable to specific applications through both the specification of the underlying VAR as well as the priors used. The dynamic latent variable for crisis detection is a useful summary of the model that does not sacrifice the inherent ability of the signals approach to detection of crises. Another novel feature of our approach is that the weights in the dynamic latent variable signaling a crisis are estimated from the data so that their relative importance is “endogenized”. Naturally, a number of indicators is found to perform poorly in terms of crisis detection. However, the ones that perform well are easy to determine via the optimal weighting scheme as well as by using the flexible neural networks specification of the composite indicator. Therefore, we can allow for very general forms of nonlinearity in predicting crises.

As the underlying model is based on a Factor-Augmented Panel Vector Autoregression (FA-PVAR) with a LASSO prior to prohibit proliferation of parameters, several aspects of the

model are easy to modify or extend in a modular way. A possible extension of great interest would be to “endogenize” the selection of thresholds for the underlying economic and financial indicators but also the critical parameter θ that is conventionally set by the user in the signals approach. This can provide substantial benefits in terms of utility (balancing type I and II errors) or NSR. As our model is semi-parametric flexible nonlinear components can be easily introduced in the composite indicator which, to the best of our knowledge, is a novel feature relative to the existing literature. The relative success of the new model over non-parametric approaches based on the bootstrap (as in EKS) owes much to the LASSO priors used in the basic FA-PVAR and the hierarchical modelling treatment of the various large covariance matrices involved in the model. Our sensitivity analysis showed that “parsimony priors” imposed on the various blocks of parameters did not affect the reported results in any substantial way. Therefore, another potentially fruitful avenue for future research might be an empirical Bayes approach to select the priors in advance, based on sample information, rather than perform sensitivity analysis. This, clearly, has both attractive and negative aspects but it might prove a more operational approach provided that it works.

References

- Abiad, A. (2003), Early Warning Systems: A Survey and a Regime-Switching Approach, *IMF Working Paper* 32.
- Alessi, L. and Detken, C. (2011), Quasi real time early warning indicators for costly asset price Boom/Bust cycles: a role for global liquidity, *European Journal of Political Economy*, 27 (3), 520–533.
- Andrieu, C. and Roberts, G. O. (2009), The pseudo-marginal approach for efficient Monte Carlo computations, *The Annals of Statistics*, 37 (2), 697–725.
- Andrieu, C., Doucet, A., & Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72, 269–342.
- Bai, J., Ng, S. (2005), Tests for skewness, kurtosis, and normality for time series data, *Journal of Business and Economic Statistics*, 23 (1), 49–60.
- Basak, G. K., P. K. Das, and A. Rohit (2017). Capital inflow-terms of trade ‘nexus’: Does it lead to financial crisis? *Economic Modelling* 65, 18–29.
- Basel Committee on Banking Supervision (2010), *Basel III: A global regulatory framework for more resilient banks and banking systems*, Bank for International Settlements, December. In: Ferguson, N. (Ed.), *The Ascent of Money*. Penguin Press, New York, 2008
- Berg, A. and Pattillo, C. (1999). Predicting currency crises: the indicators approach and an alternative, *Journal of International Money and Finance* 18 (4), 561–586.
- Bernanke, B.S. (2004), Gradualism, in: *Remarks at an Economics Luncheon* Cosponsored by the Federal Reserve Bank of San Francisco (Seattle Branch) and the University of Washington, Seattle, 20 May 2004.
- Billio, M., R. Casarin, M. Costola, and A. Pasqualini (2016). An entropy-based early warning indicator for systemic risk, *Journal of International Financial Markets, Institutions and Money* 45, 42–59.
- Borio, C., Drehmann, M., March (2009), Assessing the risk of banking Crises–Revisited, *BIS Quarterly Review*, 1–18.
- 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.
- Caggiano, G., P. Calice, L. Leonida, G. Kapetanios (2016). Comparing logit-based early warning systems: does the duration of systemic banking crises matter? *Journal of Empirical Finance* 37, 104–116.

Carlin, B.P., Louis, T.A. (2000), *Bayes and Empirical Bayes Methods for Data Analysis*, second ed. Chapman & Hall, London.

Caruana, J. (2010), *The challenge of taking macroprudential decisions: who will press which button(s)?* In: Speech at the 13th Annual International Banking Conference, Federal Reserve Bank of Chicago, in Cooperation with the International Monetary Fund, Chicago, 24 September 2010.

Casarin, R., Marin, J.-M. (2007), Online data processing: Comparison of Bayesian regularized particle filters, University of Brescia, Department of Economics, *Working Paper n. 0704*.

Cheng, X., and H. Zhao (2019). Modeling, analysis and mitigation of contagion in financial systems. *Economic Modelling* 76, 281–292.

Committee on the Global Financial System (CGFS) (2012), *Operationalising the selection and application of macroprudential instruments*, CGFS Publications, No.48.

Caprio, C. and Klingebiel, D. (1996), Bank Insolvencies: Cross-Country Experience, World Bank Policy Research Working Paper 1620.

Cheng, B. and D.M. Titterton (1994), Neural Networks: A Review from a Statistical Perspective, *Statistical Science*, 9 (1), 2-30.

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

Doucet, A., Godsill, S., and Andrieu, C. (2000), On sequential Monte Carlo sampling methods for Bayesian filtering, *Statistics and Computing* 10 (3), 197–208

Doucet, A., de Freitas, N., Gordon, N. (Eds.), (2001), *Sequential Monte Carlo Methods in Practice*, Springer-Verlag.

Doucet, A., Pitt, M., Deligiannidis, G., and Kohn, R. (2014), *Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator*, arXiv preprint arXiv:1210.1871v4.

Drakos, A. A., and G. P. Kouretas (2015). The conduct of monetary policy in the Eurozone before and after the financial crisis. *Economic Modelling* 48, 83–92.

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

El-Shagi, M., T. Knedlik, and G. von Schweinitz (2013) Predicting financial crises: The (statistical) significance of the signals approach, *Journal of International Money and Finance*, (35), 76–103.

Fearnhead P. and Clifford P. (2003), Online inference for hidden Markov models via particle filters, *Journal of the Royal Statistical Society, Series B* (65), 887–899

Fearnhead, P., Papaspiliopoulos, O., and Roberts, G. O. (2008), Particle filters for partially observed diffusions, *Journal of the Royal Statistical Society, Series B*, 70, 1–28.

Frankel, J.A., Rose, A.K. (1996), Currency crashes in emerging markets: an empirical treatment, *Journal of International Economics* 41 (3–4), 351–366.

Fuertes, A.-M. , and E. Kalotychou (2006). Early warning systems for sovereign debt crises: the role of heterogeneity *Computational Statistics & Data Analysis* 51, 1420–1441.

Geweke J, Amisano G (2010). Evaluating the predictive distributions of Bayesian models of asset returns. *International Journal of Forecasting* 26: 216-230.

