

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

Model-based approach for scenario design: stress test severity and banks' resiliency

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:

Barbieri, P.N., Lusignani, G., Prospero, L., Zicchino, L. (2022). Model-based approach for scenario design: stress test severity and banks' resiliency. *QUANTITATIVE FINANCE*, 22(10), 1-28 [10.1080/14697688.2022.2090420].

Availability:

This version is available at: <https://hdl.handle.net/11585/890209> since: 2022-07-08

Published:

DOI: <http://doi.org/10.1080/14697688.2022.2090420>

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).
When citing, please refer to the published version.

(Article begins on next page)

Model-based approach for scenario design: stress test severity and banks' resiliency¹

Paolo Nicola Barbieri,²Giuseppe Lusignani,³Lorenzo Proserpi,⁴Lea Zicchino⁵

June 10, 2022

Abstract

After the financial crisis, evaluating the financial health of banks under stressed scenarios has become common practice among supervisors. According to supervisory guidelines, the adverse scenarios prepared for stress tests need to be severe but plausible. The first contribution of this paper is to propose a model-based approach to assess the severity of the scenarios. To do so, we use a Large Bayesian VAR model estimated on the Italian economy where potential spillovers between the macroeconomy and the aggregate banking sector are explicitly considered. We show that the 2018 exercise has been the most severe so far, in particular due to the path of GDP, the stock market index and the 3-month Euribor rate. Our second contribution is an evaluation of whether the resilience of the Italian banking sector to adverse scenarios has increased over time (for example, thanks to improved risk management practices induced by greater awareness of risks that come with performing stress test exercises). To this scope, we construct counterfactual exercises by recalibrating the scenarios of the 2016 and 2018 exercises so that they have the same severity as the 2014 exercise. We find that in 2018 the economy would have experienced a smaller decline in loans compared to the previous exercises. This implies that banks could afford to deleverage less, i.e. maintain a higher exposure to risk in their balance sheets. We interpret this as evidence of increased resilience.

Keywords: Bayesian VAR, forecasting, bank stress tests, scenario analysis & design, banking system

JEL: E30, C11, C33, C53, G21

¹We are thankful to Emanuele de Meo and Federica Orsini for their past work on this project. Inês Drumond, Andrea Nobili, Ivan Petrella, Maurizio Pierigé, Rita Romeo, and Giacomo Tizzanini for insightful comments. We also thank participants to seminars at Banca d'Italia, the EBA Research Workshop on *The future of stress tests in the banking sector*

²University of Gothenburg and Prometeia SpA

³University of Bologna and Prometeia SpA

⁴Prometeia SpA, corresponding author: lorenzo.proserpi@prometeia.com

⁵Prometeia SpA

1 Introduction

Given the experience gathered during the financial crisis, the European Banking Authority (EBA) decided to update its guidelines on stress testing to assist institutions in identifying risks as well as their potential mitigants. In particular, supervisors observed that in many instances stress testing did not appear to be sufficiently integrated into institutions' risk management frameworks or senior management decision-making. In other instances, supervisors observed that risk concentrations and second-round effects were not considered in a meaningful fashion BCBS (2009a). One of the consequences of the lessons learned from the financial crisis was that the scenarios provided by EBA for the stress testing exercises became more severe. Or at least they were meant to be more severe. One fundamental issue is, in fact, that it may be difficult to evaluate how severe a scenario is, given the large number of variables involved. Moreover, since the start of the stress testing rounds banks have taken risk management actions to reduce their sensitivity to adverse shocks. Were these actions successful in making banks more resilient?

We propose an operational approach for scenario design and banks' resiliency evaluation consistent with the guidelines of the regulators for the construction of stress scenarios. In particular, our contribution is twofold. First, we propose an approach to measure the severity of a stress test scenario defined by EBA and apply it to the stress tests conducted in 2014, 2016, 2018. We focus on a restricted set of variables of scenarios that are specific to the Italian economy. Second, we investigate whether the resiliency of the Italian banking sector to adverse shocks has increased over time (from the 2014 stress test to the 2018 one).

In our work, we give an interpretation of the guidelines of the Regulator, and therefore our definitions need to be clarified.

In the 2017 stress test guidelines, the European Banking Authority (EBA) defines the severity of a stress test scenario as

"[...] the degree of severity of the assumptions or the deterioration of the scenario (from baseline to an adverse scenario) expressed in terms of the underlying macroeconomic and financial variables (or any other assumptions) (EBA, 2018)

A stress test scenario also needs to be plausible in order to allow the evaluation of a bank's financial position under a severe scenario that has some probability to occur BCBS (2009a). Therefore, not only a scenario has to be severe in terms of deviation of its main variables, but it has also to be *regarded as likely to materialize in respect to the consistency of the relationship of each macro-financial input variables with their current trend* (EBA, 2018). In designing our approach to evaluate stress test scenarios we stayed consistent with the Regulators' definition and tried to develop an operational approach by measuring the severity of an alternative scenario based on its likelihood of realization. Such probability of realization depends on the baseline scenario associated with each stress test and on the uncertainty surrounding such scenario. In our approach, we define a scenario on a variable as more or less severe according to the probability of realization of the path of the macro-financial input variables. This means that in our definition the greater the severity

of the scenario the lower will be its overall probability, consistently with the guidelines of the regulators¹. The EBA guidelines are clear in defining the severity of a variable but give no indication on how to evaluate the overall severity of the scenario. We therefore provide an approach to assess the severity of the stress test scenario as a whole.

Resiliency on the other hand is generally defined as the

"[...] ability to absorb shocks arising from financial and economic stress, whatever the source, thus reducing the risk of spillover from the financial sector to the real economy" (BCBS, 2009b)

Resiliency is therefore the ability to absorb and adapt to shocks, rather than contribute to them. As such, the ultimate goal of stress testing should be to evaluate if a banking system becomes more or less resilient in terms of: (1) the recovery after a negative shock and (2) its ability to dampen such negative shock instead of amplifying it. However, differently from severity, the Regulators' guidelines on resiliency are much less precise and usually refer to the capital shortfall after the shock. In other words, a more resilient banking sector is usually one with adequate capital position or lower capital depletion after a shock. As with severity, we provide an operational approach to evaluate resiliency, by measuring the impact on banks' balance sheets of different stress tests that are equally severe. In particular, in order to understand how the capability of the Italian banking sector to absorb shocks evolved over time we sterilize the effect of the different severity of the stress test exercises conducted since 2014 by building a counterfactual stress test with the same degree of severity. We are therefore able to compare the responses of the banking variables across time. Moreover, we claim that the impact on capital ratios is not a sufficient statistic to evaluate banks' resiliency. Instead, banks should not contribute to increase the size of the shock. As such, it is important also to evaluate other factors, like the ability to continue to lend in adverse scenarios. Indeed, a bank that strongly reduces credit to the private sector to meet capital requirements might generate "second-round" effects due to the impact of its response on the real economy, exacerbating the size of the shock. Therefore, we claim that a more resilient banking sector is one which not only has low capital shortfall in an adverse scenario (i.e. absorption of shock) but is also able to sustain the domestic credit market (i.e. reduce spillover effects).

We investigate the issues of severity and resilience with Italian Banking System Scenario Evaluation (IBASE) model. The IBASE is a Large Bayesian VAR model (LBVAR) that includes macro-financial variables and aggregate bank balance sheet variables. In the model bank and macro-financial variables are closely interconnected. In particular, bank variables affect macro-financial variables, a feature that is absent in traditional stress testing models. Despite the limited scale of our model, any potential feedback among variables is allowed and we show that it is a valid tool for scenario design in macroprudential stress testing.

The model allows us to compare all the recent EBA stress test scenarios in terms of severity, i.e. in terms of likelihood of realization. To do so we develop a novel contribution which is an aggregate measure of severity, defined as the weighted average of the severity of each variable², where the weights are given by the

¹In the remainder of the paper, with the term "stress test severity" we refer to the severity of the scenario

²We measure severity by computing the percentile of the marginal simulated distribution in which the path of each variable in the adverse scenario is located.

contribution of the shock of each variable to the variance of the bank aggregates. To anticipate the results, we find that the 2018 exercise was the most severe of the three considered, in particular due to the profiles of GDP, stock market index and 3-month Euribor rate in the adverse scenario. Moreover, we find that the profile of GDP in the last exercise is comparable in terms of severity with previous crisis episodes (Subprime and Sovereign debt crisis) while the drop in stock prices is more severe than in past exercises.

The other issue we want to investigate is whether the resiliency of the Italian banking sector has increased over time. To measure resiliency, we build counterfactual exercises by recalibrating the scenarios of the 2016 and 2018 exercises so that they have the same severity of the 2014 exercise. Under these new artificial scenarios, we can evaluate the response of the banking system at different points in time to now comparable stress scenarios. We find that in 2018 the economy would have experienced a lower decline in loans compared to the previous exercises. This comes at the expense of capital ratios. The reduction in Tier 1 ratio in 2018 would have been lower than in the official scenario but still larger than in the previous exercises. However, derisking from banks (i.e. reduction in lending) would strongly mitigate the impact on banks' capital and on Tier 1 ratio suggesting that the banking sector would be sufficiently capitalized. We take this result as evidence that the resiliency of the banking sector to adverse stress scenarios has increased: banks can accommodate a lower decline in loans, i.e. maintain a higher exposure to risk in their balance sheets while keeping capital adequacy levels under control.

The paper is organized as follows. In section 2 we briefly review the existing literature. In section 3 we describe the data used for estimating the model. In section 4 we present the methodology, in particular the model estimation strategy, the analysis conducted to validate the model and the approach used to measure severity and to construct counterfactual stress tests. The results of the different exercises are presented in section 5 and compared to previous literature in section 6. Section 7 concludes.

2 Related Literature

To analyse the issues at hand we draw from several strands of the literature.

First, the Prometeia IBASE model is closely related to the recent literature on medium- and large-scale Bayesian VAR models designed to address and overcome the *curse of dimensionality* (Bańbura et al., 2015, 2010, De Mol et al., 2008, Giannone et al., 2012). More specifically, this paper relates to other attempts to use this class of model to study the interaction between monetary policy, the real economy and the banking sector (Giannone et al., 2012, Altavilla et al., 2015) and to other contributions that analyse the same issues in a data-rich-environment (Boivin et al., 2018, Von Borstel et al., 2016, Dave et al., 2013). Large BVAR has been applied to the Italian credit market in Conti et al. (2018) to study the impact of bank capital shocks on credit supply and economic activity.

Secondly, there is a growing literature providing alternative approaches for stress testing as, according to many, traditional methodology might not represent the most efficient early warning device, but is more effective as a crisis management and resolution tool (Borio et al., 2014, Arnold et al., 2012). According

to Dees and Henry (2017), traditional stress tests have several limitations. First of all, the static balance sheet hypothesis makes the exercises too conservative. In practice banks react to adverse conditions by deleveraging, increasing capital or working out non-performing loans. Moreover, macro-financial shocks are treated as exogenous to the banking sector, ignoring any potential feedback. Finally, traditional approaches ignore any interactions between different sectors in the economy. There have been several attempts to develop novel methodology to give supporting information to the supervisor authority. Hoggarth et al. (2005) was one of the first paper to suggest a VAR approach to run a stress test on banks, accounting not only for the transmission of macroeconomics shocks to bank variables but also for potential feedback effects from banks' balance sheets to the macroeconomy. To address the interaction between German banks and the real sector Dovern et al. (2010) also employ a VAR analysis and find that the level of stress in the banking sector is strongly affected by monetary policy shocks, giving credit to the active behavior of central banks observed during periods of financial market crises. Hirtle et al. (2016) develop a framework for assessing the impact of macroeconomic conditions on the U.S. banking sector, using the Capital and Loss Assessment under Stress Scenario (CLASS) model. The CLASS model is a top-down capital stress testing framework that uses public data, simple econometric models, and auxiliary assumptions to project the effect of macroeconomic scenarios on U.S. banking firms. The authors find that the U.S. banking industry's vulnerability to under-capitalization (i.e. the capital gap under the crisis scenario) has declined. For the Euro Area, a similar top-down approach, the STAMP€ framework, has been proposed by Henry et al. (2013) and later by Dees and Henry (2017). Finally, Budnik et al. (2020) propose the BEAST model that combines the dynamics of a high number of euro area banks with that of the Euro Area economies. In their model bank responses feed back to the macroeconomic environment affecting credit supply conditions.

Similar to these papers, our analysis embeds a dynamic balance sheet approach and considers potential spillover effects between the banking sector, financial markets and the real economy. Although our approach focuses on a more limited set of variables than the aforementioned papers, we develop a methodological tool to provide additional information and guidance for more detailed supervisory models of components of bank revenues and expenses. Among other attempts to develop novel methodologies for stress testing, Acharya et al. (2014) compare the capital shortfall measured by the regulatory stress test to the one measured by a benchmark methodology (i.e. the V-Lab stress test) employing only publicly available market data. When capital shortfalls are measured relative to risk-weighted assets (RWAs), the ranking of financial institutions is not well correlated to the ranking obtained with their alternative methodology, while the correlation increase when required capitalization is a function of total assets. They also find that the banks that appeared to be well capitalized relative to RWAs do not fare better than the rest during the sovereign debt crisis in 2011-12.

Finally, this paper is connected with the strand of literature on scenario design and on the evaluation of the severity of stress test scenarios. According to Henry et al. (2013) there are three approaches to calibrate the exogenous shocks and can be used jointly to define the scenario.

1. **ad-hoc:** Ad-hoc calibration without recourse to any model or historical distribution of risk factors.
2. **model-free:** Shock size calibrated on historical distributions in a "model-free" non-parametric envi-

ronment³

3. **model-based:** Shock size calibrated on the residuals of a dynamic parametric model

Depending on which approach is used the definition of severity of a scenario might differ. Durdu et al. (2017), for example, measure the severity of the Federal Reserve Board’s Comprehensive Analysis and Review (CCAR) scenario from 2013 to 2017 by using a methodology which relies on a comparison of the scenarios provided in the CCAR with historical stressful episodes, in order to derive a severity score. Similarly, Yuen (2015) measures the severity associated with the 2009 stress test for U.S. bank holding companies (BHCs), known as the Supervisory Capital Assessment Program⁴ (SCAP) by comparing them with the 2008 recession and creating a score based on relative rankings. The methodology applied in these two papers can be considered as part of the model-free approach to measure severity. They don’t measure the severity as the probability of realization of the adverse scenarios, but instead as the deviation between adverse and baseline scenarios. In section 5.4 we show that this approach compared to the one proposed in this paper delivers different results. Breuer et al. (2009) proposes an alternative approach in order to rank variables according to severity across different stress tests; by also considering their joint comovement in the scenarios using the Mahalanobis distance⁵. Their paper uses a model-based approach but measures severity using a distance metric instead of looking at probabilities of realization. Mahalanobis distance is also used, along with Entropy pooling, by De Meo (2019) as approaches to measure plausibility for scenario design on economic growth, inflation and long-term interest rates in Italy using a Large Bayesian VAR. Breuer and Summer (2020) propose a procedure of systematic scenario selection where the plausibility is treated as a constraint and one searches for the worst expected loss satisfying such constraint. This systematic scenario selection is achieved by an appropriate form of worst case search over plausibility domains, finding among all equally plausible scenarios the scenario leading to the worst expected loss for any given portfolio. Antolin-Diaz et al. (2020) develops tools for constructing economically meaningful scenarios with structural VARs, and proposes a metric to assess and compare their plausibility. Lastly, Bonucchi and Catalano (2019) propose a model-based approach to calculate the joint probability of the whole scenarios starting from a general structural macroeconomic model and provide an application to assess the severity and plausibility of the 2016 and 2018 EBA stress tests for Italy. Our methodology also follows a model-based approach, but our severity measure is based on the percentile of the distribution of the variables in stress scenarios. Moreover, we propose a measure of aggregate severity by weighting different percentiles calculated on marginal distributions.

³In Henry et al. (2013) financial shocks to stock and bond prices are calibrated using the copula approach to account for the dependence across asset classes.

⁴Since conducting the SCAP in 2009, the Federal Reserve System has conducted annual stress tests on the U.S. banking system, the CCARs. In each CCAR, the Federal Reserve Board generates an adverse macroeconomic scenario and requires BHCs to submit at least one adverse scenario that is related to their own specific portfolios and risk profiles. However, at least for the 2013 CCAR analyzed by Yuen (2015), there is no precise information on what should be the appropriate severity.

⁵Mahalanobis distance is given by

$$Maha(r) = \sqrt{(r - \mu)^T \cdot Cov^{-1} \cdot (r - \mu)}$$

which is simply the distance of the test point r from the center of mass μ divided by the width of the ellipsoid in the direction of the test point. Intuitively it can be interpreted as the number of standard deviations of the multivariate move from μ to r .

3 Data

The dataset includes monthly information on the Italian banking sector from January 1999 to March 2019. The time interval covered allows us to include several recessions periods as well normal times in the estimation. In particular we are able to include three major crises: (1) the early 2000s dot-com bubble recession; (2) the Global Financial Crisis of 2008-09 and the Sovereign Debt Financial Crisis of 2011-2012.

In the benchmark model we consider 23 variables: (1) seven macroeconomics and financial variables: Italian Real GDP; consumer price inflation; one measure of short-term interest rate, the 3-months Euribor; two measure of long-term interest rate, the 5-years euro interest rate swap and the 10-years Italian government bond yield and a risk measure of the Italian economy, the 10-years BTP-Bund sovereign spread, the Italian stock market index; (2) seven variables related to the credit market: the cost and quantity of (net) loans to non-financial firms (NFC) and households (HH); the amount of (net) loans to other intermediaries⁶, the amount of bad loans and the markup on short-term loans⁷, which is supposed to be a synthetic measure on how the Italian banking sector is able to profit from the emission of new (short-term) loans; (3) six measures of bank funding and liabilities: the stock and interest rate on bonds and bank deposits, divided in current accounts and all other type of deposits and the amount of funds obtained by Italian banks from the ECB; (4) the stock of Italian public debt held by banks and the Tier 1 ratio⁸. We use the Denton-Cholette (Denton, 1971) method to get a monthly frequency version of Real GDP from R software package *tempdisagg*⁹.

