

FEDERAL RESERVE BANK OF SAN FRANCISCO

WORKING PAPER SERIES

Stress Testing with Misspecified Models

Rhys Bidder

Federal Reserve Bank of San Francisco

Raffaella Giacomini

University College London

Andrew McKenna

Federal Reserve Bank of New York

September 2016

Working Paper 2016-26

<http://www.frbsf.org/economic-research/publications/working-papers/wp2016-26.pdf>

Suggested citation:

Bidder, Rhys, Raffaella Giacomini, Andrew McKenna. 2016. “Stress Testing with Misspecified Models” Federal Reserve Bank of San Francisco Working Paper 2016-26.

<http://www.frbsf.org/economic-research/publications/working-papers/wp2016-26.pdf>

The views in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Federal Reserve Bank of San Francisco or the Board of Governors of the Federal Reserve System.

Stress testing with misspecified models[☆]

R. M. Bidder^{a,*}, R. Giacomini^b, A. McKenna^c

^a*Federal Reserve Bank of San Francisco*

^b*University College London*

^c*Federal Reserve Bank of New York*

Abstract

Stress testing has become an important component of macroprudential regulation yet its goals and implementation are still being debated, reflecting the difficulty of designing such frameworks in the context of enormous model uncertainty. We illustrate methods for responding to possible misspecifications in models used for assessing bank vulnerabilities. We show how ‘exponential tilting’ allows the incorporation of external judgment, captured in moment conditions, into a forecasting model as a partial correction for misspecification. We also make use of methods from robust control to seek the most relevant dimensions in which a regulator’s forecasting model might be misspecified - a search for a ‘worst case’ model that is a ‘twisted’ version of the regulator’s initial forecasting model. Finally, we show how the two approaches can be blended so that one can search for a worst case model subject to restrictions on its properties, informed by the regulator’s judgment. We demonstrate the methods using the New York Fed’s CLASS model, a top-down capital stress testing framework that projects the effect of macroeconomic scenarios on U.S. banking firms.

[☆]The authors thank Fred Furlong, Galina Hale, Beverly Hirtle, Oscar Jorda, Anna Kovner, Jose Lopez, Ulysses Velasquez and James Vickery for useful comments and advice. Remaining errors are our own. Disclaimer: The views expressed in this document and all errors and omissions should be regarded as those of the authors and not necessarily those of the Federal Reserve Bank of San Francisco, the Federal Reserve Bank of New York, the Federal Reserve Board of Governors or the Federal Reserve System.

*Corresponding author

Email addresses: rhys.bidder@sf.frb.org (R. M. Bidder), r.giacomini@ucl.ac.uk (R. Giacomini), andrew.mckenna@frb.gov (A. McKenna)

1. Introduction

In this paper, we contribute to the debate over the design and aims of stress testing by setting out an approach based on the methods of robust forecasting analysis (Hansen and Sargent (2008)) and on the tools of exponential tilting (Robertson, Tallman, and Whiteman (2005), Cogley, Morozov, and Sargent (2005) and Giacomini and Ragusa (2013)).

The approach takes as a starting point a particular model for risks to a portfolio or an institution or an entire financial system. This model, referred to as the ‘benchmark’, is known to be only an approximation to the true reality - as all models are. The question is, how should regulators use it to construct stress tests. One approach, nested in our techniques, is conditional forecasting. Commonly, this amounts to identifying tail events and envisaging bank performance in these extreme scenarios. The question arises, though, whether one trusts the model enough to believe that one has truly identified a tail event or has confidence in its predictions for the economy’s behavior conditional on that event.

Almost by definition, financial crises are rare and poorly understood so that standard models estimated using data from normal times or a handful of stressed episodes are likely to be heavily misspecified precisely in the dimensions most relevant for stress testing (Frame, Gerardi, and Willen (2013), Guerrieri and Welch (2012), Covas, Rump, and Zakrajsek (2013)). The fact that forecast disagreement and divergence in model predictions increases in stressed times also emphasizes the particular relevance of misspecification (Danielsson, James, Valenzuela, and Zer (2016)). The difficulties faced in designing forecasting models for stress testing are such that it has been suggested that they may be best suited to crisis resolution, rather than ‘early warning’ (Borio, Drehmann, and Tsatsaronis (2013)). Nevertheless, in practice there is a strong forward-looking element to stress testing programs (see Lehnert and Hirtle (2015), for example) and it is therefore relevant to consider how best to use forecasting models in support of these programs.