Geweke J, Amisano G (2011). Optimal prediction pools. *Journal of Econometrics* 164, 130-141.

Guerreiro, D. (2014). Is the European debt crisis a mere balance of payments crisis? *Economic Modelling* 44, Supplement 1, S50–S56.

Lopes. H.F. and R.S. Tsay (2011), Particle Filters and Bayesian Inference in Financial Econometrics, *Journal of Forecasting* 30, 168–209.

- Hans, C. (2009), Bayesian lasso regression, *Biometrika* 96 (4), 835-845.
- Hartman, E.J., J.D. Keeler and J.M. Kowalski (1990), Layered Neural Networks with Gaussian Hidden Units as Universal Approximations, *Neural Computation* 2 (2), 210-215.
- Heryán, T., and P. G. Tzeremes (2017). The bank lending channel of monetary policy in EU countries during the global financial crisis. *Economic Modelling* 67, 10–22.
- Hoggarth, G., Reis, R., and Saporta, V. (2002), Costs of Banking System Instability: Some Empirical Evidence, *Journal of Banking and Finance* 26, 825–855.
- Hornik, K., M. Stinchcombe, H. White (1989), Multilayer feedforward networks are universal approximators, *Neural Networks* 2 (5), 359–366.
- Kamin, S., Schindler, J., Samuel, S. (2001), The Contribution of Domestic and External Factors to Emerging Market Devaluation Crises: An Early Warning Systems Approach, *FRB International Finance Discussion Paper* 711.
- Kaminsky, G.L., Lizondo, S. and Reinhart, C.M. (1998) Leading indicators of currency crises, *IMF 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.
- Kass, R. E., and Raftery, A. E. (1995). Bayes Factors . *Journal of the American Statistical Association* 90 (430): 773-795.
- Kittlmann, K., Tirpak, M., Schweickert, M., Vinhas De Souza, L. (2006), From transition crises to macroeconomic stability? Lessons from a crises early warning system for Eastern European and CIS countries, *Comparative Economic Studies* 48 (3), 410–437.
- Knedlik, T., Scheufele, R. (2008), Forecasting currency crises: which methods signaled the South African crisis of June 2006?, *South African Journal of Economics* 76 (3), 367–383.
- Knedlik, T., von Schweinitz, G. (2012), Macroeconomic imbalances as indicators for debt crises in Europe, *Journal of Common Market Studies* 50 (5), 726–745.
- Koop, G. and Poirier, D. J. (2004), Bayesian Variants of Some Classical Semiparametric Regression Techniques, *Journal of Econometrics*, 123: 259–282.
- Koop, G., Poirier, D. J., and Tobias, J. (2005), Bayesian Semiparametric Inference in Multiple Equation Models, *Journal of Applied Econometrics*, 20: 723–747.
- Laeven, L. and Valencia, F., “Systemic Banking Crises Database: An Update”, *IMF Working Paper*, No WP/12/163, IMF, June 2012.
- Lawrence, M., Goodwin, P., O’Connor, M. and Onkal, D. (2006), Judgmental forecasting: a review of progress over the last 25 years, *International Journal of Forecasting* 22, 493–518.
- Liu, J., West, M. (2001), Combined parameter and state estimation in simulation-based filtering, in: Doucet, A., de Freitas, N., Gordon, N. (Eds.), *Sequential Monte Carlo Methods in Practice*, Springer-Verlag.
- Nemeth, C., C. Sherlock, P. Fearnhead (2016), Particle Metropolis-adjusted Langevin algorithms, *Biometrika* 103, 701-717.
- Mariano, S., Gultekin, B., Ozmucur, S., Shabbir, T., Alper, C. (2004), Prediction of currency crises: case of Turkey, *Review of Middle East Economics and Finance* 2 (2), 87–107.
- Önkal, D., Thomson, M.E., Pollock, A.A.C. (2002), Judgmental forecasting, in: Clements, M.P., Hendry, D.F. (Eds.), *A Companion to Economic Forecasting*, Blackwell Publishers, Malden and Oxford.
- Park, T., and G. Casella (2008), The Bayesian Lasso, *Journal of the American Statistical Association* 103 (482), 681-686.
- Pitt, M.K. and Shephard, N. (1999), Filtering via simulation: Auxiliary particle filters, *Journal of the American Statistical Association*, 94, 590-599.
- Pitt, M. K., Silva, R., Giordani, P., and Kohn, R. (2012), On some properties of Markov

chain Monte Carlo simulation methods based on the particle filter, *Journal of Econometrics*, 171(2), 134–151.

Qin, X., and C. Luo (2014). Capital account openness and early warning system for banking crises in G20 countries. *Economic Modelling* 39, 190–194.

Ristic, B., Arulampalam, S., Gordon, N. (Eds.), (2004), *Beyond the Kalman Filter: Particle Filters for Tracking Applications*, Artech House, Boston.

Robert, C.P. (2001), *The Bayesian Choice*, second ed. Springer-Verlag, New York.

Roberts, G. O., Gelman, A., and Gilks, W. (1997), Weak Convergence and Optimal Scaling of the Random Walk Metropolis Algorithms, *The Annals of Applied Probability*, 7(1), 110–120.

Roberts, G. O. and Rosenthal, J. S. (1998), Optimal scaling of discrete approximations to Langevin diffusions, *Journal of the Royal Statistical Society: Series B*, 60(1), 255–268.

Tibshirani, R. (1996), Regression Shrinkage and Selection via the Lasso, *Journal of the Royal Statistical Society: Series B*, 58, 267–288.

Tierney, L. (1994), Markov chains for exploring posterior distributions, *Annals of Statistics* 22, 1701–1762.

Ureche-Rangau, L., and A. Burietz (2013). One crisis, two crises...the subprime crisis and the European sovereign debt problems. *Economic Modelling* 35, 35–44.

Wahba, G. (1990), *Spline models for observational data*, SIAM, Philadelphia, PA.

Wasserman, L. (2004), *All of Statistics: A Concise Course in Statistical Inference*, Springer-Verlag, New York.

TECHNICAL APPENDIX A. Derivation of likelihood / posterior

We summarize the model below.

Conditional on $1 \leq G < n$ global factors \mathbf{f}_t we have:

$$\mathbf{X}_t^{(n)} = \mathbf{A}_n(L)\mathbf{f}_t + \mathbf{v}_t^{(n)}, \forall n \in \mathbb{N}, \quad (\text{A.1})$$

For the error term we assume:

$$\xi_t \sim \mathcal{N}(0, \Sigma_\xi) \forall t \in \mathbb{T}. \quad (\text{A.2})$$

For the individual country-specific components we assume:

$$\mathbf{v}_t^{(n)} = \mathbf{D}(L)\mathbf{v}_t^{(n)} + \mathbf{c}_t^{(n)} + \boldsymbol{\varepsilon}_t^{(n)}, \quad (\text{A.3})$$

These country-specific effects have the following structure:

$$\mathbf{c}_t^{(n)} = \mathbf{c}_{t-1}^{(n)} + \mathbf{1}_M \cdot \gamma_b^{(n)} D_{b,t}^{(n)} + \boldsymbol{\zeta}_t^{(n)} \forall t \in \mathbb{T}, \forall n \in \mathbb{N} \quad (\text{A.4})$$

where $D_{b,t}^{(n)} = 1$ if there is a structural break at date t for country n , and zero otherwise.

