ANNALS OF ECONOMICS AND FINANCE **19-2**, 443–471 (2018)

Backtesting Stress Tests: A Guide for M2 Forward Guidance*

Kaihua Deng and Dun Jia†

We propose a simple procedure to gauge the reliability of a macroeconomic stress test model by positioning policy makers' projection of the economy in historical episodes that mirror the patterns of key economic variables under relevant test scenarios. The stress test model is backtested based on a weighted average measure of forecast errors. We justify the choice of scenario weights in two ways: Rational Inattention theory and calibration using historical data of economic disasters. The evaluation framework can be fruitfully applied to M2 forward guidance and is potentially valuable for Chinese monetary authorities.

Key Words: Cointegration; M2 components; Mean absolute error; Supervisory scenarios.
JEL Classification Numbers: E47, E58, G28.

1. INTRODUCTION

Stress testing has become an indispensable toolbox in the post-crisis global macro-prudential era. The macro version of it subjects a collection of key economic variables to high levels of stress and aims to evaluate the stability and resilience of the financial system to external shocks. The results of a stress test are then summarized by how well the financial system or certain variables are able to withstand a highly stressful operating environment (Drehmann, 2009), and it is standard practice to report only point forecasts under each scenario.

In mapping the external shocks to an observable outcome, one common criticism is that models are most needed when they are least reliable (Danielsson, 2008). Thus in a stress testing context, is there a way to draw inference about the robustness of models? Unlike traditional forecasting

* Kaihua Deng acknowledges support by the Fundamental Funds for the Central Universities and the Research Funds of Renmin University of China, 17XNB005.

† Deng: Co-corresponding author. Hanqing Advanced Institute of Economics and Finance, Renmin University of China, Beijing 100872, China. Email: dengk@ruc.edu.cn; Jia: Co-corresponding author. Hanqing Advanced Institute of Economics and Finance, Renmin University of China, Beijing 100872, China. Email: dun.jia@ruc.edu.cn.

1529-7373/2018

exercises in which horse races of candidate models can be conducted based on, say, recursive out-of-sample forecast errors, the supervisory scenario in a stress test, by contrast, is a one-time conjectural path and it is far from straightforward to find appropriate realized values against which the forecasts can be evaluated. This limitation has largely prevented a disciplined thinking about the choice of models in stress testing. In practice, stress tests implemented by the IMF and central banks rely on a combination of the monetary authorities' internal models and the financial institutions' in-house models that satisfy a number of regulatory requirements and reporting standards. Examples include the Supervisory Capital Assessment Program (Board of Governors, 2009), the Comprehensive Capital Analysis and Review (Board of Governors, 2014) and the European Banking Authority's EU-wide stress test (EBA, 2011). Whichever model is used, the implicit assumption is that the selected model correctly represents the underlying relationship in some exact sense. It turns out that this assumption was constantly refuted in the wake of the 2007-2009 crisis, yet researchers in recent years have somewhat glossed over the issue of scenario-based model evaluation. Intuitively, a model that performs well in normal times need not give a good forecast in hard times-from an internal risk management perspective, the opportunity cost for the financial industry to be either too conservative or too aggressive is high.

In view of this challenge, the paper suggests a way to backtest the performance of models by recasting each test scenario in a historical episode that bears similar patterns of the defining variables such as output growth, CPI inflation and interest rate. The contribution is less theoretical; rather, our goal is to provide a simple model evaluation tool in a stress testing context. We elaborate on the method using US data and extract the scenarios from the Board of Governors of the Federal Reserve System (FRB)'s 2015 Supervisory Scenarios for Annual Stress Tests Required under the Dodd-Frank Act Stress Testing Rules and the Capital Plan Rule (henceforth 2015 Supervisory Scenarios). The first step is to identify the relevant time windows in history that best approximate the test scenarios in question. Then, the researcher estimates the model using the in-sample data and conducts pseudo out-of-sample forecasts in the selected time windows. The next step is to compute the mean absolute forecast error (MAFE) or mean squared forecast error (MSFE) for each time window. Finally, to make the results under different scenarios comparable across different models, we draw upon the Rational Inattention theory à la Sims (2003) along with a historical calibration (Barro, 2006) to construct a weighted average measure of forecast performance. In doing so, the weighting scheme is a parsimonious way to incorporate the supervisors' optimal loss function by taking into account of the likelihood of scenario occurrence, the severity of damage, and possibly the best reaction of the central bank under each scenario. The loss is measured by the expected negative impacts on the economy due to a deviation of the forecast from the realized values.

In designing the stress scenarios, a recurring theme is that the external shocks "should be severe enough to be meaningful yet plausible enough to be taken seriously" (Borio et al., 2014). To determine the plausibility of scenarios, the reasoning often boils down to a combination of judgment and past experience, which lends some support to the use of historical data to backtest the model before moving on to the next stage. Granted, history is not a perfect test field for the future and the extent to which the selected historical episodes resemble the stress scenarios is open to debate, but on balance, the past does shed light on the future. In particular, we select three time windows that bear good resemblance to the three Supervisory Scenarios identified by the FRB: the burst of the internet bubble and the subsequent rebound (2001Q1-2003Q4), a robust boom period (2004Q1-2006Q4) and a broader Great Recession period (2008Q1-2010Q4) Another critical aspect in discerning relevant historical windows is that the model itself should encompass a stable relationship between the outcome variable of interest and the inputs (risk exposures) for a given time interval. Over time, a structural break may occur due to exogenous shocks, and the researcher must check whether the model still holds.

To illustrate, we work out a full example of forecasting M2 components under different scenarios. Forecasting the dynamics of aggregate M2 has a long tradition in the practice of monetary policy as it contains useful information about economic growth, and was once a critical instrument for the Fed; see Lown et al. (1999). Nowadays, although the more "conventional" monetary policy in the U.S. is the interest rate based management, we find it particularly informative to look into the movements of M2 components when the task is to better design a scenario-specific stress test model to conduct M2 forward guidance. For individual banks, a pertinent question is how M2 components as correspond to different types of deposits respond to varying market conditions, the answer to which is then used to facilitate analysis in many areas such as balance sheet management.

Sections 2 through 4 constitute a detailed analysis of conducting M2 forward guidance under different scenarios. Finally, we briefly discuss the merits of this evaluation framework for the Chinese economy where the monetary policy—viewed through the lens of M2—reveals some very interesting patterns.

2. DATA

The Dodd-Frank Act requires the Board of Governors of the Federal Reserve System (FRB) to conduct an annual supervisory stress test of bank holding companies (BHCs) with \$50 billion or greater in total consolidated assets. Three test scenarios are typically implemented, i.e., baseline, adverse and severely adverse. We extract the scenarios from the Board of Governors' 2015 Supervisory Scenarios for Annual Stress Tests Required under the Dodd-Frank Act Stress Testing Rules and the Capital Plan Rule (henceforth, 2015 Supervisory Scenarios). According to the FRB's 2015 Supervisory Scenarios, the baseline scenario is not the forecast of the Federal Reserve, but "follows a similar profile to the average projections from surveys of economic forecasters." In contrast, the adverse and severely adverse scenarios are not forecasts and they "describe hypothetical sets of conditions designed to assess the strength of banking organizations and their resilience to adverse economic environments."

The three scenarios start in the fourth quarter of 2014 and extend through the fourth quarter of 2017. Each scenario is defined by a total of sixteen variables deemed to be the most important barometers of the economic environment. These variables span three categories: six measures of aggregate output and prices (real and nominal GDP, real and nominal disposable income, unemployment rate, and CPI inflation rate), six measures of interest rates (rate on the 3-month Treasury bill, 5- and 10-year Treasury yield, 10 year BBB corporate bond yield, fixed-rate 30-year mortgage rate, and prime rate), and four measures of asset prices and financial conditions (Dow Jones stock market index, house price index, commercial real estate price index, and VIX). Historical data are obtained from the FRB data base and public sources. For more details on the definition of each variable, see FRB (2014). A snapshot of some of the domestic state variables that define the severely adverse scenario is given in Table 1. As it shows, this severe scenario is marked by the sizable deteriorations of all major aggregate variables followed by a steady recovery phase. Figure 1 and 2 illustrate the FRB scenarios in terms of the real GDP and the patterns of some other defining variables.

In Figure 1, we split the observation window into three periods by the president of the Federal Reserve System. The three 2015 FRB supervisory scenarios (2014Q4-2017Q4) are annexed to the historical real GDP series. The baseline scenario represents a sustained, moderate expansion in economic activities with a gradual normalization in the treasury yields across the term structure. Asset prices are assumed to increase modestly in the baseline case. The adverse scenario is characterized by a mild recession accompanied by falling asset prices and an increase in the U.S. inflationary pressures that result in a rapid increase in both short- and long-term Treasury yields. The severely adverse scenario features a deep and prolonged recession in which unemployment rate rises sharply and a broad-based contraction in asset prices takes place.