In the context of model misspecification one might not wish to put faith entirely in forecasts based on a particular conditioning scenario.¹ In fact one might only wish to impose a looser form of conditioning.² If one thinks of a traditional conditional forecast as a restriction on

¹Breuer, Jandacka, Rheinberger, and Summer (2008) refer to stress scenarios as ‘partial’ when some risk factors are held fixed and argue they should be set equal to their conditional expectations.

²As Pritsker (2012), there are also concerns that if one concentrates on only a handful of selected scenarios, important and still plausible vulnerabilities can be obscured. Christensen, Lopez, and Rudebusch (2015)

the pointwise mean of a variable on whose path one conditions *combined with* a pointwise restriction on the variance (to be zero) then a natural loosening is to impose a mean path, but allow for a non-degenerate distribution around it. This sort of restriction may capture the fact that regulators have extra-model information or ‘judgment’ with which they wish to adjust the benchmark but (reflecting their modesty about their understanding of the economy) they may only want to incorporate in an unassertive manner.

Exponential tilting provides a formal approach for incorporating judgment in this way, where judgment is characterized by a limited set of moment conditions. Exponential tilting takes the probability distribution implied by the benchmark and adjusts it to impose certain moment restrictions that the benchmark does not initially satisfy. Importantly, it does this in a parsimonious way by minimizing the entropic distance between the benchmark distribution and the new ‘tilted’ distribution that satisfies the restrictions. Thus, formally, the method ensures that the least information possible is introduced to the new model subject to ensuring the moment restrictions are satisfied. The term ‘exponential’ reflects the fact that (because of the choice of relative entropy as a distance measure) the tilt is characterized by an exponential function of the moment conditions, appropriately weighted.

In the robustness approach of Hansen and Sargent (2008), one also begins with a benchmark model that is thought to be misspecified in some unknown way. In this case, however, one explicitly seeks vulnerability to misspecifications in the benchmark that would be particularly damaging. The regulator’s desire to explore vulnerability to model misspecification, together with her desire only to consider ‘plausible’ models is captured in an optimization problem where the regulator seeks an alternative distribution that is the most damaging, subject to a requirement that the distribution not be too ‘distant’ from the benchmark. Distance is measured in entropic terms and the distance requirement is meant to capture the plausibility of the alternative distribution - more distant means less plausible. The output of this method is a ‘worst case’ model that is a ‘twisted’ version of the benchmark.

Since the distance measure used in robustness analysis is the entropic distance between the benchmark and an alternative distribution, it turns out that robustness also leads to an exponential tilt. That is, the worst case model’s probability distribution, is an exponentially tilted version of that implied by the benchmark. In this case, the tilt is characterized by

construct a probability-based model to stress test the Fed’s balance sheet, which is an example of an approach to testing less focused on a single scenario or a small set thereof.

an exponential function of the value function of the regulator, appropriately weighted. Thus, there are close similarities between exponential tilting to respect moment conditions, discussed above, and robustness analysis. In fact, the two methods can be blended in an interesting and useful manner.

In the most basic form of robustness analysis, the plausibility of alternative distributions (models) that might concern the regulator is simply defined in terms of entropic distance from the benchmark. Thus, two models that are the same entropic distance from the benchmark are regarded as equally legitimate for consideration by the regulator. However, one might think that there are certain properties, other than entropic distance, that distinguish models by plausibility in the mind of the regulator. That is, she may feel that in any reasonable description of the world, certain statistical regularities should hold, even as she is unsure of the true data generating process more broadly. These restrictions, capturing her judgment and/or extra-model information, will be expressed as moment conditions. These conditions will restrict the set of models she considers so that her ranking of models by plausibility are not simply based on entropic distance.