For the error terms we assume:

$$\boldsymbol{\zeta}_t^{(n)} \sim \mathcal{N}(\mathbf{0}, \Sigma_\zeta^{(n)}), \boldsymbol{\varepsilon}_t^{(n)} \sim \mathcal{N}(\mathbf{0}, \Sigma_\varepsilon^{(n)}) \forall n \in \mathbb{N}. \quad (\text{A.5})$$

For the aggregation function we have a *stochastic weighting scheme*:

$$X_t^* = \sum_{n \in \mathbb{N}} \alpha_{t,n} X_t^{(n)} + u_t =: \boldsymbol{\alpha}_t' \mathbf{X}^{(t)} + u_t \forall t \in \mathbb{T}, \quad (\text{A.6})$$

where $\mathbf{X}^{(t)} := \{X_t^{(n)}, t \in \mathbb{T}\}$, the weights $\alpha_{t,n} \geq 0, \forall n \in \mathbb{N}$ and $\sum_{n \in \mathbb{N}} \alpha_{t,n} = 1 \forall t \in \mathbb{T}$. Dynamics in (12) can be accommodated using an AR(R) process for the error term:

$$\mathbf{u}_t = \sum_{r=1}^R \rho_r \mathbf{u}_{t-r} + \epsilon_t, \epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2) \forall t \in \mathbb{T}. \quad (\text{A.7})$$

For the weights $\boldsymbol{\alpha}_t = (\alpha_{t,n}, n \in \mathbb{N})$ we assume that they are time-varying:

$$\boldsymbol{\alpha}_t = 2\boldsymbol{\alpha}_{t-1} - \boldsymbol{\alpha}_{t-2} + \mathbf{v}_{\boldsymbol{\alpha},t}, \quad (\text{A.8})$$

where

$$\mathbf{v}_{\boldsymbol{\alpha},t} \sim \mathcal{N}_n(\mathbf{0}, \lambda^2 \mathbf{I}), \forall t \in \mathbb{T}. \quad (\text{A.9})$$

We derive the posterior for date t as we will apply SMC to compute the likelihood / posterior.

From (6) and (9) we have:

$$\mathbf{X}_t^{(n)} - \mathbf{D}(L)\mathbf{X}_t^{(n)} = (\mathbf{I} - \mathbf{D}(L))\mathbf{A}_n(L)\mathbf{f}_t + \mathbf{c}_t^{(n)} + \boldsymbol{\varepsilon}_t^{(n)}, \forall n \in \mathbb{N}. \quad (\text{A.10})$$

Therefore

$$p(\mathbf{X}_t^{(n)} | \mathbf{X}_{t,\text{lags}}^{(n)}, \mathbf{f}_t, \mathbf{c}_t^{(n)}, \boldsymbol{\theta}) \propto |\boldsymbol{\Sigma}_\varepsilon^{(n)}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \mathbf{U}_t^{(n)'} \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{U}_t^{(n)}\right), \quad (\text{A.11})$$

where $\mathbf{X}_{t,\text{lags}}^{(n)}$ denotes lagged values of $\mathbf{X}_t^{(n)}$ and $\boldsymbol{\theta}$ denotes the ‘‘structural’’ parameters in the following matrices / vectors / scalars:

$$\mathbf{D}, \{\mathbf{A}_n\}, \boldsymbol{\Sigma}_\varepsilon, \mathbf{B}, \mathbf{C}, \sigma_\varepsilon^2, \boldsymbol{\rho}, \lambda, \{\gamma_b^{(n)}\}, \boldsymbol{\Sigma}_\zeta,$$

and

$$\mathbf{U}_t^{(n)} = \mathbf{X}_t^{(n)} - \mathbf{D}(L)\mathbf{X}_t^{(n)} - (\mathbf{I} - \mathbf{D}(L))\mathbf{A}_n(L)\mathbf{f}_t - \mathbf{c}_t^{(n)}. \quad (\text{A.12})$$

If we define $\mathbf{X}^{(t)} = [\mathbf{X}_t^{(n)}, \forall n \in \mathbb{N}]$, we have

$$p(\mathbf{X}^{(t)} | \mathbf{X}_{t,\text{lags}}^{(t)}, \mathbf{f}_t, \{\mathbf{c}_t^{(n)}, \forall n \in \mathbb{N}\}, \boldsymbol{\theta}) = \prod_{n \in \mathbb{N}} p(\mathbf{X}_t^{(n)} | \mathbf{X}_{t,\text{lags}}^{(n)}, \mathbf{f}_t, \mathbf{c}_t^{(n)}, \boldsymbol{\theta}). \quad (\text{A.13})$$

From (7) we obtain:

$$p(\mathbf{f}_t | \mathbf{f}_{t,\text{lags}}, X_t^*, \boldsymbol{\theta}) \propto |\boldsymbol{\Sigma}_\xi|^{-\frac{n}{2}} \exp\left(-\frac{1}{2} \sum_{n \in \mathbb{N}} \mathbf{U}_{f_t}^{(n)'} \boldsymbol{\Sigma}_\xi^{-1} \mathbf{U}_{f_t}^{(n)}\right), \quad (\text{A.14})$$

where $\mathbf{U}_{f_t}^{(n)} = [\mathbf{I} - \mathbf{B}(L)]\mathbf{f}_t - \mathbf{C}(L)X_t^*$.

From (12) we obtain:

$$p(X_t^* | \boldsymbol{\alpha}_t, X_{t,\text{lags}}^*, \mathbf{X}^{(t)}, \boldsymbol{\theta}) \propto \sigma_\varepsilon^{-1} \exp\left(-\frac{1}{2\sigma_\varepsilon^2} [X_t^* - \boldsymbol{\alpha}_t' \mathbf{X}^{(t)} - \rho(L)X_t^* + \rho(L)\boldsymbol{\alpha}_t' \mathbf{X}^{(t)}]\right), \quad (\text{A.15})$$

where $\rho(L)$ represents the lag polynomial in (13).