Including the Tier1 ratio as an additional endogenous variable in the system is relevant to capture the effects of capital constraints on financial intermediation. An alternative approach is to calculate the evolution of this variable using the bank variables that enter into its definition. We choose not to use this approach for several reasons. The first is the lack of data for all the balance sheet items necessary for the computation of the ratio. Moreover, we followed the recent contribution from Conti et al. (2018) who run a counterfactual analysis on the effect of regulatory shocks on bank capitalization on Italian data, validating the use of the Tier 1 ratio as an endogenous variable in these types of analysis. Conti et al. (2018) not only employ an estimation strategy very similar to ours (i.e. LBVAR) but their results on Tier 1 capital are robust to several model specification and different estimation procedures, giving us confidence in using the same strategy. We also investigated the forecasting explanatory power of the IBASE for the Tier 1 ratio: it was among the best-performers, thus supporting our choice (see Table 7). Finally, according to variance decomposition analysis, more than half of the variance of the Tier 1 ratio at 3-year horizon is explained by shocks from other variables. This suggests that the variables included in the system help in explaining the dynamics of the Tier 1 ratio.

Except for macroeconomic variables and the Italian stock market index, the source for all the other variables is the Bank of Italy¹⁰. In the model all variables enter in log-levels with the exception of inflation,

⁶Including, among others, financial and insurance firms, central government.

⁷Measured as the differential between the average rate on loans under one year and the minimum rate on the stock of new loans under 1 year.

⁸Measured as the amount of Tier 1 equity of Italian banks divided by Risk Weighted Assets (RWA). Since there is no monthly public time series of Tier 1 ratio we estimated a monthly version using different sources. See the Appendix A for details.

⁹See <http://CRAN.R-project.org/package=tempdisagg>

¹⁰More specifically from the "Bank and Money: National Data", a monthly publication of statistics on monetary policy, banks'

interest rates, unemployment rate and the Tier 1 ratio which are expressed in percentages. Figures 1 and 2 provide a graphical representation of all the variables in the model, while Table 1 contains the variables names and the labels used throughout the paper.

[Figure 1 about here.]

[Figure 2 about here.]

[Table 1 about here.]

4 Methodology

4.1 The Large Bayesian VAR model

Let $Y_t = (y_{1,t}, y_{2,t} \dots y_{n,t})'$ be the (large) vector including the n (with $n = 23$) variables. We consider the following $VAR(p)$ model:

$$Y_t = c + A_1 Y_{t-1} + \dots + A_p Y_{t-p} + u_t \quad (1)$$

where Y_t is an $n \times 1$ vector of endogenous variables; A_1, \dots, A_p are $n \times n$ matrices of coefficients; $c = (c_1, \dots, c_n)'$ is an n dimension vector of constants and u_t is a normally distributed multivariate white noise with covariance matrix Ψ (i.e. $u_t \sim \mathcal{N}(0, \Psi)$).

Since we are dealing with a large dimensional vector of information, standard econometric techniques could incur in a problem of over-fitting due to the large number of parameters. The problem of parameter proliferation prevents us from conducting reliable inference with large dimensional system (i.e. *the curse of dimensionality*).

Therefore, we choose to estimate the IBASE in a Bayesian VAR (BVAR) framework, which is particularly useful when dealing with a large number of variables since it helps to overcome the curse of dimensionality via the imposition of prior beliefs on the parameters (Bańbura et al., 2010, 2015, Giannone et al., 2012). The idea behind such method is to combine the likelihood coming from the complex and highly parametrized VAR model with a prior distribution for the parameters which is naive but parsimonious.

In setting the prior for our baseline estimation, we start from the so-called Minnesota prior, (Litterman et al., 1979), which has been shown to be a reliable approach to reduce the estimation uncertainty without introducing substantial bias (De Mol et al., 2008, Bańbura et al., 2010). The basic principle behind the Minnesota prior (Litterman et al., 1979) is that each endogenous variable follows an independent random walk process with drift, such as

$$Y_t = c + Y_{t-1} + u_t \quad (2)$$

In each equation the prior mean of all coefficients is equal to zero except for the first own lag of the dependent variable, which is equal to one. Such prior specification incorporates the belief that more recent lags should

balance sheets and bank interest rates. Most of the statistics are harmonized at the Eurosystem level.

provide more reliable information than the more distant ones and that own lags should explain more of the variation of a given variable than the lags of some other variables in the equation.

However, since we are setting the IBASE to perform structural analysis, we need to consider the possible correlation among the residuals of different variables. This means that Litterman's assumption of a fixed and diagonal covariance matrix is somewhat problematic. To overcome this problem, we impose a Normal inverted Wishart prior, retaining the principles of the Minnesota ones (Kadiyala and Karlsson, 1997, Bańbura et al., 2010). We consider a conjugate prior distribution that belong to the Normal-Wishart family, where the prior for the vector of coefficients is Normal while the prior for the variance-covariance matrix is inverse-Wishart. This allows us to account for correlations among the residuals of different equations and depart from the Litterman's standard assumption of a fixed and diagonal covariance matrix.

The prior parameters are chosen so that

$$E(A_{ij,k}) = \begin{cases} \delta_i, & \text{if } j = 1, k = 1 \\ 0, & \text{otherwise} \end{cases} \quad V(A_{ij,k}) = \begin{cases} \frac{\lambda^2}{k^2}, & j = i, \\ \nu \frac{\lambda^2}{k^2} \frac{\sigma_i^2}{\sigma_j^2}, & \text{otherwise} \end{cases} \quad (3)$$

The hyperparameter δ_i characterises the integration order of the associated variable i . The hyperparameter λ controls the overall tightness of the prior distribution around the random walk and governs the relative importance of the prior beliefs with respect to the information contained in the data. For $\lambda = 0$ the posterior equals the prior and the data do not influence the estimates. If $\lambda = \infty$, on the other hand, posterior expectations coincide with the Ordinary Least Squares (OLS) estimates. To calibrate such hyperparameter we follow a method similar to that of Bańbura et al. (2010), applied to all the banking variables contained in the dataset, but differently from Bańbura et al. (2010) we validate the λ value from an out-of-sample analysis in the 60-months after the last observation available, and choose the value corresponding to the lowest relative Means Square Error (MSE).

The factor k^2 is the rate at which prior variance decreases with increasing lag length and σ_i^2/σ_j^2 accounts for the different scale and variability of the data. The coefficient $\nu \in (0, 1)$ governs the extent to which the lags of other variables are less important than the own lags. To ensure that the prior parameters of the Normal inverted Wishart are chosen so that the prior expectations and the variance of B coincide with the one of a Minnesota we impose $\nu = 1$.

We include an additional prior. To better understand it, it is useful to rewrite the VAR equation in error correction form

$$\Delta Y_t = c + \Pi Y_{t-1} + B_1 \Delta Y_{t-1} + \dots + B_{p-1} \Delta Y_{t-p+1} + \varepsilon_t \quad (4)$$

with $B_s = -A_{s+1} - \dots - A_p$ and $\Pi = A_1 + \dots + A_p - I_n$. We set a prior that shrinks Π to zero. More precisely, following Bańbura et al. (2015), we set a prior centered around one for the sum of the coefficients on own lags for each variable, and equal to zero for the sum of coefficients on other variables' lags. The tightness of this prior on the "sum of coefficients" is controlled by the hyperparameter κ . As κ goes to infinity the prior becomes flat while, as it goes to zero, we approach the case of exact differencing, which implies the presence

of a unit root in each equation.

We implement these priors by augmenting the system with dummy observations which is equivalent to imposing the Normal inverted Wishart prior (see Bańbura et al. (2010) for methodological details). In particular, the estimation of the IBASE in a multivariate regression setting using dummy observations is done in the following way:

$$Y_* = X_*\hat{B} + U_* \quad U \sim \mathcal{N}(0, \hat{\Psi}) \quad (5)$$

where Y_* is the matrix of original endogenous variable augmented with dummy observations and X_* the vector of exogenous variables augmented with dummy observations, \hat{B} are estimated coefficients in the model and $\hat{\Psi}$ is the estimated variance covariance matrix.

To motivate the use of our model we perform different analyses. In the next section we show that the response of the model to shocks from input variables is consistent with a-priori belief. Moreover we evaluate the out of sample performance of the IBASE model in forecasting the variables. This is an important analysis that shows that the model generates small forecast errors. This implies that the simulated paths of the model and consequently the severity measures proposed in this paper are reliable. We show that the model performs relatively well both in terms of RMSE and direction especially for bank stock variables and the Tier 1 ratio (see the appendix B).

4.2 Impulse-Response Analysis

To better understand the ability of the model to capture the effects of adverse shocks on banking variables, we implement an impulse response analysis. An impulse response function measures the time profile of the effect of a shock at a given point in time on the (expected) future values of variables in a dynamical system. In particular, we investigate the responses of the variables of the systems to adverse shocks on: 1) BTP-Bund spread 2) GDP 3) Italian stock Market Index, 4) 5-years Euro Area interest rate swap. These are the most relevant variables that determines the EBA scenarios. The analysis is carried on in two different ways:

- **Generalized Impulse Response**, as in Pesaran and Shin (1998). In contrast with standard impulse response analysis, which requires orthogonalization using the Cholesky decomposition, this approach is invariant to the ordering of the variables. More precisely, when the economy is hit by a shock on a specific variable, the immediate reaction on the other variable is extracted by using the historical distributions of the errors among variables.
- **Conditional Impulse Response**. In impulse response analysis, each variable in the system is endogenous and responds to exogenous shocks from all the other variables. As a result, the 3-month Euribor rate and 5-year Euro Swap, which are aggregate euro area variables, might respond to shocks arising in Italy. For example, if Italian GDP slows down, short term interest rates might decline because of an expansionary monetary policy measure from the ECB. To better gauge the response of the variables without any potential intervention from the monetary authority we provide an alternative approach to

estimate impulse responses by sterilizing the impact on the 3-month Euribor and 5-year Swap. Following the approach of Bańbura et al. (2015) and Conti et al. (2018), for each adverse shock we produce the conditional impulse responses in the following way:

- Estimate the full-sample model;
- Generate an H -period ahead unconditional forecast to generate a "baseline" scenario;
- Generate a conditional forecast in which the "adverse shock variable" is set equal to the baseline scenario plus one standard deviation of residuals for this variable at initial time and 3-month Euribor and 5-year Swap are equal to the baseline scenario for $h \in \{1, \dots, H\}$. Conditional forecasts are generated by applying the Kalman Smoother following the approach in Durbin and Koopman (2002) as in Bańbura et al. (2015).

To better assess the importance of each single shock calibrated in the adverse EBA scenario on the banking sector, we provide some evidence using Variance-Decomposition. In particular Forecast Error Variance Decomposition (FEVD) tells us how much of the forecast error variance of each of the variables can be explained by exogenous shocks to the other variables. More precisely, we define $\theta_{i,j}(h)$ as the proportion of the h -step ahead forecast error variance of variable i which is accounted for by the innovations in variable j in the VAR.

$$\theta_{i,j}(h) = \frac{\sigma_{ii}^{-1} \sum_{l=0}^h (\mathbf{e}'_i \Phi_l \Psi \mathbf{e}_j)^2}{\sum_{l=0}^h (\mathbf{e}'_i \Phi_l \Psi \Phi'_l \mathbf{e}_i)} \quad i, j = 1, \dots, n$$

where σ_{ii} is the $i - th$ element of the diagonal of Ψ , \mathbf{e}_i is a $n \times 1$ vector of zero elements except for a one in position i , Φ_l is the dynamic multiplier calculated from the vector of lag coefficients A_1, \dots, A_p . The output of FEVD is also used to create the aggregate synthetic severity index, as described in section 4.3.1.

4.3 Stress test analysis

4.3.1 Evaluation of the severity

In this section we describe in detail the approach that has been used to assess the severity of the three stress test scenarios. In a nutshell, we measure the severity of the adverse scenarios by calculating their probabilities conditioning the first moment of the distribution of the variables to the path in the baseline scenario. To better understand the output of this process, in figure 3 we reproduce the fancharts of the forecast of the model conditional to the EBA baseline scenario for each input variable and the corresponding adverse scenario. Calculating severity for a specific variable corresponds to identifying the interval where the adverse scenario is located.

This approach differs from other contributions in the literature, which measure severity in a model-free environment. We claim that relying on model-based implied probabilities is justified for two reasons:

1. A stress test is commonly described as the evaluation of a bank's financial position under a severe but plausible scenario to assist in decision making within the bank (BCBS, 2009a). In this definition the

severity is strictly related to the notion of likelihood.

2. While two scenarios might be equally severe when looking at deviations from the baseline scenario, they might be different in terms of probability of realization. This might happen because if the two scenarios are produced in different periods, the parameters of the model change in the two estimation window.

Our approach has also drawbacks. A severity measure based on the probability of realization is model dependent and, depending on the complexity of the model, cannot be easily replicated. Moreover, EBA stress test scenarios are defined as shocks on a wide set of global and country-specific macrofinancial variables. The IBASE is estimated on the variables of the Italian economy with additional European financial variables. Consequently, we can only feed the shocks on this restricted set of variables, ignoring the additional contribution of shocks originating from other countries or sectors. To be more specific, in our exercise we feed the EBA scenarios defined on the following set of variables: *GDP*, *EUR3M*, *BTP.10Y*, *SWAP.5Y*, *UN*, *INFL*, *STOCK.MARKET* (See Table 1 for definitions). Through the paper, we refer this set of variables as “input variables”, while banking variables are sometimes defined as “output variables”.

We now describe in details each step involved in the methodology adopted to measure the severity of the three stress test considered in the analysis (2014, 2016 and 2018). For $T \in \{2014, 2016, 2018\}$

1. **Estimate the model up to the date T of the exercise:** We estimate the IBASE from T_0 to $T - 1$, where T_0 is the starting year of the data (1999).
2. We produce $S = 10000$ 3-year ahead simulations on the variables of the systems conditional on the baseline scenario provided by the EBA. In particular for $s \in \{1, \dots, S\}$
 - (a) **Sample coefficients from posterior distribution:** Consider the estimation of the IBASE in a multivariate regression setting using dummy observations.

$$Y_* = X_* \hat{B}_{|T-1} + U_* \quad U \sim \mathcal{N}(0, \hat{\Psi}) \quad (6)$$

Using a Gibb’s sampler, we sample a vector of coefficient B^s and a variance-covariance matrix Ψ^s , from Normal/Inverted Wishart prior distribution

$$vec(B_{|T-1}^s) | \Psi \sim \mathcal{N}(vec(\hat{B}_{|T-1}), \Psi^s \otimes (X_*' X_*)^{-1}), \Psi^s \sim iW(\tilde{S}, T_d + T + 2 - k)$$

where T_d is the number of artificial time dummy observation, k is the number of endogenous variables in the system.

- (b) **Conditional Forecasts on the EBA baseline Scenario:** Using $B_{|T-1}^s$ and Ψ^s , we run forecast exercise by conditioning the central path to the baseline EBA scenario as in (Bańbura et al., 2015). Following their approach, the VAR system can be cast in state space representation as below
Measurement Equation

$$Z_t = CS_t + \nu_t \quad \nu_t \sim \mathcal{N}(0, R) \quad (7)$$

Transition Equation

$$S_{t+1} = GS_t + \omega_t \quad \omega_t \sim \mathcal{N}(0, H) \quad (8)$$

where $Z_t = Y_t$, $S_t = [Y_t, \dots, Y_{t-p+1}, c]$, $C = [I_k, 0_{k \times kp}]$, $R = 0_k$, $G_{[1:k]} = [A_1^s, \dots, A_p^s, I_n]$, $G_{[(k+1):(k+1)(p-1), 1:k(p-1)]} = I_k$, $G_{[(kp+1):k(p+1), (kp+1):k(p+1)]} = I_k$, $H_{[1:k, 1:k]} = \Psi^s$. In this way, the problem of finding the path of specific variables conditioned on the path of other variables can be understood as a Kalman filtering problem, i.e. estimating the states of the system when a set of observations are observed. We followed the standard approach of (Durbin and Koopman, 2002), that applies the Kalman filter to a modified state space representation. We can therefore derive point conditional forecast based on sampled coefficients and variance-covariance matrix $B_{|T-1}^s$ and Ψ^s .

- (c) **Draw disturbances around the baseline scenario:** To include additional sources of uncertainty deriving from future shocks, we also draw one set of disturbances for 3 years from the measurement equation from multivariate normal distribution¹¹. Finally, we generated a set of alternative paths, $Y_{T-1+h,s}$, of the variables in the system

3. **Measure of Severity:** We compute the severity of the stress test exercise on each single input variable as the probability at $T-1$ that the target variable is equal or below (or above depending on the variable considered) the level in the stress scenarios at horizon h . More precisely.

$$SEV_{T-1+h}^k = \text{Prob} \left(Y_{T-1+h}^k < \tilde{Y}_{T-1+h}^k \right) \quad k \in \underline{\mathcal{I}} = \{GDP, STOCK.MARKET, INFL\} \quad (9)$$

$$SEV_{T-1+h}^k = \text{Prob} \left(Y_{T-1+h}^k > \tilde{Y}_{T-1+h}^k \right) \quad k \in \bar{\mathcal{I}} = \{EUR3M, BTP.10Y, SWAP.5Y, UN\}$$

4. **Conditional Forecasts on the EBA adverse Scenario:** Using the estimated coefficients of the model, we run forecast exercises by conditioning the central path to the adverse EBA scenario as in (Bańbura et al., 2015), in order to track the response of the output variables.

As described above, uncertainty around the baseline EBA scenario is generated from two different sources: (1) the uncertainty connected with the parameters' realization (2) future realization of the shocks (Monte Carlo). As a robustness check, in the appendix C we show the result of the analysis when we only consider uncertainty arising from future realization of the disturbances.

[Figure 3 about here.]