One can incorporate these moment restrictions into the search for a worst case. This yields an amended exponential tilt that, *ceteris paribus*, adjusts the benchmark in the direction of realizations that imply lower welfare, as in the standard robustness case, but modified by the moment restrictions. Consequently, one nests several useful approaches to exploring the benchmark model - conditional forecasting, the incorporation of extra-model judgment and the search for a worst case distribution. Once one has adjusted the desired forecasting model using some combination of exponential tilting or worst case analysis, one can then, of course, calculate the same sort of objects that are traditionally used in stress test analysis, such as paths based on exploring the tails of the forecast distributions. Alternatively, one can calculate VaR-type statistics under the adjusted probability distribution. Here, however, we concentrate on the adjustment to the distributions themselves, as these are the primitives of such analyses.

We illustrate our methods by taking as a benchmark an augmented version of the New York Fed's 'Capital and Loss Assessment under Stress Scenarios' (CLASS) model, described in Hirtle, Kovner, Vickery, and Bhanot (2015). CLASS is designed to trace through the implications of a given macroeconomic scenario for U.S. banks so, when paired with a simulation model that generates such scenarios, CLASS provides draws from a joint distribution over bank performance and the broader economy. We take these draws as a proposal density for

our various tilts. Our results show that what constitutes the worst case model will differ according to the bank of interest due to variations in banks' exposures to different lines of business, which in turn load differently on macroeconomic conditions. We also show how imposing judgment to align the model's forecasts for the short rate with those obtained from the Survey of Primary Dealers will have different (possibly positive) implications for different banks. Finally, we explore the worst case model where we require it to satisfy restrictions on forecasts for the term structure, showing how this moderates the worst case.

In comparison with existing literature the worst case-type analysis is akin to that discussed in Breuer, Jandacka, Rheinberger, and Summer (2009) and, given our use of entropic measures of plausibility, to Breuer and Csiszr (2013), Glasserman and Xu (2014) and Glasserman and Xu (2013). Breuer and Csiszr (2013), however, concentrates on simplified loan-portfolio/credit-risk applications or employs linear-Gaussian approximations to implement their methods. In our case we work directly with an elaborate and realistic model of bank performance under stress and we eschew Linear-Gaussian approximations, illustrating how flexible the approach can be in working with draws from the benchmark and resampling/reweighting to obtain the worst case distribution. This is important as, in practice, different sections of a regulatory institution may be tasked with model development and simulation, while other sections may only receive simulations as input to their work. In addition, we explicitly blend the methodology with tilting to impose 'judgment' which will always be an important element of stress testing in practice.

Glasserman and Xu (2014) and Glasserman and Xu (2013) also explore robust tilts to forecast distributions, including those that impose expectations constraints as additional restrictions on plausibility. However, their work is again more theoretical and portfolio-oriented (rather than employing a practical example of an extended macro-stress testing model such as CLASS) and does not explore the usefulness of exponential tilting as an approach to introducing judgment in its own right, separate from any worst case analysis.

2. Exponential Tilting

Models used by regulators and banks imply probability distributions over various objects of interest. These distributions are characterized by a set of moments (means, variances, correlations, *etc.*) that they imply. Suppose a regulator judges that certain moment restrictions should be respected by a model, where these moment restrictions reflect judgment about the

nature of the economy or simply hunches that the regulator wishes to explore. Now suppose that the model a regulator possesses (perhaps based on historical data or provided by the regulated banks) does not initially satisfy these moment restrictions. For example, the regulator may feel that tail risks are underestimated or correlations among banks in a crisis are too low. Indeed, computational requirements may necessitate linearity and Gaussianity. Alternatively, perhaps the regulator is aware of a structural break but does not yet have sufficient data to re-specify and re-estimate an elaborate model.

In these situations, the regulator will want to adjust the probability distribution implied by her prevailing model so that it does satisfy the moment conditions. However, the regulator, realizing the dangers of working with an elaborate but misspecified model, will naturally want to do this in a way that incorporates the ‘least’ extra-model information, thus minimizing the manipulation of the distribution and retaining parsimony. Exponential tilting is a way to incorporate judgment in this disciplined way.