From (14) and (15)

$$p(\boldsymbol{\alpha}_t | \boldsymbol{\alpha}_{t,\text{lags}}, \lambda) \propto \lambda^{-\frac{n}{2}} \exp\left(-\frac{1}{2\lambda^2 \sum_{n \in \mathbb{N}} [\alpha_{t,n} - 2\alpha_{t-1,n} + \alpha_{t-2,n}]^2}\right). \quad (\text{A.16})$$

From (10) we obtain:

$$p(\mathbf{c}_t^{(n)} | \mathbf{c}_{t,\text{lags}}^{(n)}, \gamma_b^{(n)}, D_{b,t}^{(n)}, \boldsymbol{\Sigma}_\zeta^{(n)}) \propto |\boldsymbol{\Sigma}_\zeta^{(n)}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \mathbf{U}_{c_t}^{(n)'} \boldsymbol{\Sigma}_\zeta^{-1} \mathbf{U}_{c_t}^{(n)}\right), \quad (\text{A.17})$$

where $\mathbf{U}_{c_t}^{(n)} = \mathbf{c}_t^{(n)} - \mathbf{c}_{t-1}^{(n)} - \mathbf{1}_M \gamma_b^{(n)} D_{b,t}^{(n)}$. Suppose $\mathbf{c}^{(t)} = [\mathbf{c}_t^{(n)} \forall n \in \mathbb{N}]$. Then, we have:

$$p(\mathbf{c}^{(t)} | \mathbf{c}^{(t,\text{lags})}, \{\gamma_b^{(n)} \forall n \in \mathbb{N}\}, \{D_{b,t}^{(n)} \forall n \in \mathbb{N}\}, \{\boldsymbol{\Sigma}_\zeta^{(n)} \forall n \in \mathbb{N}\}) \propto \prod_{n \in \mathbb{N}} p(\mathbf{c}_t^{(n)} | \mathbf{c}_{t,\text{lags}}^{(n)}, \gamma_b^{(n)}, D_{b,t}^{(n)}, \boldsymbol{\Sigma}_\zeta^{(n)}). \quad (\text{A.18})$$

The likelihood function is

$$\begin{aligned} \ell(\boldsymbol{\theta}; \mathbf{Y}^{(t)}, \boldsymbol{\Lambda}_t) &\propto p(\mathbf{X}^{(t)} | \mathbf{X}_{t,\text{lags}}^{(t)}, \mathbf{f}_t, \{\mathbf{c}_t^{(n)}, \forall n \in \mathbb{N}\}, \boldsymbol{\theta}) \\ &\quad \cdot p(\mathbf{f}_t | \mathbf{f}_{t,\text{lags}}, X_t^*, \boldsymbol{\theta}) \\ &\quad \cdot p(X_t^* | \boldsymbol{\alpha}_t, X_{t,\text{lags}}^*, \mathbf{X}^{(t)}, \boldsymbol{\theta}) \end{aligned}$$

$$\cdot p(\boldsymbol{\alpha}_t | \boldsymbol{\alpha}_{t, \text{lags}}, \lambda) p(\mathbf{c}^{(t)} | \mathbf{c}^{(t, \text{lags})}, \{\gamma_b^{(n)} \forall n \in \mathbb{N}\}, \{D_{b,t}^{(n)} \forall n \in \mathbb{N}\}, \{\boldsymbol{\Sigma}_\zeta^{(n)} \forall n \in \mathbb{N}\}), \quad (\text{A.19})$$

where all densities have been defined above up to normalizing constants and $\mathbf{Y}^{(t)}$ denotes all available data up to time t .

To compute the likelihood, we have to integrate out the latent variables $\boldsymbol{\Lambda}_t = \{\mathbf{f}_t, \boldsymbol{\alpha}_t, \mathbf{c}_t \forall t \in \mathbb{T}\}$. The latent variables are integrated out using SMC – particle-filtering. Therefore, we can obtain a consistent (as the number of particles increases) estimator of $\ell(\boldsymbol{\theta}; \mathbf{Y}^{(t)})$. The overall posterior density is

$$p(\boldsymbol{\theta} | \mathbf{Y}) = p(\boldsymbol{\theta}) \cdot \prod_{t \in \mathbb{T}} \ell(\boldsymbol{\theta}; \mathbf{Y}^{(t)}), \quad (\text{A.20})$$

where $p(\boldsymbol{\theta})$ is the prior, and $\mathbf{Y} = \{\mathbf{Y}^{(t)} \forall t \in \mathbb{T}\}$.

To implement MCMC we use $\ell(\boldsymbol{\theta}; \mathbf{Y}^{(t)}, \boldsymbol{\Lambda}_t)$ treating $\boldsymbol{\Lambda}_t$ as a block. Moreover, parameters $\boldsymbol{\theta}$ are drawn using the MCMC described in Appendix B.

It remains to determine the number of factors G and the structural breaks. Given the dates of structural breaks we compute the posterior $p(\boldsymbol{\theta} | \mathbf{Y})$ for several values of $G \in \{1, \dots, \bar{G}\}$, where \bar{G} is set to 5. In turn, we compute the marginal likelihood $p(\mathbf{Y} | G)$ which is a by-product of MCMC.

Procedure DBR: Determination of Breaks

To determine the dates of structural breaks we proceed as follows. The number of breaks we allow is $b \in \mathbb{B} = \{1, \dots, \bar{B}\}$ where $\bar{B} \ll T$. We need to determine whether $D_{b,t} = 1$ or zero for a given date t^* . We compute the marginal likelihood for $t^* \in \{T_o, T_o + 1, \dots, T\}$ where $T_o > 1$ is the minimal starting value for t given our lag structure and the maximum value of the marginal likelihood determines t^* . Then we apply SMC again, to determine two dates $t_1^* < t_2^*$ and we recompute the marginal likelihood. If it is lower than the marginal likelihood for a single break at t^* we stop, otherwise we move on to the case of three breaks $t_1^* < t_2^* < t_3^*$. In all cases we examined, we have one or two breaks. This procedure, determines simultaneously the optimal number of factors G as well and the timings of breaks.

Below we provide pseudo code that implements the approach.

Table A1. Pseudo Code

-
- 1) For $G = \{1, \dots, \bar{G}\}$ do:
 - 2) Determine number of breaks and their timings using procedure DBR above.
 - 3) Integrate out the latent variables, $\{\boldsymbol{\Lambda}_t, t = 1, \dots, T\}$, as a block using SMC following the steps in Table B1.
 - 4) Obtain an unbiased estimator of $p(\boldsymbol{\theta} | \mathbf{Y})$ in (A.20).
 - 5) Use MCMC to obtain draws $\{\boldsymbol{\theta}^{(s)}, s = 1, \dots, S\}$ from (A.20) using the Particle Metropolis-adjusted algorithm in (B.17) – (B.17).
 - 6) Increase G by one, until the marginal likelihood falls.
-