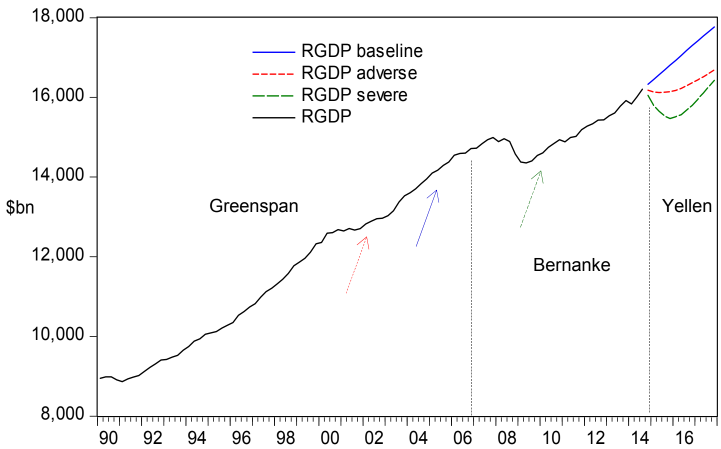
Figure 2 plots the real GDP growth rate together with the Treasury yields of three maturities. The yields series have broadly similar patterns,

TABLE 1.

Domestic state variables under the severely adverse scenario, 2014Q4-2017Q4.

Date	RGDP growth	NGDP growth	Real disp. income growth	Nominal disp. income growth	Unmpl. rate	CPI inflation	3-month Treasury yield	10-yr Treasury yield
2014Q4	−3.9	−2.8	−3	−0.1	6.9	4.3	0.1	0.9
2015Q1	−6.1	−4.7	−4.4	−2.3	8	3	0.1	1
2015Q2	−3.9	−2.4	−3.4	−2.2	8.8	1.7	0.1	1.2
2015Q3	−3.2	−1.7	−2.4	−1.4	9.5	1.3	0.1	1.3
2015Q4	−1.5	0	−1.5	−0.7	9.9	1.1	0.1	1.5
2016Q1	1.2	2.4	0.2	1.5	10	1.6	0.1	1.5
2016Q2	1.2	2.5	0.4	1.8	10.1	1.9	0.1	1.6
2016Q3	3	4.4	1.2	2.8	10	2	0.1	1.8
2016Q4	3	4.3	1.8	3.3	9.9	1.9	0.1	1.9
2017Q1	3.9	5.2	2.7	4.2	9.7	1.9	0.1	2
2017Q2	3.9	5.2	2.8	4.1	9.5	1.7	0.1	2.1
2017Q3	3.9	5.1	2.9	4.2	9.3	1.6	0.1	2.2
2017Q4	3.9	5.1	3	4.3	9.1	1.6	0.1	2.3

FIG. 1. FRB scenarios and real GDP: 2014Q4-2017Q4.



although the 3-month Treasury bill had experienced the most significant ups and downs by magnitude. Each negative spike in growth rate (1990, 2001, 2008) triggered a round of dropping interest rates across the board. Short-term interest rates in 2014 were just trivially above the zero lower bound. From the perspective of individual banks' funds management, the aggregate M2 series, by construction, can be decomposed into five compo-

FIG. 2. Treasury yields and log-differenced real GDP from 1990Q1 to 2014Q3.

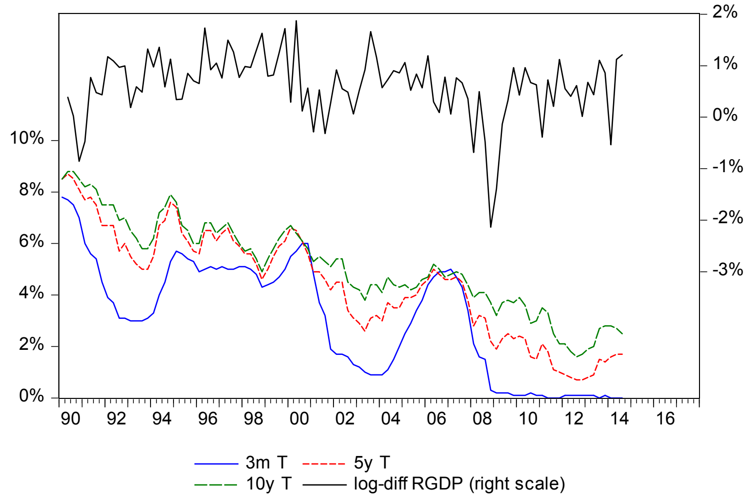
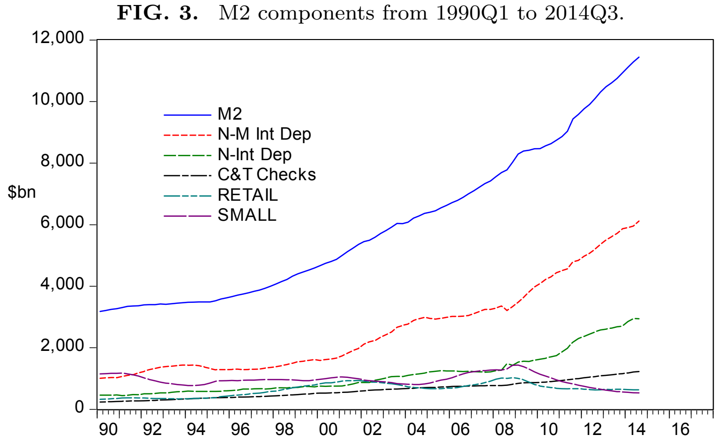
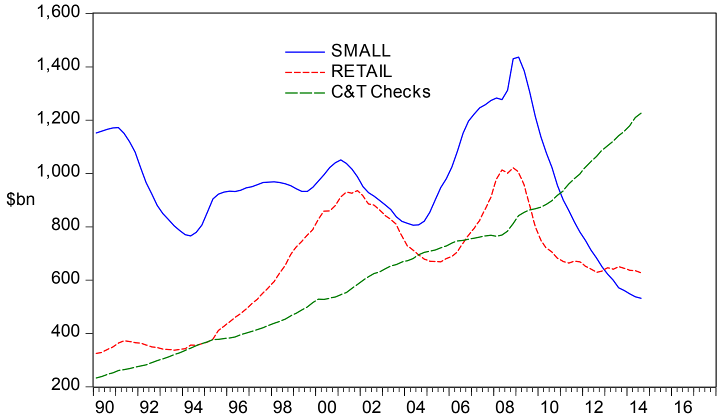


FIG. 3. M2 components from 1990Q1 to 2014Q3.



nents, namely, non-maturity interest bearing deposits (NM), non-interest bearing deposits (NI), currency and travelers checks (CT), retail money funds (RM) and small time deposits (ST). Figure 3 and 4 plot these components against the aggregate dynamics of M2. Of the five categories, the size of NM is by far the largest ever since 1992. The aggregate M2, NM, NI and CT have more or less followed some growing trends over the past thirty years, whereas RF and ST have declined since their most recent peaks in late 2008.

FIG. 4. Small time deposits, retail funds and C&T checks from 1990Q1 to 2014Q3.



3. MODEL SPECIFICATION

In this section, by augmenting the baseline model for forecasting the aggregate M2, we lay out the specifications of empirical models to generate forecasts of M2 components. Over the years, several models were proposed to capture the dynamics of M2. The FRB staff model was developed in the late 1980s and gained some popularity due to its simplicity and reasonable accuracy:

$$\begin{aligned}\Delta \ln M_t &= \beta_0 + \beta_1 t + \beta_2 \ln OPCOST_{t-1} + \beta_3 \ln(V_{t-1}) + \beta_4 \Delta \ln C_t \\ &+ \beta_5 \Delta \ln M_{t-1} + \Delta \ln OPCOST_t + \text{lags of } \Delta \ln C_t + \varepsilon_t, \end{aligned} \quad (1)$$

where the variables are opportunity cost ($OPCOST$), velocity (V), consumption (C), and M2 (M). Another commonly used model is the VAR framework:

$$y_t = \mu + \Pi_1 y_{t-1} + \Pi_2 y_{t-2} + \cdots + \varepsilon_t, \quad (2)$$

where $y_t = (\Delta \ln NGDP_t, \Delta \ln M_t, r_t)$ and may contain other variables; see Favara and Giordani (2009) and references therein.

While Eq. (1) is primarily focused on the short-run relationship, recent studies investigating the dynamics of M2 also support the classical quantitative theory of money which prescribes a cointegrating relationship among M2, real output, interest rate and price level (e.g., Stock and Watson, 1993; Friedman and Kuttner, 1992):

$$\ln M_t = \mu + \theta_0 t + \theta_1 \ln P_t + \theta_2 \ln OUTPUT_t + \theta_3 r_t + \varepsilon_t, \quad (3)$$

which may be re-written as

$$m_t = \mu + \theta_0 t + \theta_1 p_t + \theta_2 y_t + \theta_3 r_t + \varepsilon_t, \tag{4}$$

The presence of a long-run cointegrating relationship among the variables implies that a deviation is short-lived and the variables eventually return to their long-run equilibrium. Stock and Watson (1993) also documented a battery of integration and cointegration tests which are consistent with the long-run relationship assumption.

A refinement can be achieved by incorporating leads and lags of the first-differenced regressors in Eq. (4). The resulting OLS estimator of the augmented regression is known as the DOLS estimator which is consistent, asymptotically normally distributed and efficient under certain assumptions; see Stock and Watson(1993). In most cases, the leading terms are foundto be insignificant (with p-values as large as 0.98), so only contemporary and lagged differences are included in the discussion:

$$m_t = \mu+\theta_0 t+\theta_1 p_t+\theta_2 y_t+\theta_3 r_t+d_y(L)\Delta y_t+d_r(L)\Delta r_t+d_p(L)\Delta p_t+\varepsilon_t, \tag{5}$$

where $d(L)$ is the lead/lag polynomial.