The definition of severity proposed above corresponds to the probability of realization of a variable calculated from its marginal distribution. If we had to measure the severity of the whole scenario, the best

¹¹While this looks as a separate step from the previous one, it is actually implemented in the Kalman filter procedure.

approach would be to calculate the joint realization probability as

$$SEV_{T-1+h} = \left(\bigcap_{k \in \mathcal{I}} \text{Prob} \left(Y_{T-1+h}^k < \tilde{Y}_{T-1+h}^k \right) \right) \cap \left(\bigcap_{k \in \bar{\mathcal{I}}} \text{Prob} \left(Y_{T-1+h}^k > \tilde{Y}_{T-1+h}^k \right) \right) \quad (10)$$

However, the computation of this joint probability poses serious practical issues. First, the larger the set of variables included in the stress test input scenarios, the lower the probability of realization, if the variables are not perfectly correlated. Second, the model generates scenarios in which variables are somewhat correlated. As an example, we produced a scatterplot of generated scenarios for Real GDP and 3-month Euribor at 3-year horizon from the 2018 exercise (see figure 4). The model generates scenarios in which these two variables are positively correlated: this happens essentially because in the estimation sample the ECB responded to a slowdown in GDP and inflation by cutting rates. In the same scatter we draw a line where we included the adverse scenarios defined by the EBA for these two variables. It can be noticed that there are no scenarios generated by the model where Real GDP is lower and Euribor is larger than the levels provided by EBA for these variables. This implies that using this approach to measure aggregate severity would deliver inconclusive results for the comparison of severity across different exercises.

[Figure 4 about here.]

To overcome this limitation, we propose an alternative approach. We measure a synthetic aggregate severity index as a weighted average of single-variable marginal probabilities of realization defined in 9. In particular, we weigh the probability of each variable by considering its contribution to the aggregate variance of banking variables. In particular, we define the weight of an input variable k at horizon h , $\omega_k(h)$ as

$$\omega_k(h) = \frac{\tilde{\omega}_k(h)}{\sum_{m \in \mathcal{I}} \tilde{\omega}_m(h)} \quad (11)$$

$$\tilde{\omega}_k(h) = \frac{\sum_{i \in \mathcal{B}} \theta_{i,k}(h)}{\sum_{m \in \mathcal{B}, n=1, \dots, n} \theta_{m,n}(h)} \quad (12)$$

Finally, the synthetic aggregate severity is computed as

$$SSEV_{T-1+h} = \sum_{k \in \mathcal{I}} \omega_k(h) SEV_{T-1+h} \quad (13)$$

4.3.2 Resiliency evaluation: balance sheet effects from a counterfactual stress test

The strategy implemented in this paper to check whether the resiliency of the Italian banking sector has increased over time relies on two steps. The first one is to estimate the model up to the date prior to the start of the exercise, thus using a recursive sample. In this way, since the exercises are run every two years, the model parameters incorporate any potential effort that the banking system has implemented to make

their balance sheets more robust to adverse shocks. Previous studies confirm that banks reacted to stress test exercises by strengthening their balance sheet. Using U.S. bank-firm matched data from regulatory filings Berrospide et al. (2019) find that there is a pass-through effect between stress test results and the credit market, while investment and employment are largely unaffected by the capital buffers required as a result of stress tests. The author find that larger stress-test capital buffers are passed to the credit market, leading to substantial reductions in bank Commercial and industrial loan lending. Lastly, Kok et al. (2019) find that banks that participated in the 2016 stress test reduced their average risk weight density (RWD) by about 4.2 percentage points relative to banks that were not tested, indicating that banks' participation in the stress test led to a substantial reduction in their risk taking. These results confirm the existence of a credit supply channel.

Second, we constructed a counterfactual stress test scenario with the same severity and deviations from baseline (i.e. we use the 2014 stress test as a benchmark). This allow us to track the response of balance sheet variables at different dates to scenarios that are comparable in terms of severity. In this way we can evaluate whether the banking sector became more or less resilient (according to the definition we presented in Section 1)¹².

With simulated scenarios and probability of realization at hand, we can easily generate counterfactual stress tests. The intuition is to reproduce an artificial path of the input variables in 2016 and 2018 that are consistent with the probability of realization of the 2014 scenarios. To do so we proceed as follows for scenario $T \in \{2016, 2018\}$

1. For each input variable k and each horizon h , we calculate an alternative path \hat{Y}_{T-1+h}^k as the percentile of the simulated distribution corresponding to the probability level in 2014. In particular,

$$\hat{Y}_{T-1+h}^k = \left\{ y \in \left\{ Y_{T-1+h,s}^k \right\}_{s=1,\dots,S} \mid \text{Prob} \left(Y_{T-1+h,s}^k \leq y \right) = \text{Prob} \left(Y_{2014-1+h}^k \leq \tilde{Y}_{2014-1+h}^k \right) \right\} \quad k \in \underline{I} \quad (14)$$

$$\hat{Y}_{T-1+h}^k = \left\{ y \in \left\{ Y_{T-1+h,s}^k \right\}_{s=1,\dots,S} \mid \text{Prob} \left(Y_{T-1+h,s}^k \geq y \right) = \text{Prob} \left(Y_{2014-1+h}^k \geq \tilde{Y}_{2014-1+h}^k \right) \right\} \quad k \in \bar{I} \quad (15)$$

2. **Conditional Forecasts on the Counterfactual Adverse Scenario:** Using the estimated coefficients of the model, we run forecast exercises by conditioning the central path to the counterfactual adverse scenario calculated in the previous step following the methodology in (Bańbura et al., 2015), in order to track the response of the output variables.

¹²Since the model is estimated on different samples, not only the sensitivity of the variables to the lags but also the variance-covariance matrix of the disturbances might change. This implies that, for example, a 5% probability shock on GDP can correspond to different deviations from the baseline across different exercises. Since in the counterfactual exercise we want to interpret the different responses of banking variables across exercises as the effect of the derisking/hedging/capital adequacy decisions of the sector, we decided to use re-estimated lag parameters A_1, \dots, A_p as the posterior mean to simulate parameters, while the posterior mean of the variance-covariance matrix of disturbances is the one estimated in 2018

5 Results

5.1 Model estimation

Before presenting the results on the IBASE performance, on the severity of the stress test and on the counterfactual analysis, we are going to discuss all the main parameters used to estimate the LBVAR. First, the out-of-sample calibration of the λ parameter gives us a value equal to 0.05, which is a consistent value according to the LBVAR related literature (Conti et al., 2018, Sims and Zha, 1998). Given such value for λ we impose $\kappa = \kappa \times \lambda$ with $\kappa = 10$ (such value choice will be discussed in the appendix B, based on several forecasting performances exercises). In order to set the prior on δ_i we estimate an autoregressive integrated moving average (ARIMA) univariate model for each endogenous variable and for different orders of integration, choosing the one that minimizes the conditional sum-of-squares (CSS). The order of integration for each endogenous variable is presented¹³ in Table 1.

Regarding the lag order of the VAR, we choose a parsimonious value equal to one, which was the value minimizing the Bayes information criterion (BIC) (Schwarz et al., 1978) according to the formula:

$$BIC = n \cdot \ln(\widehat{\sigma}_e^2) + k \cdot \ln(n) \quad (16)$$

where $\widehat{\sigma}_e^2$ is the error variance, n is the number of observations and k the number of endogenous variables. As an additional robustness check we run the results from the estimation of the IBASE with a different lag value (i.e. four) which was selected applying a variety of lag selection processes (see appendix D).

5.2 Impulse-Response and Variance Decomposition

We now comment the response of the variables to each shock. Following a one-off 15 bp increase in the BTP-Bund spread (see figure 5), the Italian economy slows down and Real GDP is 0.2% lower compared to the baseline scenario after 3 years from the shock. Net loans to households and firms also decrease by almost 0.5% and 1% respectively, while bad loans increase. For bank interest rates, results differ when considering conditional vs generalized impulse responses. In the case of GIRF, interest rates on loans and bank bonds decline. This happens essentially because short term interest rates decline as the model captures a response from the monetary policy to a slowdown in economic activity. When these effects are shut down, interest rates on loans go up. This happens because the cost of funds for banks is now higher and the probability of default, especially for firms, increases. Therefore, banks charge higher interest rates to compensate for the higher cost of funds and the additional risk.

[Figure 5 about here.]

In figure 6 we present the results for a positive shock of 0.1% to Real GDP. In this case, net loans to firms increase, while the impact on household lending is muted. Moreover the stock of bad loans is around 0.5%

¹³When the procedure selects orders of integration greater or equal to one, we set $\delta_i = 1$ and zero otherwise.

lower at the through. Instead interest rates on loans are higher even when the effects on euro area interest rates are sterilized. This result is consistent with a credit demand shock, where higher demand for loans from firms and households results in higher interest rates charged by banks.

[Figure 6 about here.]

When stock prices are hit by a positive shock (5%), the spread between Italian and German bond yield decreases (see figure 7). The shock also stimulates lending, especially to firms, and fosters a reduction in the stock of bad loans. The effect on bank interest rates are different, depending on the methodology. When standard GIRF are used, interest rates increase after the shock because of the response of the monetary authority to improved macroeconomic conditions resulting in higher short-term interest rates. In the conditional case, bank interest rates decline because of improved creditworthiness of banks and firms that is reflected in lower risk premia.

[Figure 7 about here.]

Similar to a stock market shock, an increase in the 5-year swap rate induces a positive reaction of loans and reduces risk premia on Italian sovereign bonds and bank interest rates (see figure 8). This happens because a positive shock on swap rates carries information of improved macroeconomic conditions. In the calibrated adverse scenario, higher swap rates compared with the baseline scenario are supposed to be harmful for the banking sector, a result that seems not to be supported by the impulse response analysis. However, a positive shock on interest rates when the economy is improving (as implicit in the impulse response analysis) or when the economy slows down has a very different effect on the banking sector as it is evident when we feed the EBA scenario to the model.

[Figure 8 about here.]

In table 2 we report the FEVD for a selection of bank variables in response to shocks from the variables in the EBA stress test scenarios at different time horizons (1 to 3 years ahead), with the model estimated on the full sample. Interestingly shocks on market interest rates play a dominant role in determining the variability of response variables. In particular market rates explain more than 50% of the variability of loan rates and bank bond rates at 3-year horizon, but also explain a large portion of variability (above 10%) in loans. More generally stock variables and the Tier 1 Ratio seem to be less affected by shock variables, essentially because they are more strongly influenced by their own variability. Finally, stock market shocks are also relevant, especially for loans.

[Table 2 about here.]

5.3 Severity evaluation

Table 3 presents the deviation from the baseline (first lines) and the probabilities of realization for the three stress tests considered, over the 3-years horizon (second line). It is interesting to note that for GDP the deviation from baseline is more severe in the 2018 stress test, which has the lowest probability of realization. Similarly to what happens with GDP, for many of the other macro-financial input variables the 2018 exercise produces the highest deviations from the baseline scenario; however the associated probabilities of realization are not as low as for GDP. Moreover, we find that the 2018 exercise had been the most severe also for the 3-month Euribor and stock market, however the scenario for these variables seems more likely than the one for GDP. Instead, the 2014 exercise was the most severe for the BTP-Bund spread (at least on impact).

Looking at output variables it can be noticed that for loans the deviations from the baseline calculated by the model, are very similar for the three stress tests while the probability of realization is lower (and very similar in size) for the 2016 and 2018 stress tests. For bad loans the latest stress test entails the highest deviation from baseline associated with very high probability. This could be explained by the recent disposal of bad loans (from 2016 onward) by the Italian banking sector, which decreased the amount of bad loans; hence the 2018 deviations are the highest, although associated with a decreasing and smaller stock. As we might expect from an adverse scenario, holdings of sovereign bonds increase (with a very high probability), especially for the last stress test. This is in line with the empirical findings by Acharya and Steffen (2015), which documents the large increase in sovereign exposures among undercapitalized banks during the European Debt Crisis and offers different explanations for this phenomenon. Bank interest rates (on loans and deposits) increase in all three scenarios, although more strongly in the 2018 ones, as a direct consequence of the increase (by scenario design) in the 3-month Euribor and the 5-year swap rate. Lastly, as expected, the Tier 1 ratio decreases in all three adverse scenarios, although more strongly in the 2018 one. The larger fall in the capital ratio in the 2018 exercise is associated with the strongest increases in bad loans and interest rates. Since the interest rate on loans increases but the stock of loans decreases and the bad loans increase, the net effect of such scenario is a reduction in bank capitalization and hence in the Tier 1 ratio. The reduction in the Tier 1 ratio in the adverse scenario is however very limited compared to the average results across different banks in traditional stress tests. The main reason for this difference is in the hypothesis of a dynamic instead of a static balance sheet. According to our exercise, in response to negative shocks banks reduce their risky exposures instead of raising new capital, a result that has also been found in other macroprudential stress test approaches (Dees and Henry, 2017).

[Table 3 about here.]

To conclude on severity evaluation analysis, in Figure 9 we show the graph of the synthetic aggregate severity (SSEV) presented in Section 4.3.1. As expected, the 2018 stress test has the lowest SSEV, highlighting the fact that the probability of realization of the 2018 stress test is the lowest. Figure 9 also gives the rationale to run a counterfactual analysis standardizing the severity of all three stress tests. This is because, even when

weighted, none of the three aggregate "severities" are closely comparable; with the only exception for the second year deviations from the 2014 and 2016 stress test¹⁴.

[Figure 9 about here.]

5.4 Comparing the severity of the EBA adverse scenarios with previous crises

In this section we apply our approach to measure the severity of stress episodes occurred in the past. In particular we focus on two crises: the Financial Crisis of 2007-2008 and the European Debt crisis of 2011. We limit our severity comparison to only these two crisis instances given the similarity that they bear with a stress-test exercise which were designed to mimic them in order to "test" the financial soundness of the European banking sector, by assuming long-lasting negative financial and macroeconomics shocks. We decided to avoid any comparison with the COVID-19 pandemic stress given its structural difference that is reflected in a completely different dynamic response of the variables (i.e. a single quarter negative shock almost entirely re-asorbed after one-year), since such exercise might be not too informative. Given the relevance of the pandemic crisis future research should be devoted in order to comparing it with EBA's future stress test in order to understand if and who those excercises were updated (in magnitude and shape) after the Covid-19. The purpose of this exercise is to compare the severity of EBA stress test scenarios with previous events to better assess their plausibility and consistency. Comparing stress test scenarios with previous episodes has been proposed as a possible approach to measure severity (see Durdu et al. (2017), Yuen (2015)). However, our approach to assess the severity relies on the existence of both a baseline and an adverse scenario. Instead, when looking at the past we don't observe alternative paths, but only the realization of an adverse scenario. To be clear, we don't know how macro-financial variables would have evolved if the Financial Crisis "shock" did not occur. As a consequence, we don't observe a baseline scenario for the Italian economy under which we can generate random paths that are needed to measure the "ex-ante" probability of the realization of the adverse scenario. To solve this issue, we use the IBASE to generate a counterfactual baseline scenario by generating an unconditional forecast of the variables of the model starting from the moment the crisis began. The unconditional forecast does not incorporate the realization of the adverse shocks, as such it can be interpreted as a baseline scenario. This approach is consistent with other contributions in the literature. For example, Durdu et al. (2017), measure the severity by comparing the drop in GDP in the adverse scenario with the drop in GDP occurred during the Financial crisis. This approach corresponds to assuming that the baseline "counterfactual" growth scenario for GDP during the 2007-2009 crisis coincides with the stress test baseline scenario. This assumption can potentially lead to underestimating (overestimating) the severity of a stress test depending on whether "counterfactual" baseline growth during the Financial Crisis scenario is lower (higher) than the regulatory baseline scenario.

To compute the probability of realization, we need to select two starting dates for the two recessions considered. We took these dates from business cycle chronologies reported by the Economic Cycle Research

¹⁴We also computed (available upon request) the same severity index using simple un-weighted average of the single-variable marginal probabilities of realization and the results are qualitatively the same.

Institute for the Italian economy. According to this source, the recession generated by the Subprime Crisis started in August 2007, while the European debt crisis in April 2010. Consequently the model has been estimated with data ending in July 2007 and March 2010 and the baseline scenarios are calculated as the unconditional forecasts of the variables of the model from these dates.

[Table 4 about here.]

The results are reported in table 4. In particular we report the deviation from the adverse “realized” scenario and the baseline “counterfactual” scenario for each input variable at different horizons and the corresponding realization probability. For the sake of comparison, we also include the same variables for the 2018 stress test exercise. A first striking result is for GDP scenarios. While the deviation between adverse and baseline scenarios in 2018 are similar in magnitude to what we find in the past crises, the probability of realization differs remarkably. According to the model simulated distributions, the output drop experienced during the Financial crisis was very unlikely. This is because the probability is calculated from the model estimated on data before the beginning of the crisis, when the Italian economy enjoyed a prolonged period of stable, although low, growth (the Great Moderation). In contrast the 2011 crisis was more likely. This happens because the financial crisis is now included in the estimation sample resulting in a much higher variability in the residuals and posterior distribution of coefficients. This evidence suggests that the severity measure proposed in this paper might change over time not because of the postulated effects on macroeconomic variables, but because of the time-varying uncertainty surrounding the baseline scenarios.

In general GDP, unemployment rate, the BTP-Bund Spread and the Stock market index during previous crises are consistent with the EBA scenario in terms of direction. However, this is not true for the inflation rate, swap rates and the 10 years-BTP. In the past crises, ECB policy measures triggered a drop in the 3-month Euribor rate, while swap rates and the German 10 year Bund¹⁵ benefited from unconventional monetary policies and flight-to-quality phenomena. In contrast, in the 2014-2016-2018 EBA scenarios, monetary policy reaction was excluded by design and therefore interest rates increases. This result is connected with the argument proposed in section 4.3.1, where we find that the model generates scenarios where short-term rates and GDP are usually positively correlated, consistently with the reaction of the ECB during the more recent crises. This clearly raises a question about the plausibility of the EBA scenarios.

For stock market index, we find that the initial drop in the price level was very large in the EBA stress test compared to the previous crisis episodes. Finally, the postulated increase in the spread in the 2018 scenario is actually less severe compared than during the past crisis.

5.5 Resilience evaluation: balance sheet effects from a counterfactual stress test

The results shown in Table 3, confirm that for some but not all input and output variables the 2018 deviations were the largest and are associated with a lower probability of realization. When comparing the

¹⁵The path of the German government bond yield can be recovered by taking the difference between the 10-years BTP and the BTP-Bund spread.

scenarios from the three stress tests it seems that all of them were designed to include very unlikely and extreme GDP deviations, while the probability of realization for all the other input variables were higher. It is important to remark that we can only compare the three stress tests in terms of deviation from the baseline and probability of realization but, without further work, we are not able to draw any conclusion on whether the resilience of the Italian banking sector has increased over time.