TECHNICAL APPENDIX B

Particle filtering

The particle filter methodology can be applied to state space models of the form:

$$y_T \sim p(y_t|s_t), \quad s_t \sim p(s_t|s_{t-1}), \quad (\text{B.1})$$

where s_t is a state variable. For general introductions see [Andrieu et al. \(2010, pp. 272, 277\)](#), [Gordon et al. \(1993\)](#), [Doucet et al. \(2001\)](#) and [Ristic et al. \(2004\)](#). Given the data Y_t the posterior distribution $p(s_t|Y_t)$ can be approximated by a set of (auxiliary) particles $\{s_t^{(i)}, i = 1, \dots, N\}$ with probability weights $\{w_t^{(i)}, i = 1, \dots, N\}$ where $\sum_{i=1}^N w_t^{(i)} = 1$. The predictive density is approximated by

$$\begin{aligned} p(s_{t+1}|Y_t) &= \int p(s_{t+1}|s_t)p(s_t|Y_t)ds_t \\ &\simeq \sum_{i=1}^N p(s_{t+1}|s_t^{(i)})w_t^{(i)}, \end{aligned} \quad (\text{B.2})$$

where $w_t^{(i)} = \frac{p(s_t^{(i)}|Y_t)}{\sum_{j=1}^N p(s_t^{(j)}|Y_t)}$.

An approximation for the filtering density is

$$\begin{aligned} p(s_{t+1}|Y_t) &\propto p(y_{t+1}|s_{t+1})p(s_{t+1}|Y_t) \\ &\simeq p(y_{t+1}|s_{t+1}) \sum_{i=1}^N p(s_{t+1}|s_t^{(i)})w_t^{(i)}. \end{aligned} \quad (\text{B.3})$$

Particle filtering propagates $\{s_t^{(i)}, w_t^{(i)}, i = 1, \dots, N\}$ to the next step, viz. $\{s_{t+1}^{(i)}, w_{t+1}^{(i)}, i = 1, \dots, N\}$ but this often suffers from a weight degeneracy problem. If parameters $\theta \in \Theta \in \mathfrak{R}^k$ are available, as is often the case, we follow [Liu and West \(2001\)](#). In this context, parameter learning takes place via a mixture of multivariate normal distributions:

$$p(\theta|Y_t) \simeq \sum_{i=1}^N w_t^{(i)} N(\theta|a\theta_t^{(i)} + (1-a)\bar{\theta}_t, b^2V_t), \quad (\text{B.4})$$

where $\bar{\theta}_t = \sum_{i=1}^N w_t^{(i)} \theta_t^{(i)}$, and $V_t = \sum_{i=1}^N w_t^{(i)} (\theta_t^{(i)} - \bar{\theta}_t)(\theta_t^{(i)} - \bar{\theta}_t)'$. The constants a and b are related to shrinkage and are determined via a discount factor $\delta \in (0,1)$ as $a = (1 - b^2)^{1/2}$ and $b^2 = 1 - [(3\delta - 1)/2\delta]^2$. See also [Casarin and Marin \(2007\)](#). [Andrieu and Roberts \(2009\)](#), [Flury and Shephard \(2011\)](#) and [Pitt et al. \(2012\)](#) suggested the Particle Metropolis-Hastings (PMCMC) technique which uses an unbiased estimator of the likelihood function $\hat{p}_N(Y|\theta)$ as $p(Y|\theta)$ as the latter is often not available in closed form. Given the current state of the parameter $\theta^{(j)}$ and the current estimate of the likelihood, say $L^j = \hat{p}_N(Y|\theta^{(j)})$, a candidate θ^c is drawn from $q(\theta^c|\theta^{(j)})$ yielding $L^c = \hat{p}_N(Y|\theta^c)$. Then, we set $\theta^{(j+1)} = \theta^c$ with the Metropolis - Hastings probability:

$$A = \min \left\{ 1, \frac{p(\theta^c)L^c q(\theta^{(j)}|\theta^c)}{p(\theta^{(j)})L^j q(\theta^c|\theta^{(j)})} \right\}, \quad (\text{B.5})$$

otherwise we repeat the current draws: $\{\theta^{(j+1)}, L^{j+1}\} = \{\theta^{(j)}, L^j\}$. [Hall, Pitt and Kohn \(2014\)](#) propose an auxiliary particle filter which rests upon the idea that adaptive particle filtering ([Pitt](#)

et al., 2012) used within PMCMC requires far fewer particles **than** the standard particle filtering algorithm to approximate $p(Y|\theta)$. From Pitt and Shephard (1999) we know that auxiliary particle filtering can be implemented easily once we can evaluate the state transition density $p(s_t|s_{t-1})$. When this is not possible, Hall, Pitt and Kohn (2014) present a new approach when, for instance, $s_t = g(s_{t-1}, u_t)$ for a certain disturbance. In this case we have

$$p(y_t|s_{t-1}) = \int p(y_t|s_t)p(s_t|s_{t-1})ds_t, \quad (\text{B.6})$$

$$p(u_t|s_{t-1}; y_t) = p(y_t|s_{t-1}, u_t)p(u_t|s_{t-1})/p(y_t|s_{t-1}). \quad (\text{B.7})$$

If we can evaluate $p(y_t|s_{t-1})$ and simulate from $p(u_t|s_{t-1}; y_t)$ the filter would be fully adaptable (Pitt and Shephard, 1999). One can use a Gaussian approximation for the first-stage proposal $g(y_t|s_{t-1})$ by matching the first two moments of $p(y_t|s_{t-1})$. So in some way we find that the approximating density $p(y_t|s_{t-1}) = N(\mathbb{E}(y_t|s_{t-1}), \mathbb{V}(y_t|s_{t-1}))$. In the second stage, we know that $p(u_t|y_t, s_{t-1}) \propto p(y_t|s_{t-1}, u_t)p(u_t)$. For $p(u_t|y_t, s_{t-1})$ we know it is multimodal so suppose it has M modes are \hat{u}_t^m , for $m = 1, \dots, M$. For each mode we can use a Laplace approximation. Let $l(u_t) = \log[p(y_t|s_{t-1}, u_t)p(u_t)]$. From the Laplace approximation we obtain:

$$l(u_t) \simeq l(\hat{u}_t^m) + \frac{1}{2}(u_t - \hat{u}_t^m)' \nabla^2 l(\hat{u}_t^m)(u_t - \hat{u}_t^m). \quad (\text{B.8})$$

Then we use a a mixture approximation:

$$g(u_t|x_t, s_{t-1}) = \sum_{m=1}^M \lambda_m (2\pi)^{-d/2} |\Sigma_m|^{-1/2} \exp \left\{ \frac{1}{2} (u_t - \hat{u}_t^m)' \Sigma_m^{-1} (u_t - \hat{u}_t^m) \right\}, \quad (\text{B.9})$$

where $\Sigma_m = -[\nabla^2 l(\hat{u}_t^m)]^{-1}$ and $\lambda_m \propto \exp\{l(\hat{u}_t^m)\}$ with $\sum_{m=1}^M \lambda_m = 1$. This is done for each particle s_t^i . This is known as the Auxiliary Disturbance Particle Filter (ADPF).