2.1. Theory

Let X denote a (possibly vector valued) random variable with density function, π , under the regulator’s ‘benchmark’ model. Let us express a desired moment restriction as

$$E[g(X)] = 0 \tag{1}$$

where g is a vector valued function. We could choose X to be a sequence of target and state variables (or, in richer models, regimes and parameters). We concern ourselves with the case where this condition will not hold if the expectation is taken with respect to π . We seek, instead, the (unique) density, π^* , that a) exists, b) satisfies the moment conditions and c) is closest to π in the Kullback-Leibler sense. Formally, following the analysis of Robertson, Tallman, and Whiteman (2005), we solve the following problem

$$\min_{\pi^*} \int \log \frac{\pi^*(X)}{\pi(X)} \pi^*(X) dX \tag{2}$$

such that

$$\int g(X) \pi^*(X) dX = 0 \tag{3}$$

which yields a solution, π^* , where

$$\pi^*(X) \propto \pi(X) \exp\{\tau'g(X)\} \tag{4}$$

for appropriate Lagrange multipliers, τ , on the moment conditions. That is, we obtain an exponentially tilted distribution.

In the Gaussian case where only first and second moments are restricted by g , the tilt can be derived analytically and preserves Gaussianity. However, in more general cases, the method is implemented in terms of empirical approximations to the distributions in question. That is we use approximations that are given by a set of draws $\{X_i\}_{i=1:N}$ and a set of corresponding weights, $\{\pi_i\}_{i=1:N}$, under the benchmark, and a different set of weights, $\{\pi_i^*\}_{i=1:N}$, under the tilted distribution. In general we will have $\sum \pi_i g(X_i) \neq 0$ and we wish to construct tilted weights such that $\sum \pi_i^* g(X_i) = 0$. We do this by finding an appropriate vector of multipliers, τ which solves

$$\tau = \arg \min_{\tau} \sum_{i=1}^N \pi_i \exp\{\tau'g(X_i)\} \tag{5}$$

and set

$$\pi_i^* = \frac{\pi_i \exp\{\tau'g(X_i)\}}{\sum_{j=1}^N \pi_j \exp\{\tau'g(X_j)\}} \propto \pi_i \exp\{\tau'g(X_i)\} \tag{6}$$

2.2. Moment restrictions

One could make the desired moment conditions arbitrarily complicated. At one extreme, one could imagine a fully specified structural model, capturing infinitely many restrictions on various parameters and states. However, since we are concerned with contexts in which the regulator is facing substantial ambiguity, it is more realistic to envisage a more limited set of moments. We will concentrate on some simple conditions that allow us to construct a wide variety of useful tilts:

Mean Restrictions The moment restriction function to impose that X has mean, μ , is

$$g_1(X) \equiv X - \mu \tag{7}$$

Variance restrictions In the case of a scalar random variable, X , whose mean is being restricted to be μ , the moment restriction function to impose a certain variance, v upon

it is

$$g_2(X) \equiv (X - \mu)^2 - v \tag{8}$$

where we note that without an associated mean restriction (unless in a linear Gaussian model) the interpretation of the restriction imposed by g_2 may be misleading. In isolation, it simply imposes the requirement that squared deviations from ‘some number’, μ , are equal to v in expectation. If μ is not the expected value *after* the tilt then this will not restrict the variance as desired and, indeed, in the general case the squared deviation restriction in isolation will typically affect the mean.

Percentile Restrictions In the case of a scalar random variable, X , whose k^{th} percentile we wish to restrict to be \hat{X}_k , the moment restriction function is

$$g_k(X) \equiv \frac{k}{100} - I\left(X \leq \hat{X}_k\right) \tag{9}$$

where $I(cond)$ is the indicator function equal to unity if the condition, $cond$, is satisfied and zero otherwise.