A moment of inspection suggests that the three largest components follow very similar patterns as illustrated by Figure 5. Later we show that the results of the augmented Dickey-Fuller test for cointegration are consistentwith Eq. (5). Let $m_{1,2,3t}$ standfor NM, NI and CT, then the model for the first three component

$$m_{1,2,3t} = \mu+\theta_0 t+\theta_1 p_t+\theta_2 y_t+\theta_3 r_t+d_y(L)\Delta y_t+d_r(L)\Delta r_t+d_p(L)\Delta p_t+\varepsilon_t. \tag{6}$$

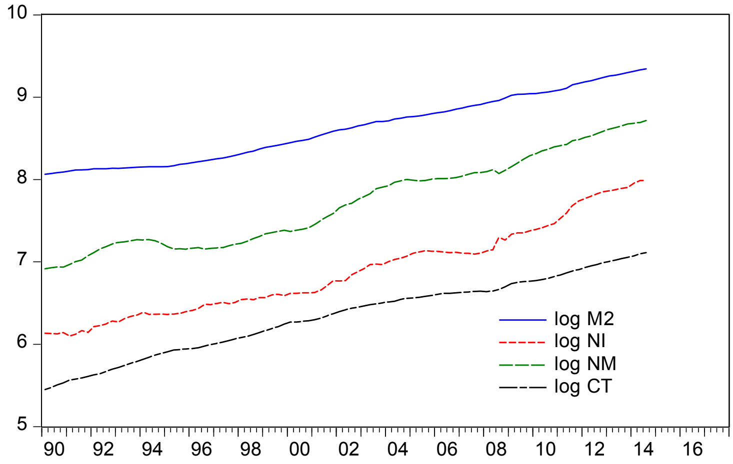
A restricted version of Eq. (6) is the real money balance case ($\theta_1 = 1$):

$$m_{1,2,3t} - p_t = \mu + \theta_0 t + \theta_2 y_t + \theta_3 r_t + d_y(L)\Delta y_t + d_r(L)\Delta r_t + \varepsilon_t. \tag{7}$$

In the most general setting, the parameters are estimated freely as they appear in Eq. (6) and Eq. (7) is invoked only when the restriction is not rejected. Indeed, the choice of variables and specifications are fairly large but we will focus on Eq. (6) to be consistent with economic theory and the money demand literature.

For RF and ST, the above cointegrating relationship is inappropriate as they behave more like I(2) processes. Therefore we take a different approach and focus on the growth rates. The data are first transformed into log-differenced growth rates and then examined with respect to interest rate, price and output. Figure 6 plots the RF and ST growth rates against the term spread (10-year minus 3-month Treasury yield) and reveals an

FIG. 5. Trend behavior of NI, NM and CT plotted against M2.



interesting countercyclical pattern. In fact, the next section shows that neither output nor price levels are significantin Eq. (6), and a more informative model—using $\Delta m_{4,5t}$ to stand for RF and ST growth rates—takes the form:

$$\begin{aligned}\Delta m_{4,5t} &= \mu + \theta_0 t + \theta_1 p_t + \theta_2 y_t + \theta_3 r_{1t} + \theta_4 r_{2t} + d_y(L)\Delta y_t \\ &+ d_p(L)\Delta p_t + d_1(L)\Delta r_{1t} + d_2(L)\Delta r_{2t} + \varepsilon_t, \qquad (8)\end{aligned}$$

where r_{1t} and r_{2t} are the 3-month Treasurybill rateand 5-year Treasury yield; or the restricted version:

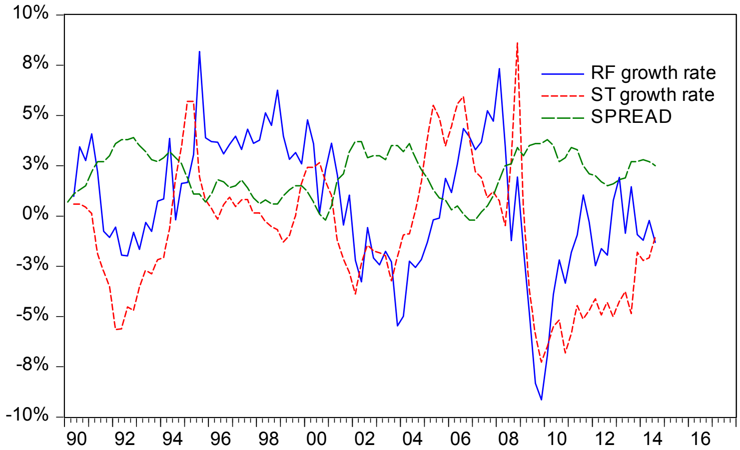
$$\Delta m_{4,5t} = \mu+\theta_0 t+\theta_1 p_t+\theta_2 y_t+\theta_3 Spr_t+d_y(L)\Delta y_t+d_p(L)\Delta p_t+d_s(L)\Delta Spr_t+\varepsilon_t, \qquad (9)$$

where *spr* is the term spread. In the empirical analysis, we proxy for p_t using GDP deflator defined as 100 $\times$ (nominal GDP/real GDP).

4. SCENARIO WEIGHTS

We now propose two ways to select the scenario weights that pin down the weighted average of scenario forecasting errors as an overall measure of the forecasting loss. First, we adopt the analytical framework of Rational Inattention theory à la Sims (2003) to optimize over weights by exploiting the relative sizes of "optimal" forecast errors across scenarios. Second, we dive into a calibration exercise to set our parameters to match the frequencies and magnitudes of damages of rare disasters observed in a large

FIG. 6. Relationship between retail funds/small time deposits and term spread.



dataset for many countries (Barro, 2006). We conclude that our backtesting performance is insensitive to the selection scheme for scenario weights.

4.1. Optimal Central Bank Monitoring Under Limited Attention

We use a simple model of Rational Inattention with frictions of imperfect information to illustrate the key mechanism through which the central bank optimally allocates its limited attention to be prepared for the realization of different economic scenarios. Solving for the attention allocation problem is eventually reduced to determine a set of scenario weights that construct the "optimal" loss function that a central bank should care about.

The model economy is static and stochastic. The economy can realize one of the three distinct scenarios as indexed by $i \in \{b, a, s\}$, respectively denoting the scenario of baseline (normal times with steady economic growth), adverse (period of temporary busts with relatively quicker rebound), and severe (events of dramatic and rare economic disasters). Each scenario is attached with a probability of occurrence $p_i \in (0, 1)$ with the assumption that the normal times along with temporary downturns are more often seen than the disasters such that $p_b > p_a > p_s$.

The central bank (CB) as monetary authority closely monitors the economy and has a payoff function that depends on the realized scenario i given by $U^i(a_i^{CB}, a_i, z_i)$. Its payoff depends on both actions of other sectors of the economy taken in scenario i, a_i, and the CB's own actions, a_i^{CB}, aiming to stabilize the economy in the form of monetary policy implementations.[1]

[1]Without loss of generality, the payoff function here does not have to be fully specified to take any particular form of monetary policy. Though in principle, the monetary policy

In addition, the scenario fundamental z_i that affects these actions enters to shift the central bank's payoff directly. Intuitively, for an altruistic and independent central bank, this payoff function may be interpreted as some aggregate welfare measure of the economy that the central bank monitors and acts to manage accordingly to changes in the economic fundamentals. The payoff thus can differ across scenarios.

We further assume that the payoff functions for all scenarios are quadratic. In addition, the first and second derivatives with respect to the first argument a_i^{CB} satisfy $U_1^i > 0$ and $U_{11}^i < 0$. We show in Appendix A that a second order approximation of payoff functions gives the optimal monetary policy taken by the central bank conditional on scenario i:

$$a_i^{CB,*} = \alpha_i + \phi_i a_i + (1 - \phi_i) z_i. \tag{10}$$

ϕ_i captures the relative magnitude of response of the central bank's reactions to actions of other sectors and to the economic fundamentals directly. ϕ_i may differ across scenarios so as to accommodate the situation that the central bank takes unconventional measures during the unusual times, along with the asymmetric responses or regime-switching reactions of the private sectors over different phases of the business cycle.[2] For tractability, impose the assumption that actions optimally taken by the private sectors are linear in the fundamentals, $a_i = \chi_i z_i$, we thus have the following linearized system that characterizes the central bank's payoffs

$$U^i(a_i^{CB}, a_i, z_i) = U^i(a_i^{CB,*}, a_i, z_i) - \gamma_i (a_i^{CB} - a_i^{CB,*})^2, \tag{11}$$

$$a_i^{CB,*} = \alpha_i + \beta_i z_i, \tag{12}$$

where $\beta_i = 1 - \phi_i + \phi_i \chi_i$ and $\gamma_i = -U_{11}^i/2 > 0$ is evaluated at the optimal action of the central bank $a_i^{CB,*}$. Despite that the objective function abstracts from many details, Eq. (11) resembles the quadratic form of the welfare criterion often seen in the class of New Keynesian models (Woodford, 2003).

Information Structure. We then impose the information frictions that when the monetary authority has to rely on imperfect information to learn which scenario is realized and then takes the corresponding reaction

---

instruments a_i^{CB} can be the monetary aggregates, interest rates, or any other liquidity tools. Actions of other sectors may be interpreted as the firms' production, household savings, government public spending, international trades among others.