An alternative is the independent particle filter (IPF) of Lin et al. (2005). The IPF forms a proposal for s_t directly from the measurement density $p(y_t|s_t)$ although Hall, Pitt and Kohn (2014) are quite right in pointing out that the state equation can be very informative. In standard particle filtering, particles are simulated through the state density $p(s_t^i|s_{t-1}^i)$ and they are re-sampled with weights determined by the measurement density evaluated at the resulting particle, viz. $p(y_t|s_t^i)$.

The ADPF is simple to construct and rests upon the following steps:

Table B1. Pseudo Code for Particle Filtering

For $t = 0, \dots, T - 1$ given samples $s_t^k \sim p(s_t|Y_{1:t})$ with mass π_t^k for $k = 1, \dots, N$.

- 1) For $k = 1, \dots, N$ compute $\omega_{t|t+1}^k = g(y_{t+1}|s_t^k)\pi_t^k, \pi_{t|t+1}^k = \omega_{t|t+1}^k / \sum_{i=1}^N \omega_{t|t+1}^i$.
- 2) For $k = 1, \dots, N$ draw $\tilde{s}_t^k \sim \sum_{i=1}^N \pi_{t|t+1}^i \delta_{s_t^i}^i(ds_t)$.
- 3) For $k = 1, \dots, N$ draw $u_{t+1}^k \sim g(u_{t+1}^k|\tilde{s}_t^k, y_{t+1})$ and set $s_{t+1}^k = h(s_t^k; u_{t+1}^k)$.
- 4) For $k = 1, \dots, N$ compute

$$\omega_{t+1}^k = \frac{p(y_{t+1}|s_{t+1}^k)p(u_{t+1}^k)}{g(y_{t+1}|s_t^k)g(u_{t+1}^k|\tilde{s}_t^k, y_{t+1})}, \pi_{t+1}^k = \frac{\omega_{t+1}^k}{\sum_{i=1}^N \omega_{t+1}^i}. \quad (\text{B.10})$$

The estimate of likelihood from ADPF is

$$p(Y_{1:T}) = \prod_{t=1}^T \left(\sum_{i=1}^N \omega_{t-1|t}^i \right) \left(N^{-1} \sum_{i=1}^N \omega_t^i \right). \quad (\text{B.11})$$

Particle Metropolis adjusted Langevin filters

Nemeth, Sherlock and Fearnhead (2016) provide a particle version of a Metropolis adjusted Langevin algorithm (MALA). In Sequential Monte Carlo we are interested in approximating $p(s_t|Y_{1:t}, \theta)$. Given that

$$\begin{aligned} & p(s_t|Y_{1:t}, \theta) \\ & \propto g(y_t|x_t, \theta) \int f(s_t|s_{t-1}, \theta) p(s_{t-1}|y_{1:t-1}, \theta) ds_{t-1}, \end{aligned} \quad (\text{B.12})$$

where $p(s_{t-1}|y_{1:t-1}, \theta)$ is the posterior as of time $t - 1$. If at time $t - 1$ we have a **set** of particles $\{s_{t-1}^i, i = 1, \dots, N\}$ and weights $\{w_{t-1}^i, i = 1, \dots, N\}$ which form a discrete approximation for $p(s_{t-1}|y_{1:t-1}, \theta)$ then we have the approximation:

$$\hat{p}(s_{t-1}|y_{1:t-1}, \theta) \propto \sum_{i=1}^N w_{t-1}^i f(s_t|s_{t-1}^i, \theta). \quad (\text{B.13})$$

See Doucet et al. (2001, 2014) for reviews. From (A.13) Fearnhead et al. (2008) make the important observation that the joint probability of sampling particle s_{t-1}^i and state s_t is:

$$\omega_t = \frac{w_{t-1}^i g(y_t|s_t, \theta) f(s_t|s_{t-1}^i, \theta)}{\xi_t^i q(s_t|s_{t-1}^i, y_t, \theta)}, \quad (\text{B.14})$$

where $q(s_t|s_{t-1}^i, y_t, \theta)$ is a density function amenable to simulation,

$$\xi_t^i q(s_t|s_{t-1}^i, y_t, \theta) \simeq c g(y_t|s_t, \theta) f(s_t|s_{t-1}^i, \theta), \quad (\text{B.15})$$

and c is the normalizing constant in (A.12).

In the MALA algorithm of Roberts and Rosenthal (1998)¹³ we form a proposal

$$\theta^c = \theta^{(s)} + \lambda z + \frac{\lambda^2}{2} \nabla \log p(\theta^{(s)}|Y_{1:T}), \quad (\text{B.16})$$

where $z \sim N(0, I)$ which should result in larger jumps and better mixing properties, plus lower autocorrelations for a certain scale parameter λ . Acceptance probabilities are

$$\begin{aligned} a(\theta^c|\theta^{(s)}) &= \min \left\{ 1, \frac{p(\theta^c|Y_{1:T}) q(\theta^{(s)}|\theta^c)}{p(\theta^{(s)}|Y_{1:T}) q(\theta^c|\theta^{(s)})} \right\} \\ &= \min \left\{ 1, \frac{p(Y_{1:T}|\theta^c) p(\theta^c) q(\theta^{(s)}|\theta^c)}{p(Y_{1:T}|\theta^{(s)}) p(\theta^{(s)}) q(\theta^c|\theta^{(s)})} \right\}. \end{aligned} \quad (\text{B.17})$$

Using particle filters it is possible to create an approximation of the score vector using Fisher's identity:

¹³The benefit of MALA over Random-Walk-Metropolis arises when the number of parameters n is large. This happens because the scaling parameter λ is $O(n^{-1/2})$ for Random-Walk-Metropolis but it is $O(n^{-1/6})$ for MALA, see Roberts et al. (1997) and Roberts and Rosenthal (1998).

$$\nabla \log p(Y_{1:T}|\boldsymbol{\theta}) = E[\nabla \log p(s_{1:T}, Y_{1:T}|\boldsymbol{\theta})], \quad (\text{B.18})$$

which corresponds to the expectation of

$\nabla \log p(s_{1:T}, Y_{1:T}|\boldsymbol{\theta}) = \nabla \log p(s_{1:T-1}, Y_{1:T-1}|\boldsymbol{\theta}) + \nabla \log g(y_T|s_T, \boldsymbol{\theta}) + \nabla \log f(s_T|s_{T-1}, \boldsymbol{\theta})$, over the path $s_{1:T}$. The particle approximation to the score vector results from replacing $p(s_{1:T}|Y_{1:T}, \boldsymbol{\theta})$ with a particle approximation $\hat{p}(s_{1:T}|Y_{1:T}, \boldsymbol{\theta})$. With particle i at time $t-1$ we can associate a value $\alpha_{t-1}^i = \nabla \log p(s_{1:t-1}^i, Y_{1:t-1}|\boldsymbol{\theta})$ which can be updated recursively. As we sample κ_i in the APF (the index of particle at time $t-1$ that is propagated to produce the i th particle at time t) we have the update:

$$\alpha_t^i = \alpha_{t-1}^{\kappa_i} + \nabla \log g(y_t|s_t^i, \boldsymbol{\theta}) + \nabla \log f(s_t^i|s_{t-1}^{\kappa_i}, \boldsymbol{\theta}). \quad (\text{B.19})$$