To say something about this last issue, we rescale the 2016 and 2018 stress tests to the same level of severity of the 2014 exercise and look at what would have happened to the output variables. By having the same degree of severity across all three exercises and comparing the results with those in Table 3, we are able to verify whether the Italian banking sector has become more or less resilient to the same adverse shocks. Model coefficients are estimated over the period until the starting point of the stress test exercise. This implies that, for example, in 2018 we evaluate the impact of the counterfactual shock with a model that covers more recent years, with risk management actions taken by banks to improve their resilience. Table 5 presents the deviations from the baseline (first lines) and the probabilities (only for the 2014 stress test) of the counterfactual stress test exercises. We are not discussing the counterfactual changes in the input variables since they were substituted, in the 2016 and 2018, by the deviations of the 2014 adverse scenario and we already discussed the rankings of such scenarios before. We focus instead on the output variables. Looking at loans it is clear that, by imposing the same degree of severity, we obtain smaller deviations, especially for the 2018 stress test. Therefore, we conclude that there has been an increase in the resilience of the Italian banking sector to the adverse shock. In other words, with shocks that are comparable to those in 2014, Italian banks can afford to derisk their balance sheets less. Bad loans instead increase more in these new scenarios than in 2014. As for government bonds the counterfactual analysis suggests that they would have increased more; this can be easily reconciled with the fact that in 2014 the deviation for the 10-years Italian bonds yields, and the sovereign spread, were the highest. For the interest rate on loans the counterfactual analysis produces higher rates for both NFCs and households, as a consequence of the increase in long-term market rates. Lastly, we turn our attention to the Tier 1 capital ratio. The counterfactual analysis shows that resilience measured in terms of Tier 1 ratio slightly declined in 2018. Unfortunately in the IBASE we don't model all the balance sheet items necessary to interpret the impact on the Tier 1 ratio. One possible explanation for the lower resilience in 2018 for this variable is that banks reduce their loans by less as discussed above, which means that we have a higher denominator for the ratio. However the decline in the Tier 1 ratio is small and should not eat significantly into the large capital buffers that banks have accumulated over time¹⁶. We claim then the even if banks reduce less the risk in their balance sheet, they are still sufficiently capitalized.

To conclude, by relying on counterfactual adverse scenarios that are comparable in terms of severity to the first EBA exercise, we find that Italian banks respond by allowing for larger decline in the Tier 1 ratio and a smaller decline in net loans. We interpret this as an increase in resilience over time. Given the fact that

¹⁶When looking at the capital buffer over minimum capital requirements of 8 Italian banking groups we notice that the capital depletion under the adverse counterfactual stress is way smaller than the capital buffer. This means that even if the 2018 exercise is the one with the largest capital depletion, the effect on the capital ratio is not sufficient to bring the Tier 1 below the capital requirements (defined as the sum of minimum Tier 1 ratio, Pillar 2 requirement, Capital Conservation Buffer).

we are unable to isolate the single contribution of macroprudential/supervisory policies and any particular individual bank actions we argue that the increases resiliency is due to a combination of the two effects.

[Table 5 about here.]

6 Discussion

In this this section we try to compare the results presented in this paper with previous literature and discuss main differences and similarities.

Starting from severity evaluation we find that the severity of EBA's regulatory stress tests has increased over time, suggesting that european regulators are "raising the bar" on banks' ability to survive to periods of economic stress. Durdu et al. (2017) propose an approach to measure the severity of Fed CCAR's from 2013 to 2017 for different scenario variables and consistently to our results they found an in increasing trend in the severity of the scenarios for GDP, while this does not hold for other variables. Bonucchi and Catalano (2019), using a novel joint probability approach, find that, for the Italian economy both the 2016 and 2018 EBA scenarios are very unlikely. Similar to our results they too find that the adverse scenarios on Italian GDP are very severe, while the ones associated with the BTP-Bund Spread and unemployment rate are more plausible and that the overall joint probabilities of the stressed variables is very low. The implausibility result on the joint probability of EBA's stress test is also supported by Breuer and Summer (2020) which state that the current stress tests (especially EBA's 2016 one) are selected in a way that make them highly implausible leaving aside many dangerous and more plausible scenarios. Therefore the evidence from the literature do support one of the main findings of our paper that is EBA's stress test scenario seem to be very (jointly) unlikely especially among the most recent ones.

In terms of resiliency assessment Georgescu et al. (2017) find that stress tests have a potential and positive impact in fostering the resiliency of European banks by not only increasing their capital ratios via mandatory capital increases but also by enhancing transparency allowing investors to better discriminate between banks. Ahnert et al. (2020) further examine how this happens by showing that on one hand stress test results reveal new information to market participants which significantly affect bank' equity but they also place follow-up requirements on banks that fall short of supervisory expectations, and that both actions are intended to foster resiliency and capital adequacy. Similar to our results Janda and Kravtsov (2019) find that the increased regulatory supervision through stress testing enhance resiliency in European banks by supporting them in reducing their portfolios risk while not affecting loans growth. Stress tests results help banks in effectively derisking their capital position, by reducing more intensively the risk-weighted assets and non-performing exposure to improve their capital position. Similarly, Berrospide et al. (2019) and Kok et al. (2019) confirm that there is an impact of stress testing practices on banks' lending and derisking decisions Kok et al. (2019). There is indeed supporting evidence of a potential positive effect on banks' resiliency, especially via derisking but still maintaining the fundamental role of banks of financing the real economy.

7 Conclusion

In this paper we propose a model-based approach to analyze the stress tests carried out by the EBA on banks. In particular, we investigate whether 1) the severity of the adverse scenarios has increased from the first exercise to the last one, and 2) if the resilience of the Italian banking sector has improved. We argue that these two concepts need to be considered separately. We define severity as the probability of realization of adverse scenarios of input variables, while resilience depends on how strongly banking variables react to these shocks and how they are able to dampen the spillover effects to the real economy.

To address these issues, we used the IBASE model, an econometric model for the Italian economy, where banking and macro-financial variables are closely interrelated. As such, our approach departs from traditional stress test exercises on several ways. First, we consider feedback effects from the banking sector to the real economy. Second, our setup is consistent with a dynamic balance sheet approach: banks can deleverage/derisk in response to negative shocks as opposed to the static balance sheet approach assumed in stress test exercises. Severity is measured on each variable included in the exercise by calculating percentiles from conditional forecasting simulations of the model. We find that, in general, the adverse scenarios on Italian Real GDP are very severe, while the scenarios on financial variables (stock market and market interest rates) are more likely. Moreover, we find that the 2018 exercise was the most severe for Real GDP, 3-month Euribor rate and stock market, while the 2014 exercise was the most severe for the BTP-Bund spread (at least on impact). The last exercise was also more severe than the Subprime and European debt crises. We also find that the joint realization probability of the EBA adverse scenarios is extremely low, as such it is not possible to give an overall assessment of the severity of the exercises. This is also due to the positive correlation between short-term rates and GDP that we find in the data when looking at past crisis episodes. We therefore propose a novel approach. We calculate a synthetic aggregate severity index by averaging the marginal probability of realization of the shocks according to their contribution on the variability of bank variables. The index so constructed shows that the last exercise was the most severe.

To measure the resilience of Italian banks we run a counterfactual stress test exercise. Since the exercises are not comparable in terms of probability of realization, we use the model to generate alternative adverse scenarios for the 2016 and 2018 that have the same severity as the 2014 ones. Resilience is assessed by feeding these shocks to the IBASE model estimated up to the starting date of the exercise. In this way, when evaluating the exercises in 2016 and 2018 we are able to capture the effect of the risk management actions put in place to reduce the exposure to the shocks. This lower exposure should be incorporated in the model's parameters. In the 2016 and 2018 counterfactual exercises we find that Italian banks respond by reducing loans to the private sector by a smaller amount allowing for a stronger decline in their Tier 1 ratio. However, derisking from banks would strongly mitigate the impact on banks' capital and on Tier 1 ratio so if banks allow the Tier 1 ratio to decrease, it means that they have a strong enough capital position. We interpret this result as an increase in the resilience of the banking sector.

Our paper contributes to the policy discussion on how to properly design stress test scenarios. First, we show that the static vs dynamic balance sheet assumption makes a lot of difference in evaluating banks

resilience as also stated in Dees and Henry (2017). In our approach, we show that banks would react to negative shocks by reducing risky exposures in their balance sheet, as such the decline in capital ratios would be lower than in EBA stress tests (with static balance sheets). Second, we show that, according to the IBASE model, the EBA scenarios are extremely unlikely when considering the joint probability of realization of the whole scenario.

While our analysis is limited to the Italian banking sector, the IBASE modelling framework proposed in this paper is flexible enough not only to be applied to other countries but also to being modified in its main component by adding more endogenous variables. We design the model with a degree of flexibility in mind to deliver a methodology that policymakers and supervisory authorities could use in a variety of instances and for multiple purposes (e.g. address the severity and resiliency of banking sector to past (or future) shocks, to address stress-test severity design, to run counterfactual impact analysis). The LBVAR model used in this paper can easily account for external factors resulting from cross-border exposure of Italian banks by either including variables measuring foreign's bank exposure as endogenous variables. Even if the IBASE is flexible enough to handle such further sophistication these possibilities are beyond the scope of the current analysis and we leave them as open points for future analysis. While our approach relies on a limited set of variables at the aggregate level compared to the detailed analysis of the EBA exercise, we argue that it can be used in a more complex environment as the STAMP€ and CLASS framework do. In particular, we can potentially consider cross-country spillovers by modeling interlinkages in a Bayesian context as in Canova and Ciccarelli (2004) or in a GVAR/IVAR fashion (Dees et al., 2007, De Meo et al., 2018). Moreover, the aggregate responses of bank variables to the shocks can be projected on single banks while still considering the potential spillovers from the banking sector to the aggregate economy as in Gross et al. (2016). Finally, the potential cross-sector interlinkages could work as an additional propagation channel of adverse macroeconomic shocks as in Castrén and Rancan (2014). We leave these extensions to future research.

References

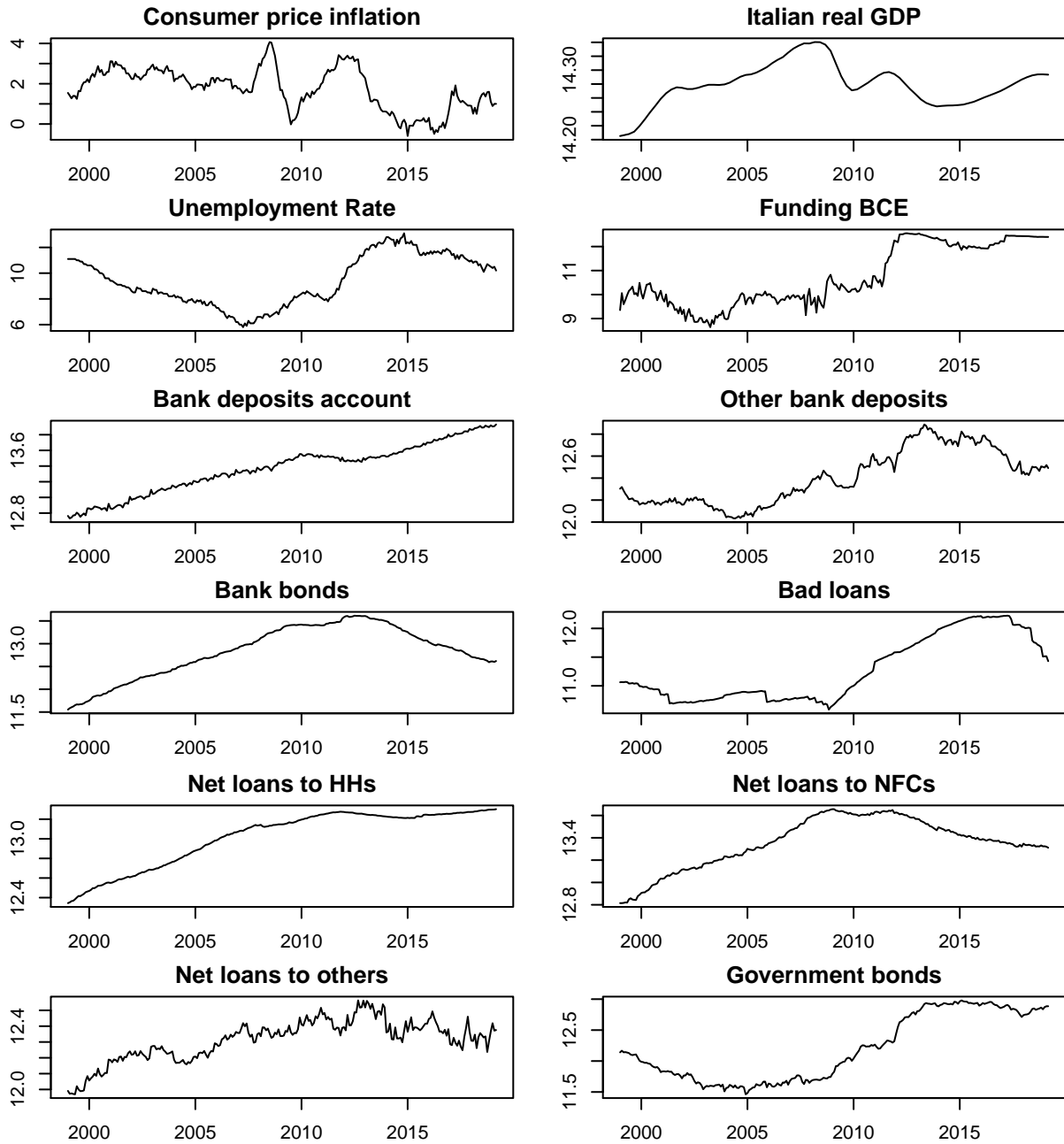
- Acharya, V., R. Engle, and D. Pierret (2014). Testing macroprudential stress tests: The risk of regulatory risk weights. *Journal of Monetary Economics* 65, 36–53.
- Acharya, V. V. and S. Steffen (2015). The "greatest" carry trade ever? Understanding eurozone bank risks. *Journal of Financial Economics* 115(2), 215–236.
- Ahnert, L., P. Vogt, V. Vonhoff, and F. Weigert (2020). Regulatory stress testing and bank performance. *European Financial Management* 26(5), 1449–1488.
- Akaike, H. (1969). Fitting autoregressive models for prediction. *Annals of the institute of Statistical Mathematics* 21(1), 243–247.
- Akaike, H. (1974). A new look at the statistical model identification. In *Selected Papers of Hirotugu Akaike*, pp. 215–222. Springer.
- Altavilla, C., M. Darracq Pariès, and G. Nicoletti (2015). Loan supply, credit markets and the euro area financial crisis. Technical report, ECB Working Paper, No. 1861.
- Antolin-Diaz, J., I. Petrella, and J. F. Rubio-Ramirez (2020). Structural scenario analysis with svars. *Journal of Monetary Economics*.
- Arnold, B., C. Borio, L. Ellis, and F. Moshirian (2012). Systemic risk, macroprudential policy frameworks, monitoring financial systems and the evolution of capital adequacy. *Journal of Banking & Finance* 36(12), 3125–3132.
- Bañbura, M., D. Giannone, and M. Lenza (2015). Conditional forecasts and scenario analysis with vector autoregressions for large cross-sections. *International Journal of Forecasting* 31(3), 739–756.
- Bañbura, M., D. Giannone, and L. Reichlin (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics* 25(1), 71–92.
- BCBS (2009a). Principles for sound stress testing practices and supervision. Technical report, Basel Committee on Banking Supervision.
- BCBS (2009b). Strengthening the resilience of the banking sector. *BIS Consultative Document*.
- Berrospide, J. M., R. M. Edge, et al. (2019). The effects of bank capital buffers on bank lending and firm activity: What can we learn from five years of stress-test results? In *Board of Governors of the Federal Reserve System, Finance and Economics Discussion Series 2019-050*.
- Boivin, J., M. P. Giannoni, and D. Stevanović (2018). Dynamic effects of credit shocks in a data-rich environment. *Journal of Business & Economic Statistics*, 1–13.

- Bonucchi, M. and M. Catalano (2019). How severe are the EBA macroeconomic scenarios for the Italian Economy? A joint probabilistic approach. *Prometeia working papers*.
- Borio, C., M. Drehmann, and K. Tsatsaronis (2014). Stress-testing macro stress testing: does it live up to expectations? *Journal of Financial Stability* 12, 3–15.
- Breuer, T., M. Jandacka, K. Rheinberger, M. Summer, et al. (2009). *How to find plausible, severe, and useful stress scenarios*. Österr. Nationalbank.
- Breuer, T. and M. Summer (2020). Systematic stress tests on public data. *Journal of Banking & Finance* 118, 105886.
- Budnik, K. B., M. Balatti, I. Dimitrov, J. Groß, M. Kleemann, T. Reichenbachas, F. Sanna, A. Sarychev, N. Siginenko, and M. Volk (2020). Banking euro area stress test model.
- Canova, F. and M. Ciccarelli (2004). Forecasting and turning point predictions in a Bayesian panel VAR model. *Journal of Econometrics* 120(2), 327–359.
- Castrén, O. and M. Rancan (2014). Macro-Networks: An application to euro area financial accounts. *Journal of banking & finance* 46, 43–58.
- Conti, A. M., A. Nobili, F. M. Signoretti, et al. (2018). Bank capital constraints, lending supply and economic activity. Technical report, Bank of Italy, Economic Research and International Relations Area.
- Dave, C., S. J. Dressler, and L. Zhang (2013). The bank lending channel: a FAVAR analysis. *Journal of Money, Credit and Banking* 45(8), 1705–1720.
- De Meo, E. (2019). Scenario design for macro-financial stress testing. *Available at SSRN 3493554*.
- De Meo, E., L. Prosperi, G. Tizzanini, and L. Zicchino (2018). Forecasting Macro-Financial Variables in an International Data-Rich Environment Vector Autoregressive Model (iDREAM). *Prometeia working papers*.
- De Mol, C., D. Giannone, and L. Reichlin (2008). Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics* 146(2), 318–328.
- Dees, S. and J. Henry (2017). Stress-Test Analytics for Macroprudential Purposes: Introducing STAMP€. *Satellite Models* 13.
- Dees, S., F. d. Mauro, M. H. Pesaran, and L. V. Smith (2007). Exploring the international linkages of the euro area: a global VAR analysis. *Journal of applied econometrics* 22(1), 1–38.
- Denton, F. T. (1971). Adjustment of monthly or quarterly series to annual totals: an approach based on quadratic minimization. *Journal of the American Statistical Association* 66(333), 99–102.