[2]To name a few, Diebold and Rudebusch (1990) and Sichel (1993) document the asymmetries and non-linearities over U.S. business cycles. Sims and Zha (2006) find regime-switching U.S. monetary policy response coefficients. Bianchi and Melosi (2017) estimate different patterns of aggregate dynamics in the regime of zero-lower bound relative to normal times.

conditional on its understanding of the economic fundamentals. The central bank is assumed to have a prior belief about the scenario fundamental z_i with a distribution given by

$$z = \begin{pmatrix} z_b \\ z_a \\ z_s \end{pmatrix} \sim N(\mu, \Sigma), \Sigma = \begin{pmatrix} \Sigma_b & 0 & 0 \\ 0 & \Sigma_a & 0 \\ 0 & 0 & \Sigma_s \end{pmatrix}, \tag{13}$$

where μ and Σ denotes the mean and the variance-covariance matrix of the economic fundamentals respectively. We abstract from assuming non-zero cross-scenario covariances for simplicity.[3] To determine the stance of scenario-contingent action, the central bank has to first learn from a set of noisy signals s to infer the economic fundamentals:

$$s = z + \epsilon = \begin{pmatrix} z_b + \epsilon_b \\ z_a + \epsilon_a \\ z_s + \epsilon_s \end{pmatrix}, \quad \epsilon \sim N(0, \theta), \ \theta = \begin{pmatrix} \theta_b & 0 & 0 \\ 0 & \theta_a & 0 \\ 0 & 0 & \theta_s \end{pmatrix}, \tag{14}$$

where θ defines the variance-covariance matrix of the noise term which is assumed to be independent of the fundamentals. It follows from the Gaussian projection rule that the central bank updates its beliefs with posterior variance-covariance matrix such that

$$\Omega = \Sigma - \Sigma(\Sigma + \theta)^{-1}\Sigma. \tag{15}$$

Learning reduces the prior uncertainty Σ by a degree of $\Sigma(\Sigma + \theta)^{-1}\Sigma$, which depends on the precision of signals θ. In addition, the theory of Rational Inattention considers the fact that the optimizing agents have limited capacity of processing information such that they are prevented from being fully attentive to every piece of information. Therefore, they have to optimally allocate the right amount of attention to best mitigate the prior uncertainty. In this spirit, the central bank's optimization problem starts with choosing the "optimized" noisiness σ_i for $i \in \{b, a, s\}$ in θ subject to an attention constraint, which then feeds into the posterior uncertainty Ω_i. Following Sims (2003), the central bank's attention constraint is modeled in the following:

$$H(z) - H(z|s) \leq \kappa, \tag{16}$$

where $H(\cdot)$ is the entropy measure of information flow. $H(z)$ denotes the prior uncertainty about the fundamental states while $H(z|s)$ captures the posterior uncertainty about z conditional on learning from the noisy signals

[3]In the spirit of Sims (2003), triggering nonzero covariances across economic states does not violate the optimality principles of allocating agents' attention.

s. The shrinkage of uncertainty $H(z)-H(z|s)$ has to be capped from above by an information processing capacity, κ. Given that the fundamentals and the information structure are Gaussian and we have three fundamental state variables, we use $\log_2$ to measure the flow of information bits such that $H(x)=\frac{1}{2}\log_2[(2\pi e)^3\det(x)]$. Hence, the attention constraint can be expressed as

$$\frac{\det(\Sigma)}{\det(\Omega)} \leq 2^{2\kappa}, \tag{17}$$

where $\det(\cdot)$ is the determinant operator for a matrix. To optimize the signal structure, the expected payoff loss of the central bank is given by

$$\max_{\Omega_b,\Omega_a,\Omega_s} -E\left(\gamma_i(E(a_i^{CB,*}|s)-a_i^{CB,*})^2\right) \tag{18}$$

subject to Eq. (12) and (17). By exploiting the orthogonality properties of the signal structure s and economic fundamentals, we can show that the central bank is to minimize the quadratic loss

$$\max_{\Omega_b,\Omega_a,\Omega_s} -p_b\gamma_b q_b\Omega_b - p_a\gamma_a q_a\Omega_a - p_s\gamma_s q_s\Omega_s, \tag{19}$$

$$s.t.\ \ \Sigma_i=\Omega_i 2^{2\kappa_i}, \tag{20}$$

$$\kappa_b+\kappa_a+\kappa_s\leq\kappa, \tag{21}$$

where $q_i=\beta_i^2$. κ_i measures the size of reduction from the prior uncertainty associated with scenario i, Σ_i about z_i to a posterior uncertainty Ω_i conditional on learning. Thus, we interpret κ_i as the amount of attention a central bank pays to scenario i in order to be best prepared for scenario-contingent action a_i^{CB}. Given κ_i, the desired signal noisiness for scenario i, σ_i^2 is optimally determined.

In Appendix B, we show that the optimal attention allocated to scenario i for $i\in\{b,a,s\}$ is given by the following

$$\kappa_i=\begin{cases} 0, & x_i\leq 2^{-\kappa} \\ \frac{1}{3}(\kappa+\log_2(x_i)), & 2^{-\kappa}<x_i\leq 2^{2\kappa}, \\ k, & x_i\geq 2^{2\kappa}\end{cases} \tag{22}$$

where $x_i=\tau_i/\sqrt{\tau_{-i}}, \tau_i=p_i\gamma_i q_i\Sigma_i$, and $\tau_{-i}=\tau_b\tau_a\tau_s/\tau_i$. The relative weighted loss due to central bank's suboptimal action in scenario i under the prior uncertainty can be denoted by x_i. In specific, x_i measures the probability-weighted loss due to the monetary authority's prior uncertainty about scenario i fundamental z_i, τ_i, relative to the multiplicative weighted losses driven by the prior uncertainties associated with other two scenarios, τ_{-i}. Cutoffs of x_i help determine the optimal attention allocated to scenario i that reduces the prior uncertainty Σ_i about z_i.

Intuitively, if the absolute weighted loss of prior uncertainty is equalized across scenarios such that $\tau_b = \tau_a = \tau_s$, a central bank will thus equally allocate its information processing capacity as $\kappa_b = \kappa_s = \kappa_a = \kappa/3$ Importantly, we see that τ_i increases in four variables including the probability of scenario occurrence, p_i, the marginal payoff loss per uncertainty changes γ_i, the optimal magnitude of central bank's policy responses to fundamental, q_i, and the prior uncertainty, Σ_i. Therefore, rises in any of these scenario-specific parameters would require the central bank to care more about that scenario with larger attention allocated κ_i relative to other scenarios. For example, the damage of the economic disasters is immensely high with big γ_s that necessitates the central bank's greater attention and better preparedness, though this has to be traded off against the fact that the relative likelihood of being in a severe situation p_s is relatively lower. Note that when a scenario is peculiar with extremes of any of these four variables, this model allows for corner solutions to exist such that κ_i may be zero, a complete ignorance of scenario i, or κ, a complete exhaustion of the processing capacity.

To map the complexity of theory to our empirical application, we consider the case when the central bank cares about all three scenarios every period such that the corner solutions are ruled out. Hence, the optimization naturally yields the following condition that with optimal learning, the weighted loss due to the central bank's suboptimal action in scenario i under the posterior uncertainty should be equalized across scenarios:

$$p_b \gamma_b q_b \Sigma_b 2^{-2\kappa_b} = p_a \gamma_a q_a \Sigma_a 2^{-2\kappa_a} = p_s \gamma_s q_s \Sigma_s 2^{-2\kappa_s}, \tag{23}$$

where κ_i is characterized by the interior solutions according to Eq. (22).

In practice, the FRB closely monitors and supervises the economy, reacts decisively, and makes the best efforts to forecast the fundamentals to infer all possible economic scenarios for policy preparations. However, it's less likely that the FRB has all the perfect and sufficient information available whenever it comes to make a prompt and right decision in response to scenario shifts.[4] Therefore, we impose the idea that conditional on imperfect information with the optimized attention, a good loss measure of the FRB's

[4]For example, at the peak of the 07-09 financial crisis, we see ferocious debates regarding the merits of unconventional monetary policy practices such as what institutions to bail out, and how large the liquidity is really needed for saving the market from turmoil. For example, Eric Rosengren, then president of Fed Boston, disagreed with the Treasury Secretary Henry Paulson and FRB chair Ben Bernanke on whether Lehman Brothers should fall. More recently, we see again large disagreements on how quick the zero lower bound should be lifted. See general discussions in Summers (2015)

supervision efforts can be approximately mapped to the optimized central bank's utility loss given by $p_b \gamma_b q_b \Sigma_b 2^{-\kappa_b} + p_a \gamma_a q_a \Sigma_a 2^{-2\kappa_a} + p_s \gamma_s q_s \Sigma_s 2^{-\kappa_s}$.

Ideally, we would like to use a squared measure of forecast error to approximate the posterior uncertainty $\Sigma_i 2^{-\kappa_i}$. Nonetheless, empirically, the MAFE, a square root-based measure of forecast error has been tested to perform well when it comes to the measurement of statistical loss. Note that by the optimality condition of Eq. (23), the square root of weighted loss across scenarios should also be equalized. Therefore, subject to such equalization constraint, we are able to construct the following empirical weights as

$$\omega_i = \frac{\mathrm{MAFE}_{-i}}{\mathrm{MAFE}_b \cdot \mathrm{MAFE}_a + \mathrm{MAFE}_b \cdot \mathrm{MAFE}_s + \mathrm{MAFE}_a \cdot \mathrm{MAFE}_s}, \tag{24}$$

where $\mathrm{MAFE}_{-i} = \frac{\mathrm{MAFE}_b \mathrm{MAFE}_a \mathrm{MAFE}_s}{\mathrm{MAFE}_i}$. It shows that when a scenario is forecasted with greater precision of lower MAFE_i, its weight is set to be relatively larger as implied by the optimality condition. Hence, the empirical loss measure can be expressed as below

$$\mathrm{Loss} = \omega_b \cdot \mathrm{MAFE}_b + \omega_a \cdot \mathrm{MAFE}_a + \omega_s \cdot \mathrm{MAFE}_s. \tag{25}$$

This loss function is then used to evaluate the performance of back testing.