To avoid problems with increasing variance of the score estimate $\nabla \log p(Y_{1:t}|\boldsymbol{\theta})$ we can use the approximation:

$$\alpha_{t-1}^i \sim N(m_{t-1}^i, V_{t-1}) \quad (\text{B.20})$$

The mean is obtained by shrinking α_{t-1}^i towards the mean of α_{t-1} as follows:

$$m_{t-1}^i = \delta \alpha_{t-1}^i + (1 - \delta) \sum_{i=1}^N w_{t-1}^i \alpha_{t-1}^i, \quad (\text{B.21})$$

where $\delta \in (0,1)$ is a shrinkage parameter. Using Rao-Blackwellization one can avoid sampling α_t^i and instead use the following recursion for the means:

$$m_t^i = \delta m_{t-1}^{\kappa_i} + (1 - \delta) \sum_{i=1}^N w_{t-1}^i m_{t-1}^i + \nabla \log g(y_t|s_t^i, \boldsymbol{\theta}) + \nabla \log f(s_t^i|s_{t-1}^{\kappa_i}, \boldsymbol{\theta}), \quad (\text{B.22})$$

which yields the final score estimate:

$$\nabla \log \hat{p}(Y_{1:t}|\boldsymbol{\theta}) = \sum_{i=1}^N w_t^i m_t^i. \quad (\text{B.23})$$

As a rule of thumb **Nemeth, Sherlock and Fearnhead (2016)** suggest taking $\delta = 0.95$. Furthermore, they show the important result that the algorithm should be tuned to the asymptotically optimal acceptance rate of 15.47% and the number of particles must be selected so that the variance of the estimated log-posterior is about 3. Additionally, if measures are not taken to control the error in the variance of the score vector there is no gain over a simple random walk proposal. The marginal likelihood is given by:

$$p(Y_{1:T}|\boldsymbol{\theta}) = p(y_1|\boldsymbol{\theta}) \prod_{t=2}^T p(y_t|Y_{1:t-1}, \boldsymbol{\theta}), \quad (\text{B.24})$$

where

$$p(y_t|Y_{1:t-1}, \boldsymbol{\theta}) = \int g(y_t|s_t) \int f(s_t|s_{t-1}, \boldsymbol{\theta}) p(s_{t-1}|Y_{1:t-1}, \boldsymbol{\theta}) ds_{t-1} ds_t, \quad (\text{B.25})$$

provides, in explicit form, the predictive likelihood.

Table 1. **Different cases**

		$C_{t,n}$	
		1	0
$S_{tn,m}$	1	A	B
	0	C	D

Notes: Combination of crisis and signaling variable. The letters A, B, C, and D denote both the sets themselves and the number of elements in the respective sets.

Source: El-Shagi, Knedlik, and von Schweinitz (2013, p. 80).

Table 2. **Currency and Banking Crises**

	Currency crisis		Banking crisis	
	P	non-linearity	P	non-linearity
Currency crisis	—	—	0.813	x
M2 multiplier	0.0017		0.765	
Domestic credit	0.215		0.873	
Real interest rate	0.417		0.720	x
Lending-deposit rate ratio	0.000		0.001	
Excess M1 balances	0.000		0.000	
M2 (%reserves)	0.000		0.122	
Bank deposits	0.000		0.000	
Exports	0.000		0.000	
Imports	0.000		0.000	
Terms of trade	0.002		0.000	
Real exchange rate	0.717		0.745	x
Output	0.887	x	0.844	x
Stock prices	0.220		0.873	
Deficit (%GDP)	0.781	x	0.045	x
Banking crisis	0.891	—	—	
Structural Breaks	3		3	

Notes: P denotes the posterior probability that an indicator is included. “Non-linearity” is denoted by “x” when present in the semi-parametric specification.

Table 3. Costly Asset Price Boom

	P	non-linearity
LRN_lev	0.515	
LRR_lev	0.000	
SRN_lev	0.000	
SRR_lev	0.000	
SPREAD_lev	0.828	x
SPREAD_-lev	0.000	
CPI_yoy	0.727	x
CONS_yoy	0.701	
INV_yoy	0.708	
HINV_yoy	0.716	
REX_yoy	0.000	
M1_yoy	0.000	
M3_yoy	0.000	
PCR_yoy	0.815	
DCR_yoy	0.000	
CPI_cum	0.033	x
CONS_cum	0.894	x
INV_cum	0.788	
HINV_cum	0.000	
REX_cum	0.000	
M1_cum	0.000	
M3_cum	0.000	
PCR_cum	0.000	
DCR_cum	0.000	
GDPR_detr	0.825	x
LRN_detr	0.904	
LRR_detr	0.000	
SRN_detr	0.000	
SRR_detr	0.000	
REX_detr	0.811	
CONStoGDP_detr	0.744	x
INVtoGDP_detr	0.509	
HINVtoGDP_detr	0.689	
M1toGDP_detr	0.008	
M3toGDP_detr	0.715	
PCNtoGDP_detr	0.000	
DCNtoGDP_detr	0.000	
GDPR_HP	0.000	
LRN_HP	0.000	
LRR_HP	0.000	
SRN_HP	0.000	
SRR_HP	0.000	
REX_HP	0.000	
CONStoGDP_HP	0.000	
INVtoGDP_HP	0.774	x
HINVtoGDP_HP	0.000	

M1toGDP_HP	0.000
M3toGDP_HP	0.000
PCNtoGDP_HP	0.000
DCNtoGDP_HP	0.000
Structural Breaks	2

Notes: P denotes the posterior probability that an indicator is included. “non-linearity” is denoted by “x” when present in the semi-parametric specification.

Table 4. **Debt Crisis**

	Actual		Counterfactual	
	P	non-linearity	P	non-linearity
Government debt (%GDP)	0.000		0.000	
Government deficit (%GDP)	0.000		0.000	
Interest payment (% gov expenditure)	0.827		0.813	
Unit labor costs	0.612		0.624	
Current account	0.085		0.091	
Share in world trade	0.000		0.000	
Unemployment rate	0.562		0.555	
Labor participation rate	0.817		0.822	
Private debt	0.826	x	0.814	x
Non-MFI debt	0.891		0.896	
Household debt	0.885	x	0.873	x
Foreign assets	0.000		0.000	
Inflation	0.000		0.000	
Asset prices	0.000		0.000	
HICP-competitiveness	0.000		0.000	
GDP-deflated competitiveness	0.005		0.000	
ULC-competitiveness	0.012		0.017	
10 continuous random variables	—		0.000	
10 binary random variables	—		0.007	
20 continuous and binary	—		0.007	
Structural Breaks	3		3	

Notes: P denotes the posterior probability that an indicator is included. “non-linearity” is denoted by “x” when present in the semi-parametric specification.