- Diebold, F. X. and R. S. Mariano (2002). Comparing predictive accuracy. *Journal of Business & economic statistics* 20(1), 134–144.
- Dovern, J., C.-P. Meier, and J. Vilsmeier (2010). How resilient is the German banking system to macroeconomic shocks? *Journal of Banking & Finance* 34(8), 1839–1848.
- Durbin, J. and S. J. Koopman (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika* 89(3), 603–616.
- Durdu, C. B., R. M. Edge, D. Schwindt, et al. (2017). Measuring the Severity of Stress-Test Scenarios. Technical report, Board of Governors of the Federal Reserve System (US).
- EBA (2018, July). *Guidelines on institution’s stress testing*. European Banking Authority.
- Georgescu, O. M., M. Gross, D. Kapp, and C. Kok (2017). Do stress tests matter? evidence from the 2014 and 2016 stress tests.
- Giannone, D., M. Lenza, H. Pill, and L. Reichlin (2012). The ECB and the interbank market. *The Economic Journal* 122(564), F467–F486.
- Giannone, D., M. Lenza, and L. Reichlin (2012). Money, credit, monetary policy and the business cycle in the euro area. *Centre for Economic Policy Research (CEPR)*.
- Gross, M., C. Kok, and D. Żochowski (2016). The impact of bank capital on economic activity-Evidence from a Mixed-Cross-Section GVAR model. *ECB Working Paper*.
- Hannan, E. J. and B. G. Quinn (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society: Series B (Methodological)* 41(2), 190–195.
- Henry, J., C. Kok, A. Amzallag, P. Baudino, I. Cabral, M. Grodzicki, M. Gross, G. Halaj, M. Kolb, M. Leber, et al. (2013). A macro stress testing framework for assessing systemic risks in the banking sector. *ECB Occasional Paper* (152).
- Hirtle, B., A. Kovner, J. Vickery, and M. Bhanot (2016). Assessing financial stability: the capital and loss assessment under stress scenarios (CLASS) model. *Journal of Banking & Finance* 69, S35–S55.
- Hoggarth, G., S. Sorensen, and L. Zicchino (2005). Stress tests of UK banks using a VAR approach. *Bank of England working paper no. 282*.
- Janda, K. and O. Kravtsov (2019). Does regulatory stress testing make banks perform better and be less risky? In *Annual Conference on Finance and Accounting*, pp. 109–120. Springer.
- Kadiyala, K. R. and S. Karlsson (1997). Numerical methods for estimation and inference in Bayesian VAR-models. *Journal of Applied Econometrics* 12(2), 99–132.

- Kok, C., C. Müller, C. Pancaro, et al. (2019). The disciplining effect of supervisory scrutiny on banks, and risk-taking: evidence from the eu wide stress test. *Macroprudential Bulletin* 9.
- Litterman, R. B. et al. (1979). Techniques of forecasting using vector autoregressions. Technical report.
- Pesaran, M. H. and Y. Shin (1998). Generalized impulse response analysis in linear multivariate models. *Economics letters* 58(1), 17–29.
- Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics* 6(2), 461–464.
- Sims, C. A. and T. Zha (1998). Bayesian methods for dynamic multivariate models. *International Economic Review*, 949–968.
- Von Borstel, J., S. Eickmeier, and L. Krippner (2016). The interest rate pass-through in the euro area during the sovereign debt crisis. *Journal of International Money and Finance* 68, 386–402.
- Yuen, K. (2015). Determining the Severity of Macroeconomic Stress Scenarios. *Supervisory Staff Reports, Federal Reserve Bank of New York*.

Figure 1: Variables for the empirical analysis



Note: All variables enter in log-levels with the exception of inflation, interest rates, unemployment rate and the Tier 1 ratio which are expressed in percentages. Banking macroeconomic variables are specific to Italian economy.

Figure 2: Variables for the empirical analysis (cont.)

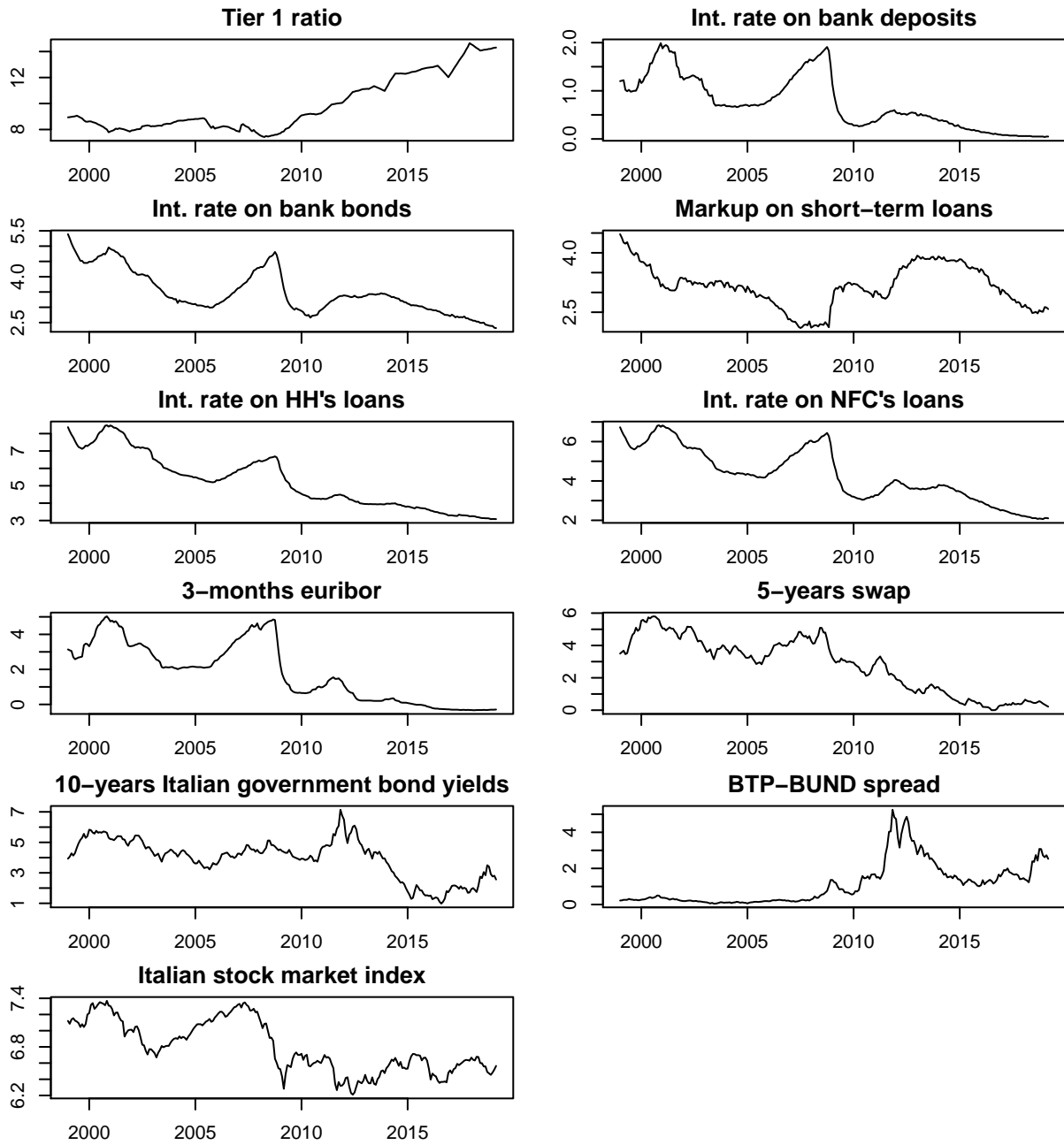
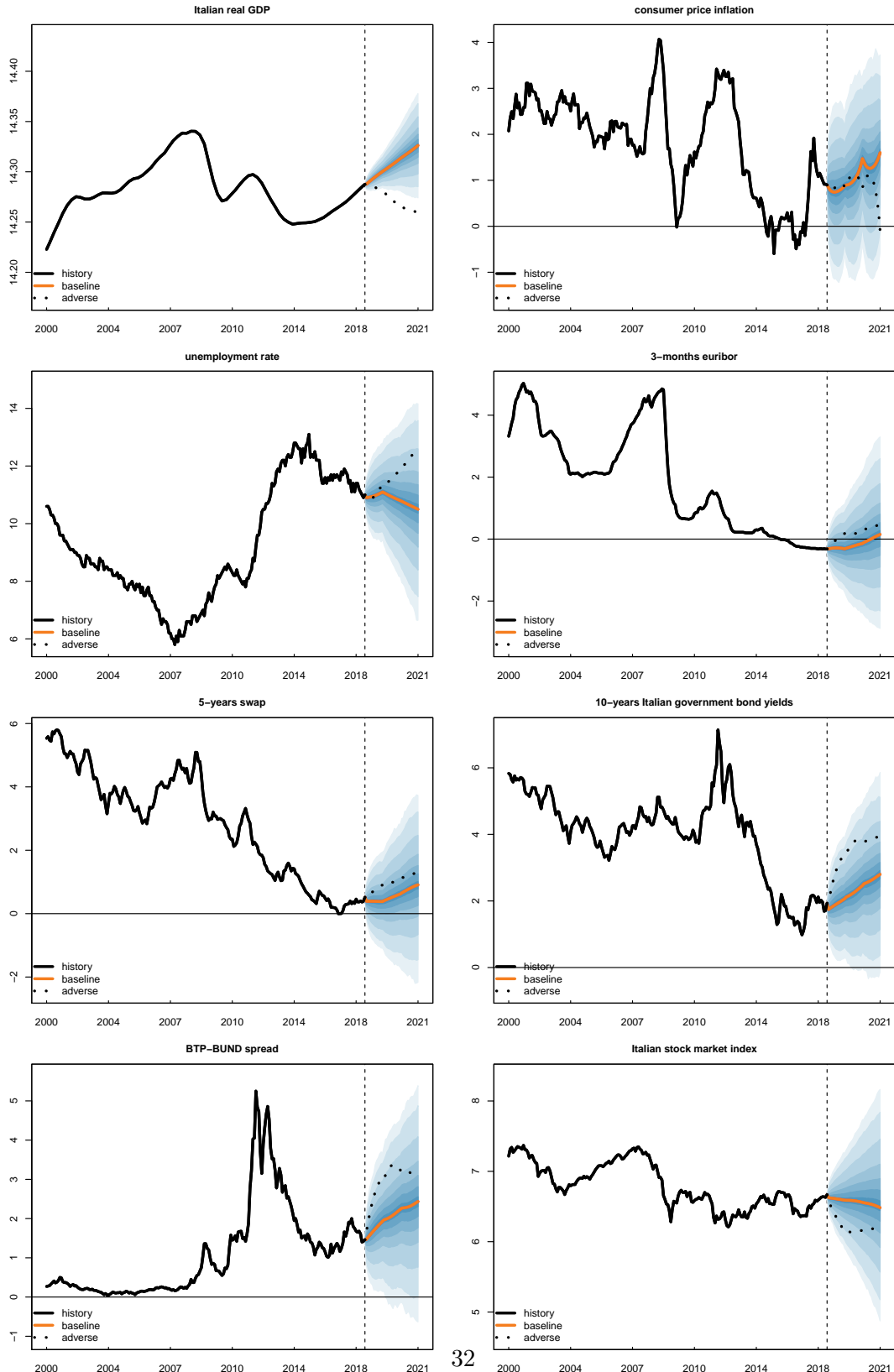
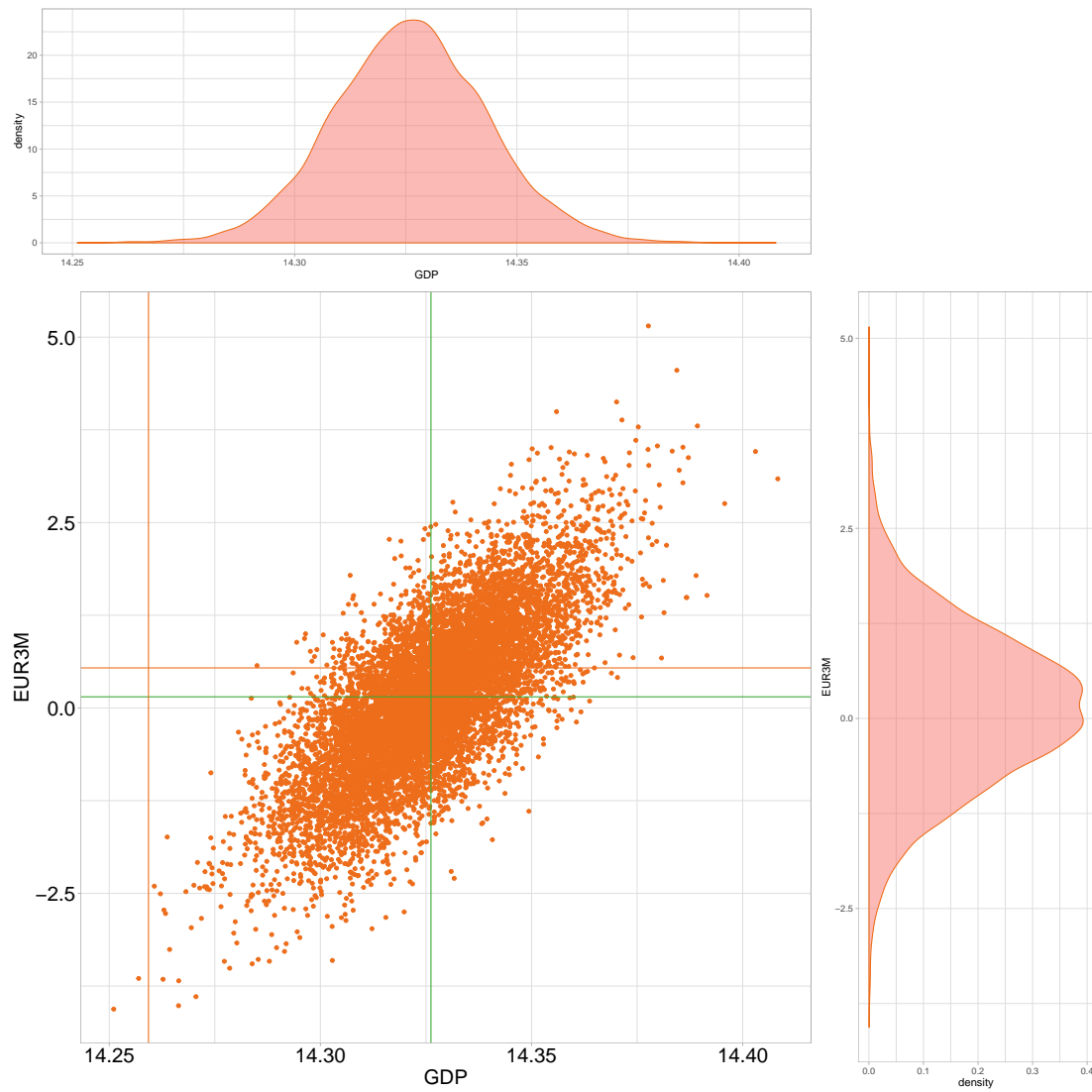


Figure 3: Fanchart of input variables in the 2018 EBA stress-test exercise: baseline scenario with confidence bands and adverse scenario



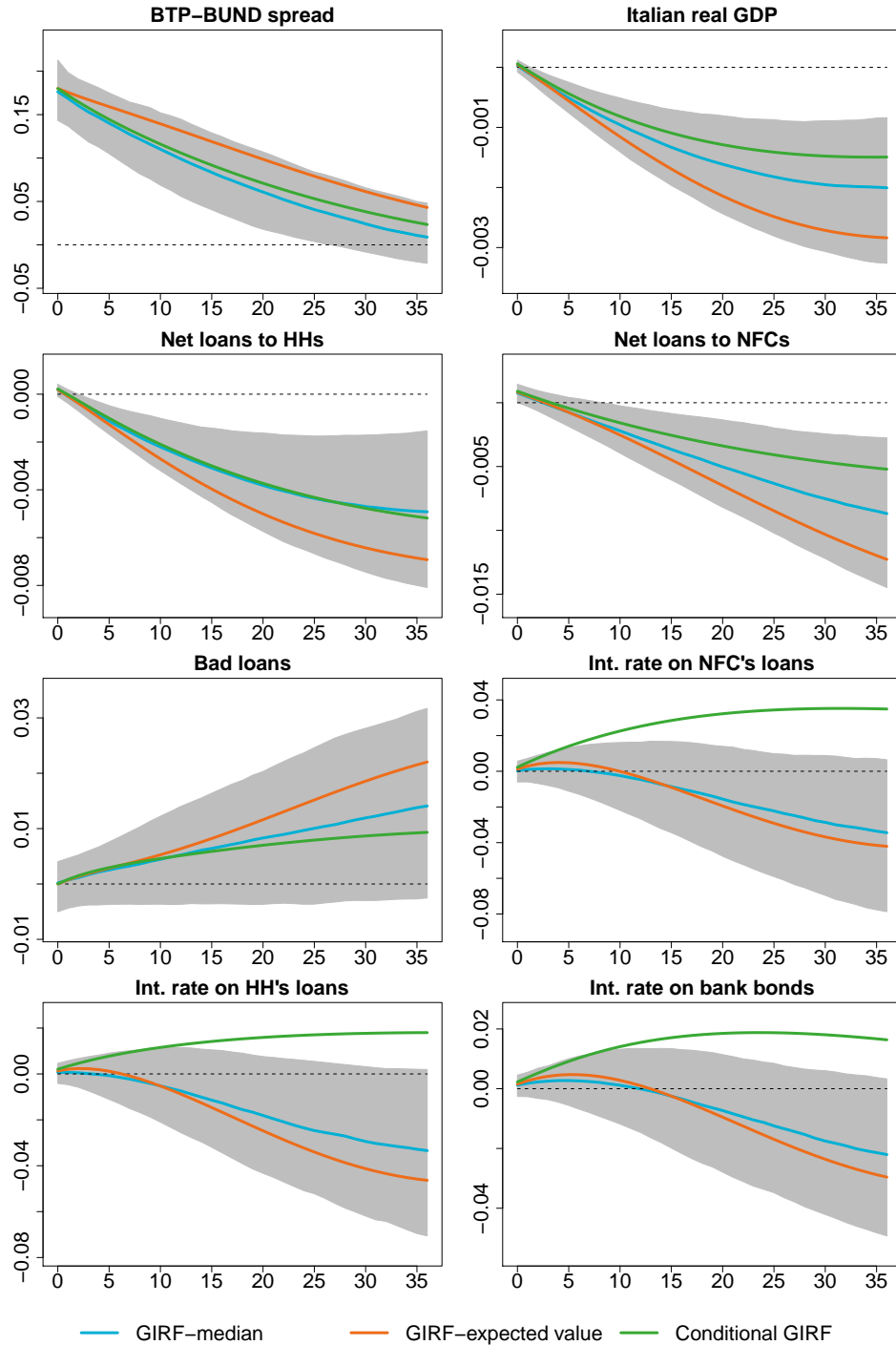
Note: The figures show the fanchart for each input variable of the 2018 EBA scenario. The orange line is the baseline scenario, while the dashed black line represents the adverse scenario. Confidence bands are calculated from simulated path around the baseline scenario as described in section 4.3.1. Confidence bands corresponds to the following confidence intervals: 10%, 30%, 50%, 70%, 90%, 98%, 99.5%