4.2. Historical Disaster Events: Calibration

Alternatively, in order to gauge the weights for different scenarios, we could also take a model-free approach by taking in the data moments that firstly approximate the likelihood of each scenario occurrences. Intuitively, having the right weight for the severely adverse scenario MAFE in our backtesting exercise is particularly important because from a central bank's perspective, experiencing an extreme economic disaster can be extremely painful for the economy. Therefore, we resort to the literature of disaster risks, which shows that these rarely seen severe episodes have very rich implications for asset prices and macro-financial stability of the economy along with other dimensions (Barro, 2006; Barro and Ursua, 2008; 2012).

In line with the literature, we stick to the definition of a disaster as an event of more than 15 % real per capita GDP drop for a window period from peak to trough (Barro, 2006). Although we do not have an exact definition of how severe is "severe," Barro and Ursua (2010) finds that economic episodes such as the Great Depression and the most recent global financial crisis can be considered as the windows for rare disaster events. Hence, by our selection of the third window period, we can take the data moments regarding rare disasters to proxy for the weights associated with our severe scenario.

To approach this, we note that the data moments regarding disasters are based on estimations of more than one hundred years of historical data for more than twenty countries. Therefore, a critical assumption is taken as given that the economic system is robust over time and across countries. Then we first take the standard estimates of the probability of disaster occurrence in a year 0.017 to imply for a quarterly probability of $p_s = 0.0043$, a verysmall number.[5] In addition, we further weight the forecast errors in each scenario with the relative size of contraction in the real per capita GDP as a measure of "severity" of the scenario occurrence. Barro (2006) gives the estimates that on average, economic disasters yield an annualized 30% per capita GDP loss. For quarterly number, we convert the size to 7.5%.

However, there is little guidance from this literature for the probability and the size of damage under other scenarios. To make the exercise feasible, we set an average size of real GDP per capita annual loss of 10% to capture those temporary busts. By fitting a hazard function $p_s = \exp(-\lambda \cdot 0.075) = 0.0043$ to get the hazard rate of λ, and using a quarterly contraction rate of 2.5% for a regular bust period, we get an estimate of $p_a = 0.162$. Hence we have the fitted objective probability for normal times $p_b = 1 - p_a - p_s = 0.833$. We also define a "damage index" and normalizeits value to one in the baseline case, $\gamma_b = 1$. For the US, the baseline scenario corresponds to roughly a 0.5% growth in quarterly real GDP. It then follows that we can take $\gamma_a = 1 + \frac{(0.5+2.5)}{0.5} = 7$ and $\gamma_s = 1 + \frac{0.5+7.5}{0.5} = 17$. The following equation thus defines the final set of scenario weights:

$$\omega_i = \frac{p_i \gamma_i}{v}, \tag{26}$$

where the denominator is given by $v = \sum_{i \in \{b,a,s\}} p_i \gamma_i$. Table 2 summarizes the weights.

TABLE 2.

Scenario weights based on historical calibration (quarterly).

	Baseline	Adverse	Severe
ω_i	0.408	0.556	0.036

5. EMPIRICAL ANALYSIS

Not all of the sixteen variables that characterize the three supervisory scenarios (Section II) can be used as predictors at the same time, and a

[5] $p_s = 1 - (1 - 0.017)^{1/4}$

"kitchen sink" regression that pools all regressors together produces bad out-of-sample results. Another issue is the near multicolinearity (correlation very close to unity) which is found in mortgage rate/BBB yield, 3-month Treasury/prime rate, GDP/disposable income, and 5-year/10-year Treasury yield. Experiments with different sets of variables show that many variables are redundant and caution must be exercised in judging the merit of a model based on the in-sample result: occasionally the model that gives the best in-sample fit is dwarfed by an alternative specification in terms of the out-of-sample performance. Overall, the loss in in-sample fit is relatively small compared to the gain in out-of-sample forecast accuracy. Since the ultimate goal of a stress test is to determine the outcome of an external shock under forward-looking scenarios, we recommend using out-of-sample performance, say MAFE, as the primary model selection criterion. For instance, the choice between real and nominal GDP or interest rate and term spread is pragmatic and depends on which M2 component is being studied.

As mentioned in the introduction, model stability is crucial to any scientific inquiry. In this respect, the strength of the cointegrating relationship (Eq. (6)) has varied over time. From 1959 until 1989, the theory yields a stable and strong result; in the early 90s, large forecast errors began to emerge as many depository institutions became increasingly capital-constrained due to shocks to the thrift industry (Lown et al., 1999). Consequently, the secular breakdown in the long-run relationship over this period prompted the FRB to downgrade M2 as a reliable indicator for monetary policy (e.g., Estrella and Mishkin, 1998). Nevertheless, starting in 1994 and followed by a decade-long deregulation in the banking sector, the relationship has resumed itself (e.g., FRB, 1998). To avoid disruptions to the model, the sample ranges from 1995Q1 for the two largest components, NM and NI and starts from 1990Q1 for CT, RF and ST.

The baseline scenario is an example of stable unemployment rate ($\sim$ 5.3%), stable CPI inflation rate ($\sim$ 2.3%) and slowly increasing interest rates; the adverse scenario sees unemployment gradually picking up and inflation rate stabilizing around 4% with interest rates rising faster than in the baseline case; the severe scenario registers high unemployment rate ($\sim$ 10%) and declining inflation rate ($\sim$ 1.6%) with extremely low interest rates (3-month Treasury at 0.1%). In the words of New Keynesian economics, the baseline reflects a mildly strong aggregate demand, the adverse is not unlike a stagflation, and the severe is what ensues after another recession hits. A moment of thought suggests a striking resemblance of the supervisory scenarios to three historical episodes as highlighted by the arrows in Figure 1: the baseline scenario vs. 2004Q1-2006Q4, the adverse scenario vs. 2001Q1-2003Q4, and the severely adverse scenario vs. 2008Q1-2010Q4. These periods are later used to backtest the stress test models.

The principle measures of forecast performance are the MAFE and MAFPE for each of the above test periods. A problem with the single-period MAFE approach is that the rankings of models based under, say, the baseline scenario are different from those under the adverse scenario. Aside from the ambiguity issue, another concern is that different scenarios occur with different probabilities and will cause damage to the economy in various degrees. A severely adverse scenario may occur with very low probability but the preponderance of its negative impact weighs heavily on the policy makers' loss function. Put together, these observations call for a compound diagnostic tool to guide empirical analysis. One simple measure is a weighted average of the MAFEs. The weights sum to one and can be viewed as loss-normalized probabilities: they are meant to reflect a combination of analytical results and shrewd judgment of policy makers. We will entertain the set of scenario weights discussed in Section IV for our empirical analysis.

We first estimate the model using the sample up to the quarter right before each of the above three test periods, then we compute the out-of-sample MAFE with respect to the true values for each period, after which a weighted average measure is constructed by putting different weights on the three MAFEs. In running the regressions, variables are selected within the given model and those that are insignificant are omitted. For illustrative purposes, we follow the historical calibration approach discussed in Section 4.2 and pick the weights according to Table 2. In rare cases, it is also possible that the policy maker is interested in certain scenario that is so unique as to have no comparable predecessors. When this happens, the backtesting procedure proposed in this paper should be combined with other means to judge the reliability of the model. Our results are robust to the use of weights determined by the Rational Inattention theory.

Eventually our loss function becomes

$$\text{Loss} = 0.408\text{MAFE}_{base} + 0.556\text{MAFE}_{adverse} + 0.036\text{MAFE}_{severe}. \quad (27)$$

The weighting scheme captures the expected loss due to a deviation of the model's forecast from the real world. We will use Eq. (27) throughout the analysis.

5.1. Backtesting NM, NI and CT

For NM, the model selected by forecast performance is

$$m_{1t} = \mu + \theta_1 y_t + \theta_2 r_t + \beta_1 \Delta r_t + \text{lags of } \Delta r_t + \varepsilon_t, \quad (28)$$

where y_t is log-nominal GDP and r_t is 3-month Treasury bill rate. Eq. (28) is not the model that gives the best in-sample fit: adding a time trend improves the fit on the margin but the coefficient is very small and one has to sacrifice the forecast accuracy by a full degree of magnitude. This logic applies to the rest of the paper. The last column of Table 3 reports the full sample (1995Q1-2014Q3) regression result; the second last row reports the p-valuesof ADF tests for integrated residuals by which we want to check if there are any unstationarity spillovers not captured by the model.

TABLE 3.

Backtesting DOLS model for non-maturity interest bearing deposits.

	Quarterly (1995q1-2014q3)									
	base (2004Q1-2006Q4)			adverse (2001Q1-2003Q4)			severe (2008Q1-2010Q4)			full
	MAFE	MAFPE	R^2-adj.	MAFE	MAFPE	R^2-adj.	MAFE	MAFPE	R^2-adj.	R^2-adj.
	0.056	0.70%	0.992	0.047	0.62%	0.989	0.043	0.52%	0.987	0.994
T	36			24			52			79
ADF	0.001			0.000			0.002			0.001
Loss				0.051						

Note: The model, $m_{1t} = \mu + \theta_1 y_t + \theta_2 r_t + \beta_1 \Delta r_t$ + lags of$\Delta r_t + \varepsilon_t$, is estimated using the sample up to the point right before each of the three test periods, then the out-of-sample MAFE and MAFPE are computed with respect to the true values for each period. A weighted average measure is constructed afterward by putting different weights (0.408, 0.556 and 0.036) on the MAFEs. Loss $= 0.408\text{MAFE}_{base} + 0.556\text{MAFE}_{adverse} + 0.036\text{MAFE}_{severe}$.