Histogram Restrictions Given a histogram characterized by a collection of buckets and relative frequencies we can derive a vector valued moment restriction function by stacking scalar percentile restrictions.

3. Robustness

In this section we lay out another approach to exploring misspecification in a benchmark model based on the robustness approach of Hansen and Sargent (2008). A robust agent (in our case, a regulator) is endowed with a ‘benchmark’ model but fears that it is misspecified. She is concerned that the world is actually described by a model that is similar to the benchmark but distorted in some way. The agent expresses her doubts of her model by considering alternative distributions that are distorted or ‘twisted’ versions of the distribution implied the benchmark. In order to explore the fragility of her model, the agent considers adverse distributions and balances the damage that an implicit misspecification would cause her, against the plausibility of the misspecification. The distribution that emerges from this problem is typically referred to as a ‘worst case distribution’ that encodes these concerns and allows insight into the fears that inform the agent.

The worst case distribution is a natural object of interest in the construction of stress tests in that it can be used to generate specific scenarios or to provide simulations that emphasize the dimensions in which the system (such as a collection of banks and their interaction with each-other and the wider economy) is particularly vulnerable (see, also, Breuer and Csiszr (2013) and Glasserman and Xu (2014) on similar points). We here lay out the abstract formalism of a robust decision problem before relating it specifically to our stress testing applications below.

3.1. Standard robustness

An agent derives utility according to the realization of a random variable, X , and a function, v , that maps a realization of X into a payoff for the agent. The agent’s ‘benchmark’ model implies a distribution for X , π . Were the agent to fully trust this model, she would evaluate her welfare on the basis of expected utility. However, the agent distrusts her model and evaluates welfare using a pessimistic twist to π . The twist is captured by a likelihood-ratio, m , with respect to which the agent minimizes her expected payoff, under the twisted measure. This minimization is subject to a penalty for twisting the benchmark distribution, controlled by θ and related to the relative entropy of the twisted distribution to the benchmark

$$W = \min_m \underbrace{\int v(X) m(X) \pi(X) dX}_{\text{Expected Payoff}} + \theta \underbrace{\int m(X) \log(m(X)) \pi(X) dX}_{\text{Relative Entropy}} \quad (10)$$

s.t.

$$1 = \int m(X) \pi(X) dX \quad (11)$$

$$m \geq 0 \quad (12)$$

$$m < \infty \quad (13)$$

Clearly, we can re-express the problem in terms of a minimization with respect to an alternative distribution, π^*

$$W = \min_{\pi^*} E_{\pi^*} [v(X)] + \theta KL(\pi, \pi^*) \quad (14)$$

s.t.

$$1 = \int \pi^*(X) dX \quad (15)$$

$$\pi^* \geq 0 \quad (16)$$

$$\pi^* \ll \pi \quad (17)$$

where $KL(\cdot, \cdot)$ is the Kullback-Leibler (KL) distance between distributions (we will use relative entropy and KL interchangeably). Solving problem (10) yields a particular minimizing likelihood ratio

$$m^*(X) = \frac{\exp\left\{-\frac{v(X)}{\theta}\right\}}{E\left[\exp\left\{-\frac{v(X)}{\theta}\right\}\right]} \quad (18)$$

The degree of robustness is controlled by the penalty parameter, $\theta > 0$, that enters the objective by multiplying the conditional relative entropy associated with a given distortion. The penalty reflects the intuition that the agent considers models that, although different, are somehow ‘near’ the benchmark. Thus, a particular m that might imply an extremely pessimistic evaluation of welfare may not solve the minimization problem, due to it implying an excessive relative entropy penalty.