Figure 4: Multivariate distribution of Italian Real GDP and Euribor 3m at 3 years horizon in the 2018 exercise



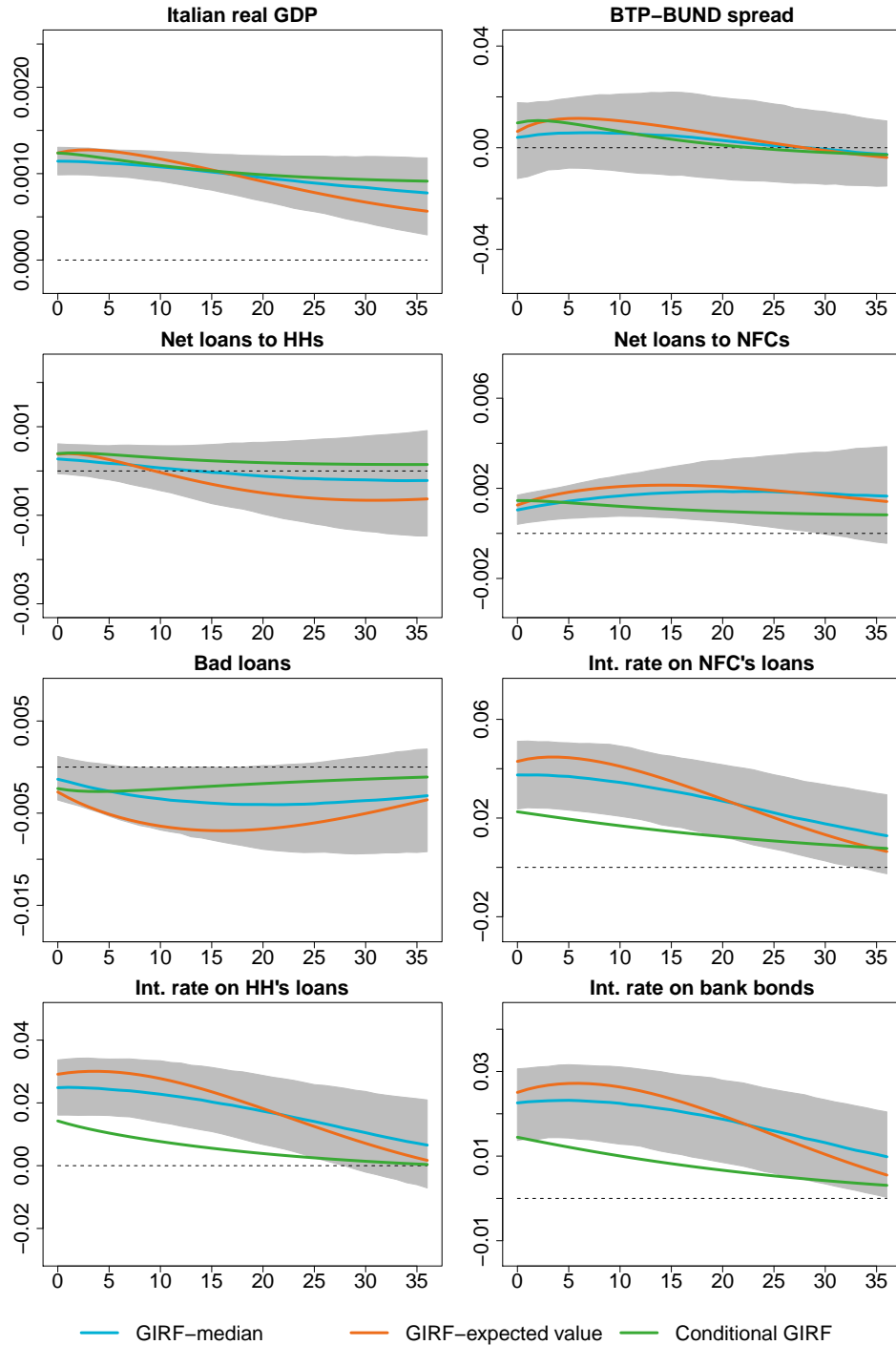
Note: The figure shows the scatter plot of simulated scenarios for Italian Real GDP and 3-month Euribor for the 2018 stress tests at 3-year horizon. The vertical line represents the adverse scenario of the 2018 stress test for Real GDP while the horizontal one for the 3-month Euribor.

Figure 5: Responses of to a shock in the BTP Bund spread



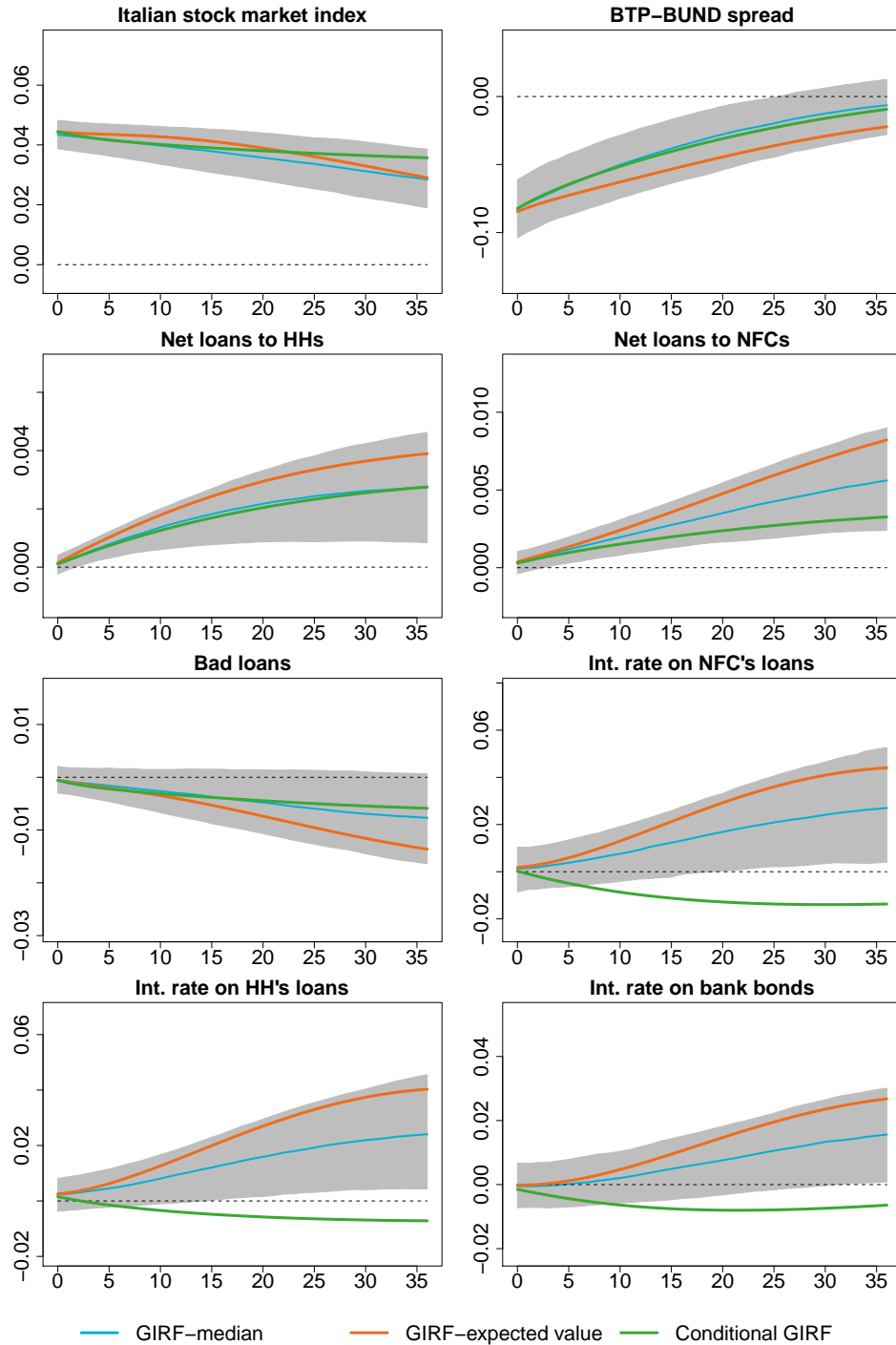
Note: The figure represents the impulse response function to a 15bp increases in the BTP-BUND spread and the corresponding 90% confidence interval from bootstrap (1000 draws) for BTP-BUND spread, Real GDP, net loans stock and interests rate to HHs and NFCs, bad loans and bank bonds' interest rate. The prior on the sum of coefficients has been added with the hyperparameter $\kappa = 10\lambda$. The blue line represents the median of all bootstrapped responses; the orange line is the GIRF from the estimated model and the green line is conditional GIRF as explained in Section 4.2.

Figure 6: Responses to a shock in Italian Real GDP



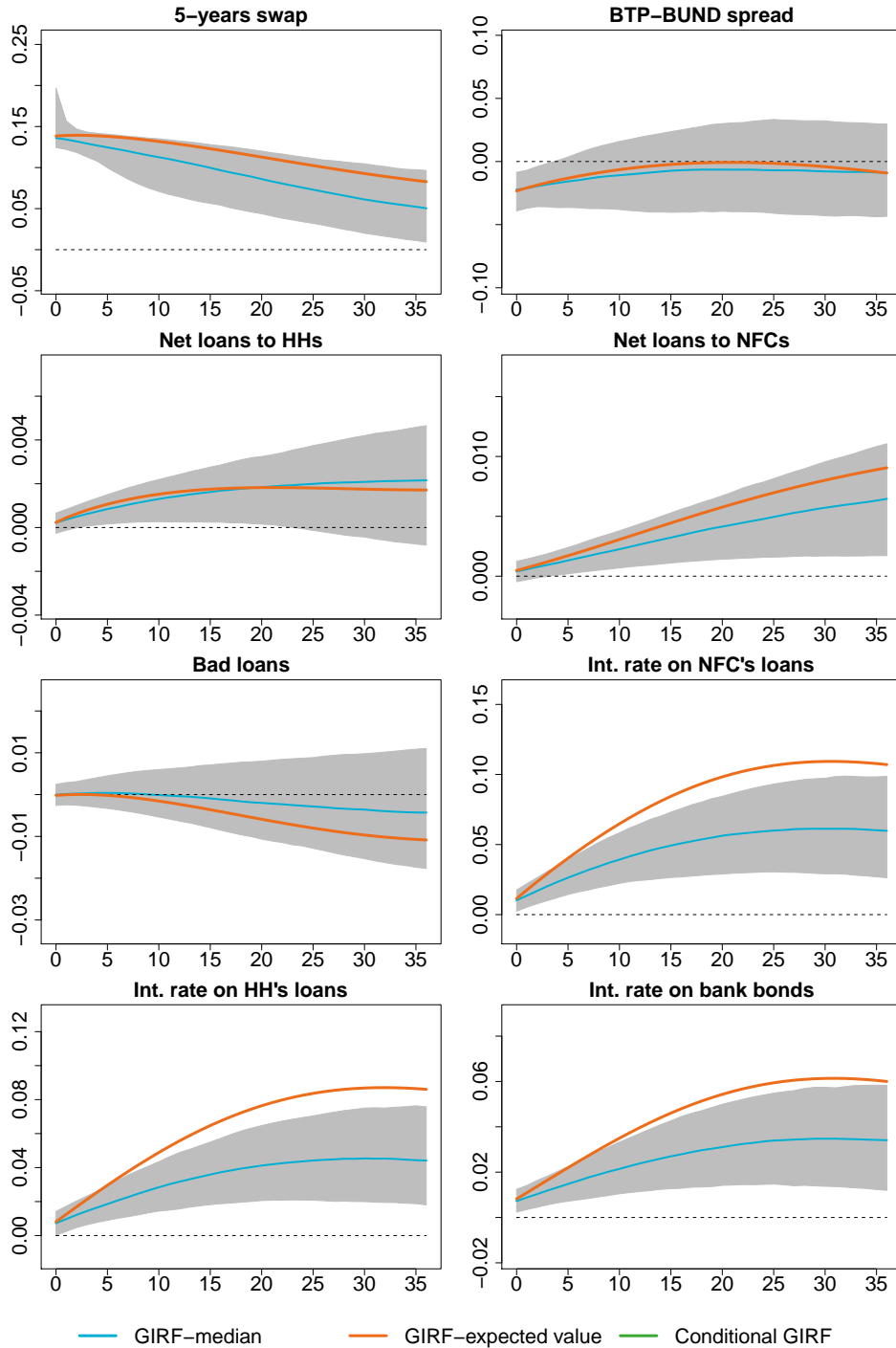
Note: The figure represents the impulse response function to a 0.1% drop in Italian Real GDP and the corresponding 90% confidence interval from bootstrap (1000 draws) for BTP-BUND spread, Real GDP, net loans stock and interests rate to HHs and NFCs, bad loans and bank bonds' interest rate. The prior on the sum of coefficients has been added with the hyperparameter $\kappa = 10\lambda$. The blue line represents the median of all bootstrapped responses; the orange line is the GIRF from the estimated model and the green line is conditional GIRF as explained in Section 4.2.

Figure 7: Responses to a shock in Italian Stock Price index



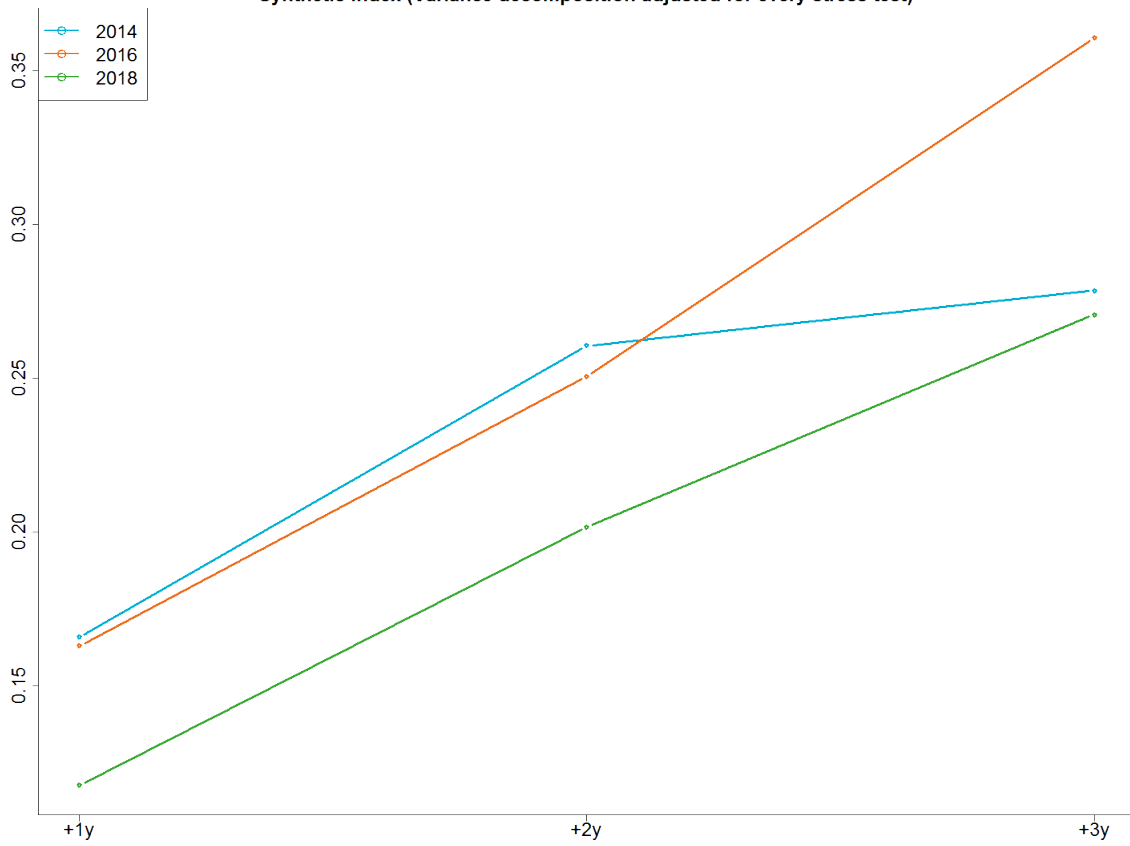
Note: The figure represents the impulse response function to a 5% increase in Italian stock market prices and the corresponding 90% confidence interval from bootstrap (1000 draws) for BTP-BUND spread, Real GDP, net loans stock and interests rate to HHs and NFCs, bad loans and bank bonds' interest rate. The prior on the sum of coefficients has been added with the hyperparameter $\kappa = 10\lambda$. The blue line represents the median of all bootstrapped responses; the orange line is the GIRF from the estimated model and the green line is conditional GIRF as explained in Section 4.2.