For non-interest bearing deposits, the model is

$$m_{2t} = \mu+\theta_1 p_t+\theta_2 y_t+\theta_3 r_t+\beta_1 \Delta p_t+\beta_2 \Delta y_t+\beta_3 \Delta r_t+\text{lags of}\Delta r_t \text{ and } \Delta y_t+\varepsilon_t, \tag{29}$$

where y_t is log-real GDP, r_t is 3-month Treasury bill rate, and p_t is the GDP deflator defined earlier. Results are collected in Table 4.

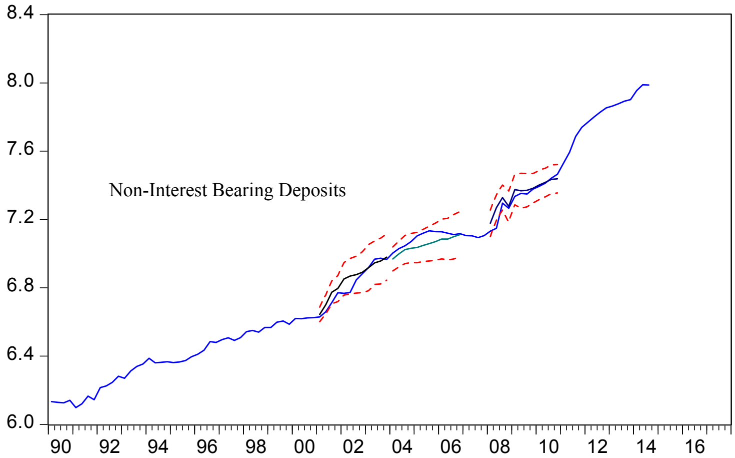
TABLE 4.

Backtesting DOLS model for non-interest bearing deposits.

	Quarterly (1995Q1-2014Q3)									
	base (2004Q1-2006Q4)			adverse (2001Q1-2003Q4)			severe (2008Q1-2010Q4)			full
	MAFE	MAFPE	R^2-adj.	MAFE	MAFPE	R^2-adj.	MAFE	MAFPE	R^2-adj.	R^2-adj.
	0.041	0.58%	0.978	0.034	0.51%	0.974	0.028	0.38%	0.989	0.971
T	36			24			52			79
ADF	0.000			0.000			0.000			0.032
Loss				0.037						

Note: The model, $m_{2t} = \mu+\theta_1 p_t+\theta_2 y_t+\theta_3 r_t+\beta_1 \Delta p_t+\beta_2 \Delta y_t+\beta_3 \Delta r_t$+lags of$\Delta r_t$ and $\Delta y_t+\varepsilon_t$, is estimated using the sample up to the point right before each of the three test periods, then the out-of-sample MAFE and MAFPE are computed with respect to the true values for each period. A weighted average measure is constructed afterward by putting different weights (0.408, 0.556 and 0.036) on the MAFEs. Loss $= 0.408\text{MAFE}_{base} + 0.556\text{MAFE}_{adverse} + 0.036\text{MAFE}_{severe}$.

FIG. 7. NI forecast and confidence bands in test periods.



For currency and travelers checks, the model is

$$m_{3t} - p_t = \mu + \theta_1 y_t + \theta_2 r_t + \beta_1 \Delta y_t + \beta_2 \Delta r_t + \beta_3 \Delta r_{t-1} + \varepsilon_t, \qquad (30)$$

where the variables are the same as in Eq. (29). This takes the real balance form of Eq. (7) because the estimated coefficient on p_t is not significantly different from one. In Table 5, the forecast is made for CT only.

TABLE 5.

Backtesting DOLS model for currency and travelers checks.

	Quarterly (1990Q1-2014Q3)									
	base (2004Q1-2006Q4)			adverse (2001Q1-2003Q4)			severe (2008Q1-2010Q4)			full
	MAFE	MAFPE	R^2-adj.	MAFE	MAFPE	R^2-adj.	MAFE	MAFPE	R^2-adj.	R^2-adj.
	0.062	0.95%	0.992	0.039	0.61%	0.986	0.032	0.47%	0.987	0.983
T		54			42			70		97
ADF		0.003			0.001			0.067		0.056
Loss					0.048					

Note: The model, $m_{3t} - p_t = \mu + \theta_1 y_t + \theta_2 r_t + \beta_1 \Delta y_t + \beta_2 \Delta r_t + \beta_3 \Delta r_{t-1} + \varepsilon_t$, is estimated using the sample up to the point right before each of the three test periods, then the out-of-sample MAFE and MAFPE are computed with respect to the true values for each period. A weighted average measure is constructed afterward by putting different weights (0.408, 0.556 and 0.036) on the MAFEs. Loss $= 0.408\text{MAFE}_{base} + 0.556\text{MAFE}_{adverse} + 0.036\text{MAFE}_{severe}$.

The projection patterns for NM, NI and CT are similar, therefore we only demonstrate the superior forecasting ability of the NI model in Figure 7. From Table 3-5, the loss functions are mostly below 0.05 and the adjusted R^2 is high in all cases. Alternative cointegration tests can be conducted, but since the focus of the paper is not to find the best cointegration test, we do not pursue this issue further. The ADF test firmly supports the existence of a cointegrating relationship in most test periods with just one marginal case being CT from 2008Q1 to 2010Q4. Nevertheless, given the turbulent economic conditions during the financial crisis, a p-value of 0.067 should be taken with moderation.

5.2. Backtesting RF and ST growth rates

Retail money funds and small time deposits are modeled in terms of growth rates. For RF, the model is

$$\Delta m_{4t} = \theta_1 r_{1t} + \theta_2 r_{2t} + \beta_1 \Delta r_{1t} + \beta_2 \Delta r_{1t+1} + \beta_3 \Delta r_{1t-1} + \varepsilon_t, \qquad (31)$$

where r_{1t} is the 3-month Treasury bill rate and r_{2t} is the 5-year Treasury yield. The intercept term, being very insignificant in all regressions, detracts from forecast accuracy and hence is omitted. In Table 6 and 7. the MAFE and the loss function should be interpreted in percentage terms.

TABLE 6.

Backtesting DOLS model for retail money funds growth rate.

	Quarterly (1990Q1-2014Q3)						
	base (2004Q1-2006Q4)		adverse (2001Q1-2003Q4)		severe (2008Q1-2010Q4)		Full
	MAFE	R^2-adj.	MAFE	R^2-adj.	MAFE	R^2-adj.	R^2-adj.
	1.70	0.719	1.09	0.594	2.47	0.706	0.682
T	54		42		70		97
ADF	0.000		0.000		0.000		0.000
Loss			1.39				

Note: The model, $\Delta m_{4t} = \theta_1 r_{1t} + \theta_2 r_{2t} + \beta_1 \Delta r_{1t} + \beta_2 \Delta r_{1t+1} + \beta_3 \Delta r_{1t-1} + \varepsilon_t$, is estimated using the sample up to the point right before each of the three test periods, then the out-of-sample MAFEs are computed with respect to the true values for each period. A weighted average measure is constructed afterward by putting different weights (0.408, 0.556 and 0.036) on the MAFEs. Loss $= 0.408\text{MAFE}_{base} + 0.556\text{MAFE}_{adverse} + 0.036\text{MAFE}_{severe}$.

In Figure 4, it can be seen that small time deposits (ST) have more or less followed a pattern similar to that of RF after 1995. For a detailed discussion of the relationship between small time deposits and M2, see Wenninger and Partlan (1992). For ST, the selected model is

$$\Delta m_{5t} = \mu + \theta_1 \Delta y_t + \theta_2 Spr_t + \beta_1 \Delta y_{t+1} + \beta_2 \Delta y_{t-1} + \beta_3 \Delta Spr_t + \text{leads of } \Delta Spr_t + \varepsilon_t, \qquad (32)$$

where y_t is log-real GDP and Spr is the term spread defined in Section 3.

Unlike NM, NI and CT for which forecast errors are about evenly distributed across the three test periods (Table 3-5), those for RF and ST

FIG. 8. RF growth rate forecast and confidence bands in the test periods.

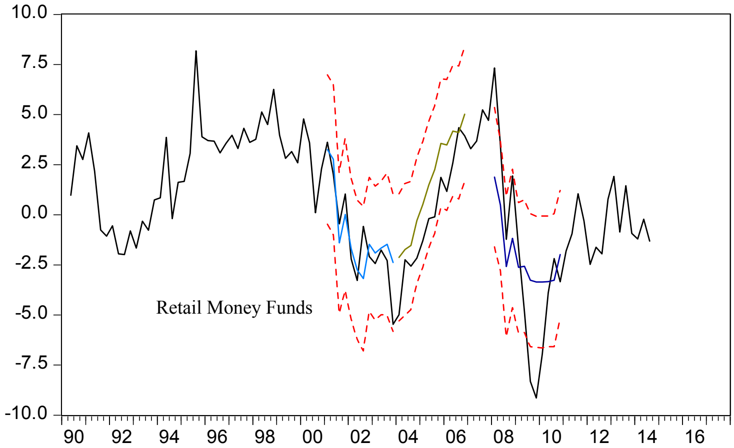


TABLE 7.