3.2. Tilted robustness

In a ‘tilted robustness’ problem the agent solves a similar problem to (10) but it is augmented by the presence of a constraint requiring the worst case distribution to respect moment conditions captured by the function, g .

$$W = \min_m \int m(X) v(X) \pi(X) dX + \theta \int m(X) \log(m(X)) \pi(X) dX \quad (19)$$

s.t.

$$1 = \int m(X) \pi(X) dX$$

$$0 = \int m(X) \pi(X) g(X) dX \quad (20)$$

Solving problem 19 yields another particular minimizing likelihood ratio

$$m^*(X) = \frac{\exp\left\{-\frac{v(X)-\varphi^*g(X)}{\theta}\right\}}{E\left[\exp\left\{-\frac{v(X)-\varphi^*g(X)}{\theta}\right\}\right]} \quad (21)$$

where φ^* is the Lagrange multiplier on the constraint (20) at the solution. This change in measure is similar to the familiar exponential tilt obtained under the standard robustness approach, $m_{RC}^*(X) \propto \exp\left\{-\frac{v(X)}{\theta}\right\}$, but differs due to the presence of the moment restrictions that the tilted distribution must satisfy. For further more formal analysis of these points, we refer the reader to Petersen, James, and Dupuis (2000) and Glasserman and Xu (2014).³

Intuitively, while the agent acknowledges enormous uncertainty over the true data generating process, she implicitly regards certain aspects of the world as well understood or worth exploring specifically, as captured by the moment conditions. The plausibility of alternative models (distributions) are no longer simply characterized by relative entropy, as in standard robustness, but the ranking of models' plausibility also depends on their satisfying certain conditions.

4. Application: The CLASS Model

To illustrate these methods we make use of the ‘Capital and Loss Assessment under Stress Scenarios’ (CLASS) model, developed at the New York Fed (Hirtle, Kovner, Vickery, and Bhanot (2015)). The CLASS model is a top-down framework for examining the implications of macroeconomic scenarios for U.S. banks. It is top-down in the sense that, rather than building the assessment up from granular loan- or position-level regulatory data, it is based on publicly available data on components of bank performance aggregated at the business line level. These components are related to macroeconomic and financial variables *via* a set of econometric models, comprising regressions of components of bank income, expense and loan performance, combined with assumptions about provisioning, dividends, asset growth and other factors. For a given scenario one can then generate a projection for bank performance - at the level of each individual bank - which can be aggregated to a system-wide measure of

³There are also interesting parallels to Kwon and Miao (2013) and Hansen and Sargent (2012) where decisions over the knowledge of a true model by a private sector can lead to restrictions on the worst case envisaged by a robust policymaker.

capital adequacy and other variables of interest. For further details of CLASS, including the regression specifications, we refer the reader to Hirtle, Kovner, Vickery, and Bhanot (2015).

Hirtle *et al.* show that the more granular and intensive DFAST/CCAR models yield projections for components of income and loss that are highly correlated with those from the CLASS model. Given the far greater simplicity of CLASS, this is a significant achievement. We therefore take the CLASS model as a reasonable and manageable model for capturing the connection between banks and the macroeconomy. Nevertheless, it is clearly a simplification and surely (unavoidably) misspecified in known and unknown dimensions. Thus it is a natural basis for the sort of benchmark model to which tilting could apply.

4.1. Data generating process

The variables constituting the underlying macroeconomic scenarios used in CLASS are: the term spread, stock market returns, real and nominal GDP growth, the change in the unemployment rate, the short rate, the level and change of the corporate bond spread, the change in the 10-year Treasury yield and house and commercial real estate price growth. These variables are augmented through non-linear transformations. Specifically, regression slopes for the change in the corporate spread and the property indices are allowed to differ according to the sign of the changes. The regressions relating these factors to banks' income also allow an interaction between the change in the corporate spread and a bank's ratio of 'risky' AFS assets.

Hirtle *et al.* consider a single scenario. However, for our purposes it is necessary to augment the CLASS framework with a simulation model to draw multiple scenario realizations. This model will define a proposal density for the importance sampling implicit in our tilts. As it is, CLASS simply maps from a given scenario to its outcome variables and does not comprise a data generating process for that scenario.