Figure 8: Responses to a shock in 5-years swap rates



Note: The figure represents the impulse response function to a 15bp increase in 5-years swap rate and the corresponding 90% confidence interval from bootstrap (1000 draws) for BTP-BUND spread, Real GDP, net loans stock and interests rate to HHs and NFCs, bad loans and bank bonds' interest rate. The prior on the sum of coefficients has been added with the hyperparameter $\kappa = 10\lambda$. The blue line represents the median of all bootstrapped responses; the orange line is the GIRF from the estimated model and the green line is conditional GIRF as explained in Section 4.2.

Figure 9: Synthetic aggregate severity measures for the three stress tests
Synthetic index (Variance decomposition adjusted for every stress test)



Note: The graph shows the synthetic aggregate severity index for the three stress tests computed according to Section 4.3.1

Table 1: Variables names, labels and orders of integration (δ_i)

name	label	δ_i
Consumer price inflation	INFL	0
Italian real GDP	GDP	1
Unemployment Rate	UN	2
Funding BCE	FUND.ECB	1
Bank deposits account	DEP.DA	1
Other bank deposits	DEP.OTH	1
Bank bonds	BBONDS	1
Bad loans	BLOANS	2
Net loans to HHs	LOANS.HH	2
Net loans to NFCs	LOANS.SNF	1
Net loans to others	LOANS.OTH	1
Government bonds	SOV.BONDS	1
Tier 1 ratio	T1.RATIO	1
Int. rate on bank deposits	DEP.DA.Y	0
Int. rate on bank bonds	BBONDS.Y	1
Markup on short-term loans	DIFF.Y	1
Int. rate on HH's loans	LOANS.HH.Y	1
Int. rate on NFC's loans	LOANS.SNF.Y	1
3-months euribor	EUR3M	1
5-years swap	SWAP.5Y	1
10-years Italian government bond yields	BTP.10Y	1
BTP-BUND spread	SPREAD	1
Italian stock market index	STOCK.MARKET	1

A Estimation of monthly time series of Tier 1 Ratio

Since there is no monthly Tier 1 capital ratio series available for the Italian banking sector we had to construct one. To do so we collected all available data on Tier 1 capital and risk-weighted assets (RWA) of the Italian banks for any semester from 2000 onward, from the statistical appendix of the Economic Bulletin of the Bank of Italy¹⁷. Then we exploit two high-frequency (i.e. monthly) proxies for Tier 1 capital and RWA in order to reconstruct the intra-semester missing months, keeping all the end-of-semester data equal to

¹⁷<https://www.bancaditalia.it/pubblicazioni/bollettino-economico/>

Table 2: Variance Decomposition

horizon	response variable	shock variable						
		GDP	EUR3M	BTP.10Y	SWAP.5Y	UN	INFL	STOCK.MARKET
H=12	LOANS.HH	0.24	0.12	1.98	5.26	0.11	0.20	6.45
	LOANS.SNF	3.51	4.77	0.24	5.17	0.01	0.13	3.21
	SOV.BONDS	1.67	3.23	1.64	1.38	0.47	0.25	3.27
	BLOANS	2.48	4.97	0.96	0.10	0.37	2.84	0.58
	LOANS.SNF.Y	8.73	22.36	3.64	11.09	0.06	1.38	0.40
	LOANS.HH.Y	6.69	21.78	2.42	10.46	0.13	0.24	0.65
	BBONDS.Y	8.41	20.12	3.25	8.33	0.10	1.77	0.12
	T1.RATIO	2.46	7.96	0.21	1.06	0.56	1.78	0.46
H=24	LOANS.HH	0.29	0.77	7.24	5.12	0.26	1.14	10.51
	LOANS.SNF	2.70	8.49	1.20	11.82	0.00	0.13	7.93
	SOV.BONDS	1.05	3.52	4.33	2.54	0.37	0.82	5.74
	BLOANS	2.59	9.89	2.00	1.05	1.02	2.38	1.88
	LOANS.SNF.Y	6.25	22.96	5.95	24.35	0.15	1.28	1.79
	LOANS.HH.Y	4.60	23.54	3.47	23.54	0.27	0.15	2.54
	BBONDS.Y	6.56	23.64	4.87	18.84	0.06	1.70	1.07
	T1.RATIO	2.30	13.07	0.95	3.63	0.69	2.58	2.15
H=36	LOANS.HH	0.35	1.25	10.94	3.99	0.36	1.90	11.81
	LOANS.SNF	1.67	8.51	2.66	15.54	0.00	0.43	11.54
	SOV.BONDS	0.70	3.41	5.78	3.46	0.79	1.22	7.21
	BLOANS	1.97	11.38	3.36	2.45	1.93	1.65	3.80
	LOANS.SNF.Y	4.53	20.04	6.42	33.77	0.25	1.02	3.66
	LOANS.HH.Y	3.15	20.82	3.27	32.53	0.51	0.12	4.88
	BBONDS.Y	5.02	22.80	4.76	27.41	0.06	1.36	2.92
	T1.RATIO	1.74	14.70	1.78	6.76	0.84	3.04	4.32

Note: The table reports the percentage share of the adverse EBA scenarios in the forecast error variance for a given forecast horizon for loans to HHs and NFCs (stock and interest rates), sovereign bank bonds, bad loans, bank deposit interest rate and Tier 1 ratio. The prior on the sum of coefficients has been added with the hyperparameter $\kappa = 10\lambda$.

Table 3: Adverse shock (deviation from baseline) and realization probabilities (% , in parentheses) for all variables

variable	2014			2016			2018		
	+1y	+2y	+3y	+1y	+2y	+3y	+1y	+2y	+3y
GDP	-0.02 (0.41)	-0.04 (0.02)	-0.06 (0.07)	-0.02 (0.07)	-0.04 (0.02)	-0.06 (0.07)	-0.02 (0.02)	-0.05 (0)	-0.07 (0.02)
INFL	0.03 (52.67)	-0.27 (32.13)	-1.2 (2.9)	-1.01 (4.16)	-1.58 (2.21)	-2.5 (0.79)	-0.02 (48.92)	-0.63 (15.99)	-1.7 (0.99)
UN	0.31 (73.67)	1.31 (94.13)	2.4 (97.13)	0.29 (69.33)	1.19 (89.9)	2.2 (95.49)	0.3 (69.23)	1.11 (88.28)	2.2 (95.71)
EUR3M	0.19 (65.2)	0.11 (55.72)	0.06 (51.77)	0.24 (69.23)	0.18 (60.65)	0.06 (52.49)	0.49 (83.51)	0.43 (71.04)	0.39 (65.35)
SWAP.5Y	1.07 (97.68)	0.8 (85.02)	0.84 (81.62)	0.48 (81.23)	0.45 (71.3)	0.24 (60.03)	0.53 (84.27)	0.46 (73.15)	0.44 (68.63)
BTP.10Y	2.08 (99.95)	1.51 (97.39)	1.49 (94.81)	1.13 (95.78)	1.07 (88.41)	0.89 (80.04)	1.33 (98.11)	1.33 (94.28)	1.17 (89.05)
SPREAD	1.23 (97.67)	0.7 (81.49)	0.38 (66.9)	0.9 (91.67)	0.72 (78.91)	0.5 (68.05)	1.11 (96.22)	1 (88.4)	0.67 (75.21)
STOCK.MARKET	-0.24 (19.87)	-0.2 (37.88)	-0.23 (40.95)	-0.36 (7.21)	-0.3 (23.4)	-0.18 (36.62)	-0.45 (3.62)	-0.39 (16.08)	-0.29 (29)
LOANS.HH	-0.02 (25.99)	-0.06 (12.01)	-0.08 (13.1)	-0.02 (20.21)	-0.06 (10.89)	-0.09 (11.87)	-0.02 (17.53)	-0.06 (10.58)	-0.08 (13.23)
LOANS.SNF	-0.01 (43.79)	-0.06 (26.25)	-0.09 (23.18)	-0.02 (34.64)	-0.06 (25.44)	-0.09 (25.09)	-0.02 (31.65)	-0.06 (25.75)	-0.08 (26.97)
BLOANS	0.04 (55.59)	0.08 (59.08)	0.12 (59.18)	0.06 (60.8)	0.17 (71.91)	0.29 (75.1)	0.09 (72.86)	0.21 (79.68)	0.33 (80.9)
SOV.BONDS	0.01 (50.93)	0.03 (52.71)	-0.01 (48.89)	0.03 (58.14)	0.06 (58.69)	0.04 (53.6)	0.06 (65.18)	0.15 (70.72)	0.21 (69.01)
LOANS.TOT.Y	0.05 (57.67)	0.07 (57.38)	-0.01 (52.25)	0.02 (51.24)	-0.01 (49.37)	-0.04 (48.83)	0.2 (69.79)	0.24 (67.08)	0.24 (63.61)
BDEPOSITS.Y	0.33 (81.55)	0.5 (84.49)	0.42 (75.09)	0.28 (84.08)	0.3 (78.5)	0.2 (66.61)	0.24 (87.22)	0.23 (79.18)	0.16 (67.39)
T1.RATIO	-0.04 (62.41)	0.02 (69.36)	-0.07 (66.09)	-0.14 (46.35)	-0.13 (52.22)	-0.14 (53.04)	-0.36 (32.24)	-0.65 (30.73)	-1.01 (27.98)

Note: The table represents the difference between the IBASE forecast conditional on the adverse EBA stress tests scenario and the baseline one for all three stress tests. The model is estimated up until the last month of the year before the stress test begins (i.e. 2013:12, 2015:12 and 2017:12). The probability of realization in brackets are derived according to Section 4.3.1.

Table 4: Adverse shock (deviation from baseline) and realization probabilities (% , in parentheses) for all variables

variable	2008-crisis			2011-crisis			2018-stress test		
	+1y	+2y	+3y	+1y	+2y	+3y	+1y	+2y	+3y
GDP	-0.02 (0.08)	-0.08 (0)	-0.1 (0.23)	-0.01 (17.76)	-0.05 (4.24)	-0.075 (6.59)	-0.02 (0.02)	-0.05 (0)	-0.07 (0.02)
INFL	2.56 (99.97)	-1.48 (12.18)	0.27 (58.71)	1.56 (95.35)	0.13 (53.01)	-1.02 (26.37)	-0.02 (48.92)	-0.63 (15.99)	-1.7 (0.99)
UN	0.74 (84.38)	1.88 (91.67)	1.98 (83.56)	2.24 (99.82)	3.18 (98.54)	3.83 (94.86)	0.3 (69.23)	1.11 (88.28)	2.2 (95.71)
EUR3M	-0.11 (43.79)	-3.66 (0.75)	-2.91 (8.22)	-0.25 (42.78)	-0.28 (45.24)	0.15 (51.92)	0.49 (83.51)	0.43 (71.04)	0.39 (65.35)
SWAP.5Y	0.32 (63.97)	-1.81 (9.48)	-2.41 (10.22)	-1.09 (14.16)	-1.49 (19.05)	-1.25 (30.62)	0.53 (84.27)	0.46 (73.15)	0.44 (68.63)
BTP.10Y	-0.07 (45.69)	-0.95 (20.83)	-1.12 (24.44)	0.05 (53.87)	-0.39 (37.34)	-2.05 (10.45)	1.33 (98.11)	1.33 (94.28)	1.17 (89.05)
SPREAD	0.04 (61.83)	0.46 (98.55)	1.26 (99.8)	1.08 (99.04)	1 (87.36)	-0.88 (22.43)	1.11 (96.22)	1 (88.4)	0.67 (75.21)
STOCK.MARKET	-0.15 (24.26)	-0.26 (27.97)	0 (48.94)	-0.13 (34.06)	-0.19 (39.25)	0.1 (54.41)	-0.45 (3.62)	-0.39 (16.08)	-0.29 (29)

Note: The table represents the difference between the IBASE forecast conditional on the adverse EBA stress tests scenario and the baseline one for all three stress tests. The model is estimated up until the last month of the year before the 2018 stress test begins (i.e. 2017:12), while for the 2008-crisis and the 2011-crisis the model is estimated a year before the first month of the beginning of the recession according to the business cycle chronologies dates of peak and trough Dates of the Economic Cycle Research Institute (ECRI) www.businesscycle.com (i.e. 2007:07 and 2010:03). This allow us to have a +2y impact at the beginning of the recession. The probability of realization in brackets are derived according to Section 4.3.1.

Table 5: Counterfactual shock (deviation from baseline, rescaled to 2014's probability level)

variable	2014			2016			2018		
	+1y	+2y	+3y	+1y	+2y	+3y	+1y	+2y	+3y
GDP	-0.02 (0.41)	-0.04 (0.02)	-0.06 (0.07)	-0.02 -	-0.04 -	-0.06 -	-0.02 -	-0.04 -	-0.06 -
INFL	0.03 (52.67)	-0.27 (32.13)	-1.2 (2.9)	0.08 -	-0.33 -	-1.83 -	0.05 -	-0.27 -	-1.35 -
UN	0.31 (73.67)	1.31 (94.13)	2.4 (97.13)	0.36 -	1.45 -	2.49 -	0.37 -	1.47 -	2.47 -
EUR3M	0.19 (65.2)	0.11 (55.72)	0.06 (51.77)	0.17 -	0.09 -	0.04 -	0.23 -	0.05 -	0.04 -
SWAP.5Y	1.07 (97.68)	0.8 (85.02)	0.84 (81.62)	1.07 -	0.82 -	0.89 -	1.09 -	0.73 -	0.83 -
BTP.10Y	2.08 (99.95)	1.51 (97.39)	1.49 (94.81)	2.21 -	1.79 -	1.8 -	2.22 -	1.71 -	1.57 -
SPREAD	1.23 (97.67)	0.7 (81.49)	0.38 (66.9)	1.4 -	1.03 -	0.75 -	1.39 -	1.16 -	0.8 -
STOCK.MARKET	-0.24 (19.87)	-0.2 (37.88)	-0.23 (40.95)	-0.22 -	-0.13 -	-0.12 -	-0.21 -	-0.13 -	-0.1 -
LOANS.HH	-0.02 (25.99)	-0.06 (12.01)	-0.08 (13.1)	-0.02 -	-0.06 -	-0.09 -	-0.01 -	-0.04 -	-0.06 -
LOANS.SNF	-0.01 (43.79)	-0.06 (26.25)	-0.09 (23.18)	-0.01 -	-0.05 -	-0.09 -	-0.01 -	-0.04 -	-0.08 -
BLOANS	0.04 (55.59)	0.08 (59.08)	0.12 (59.18)	0.05 -	0.16 -	0.26 -	0.11 -	0.25 -	0.43 -
SOV.BONDS	0.01 (50.93)	0.03 (52.71)	-0.01 (48.89)	0.03 -	0.09 -	0.08 -	0.05 -	0.16 -	0.24 -
LOANS.TOT.Y	0.05 (57.67)	0.07 (57.38)	-0.01 (52.25)	0.08 -	0.13 -	0.16 -	0.16 -	0.3 -	0.2 -
BDEPOSITS.Y	0.33 (81.55)	0.5 (84.49)	0.42 (75.09)	0.28 -	0.4 -	0.33 -	0.16 -	0.19 -	0.08 -
T1.RATIO	-0.04 (62.41)	0.02 (69.36)	-0.07 (66.09)	-0.12 -	-0.01 -	0.04 -	-0.25 -	-0.37 -	-0.5 -

Note: The table represents the difference between the IBASE forecast conditional on the adverse EBA stress tests scenario and the baseline one for the only 2014 exercise. The model is estimated up until the last month of the year before the stress test begins (i.e. 2013:12, 2015:12 and 2017:12). The probability of realization in brackets are derived according to Section 4.3.1, only for the 2014 exercise.

the official one. The high-frequency proxy for the Tier 1 capital is the monthly stock of capital and reserves from Bank of Italy statistics Banks and Money: National Data¹⁸. For the RWA we constructed a proxy for the standard RWA, using Bank of Italy data for all the sub-components in the computation. Then the standard benchmark was adjusted to account for the Italian banks that specifically stated that they moved to the internal methods for RWA’s computation¹⁹. The two proxies were used to augment the frequency of the Bank of Italy’s series using Chow et al. (1971) method for temporal disaggregation. Finally, the Tier 1 ratio was computed as the ratio between the Tier 1 capital and the RWA.

B Forecasting Performances

In this section we present the analysis on the forecasting performance of the IBASE. To do so we evaluate the model out-of-sample from December 2008 to March 2016 for a total of 88 iterations. We select a starting date, around the mid-point of the data, to guarantee a sufficiently long dataset from the first iteration. We then evaluate the model on a 3-year horizon, with the final date of the out-of-sample analysis chosen consequently and set, for the last iteration, equal to the last observation we have (i.e. March 2019). Starting from 2008 makes our forecasting exercise particularly challenging since we include two recessions period: (1) the Global Financial Crisis of 2008-09 and (2) the Sovereign Debt Financial Crisis of 2011-2012.

For each out-of-the-sample iteration we calibrate the hyperparameter λ following the same calibration method described in Section 4.1. However we change the tightness of the prior on the "sum of coefficients" setting the corresponding hyperparameter κ to five different values $\kappa = \kappa \times \lambda$ where $\kappa \in \{10, 20, 50, 100, 1000\}$, in order to support our initial choice of tightness.

For each iteration we also check the stability of the model by verifying that the unit root of the correspondent LBVAR model lies inside the unit circle. In Figure 10 we show the unit roots and the λ values, from which it is clear that the more unstable estimates are the ones associated with the two recession periods.

[Figure 10 about here.]

The performance analysis is conducted on quantitative forecast precision and direction accuracy. For quantitative forecast errors we compute the ratios of mean absolute errors (MAE) of the IBASE versus different benchmarks (Random Walk (RW), Random Walk with Trend (RWT) and AR(1)) for all values of κ on a 3-year horizon; in addition to this we also report a measure of forecasting performance that is the ability of the model to forecast the direction of the endogenous variables, measured as the number of correct directional forecasts over the total number of (stable) forecasts. Before comparing the IBASE with the benchmark models for all endogenous variables, in Table 6, we present the number of stable iterations, average MAE and direction accuracy across variables for all five values of κ in order to support our choice for $\kappa = 10$ and then present for that specific values the MAE and direction for all endogenous variables.

¹⁸<https://www.bancaditalia.it/pubblicazioni/moneta-banche/>

¹⁹In order to collect such information we looked at the balance sheets of the 10 largest Italian banks from 2013 to 2018.