Backtesting DOLS model for small time deposits growth rate.

	Quarterly (1990Q1-2014Q3)						
	base (2004Q1-2006Q4)		adverse (2001Q1-2003Q4)		severe (2008Q1-2010Q4)		full
	MAFE	R^2-adj.	MAFE	R^2-adj.	MAFE	R^2-adj.	R^2-adj.
	1.13	0.710	1.08	0.729	2.96	0.799	0.536
T	54		42		70		98
ADF	0.014		0.017		0.002		0.004
Loss			1.17				

Note: The model, $\Delta m_{5t} = \mu + \theta_1 \Delta y_t + \theta_2 Spr_t + \beta_1 \Delta y_{t+1} + \beta_2 \Delta y_{t-1} + \beta_3 \Delta Spr_t +$leads of $\Delta Spr_t + \varepsilon_t$, is estimated using the sample up to the point right before each of the three test periods, then the out-of-sample MAFEs are computed with respect to the true values for each period. A weighted average measure is constructed afterward by putting different weights (0.408, 0.556 and 0.036) on the MAFEs. Loss $= 0.408\text{MAFE}_{base} + 0.556\text{MAFE}_{adverse} + 0.036\text{MAFE}_{severe}$.

growth rates are noticeably larger over the recent crisis period. The patterns for RF and ST are quite similar so we only report the forecasts for RF growth rate in Figure 8. A closer look at the data suggests that the deviation is mostly due to the spike in the middle of 2008 followed by a sharp drop in the next year. Interestingly, the drop coincides with the aftermaths of the fall of Lehman Brothers in September 2008 and the eventual recovery in late 2009 is met with the FED's first round of quantitative easing. This is consistent with findings by Baba et al. (2009) who show that money market funds, being the largest suppliers of dollar funding to non-US banks, were subject to runs following the Lehman Brothers failure; see also Gorton (2009). This effectively let off the steam of RF and ST growth which resumed only when the FED's policy stopped the run and replaced private with public funding.

5.3. FRB Supervisory Scenarios

We now apply the previous models to the three FRB test scenarios starting from 2014Q4 up until 2017Q4. To save space, we only report projections of M2 components under the severely adverse scenario and provide confidence bands thereof. The results are shown graphically in Figure 9.

FIG. 9. M2 forward guidance under the FRB severely adverse scenario. NM, NI and CT are in logarithms; ST and RF are growth rates.

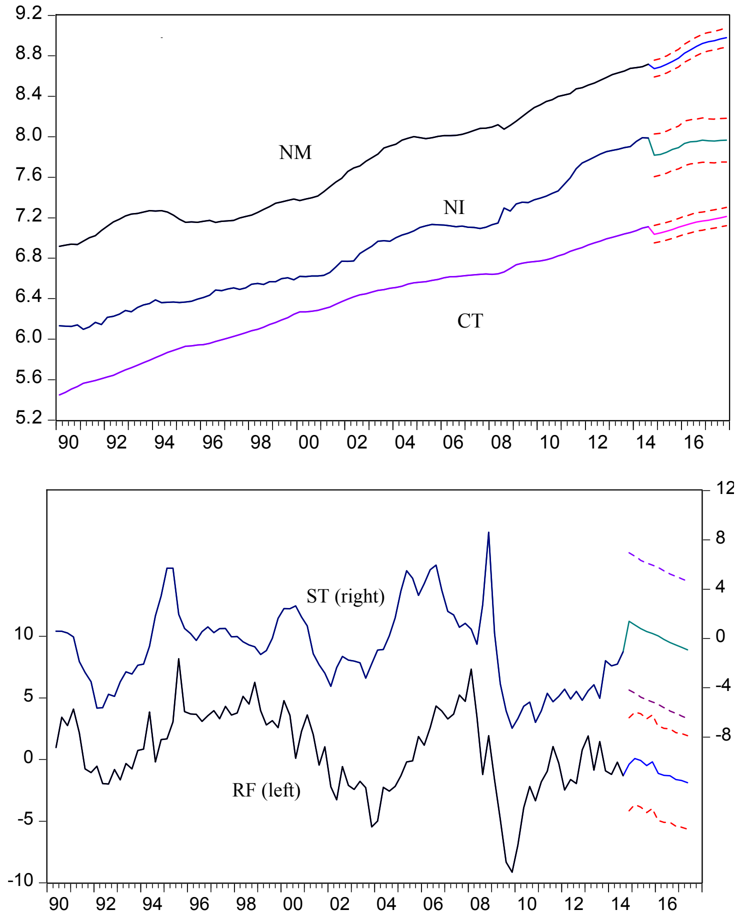
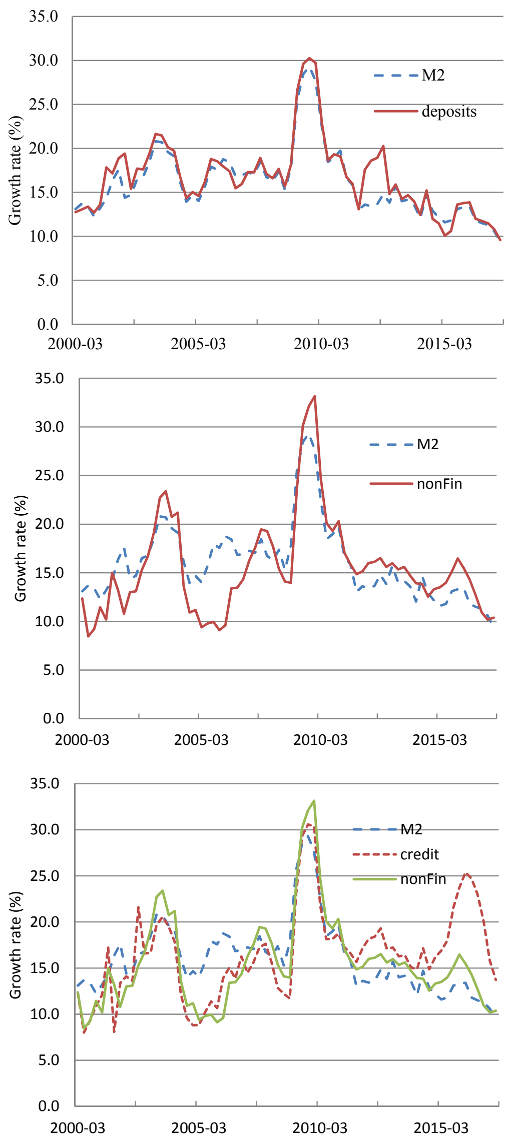


FIG. 10. Growth rates (year over year) of M2, total bank credit, bank loans to non-financial sector (non−financial institutions and households) and total RMB deposits (current account + savings account + time deposits). Source: PBC.



6. M2 IN CHINA

In this section, we take an excursion to draw implications of our backtesting methodology for China's monetary policy and highlight the critical role M2 plays in the Chinese economy. The People's Bank of China (PBC) has not officially published any stress scenarios, but the logic presented in the above analysis can be applied to China as well. In fact, the backtesting approach is potentially more valuable to policy makers in China than in the US. This is because M2 is found to be the most important policy target on the watch list of China's monetary authorities (Chen et al., 2017). The main objective of the monetary policy making in China is to accommodate government stimulus plans and smooth out growth shortfalls. The PBC achieves this by maintaining a firm control of bank credits and total deposits as illustrated by Fig. 10.

Figure 10 plots the growth rates of M2, total bank credit (credit), bank loans to non-financial sector (nonFin) and total RMB deposits on a quarterly basis. Banks are broadly defined as depository institutions and the non-financial sector includes non-financial institutions and households. Total deposits are the sum of current account, savings account and time deposits. While the M2 series closely tracks total deposits, there is some discrepancy between nonFin and M2 in the first half of the sample. For total bank credit, it moves in tandem with nonFin and there is a short-lived swing around 2016. Overall, these patterns uncover a tightly knit relationship between M2 and aggregate financing to the economy.

7. CONCLUSIONS

In this paper, we have demonstrated the use of a weighted average of scenario-specific MAFEs to evaluate the reliability of stress test models in forecasting M2 components. Each FRB supervisory scenario is matched with a historical episode that bears similar patterns of the defining economic variables. This new method can be viewed as a proxy for the regulators' loss functions and is one way to incorporate information about not only the likelihood of occurrence of certain scenario but also the severity of damage done to the economy. We provide a Rational Inattention theory about central bank's supervision of the economy and optimize the scenario weights that eventually pin down a loss measure and calibrate the weights using historical data. The PBC can come up with stress scenarios tailored to the Chinese economy to which the method of this paper can be applied.

Meanwhile, the backtesting procedure highlights a fundamental limitation of the current stress testing mind-set, i.e., a lack of readily implementable repeated experiments hence no probabilistic interpretation of the results. In fact, stress tests generally report no confidence intervals but only point estimates. Another limitation of stress testing is that backtesting is generally impossible if one is interested in shocks that have never occurred. As the simulated or conjectured stress scenarios get more and more disconnected from the past, the policy maker trades off heightened model risk for a higher degree of independence in the test scenarios. Instead of tethering stress scenarios to a path of rigid numbers, it is possible to use a projected band with lower and upper bounds. Or better still, one can think of ways to simulate the paths of key economic variables from some conditional distributions so that a Monte Carlo exercise becomes feasible. As stress testing is an invaluable tool for crisis management, efforts in this direction are worthwhile.