In order to construct a proposal density, we estimate a first order vector autoregression (VAR) in 'core' macroeconomic variables: real GDP growth, the change in the unemployment rate, the short rate, the term spread and inflation. We then independently regress the remaining 'financial' variables of interest (stock market growth, corporate spreads and house and commercial real estate price growth) on these core variables, including a lagged dependent variable in each regression. Then, in order to simulate realizations of the CLASS scenario variables we simulate the VAR for $t = 1 : n_{hor}$ periods, n_{Prop} times (with $t = 0$ being 2015Q1)

and then apply the financial variable regressions.⁴

With regard to the short rate we account for the zero lower bound (ZLB) by first replacing the three month T-bill rate used in CLASS with the shadow rate measure of Wu and Xia (2016), and then estimate the VAR.⁵ Over the following periods, Wu and Xia argue that the variable captures the stance of policy while unconventional methods were being implemented.⁶ Our regressions of the financial variables include the *shadow* rate as an independent variable. However, since the CLASS regressions do not make use of the shadow rate concept, we impose non-negativity on our simulated short rate paths. In addition, whenever we construct or simulate term spreads or implied changes in 10-year yields (from changes in the term spread and the short rate), we respect the ZLB.

Our aim is to avoid over-fitting by invoking a fairly standard small scale VAR and relating the somewhat less standard ‘financial’ variables to the core variables in a simple manner. There are significant caveats to our simple approach but since our model is explicitly intended only to be an initial proposal density that can be adjusted with judgment or to reveal the most damaging types of plausible misspecification, we regard it as a useful illustration. Our framework is arguably pertinent to the type of ‘satellite’ models used in policy institutions where individual departments have independently estimated partial models that then must be combined to generate an overall simulation.

4.2. A measure of welfare

For the tilts involving robustness, we must associate a value from the perspective of the regulator with any given draw from the benchmark. Following Hirtle *et al* we will define the input to the regulator’s value function to be the minimum tier 1 common equity ratio (tier 1 common equity over risk weighted assets) over the forecast horizon. That value is then mapped to a welfare index using the following form:

$$v(x) = (\alpha + (1 - 2\alpha)(x < \bar{x}))(x - \bar{x})^p$$

⁴Inflation is not a variable included in the class scenarios but we include it in the VAR nevertheless. We also note that some of the variables included in CLASS, such as changes in the corporate spread or changes in the ten year yield are pinned down deterministically given values of the other variables, such as the short rate and the term and corporate spreads. Finally, we use the fitted values from the financial variable regressions, rather than also drawing regression disturbances.

⁵Over most of the sample, the difference between these two series is minor but the shadow rate passes below the ZLB in 2009.

⁶We ignore the fact that the Wu and Xia shadow rate is subject to estimation error in what follows.

We set \bar{x} to be 10% and α to be 0.2. In fact, for simplicity $p = 1$ so that the regulator’s index of welfare is increasing in the capital ratio, x and, locally, welfare varies linearly in the deviation from \bar{x} but is steeper below than above. The differing slopes of the function either side of \bar{x} are intended to capture the idea that capital ratios falling below this schedule are treated more severely than surpluses of the same magnitude.⁷

A useful feature of our approach is that what constitutes the worst case distribution under the robustness approach will depend on what variables the regulator cares about. That is, if the regulator’s welfare is sensitive to the performance of the aggregate financial system then the scenarios generated from the worst case will be systemically damaging. If, however, the regulator’s value function is defined with respect to the performance of a particular bank or group of banks, then the worst case will be different, reflecting the different vulnerabilities. Therefore, if desired, one can easily tune scenarios to particular concerns, rather than constructing a ‘one size fits all’ scenario or model. This is relevant for cross checking bank-specific scenarios volunteered by banks themselves, as in the ‘reverse stress testing’ elements of Dodd-Frank reforms.⁸

In our robustness applications it is instructive to report scenarios and tilts associated with particular ‘banks’. We use inverted commas, as our banks will be artificially constructed to represent certain classes of institution, rather than being individual banks observed in reality. Specifically, weighted by relative total assets, we construct a ‘CCAR’ bank from financial data of those U.S. institutions involved in CCAR at the point when CLASS was first constructed, a ‘Small CCAR’ bank from the five smallest CCAR banks (by 2015Q1 total assets) and an ‘Investment Bank’, which combines information for Goldman Sachs and Morgan Stanley.⁹

The regressions underpinning the CLASS model are estimated by a particular pooling method, weighting banks in terms of their sizes, to obtain a set of regression coefficients that are *common* across banks. These coefficients typically relate the CLASS scenario variables to

⁷Of course, some might argue that this function should be non-monotonic, reflecting the fact that there is some internal maximally desirable capital ratio from the regulator’s perspective. However, local to the range of ratios discussed in practice, monotonicity is a reasonable assumption.