Table 5. **Posterior statistics**

parameter	Currency Crisis	Banking Crisis	Asset Price Boom	Debt Crisis
λ	0.073 (0.015)	0.042 (0.011)	0.081 (0.015)	0.045 (0.008)
λ_{β}^{-1}	0.044 (0.009)	0.032 (0.014)	0.055 (0.022)	0.030 (0.011)
ω	0.313 (0.132)	0.022 (0.008)	0.019 (0.007)	0.025 (0.006)

Notes: Posterior means are reported. Figures in parentheses are posterior standard deviations.

Table 6a. Prior Sensitivity analysis:Utility

	Currency Crisis	Banking Crisis	Asset Price Boom	Debt Crisis
% abs. change in post. mean	0.17	0.21	0.08	0.067
% abs. change in post. median	0.13	0.15	0.11	0.069
% abs. change in post. s.d.	0.16	0.13	0.07	0.055

Notes: Reported are % absolute changes in key summaries of the posterior distribution for Utility. For NSR, the respective figures are reported in parentheses. The percentage changes are computed as averages of the percentage changes in the baseline specification relative to the posterior means for each one of the 1,000 alternative prior configurations.

Table 6b. Prior Sensitivity analysis:NSR

	Currency Crisis	Banking Crisis	Asset Price Boom	Debt Crisis
% abs. change in post. mean	0.14	0.14	0.05	0.082
% abs. change in post. median	0.10	0.10	0.09	0.077
% abs. change in post. s.d.	0.11	0.07	0.05	0.031

Notes: Reported are % absolute changes in key summaries of the posterior distribution for Utility. The percentage changes are computed as averages of the percentage changes in the baseline specification relative to the posterior means for each one of the 1,000 alternative prior configurations.

Table 7. **Spurious predictive performance.**

	Currency crises	Banking crises	Asset price boom	Debt crises
linear model				
range of RPS	0.0012 – 0.092	0.0033 – 0.081	0.00012 – 0.035	0.0051 – 0.017
median	0.0071	0.0045	0.012	0.010
non-linear model				
range of RPS	0.282 – 1.023	0.133 – 1.018	0.212 – 1.044	0.372 – 0.970
median	0.734	0.655	0.877	0.710

Notes: RPS is the relative predictive score, defined as the ratio of predictive distributions evaluated at each observation in the hold-out sample. MCMC is implemented using re-weighting of the MCMC draws in the full sample, using sampling-importance-resampling (SIR). The length of the re-sample is set to 3,000.

Figure 1. Posterior distribution of utility

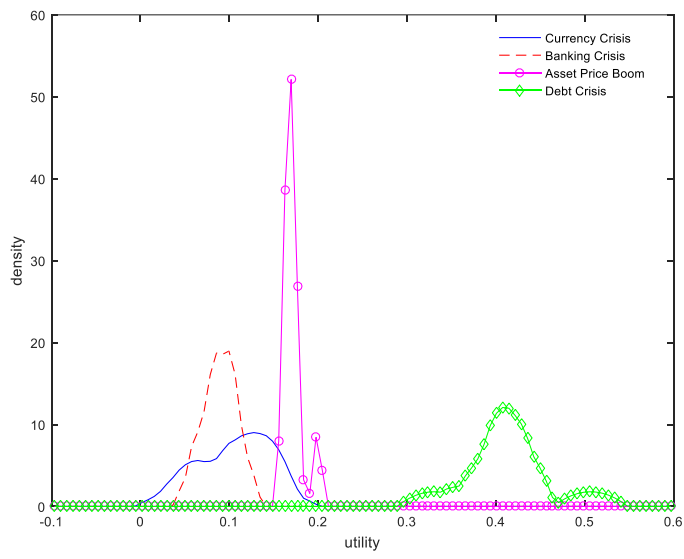


Figure 2. Posterior distribution of NSR

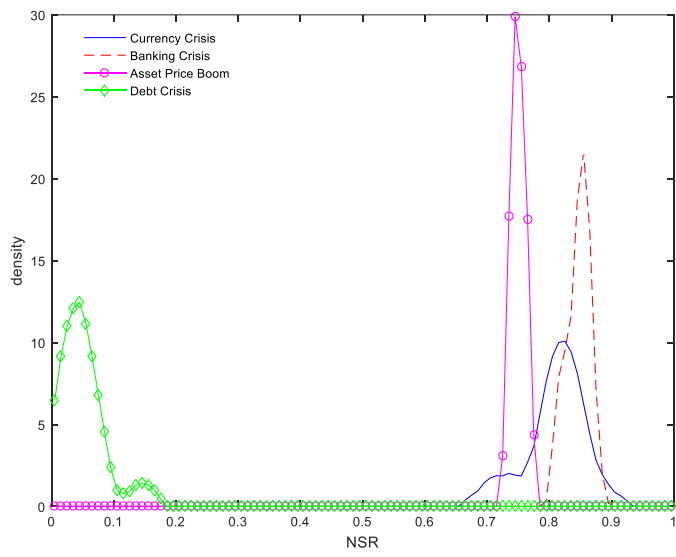


Figure 3. Posterior predictive distribution of utility

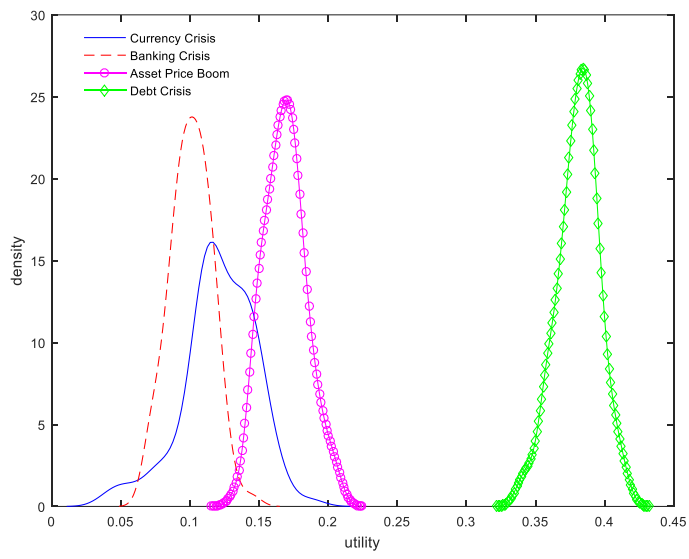


Figure 4. Posterior predictive distribution of NSR