APPENDIX A

A.1. DERIVING CENTRAL BANK'S OPTIMAL ACTION

Proof. We approximate the central bank's payoff function in regime i up to a second order around a triplets $(0, 0, 0)$ such that

$$\begin{aligned} U^i(a_i^{CB}, a_i, z_i) &= U_0^i + U_1^i a_i^{CB} + U_2^i a_i + U_3^i z_i + \frac{U_{11}^i}{2}(a_i^{CB})^2 \\ &+ \frac{U_{22}^i}{2}a_i^2 + \frac{U_{33}^i}{2}z_i^2 + U_{12}^i a_i^{CB} a_i + U_{13}^i a_i^{CB} z_i + U_{23}^i a_i z_i \end{aligned}$$

First order condition over a_i^{CB} gives

$$0 = U_{11}^i + U_{11}^i a_i^{CB} + U_{12}^i a_i + U_{13}^i z_i$$

The equality follows from the fact that third order partials and cross partials are zero due to the assumption of a quadratic payoff function. It then gives

$$a_i^{CB,*} = -\frac{U_1^i}{U_{11}^i} - \frac{U_{12}^i}{U_{11}^i}a_i - \frac{U_{13}^i}{U_{11}^i}z_i$$

Evaluating U_1^i at triplet zeros gives a constant number and it yields that

$$U_{11}^i + U_{12}^i + U_{13}^i = 0$$

such that

$$a_i^{CB,*} = \alpha_i + \phi_i a_i + (1 - \phi_i) z_i$$

where $\phi_i = -\frac{U_{12}^i}{U_{11}^i}$ and $\alpha_i = -\frac{U_1^i}{U_{11}^i}$. Also, by $U_{11}^i < 0$, it gives $\alpha_i > 0$ ∎

A.2. DERIVING OPTIMAL ATTENTION ALLOCATIONS

Proof. For interior solutions, we set up a Lagrangian such that

$$\max -p_b\gamma_b q_b\Sigma_b 2^{-2\kappa_b} - p_a\gamma_a q_a\Sigma_a 2^{-2\kappa_a} - p_s\gamma_s q_s\Sigma_s 2^{-2\kappa_s} + \lambda(\kappa - \kappa_b - \kappa_a - \kappa_s)$$

FOCs are given by

$$p_b\gamma_b q_b\Sigma_b 2^{-2\kappa_b} = p_a\gamma_a q_a\Sigma_a 2^{-2\kappa_a} = p_s\gamma_s q_s\Sigma_s 2^{-2\kappa_s}$$

which gives

$$\begin{aligned} \log_2\left[\sqrt{\frac{p_b\gamma_b q_b\Sigma_b}{p_a\gamma_a q_a\Sigma_a}}\right] &= \kappa_b - \kappa_a \\ \log_2\left[\sqrt{\frac{p_b\gamma_b q_b\Sigma_b}{p_s\gamma_s q_s\Sigma_s}}\right] &= \kappa_b - \kappa_s \end{aligned}$$

Therefore

$$\begin{aligned} \kappa_b &= \frac{1}{3}\left[\kappa + \log_2\left[\frac{p_b\gamma_b q_b\Sigma_b}{\sqrt{p_a\gamma_a q_a\sigma_a \cdot p_s\gamma_s q_s\Sigma_s}}\right]\right] \\ \kappa_a &= \frac{1}{3}\left[\kappa + \log_2\left[\frac{p_a\gamma_a q_a\Sigma_a}{\sqrt{p_b\gamma_b q_b\Sigma_b \cdot p_s\gamma_s q_s\Sigma_s}}\right]\right] \\ \kappa_s &= \frac{1}{3}\left[\kappa + \log_2\left[\frac{p_s\gamma_s q_s\Sigma_s}{\sqrt{p_b\gamma_b q_b\Sigma_b \cdot p_a\gamma_a q_a\Sigma_a}}\right]\right] \end{aligned}$$

For $\kappa_i \in (0, \kappa)$ as interior solutions, it requires that

$$\begin{aligned} \frac{\tau_b}{\sqrt{\tau_a\tau_s}} &\in (2^{-\kappa}, 2^{2\kappa}) \\ \frac{\tau_a}{\sqrt{\tau_b\tau_s}} &\in (2^{-\kappa}, 2^{2\kappa}) \\ \frac{\tau_s}{\sqrt{\tau_b\tau_a}} &\in (2^{-\kappa}, 2^{2\kappa}) \end{aligned}$$

where $\tau_i = p_i\gamma_i q_i\Sigma_i$. For corner solutions, we have $\kappa_i = 0$ if $\frac{\tau_i\sqrt{\tau_i}}{\sqrt{\prod_{i\in\{b,a,s\}}\tau_i}} \leq 2^{-\kappa}$ and $\kappa_i = 1$ if $\frac{\tau_i\sqrt{\tau_i}}{\sqrt{\prod_{i\in\{b,a,s\}}\tau_i}} \geq 2^{2\kappa}$ ▮

REFERENCES

Alfaro, Rodrigo A. and Mathias Drehmann, 2009. Macro stress tests and crises: What can we learn? *BIS Quarterly Review* **December**, 29-41.

Baba, Naohiko, Robert N. McCauley, and Srichander Ramaswamy, 2009. US dollar money market funds and non-US banks. *BIS Quarterly Review* **March**, 65-81.

Barro, Robert, 2006. Rare disasters and asset markets in the twentieth century. *The Quarterly Journal of Economics* **121(3)**, 823-866.

Barro, Robert and Jose Ursua, 2008. Macroeconomic crises since 1870. *Brookings Papers on Economic Activity* **39(1)**, 255-350.

Barro, Robert and Jose Ursua, 2012. Rare macroeconomic disasters. *Annual Review of Economics* **4(1)**, 83-109.

Basel Committee on Banking Supervision, 2009. Principles for Sound Stress Testing Practices and Supervision.

Bianchi, Francesco and Leonardo Melosi, 2017. Escaping the Great Recession. *American Economic Review* **107(4)**, 1030-58.

Blaschke, W., M. Jones, G. Majnoni,and S. Peria, 2001. Stress testing of financial systems: An overview of issues, methodologies, and FSAP experiences. IMF Working Papers 88.

Board of Governors of the Federal Reserve System, July 1998. Monetary Policy Report to the Congress Pursuant to the Full Employment and Balanced Growth Act of 1978.

Board of Governors of the Federal Reserve System, April 2009. The Supervisory Capital Assessment Program: Design and Implementation.

Board of Governors of the Federal Reserve System, March 2014. Comprehensive Capital Analysis and Review 2014: Assessment Framework and Results.

Board of Governors of the Federal Reserve System, October 2014. 2015 Supervisory Scenarios for Annual Stress Tests Required under the Dodd-Frank Act Stress Testing Rules and the Capital Plan Rule.

Chen, Kaiji, Patrick Higgins, Daniel F. Waggoner, and Tao Zha, 2016. Impacts of monetary stimulus on credit allocation and macroeconomy: Evidence from China. NBER Working Papers 22650.

Danielsson, J., 2008. Blame the models. *Journal of Financial Stability* **4**, 321-328.

Diebold, Francis X. and G. D. Rudebusch, 1990. A Nonparametric investigation of duration dependence in the American business cycle. *Journal of Political Economy* **98(3)**, 596-616.

Drehmann, M., 2009. Macroeconomic stress testing banks: A survey of methodologies. In: Quagliariello, M. (Ed.), Stress Testing the Banking System: Methodologies and Applications. Cambridge University Press, Cambridge.

Estrella, Arturo and Frederic S. Mishkin, 1997. Is there a role for monetary aggregates in the conduct of monetary policy? *Journal of Monetary Economics* **40**, 279-304.

European Banking Authority (EBA), 2011. Overview of the EBA, 2011 Banking EU-wide Stress Test.

Favara, Giovanni and Paolo Giordani, 2009. Reconsidering the role of money for output, prices and interest rates. *Journal of Monetary Economics* **56**, 419-430.

Feldstein, M. and J. H. Stock, 1993. The use of monetary aggregates to target nominal GDP. In: Mankiw, N.G. (Ed.), Monetary Policy. University of Chicago Press.

Friedman, Benjamin M. and Kenneth N. Kuttner, 1992. Money, income, prices, and interest rates. *American Economic Review* **82**, 472-92.

Gorton, G., 2009. Information, liquidity, and the (ongoing) panic of 2007. *American Economic Review* **99**, 567-572.

Lown, Cara S., Stavros Peristiani, and Kenneth J. Robinson, 1999. What was behind the M2 breakdown? FRB of New York Staff Report 83.

Sichel, Daniel E., 1993. Business cycle asymmetry: A deeper look. *Economic Inquiry* **31(2)**, 224-236.

Sims, Christopher A., 2003. Implications of rational inattention. *Journal of Monetary Economics* **50(3)**, 665-690.

Sims, Christopher A. and Tao Zha, 2006. Were there regime switches in U.S. monetary policy? *American Economic Review* **96(1)**, 54-81.

Stock, James H. and Mark W. Watson, 1993. A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica* **61**, 783-820.

Summers, Larry, 2015. What the Federal Reserve got wrong, and what it should do next. Wonkblog, Washington Post, December 15.

Wenninger, John and John Partlan, 1992. Small time deposits and the recent weakness in M2. *Federal Reserve Bank of New York Quarterly Review* **17**, 21-35.

Woodford, Michael, 2003. Interest and Prices. Princeton University Press, Princeton.