We support the choice of $\kappa = 10$ for two main reasons. First, Table 6 shows that in terms of direction of the forecast such value clearly outperform all the other values, also being among the best performers in terms of MAE; the only downside of choosing $\kappa = 10$ it is represented by its stability ratio, however, from the graphical inspection of Figure 10 we already established that the majority of unstable episodes corresponds to the two financial crises. Second, we set a value equal to 10 for the degree of shrinkage following the standard approach used by the literature (Bańbura et al., 2010).

[Table 6 about here.]

Having established our choice for the κ parameter, tables 7 presents the IBASE performances for some selected bank variables. The IBASE outperforms all benchmark models for many of the bank stock variables such as: bank bonds, sovereign bonds (against AR(1)), bad loans, net loans to non-financial corporations and households, although for the latter variable our model overperforms only the random walk with trend. As for interest rates the IBASE is not able to clearly outperform all the other benchmark models. However, the sample used for the analysis covers two major recessions that created unprecedented financial turmoil and therefore pose many challenges for the ability of an econometric model to capture interest rates' movements, that were driven by a range of unconventional monetary policy measures. Lastly the IBASE model is able to outperform all selected benchmarks for the Tier 1 ratio. This is an important results given the pivotal importance that such indicator has in the stress tests.

[Table 7 about here.]

Finally, table 8 presents the directions results for some selected bank variables. Similar to the performance analysis the IBASE is more accurate in forecasting the direction of the stock variables and the Tier 1 ratio rather than the interest rates. Overall the IBASE proves to be an efficient LVAR model, especially suited for bank-stock variables and Tier 1 ratio.

[Table 8 about here.]

C Parameters vs Shock Uncertainty

In section 4.3.1 we describe the procedure to generate uncertainty around the baseline scenario. Our benchmark results rely on two sources of uncertainty: 1) the uncertainty connected with the parameters' realization (2) and the uncertainty due to the future realization of the shocks (Monte Carlo). In this section, we show the importance of the first source of uncertainty in generating confidence bands for a selection of variables. Figure 11 presents the simulated distributions conditional on the central path of the EBA baseline scenarios for a selection of variables. Simulated paths are generated at 1- to-3-year horizon with the model estimated until December 2017. As it can be noticed, accounting for both sources of uncertainty contributes to generate more probability mass for more extreme events, especially at longer horizons. In particular, the distance between the 99.9% and the 0.1% percentile is around 50% larger for loans at the 3-year horizon and

30% larger for Real GDP. This result is explained by larger second moments of the simulated distribution when accounting for both sources of uncertainty, but also by larger leptokurtosis. This implies that simulated paths in our benchmark setup do not follow a normal distribution as suggested by Jarque Bera test statistics. We claim then that accounting for both types of uncertainty is important to better capture the likelihood of adverse and rare events.

[Figure 11 about here.]

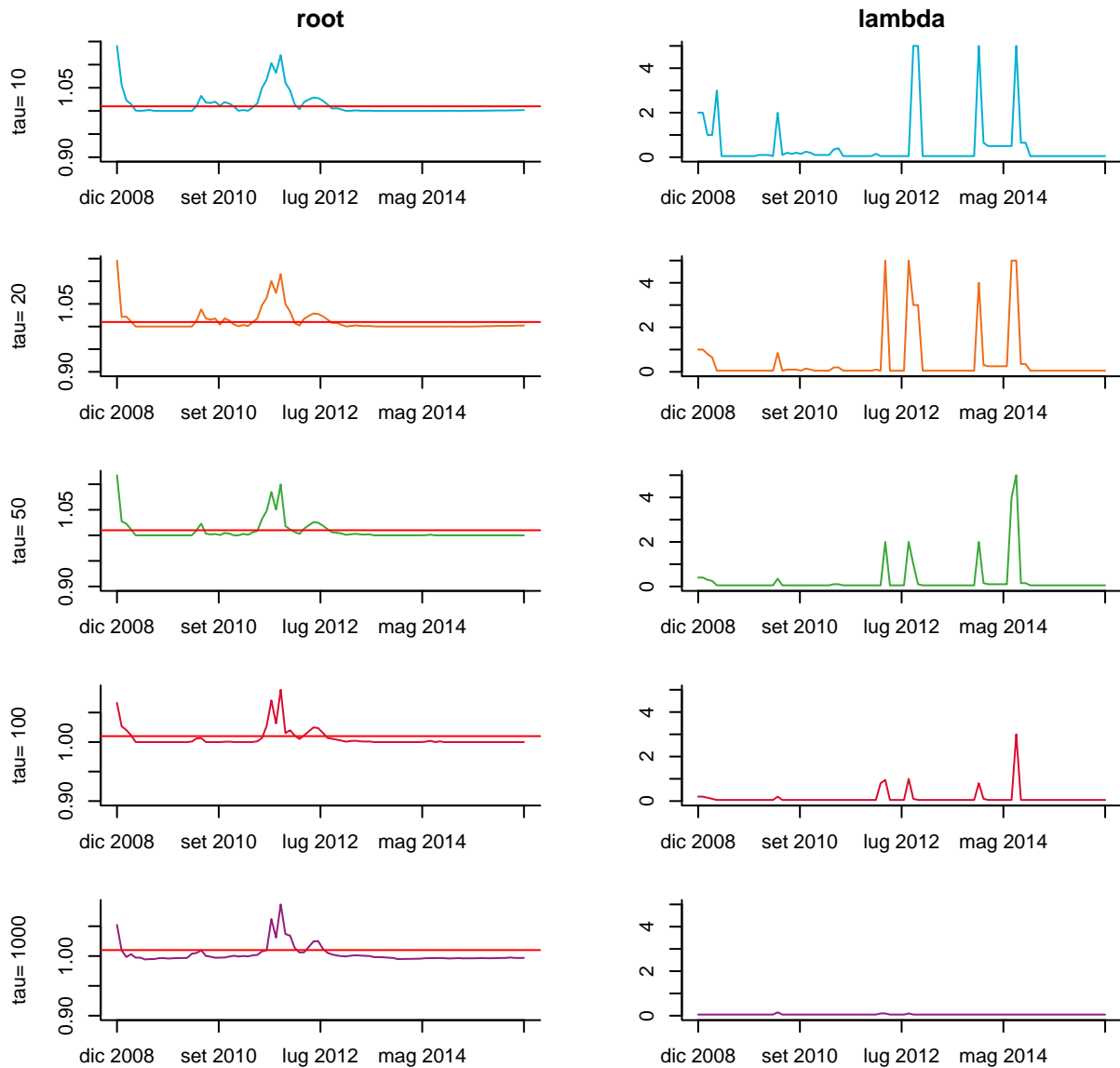
D The model with more lags

As a robustness check we present the results from the estimation of the IBASE with a different lag structure. In order to decide which lag value to use, we employed several identification criteria, different from the BIC used in the baseline model. More precisely, we used four different information criteria: (1) Akaike information criterion (AIC) (Akaike, 1974); (2) Hannan-Quinn information criterion (HQ) (Hannan and Quinn, 1979); (3) Schwarz Criterion (SC) (Schwarz et al., 1978) and (4) Final Prediction Error (FPE) (Akaike, 1969). WBy averaging the results from all these tests we set the number of lags equal to four. To reduce the computational burden we generate confidence bands around the baseline scenario only considering the uncertainty coming from shocks' realization (i.e. Monte Carlo).

Table 9 present the stress tests deviations and probabilities for the input variables from the IBASE estimated with four lags. If we compare the probability associated with the three stress test from Table 9 we notice that they are qualitatively similar to the benchmark ones of Table 3. There are some small differences in particular for unemployment rate and interest rates, for which using a higher lag structure results in a higher probability. As for the probability associated with the Real GDP and Italian stock market the results from Table 9 are smaller than those obtained by estimating the IBASE with only one lag. Overall we don't notice any major differences thus supporting the choice of using, as a benchmark model the IBASE estimated with one lag.

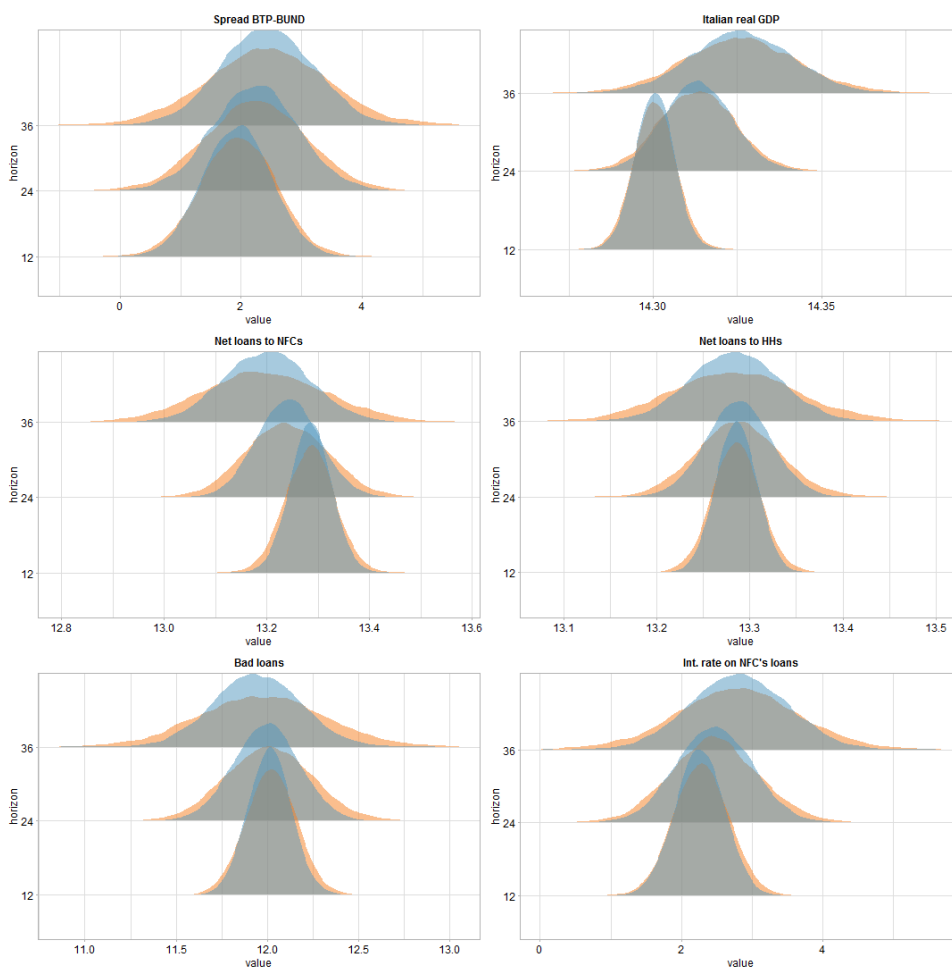
[Table 9 about here.]

Figure 10: Roots of the characteristic polynomial (left) and λ_s (right) values of the model estimated recursively for different κ



Note: Each point in time represents the value of the root of the characteristic polynomial and the λ value for different out-of-sample intervals. The red line on the left-panel graphs separates the stable model estimations from the unstable ones.

Figure 11: Simulated distribution for selected variables at different horizon with and without parameter uncertainty



Note: The graph shows for a selection of variables the simulated distribution at different horizons generated by the model with data available until 2017:12. Central forecast are conditioned on the path of available variables in the EBA baseline scenario. The orange area represents the distribution generated accounting for both shocks' and parameters' uncertainty, while the blue area is the distribution generated accounting for shocks' uncertainty only.

Table 6: IBASE performance comparison with RW and AR(1) for different values of κ

			IBASE vs RW		IBASE vs RWT		IBASE vs AR(1)	
	stability	direction	MAE	RMSE	MAE	RMSE	MAE	RMSE
10	0.69	0.52	0.96	0.89	1.25	1.14	1.17	1.06
20	0.72	0.48	0.96	0.90	1.25	1.14	1.17	1.07
50	0.78	0.47	0.95	0.90	1.25	1.16	1.15	1.05
100	0.82	0.48	0.94	0.88	1.26	1.16	1.15	1.04
1000	0.88	0.46	0.95	0.92	1.25	1.18	1.13	1.06

Note: The table represents average stability, direction, MAE and RMSE performances for different value of κ , across variables. The stability is given by the percentage of stable model iteration while the direction is measured as the number of correct directional forecasts over the total number of (stable) forecasts.

Table 7: IBASE performance comparison with RW, RWT and AR(1) for $\kappa = 10$

variable	IBASE vs RW				IBASE vs RWT				IBASE vs AR(1)			
	+1y	+2y	+3y	mean	+1y	+2y	+3y	mean	+1y	+2y	+3y	mean
BBONDS	1.57	1.61	1.63	1.59	<u>2.75</u>	<u>2.72</u>	<u>2.87</u>	<u>2.78</u>	<u>2.27</u>	<u>2.11</u>	<u>2.19</u>	<u>2.21</u>
BLOANS	1.59	1.11	0.83	1.4	1.41	1.02	0.85	1.26	1.68	1.58	<u>1.56</u>	<u>1.64</u>
LOANS.HH	0.95	0.7	0.51	0.81	1.72	<u>1.63</u>	<u>1.69</u>	<u>1.7</u>	0.82	0.58	0.57	0.77
LOANS.SNF	1.72	1.55	1.02	1.41	<u>3.75</u>	<u>3.66</u>	<u>2.55</u>	<u>3.09</u>	<u>2.88</u>	<u>2.71</u>	<u>1.84</u>	<u>2.35</u>
LOANS.OTH	0.67	0.51	0.48	0.63	0.7	0.67	0.78	0.76	0.62	0.4	0.38	0.56
SOV.BONDS	0.76	0.72	0.65	0.78	0.84	0.88	0.83	0.89	1.52	1.76	1.73	1.58
T1.RATIO	1.22	<u>1.64</u>	<u>1.51</u>	1.38	1.09	1.39	<u>1.17</u>	1.18	1.45	<u>1.63</u>	<u>1.58</u>	1.48
DEP.DA.Y	0.57	0.62	0.37	0.53	0.43	0.51	0.31	0.42	0.58	0.63	0.38	0.54
BBONDS.Y	0.6	0.77	0.64	0.66	0.46	0.66	0.52	0.54	0.59	0.74	0.61	0.65
LOANS.HH.Y	0.62	0.67	0.56	0.66	0.48	0.54	0.38	0.53	0.45	0.34	0.22	0.43
LOANS.SNF.Y	0.63	0.78	0.61	0.7	0.57	0.67	0.47	0.59	0.66	0.79	0.63	0.71
Mean	0.99	0.97	0.8	0.96	1.29	1.3	1.13	1.25	1.23	1.21	1.06	1.17

Note: The table represents the ratio between the MAE of different benchmark models (random walk, random walk with trend and AR(1)) and the IBASE model for different time horizons and for selected banking variables. Bold numbers correspond to value of the ratio above one, while underlined values corresponds to combinations of horizon and variables where we reject the null hypothesis of the Diebold-Mariano test (Diebold and Mariano, 2002) (at the 10% confidence interval) in favor of the IBASE.

Table 8: IBASE direction accuracy for $\kappa = 10$ and different time horizon

	direction			
	+3m	+1y	+2y	+3y
BBONDS	0.66	0.57	0.59	0.51
BLOANS	0.91	0.69	0.63	0.41
LOANS.HH	0.52	0.55	0.54	0.50
LOANS.SNF	0.63	0.59	0.45	0.50
LOANS.OTH	0.59	0.55	0.61	0.59
SOV.BONDS	0.59	0.54	0.42	0.44
T1.RATIO	0.56	0.57	0.42	0.44
DEP.DA.Y	0.49	0.35	0.24	0.20
BBONDS.Y	0.52	0.40	0.26	0.34
LOANS.HH.Y	0.71	0.54	0.51	0.30
LOANS.SNF.Y	0.68	0.42	0.42	0.27

Note: The table represents the direction accuracy for different time horizons. Direction accuracy is measured as the number of correct directional forecasts over the total number of (stable) forecasts.

Table 9: Adverse shock (deviation from baseline) and realization probabilities (% , in parentheses) for all variables with four lags

variable	2014			2016			2018		
	+1y	+2y	+3y	+1y	+2y	+3y	+1y	+2y	+3y
GDP	-0.02 (0.16)	-0.04 (0)	-0.06 (0)	-0.02 (0.03)	-0.04 (0)	-0.06 (0.01)	-0.02 (0)	-0.05 (0)	-0.07 (0)
INFL	0.05 (54.84)	-0.21 (32.43)	-1.2 (0.37)	-0.94 (9.28)	-1.49 (3.45)	-2.5 (0.11)	-0.03 (47.64)	-0.61 (11.65)	-1.7 (0.06)
UN	0.3 (73.23)	1.3 (95.37)	2.4 (99.09)	0.29 (72.85)	1.19 (93.65)	2.2 (97.92)	0.29 (73.06)	1.1 (92.7)	2.2 (98.59)
EUR3M	0.32 (78.27)	0.26 (66.2)	0.06 (53.51)	0.35 (79.04)	0.26 (65.14)	0.06 (53.32)	0.58 (90.29)	0.51 (79.01)	0.39 (70.54)
SWAP.5Y	1.19 (99.83)	0.9 (94.62)	0.84 (91.42)	0.55 (89.11)	0.48 (78.81)	0.24 (63.19)	0.59 (90.72)	0.52 (81.07)	0.44 (74.02)
BTP.10Y	2.1 (100)	1.55 (99.54)	1.49 (99.22)	1.14 (98.24)	1.1 (93.8)	0.89 (87.53)	1.29 (99.15)	1.31 (97.28)	1.17 (94.56)
SPREAD	1.12 (98.46)	0.69 (85.17)	0.49 (75.11)	0.81 (93.51)	0.68 (83.14)	0.5 (73.87)	0.86 (95.18)	0.74 (87.57)	0.51 (76.99)
STOCK.MARKET	-0.24 (9.97)	-0.2 (23.36)	-0.23 (25.28)	-0.36 (2.97)	-0.3 (14.54)	-0.18 (29.52)	-0.43 (1.05)	-0.38 (7.35)	-0.29 (19.06)

Note: The table represents the difference between the IBASE forecast conditional on the adverse EBA stress tests scenario and the baseline one for all three stress tests. The model is estimated up until the last month of the year before the stress test begins (i.e. 2013:12, 2015:12 and 2017:12). The probability of realization in brackets are derived according to Section 4.3.1. The model is estimated with four lags.