⁸Pritsker (2013) argues in favor of tuning stress scenarios to bank exposures although also see Hirtle, Schuermann, and Stiroh (2009) for concerns for consistency that arose when banks were allowed to posit certain scenarios in the initial SCAP framework.

⁹The ‘CCAR’ banks in our set exclude Deutsche Bank Trust Corporation, TD Group US Holdings LLC and BancWest Corporation. The ‘small’ CCAR banks are taken to be: Zions Bancorporation, Huntington Bancshares Incorporated, Comerica Incorporated, Discover Financial Services and BBVA Compass Bancshares, Inc.

a ‘ratio’, which is a flow normalized by some measure of scale. For example, CLASS contains a regression for the net C&I chargeoff *rate* given by C&I net charge-offs (flow) divided by total C&I loans (scale). To map from a scenario to a flow, which is necessary to calculate net income and, ultimately, capital, one must unwind the scaling by multiplying through by the projected scale variable (the amount of C&I loans over the forecast period, in our example).

Banks differ according to the value of the scale measures prevailing at the start of the forecast horizon and in their initial levels of capital. Further, they differ in the initial values of the ratios of interest, which is relevant as the CLASS regressions incorporate autoregressive elements. These differences underpin the heterogeneity in the implications of a given scenario for different banks’ capital ratios. For a given flow of income, initial capital positions will clearly induce different projections for capital going forward. Similarly, if one bank makes a disproportionately large number of C&I loans then the overall implications of a scenario will differ from other banks, even though it is predicted that *for any given value of loans* there should be the same effect.

Our choice of banks should not be over-interpreted - it is guided primarily by the fact that they illustrate how different worst cases may emerge for different institutions. The CCAR group is of broad interest and is a natural starting point. Within that class of institutions, it is natural to think of the two institutions with ‘investment banking’ lineage as worthy of separate analysis. It is also natural to think that some of the smaller institutions may have different business models that could lead to interesting insights. Nevertheless, we should emphasize that these results do not represent the Federal Reserve’s assessment of the institutions’ vulnerabilities and are not part of the official stress testing program.

All the information underpinning the model is public but we choose not to work with individual institutions. Even if we were to use individual institutions’ it would only be in a very limited sense. The results can to some extent be related to the constituent ‘firms’ because we do make use of the initial capital levels and relative scale of the various business lines considered at each, as of 2015Q1. However, these are very rough and incomplete characterizations of the banks’ business models. In reality they also differ in how they would respond to unfolding scenarios so the common regression coefficients assumed in CLASS are misspecified (even if they were correctly specified ‘in sample’ which is unlikely). Further more, banks might also adapt their business models so that our assumed scaling projections

will also be misspecified.^{10,11} Finally, while the CLASS model’s results correlates well with those of the CCAR analysis, that is not to say at the level of any given bank the results will be a reliable indication of what the full CCAR analysis would imply. Nevertheless, some interesting insights can be gleaned and the application is useful for our purposes.

5. Results

In this section we discuss the numerical results of our analysis - illustrating each of our methods in turn. We begin with a simple application of exponential tilting before using standard and tilted robustness to seek out ‘stress’. We refer to the CCAR composite bank as CCAR-B, below, the investment bank composite as IB-B and the ‘smaller’ CCAR banks composite as CCAR-SB.

5.1. Exponential tilting

It is often of interest to take market participants’ opinions into account as an alternative view of the state of the economy. Thus, we consider incorporating information from the 2015Q1 Survey of Primary Dealers.¹² Specifically we impose the distribution of the short rate at the end of 2015 to be in-line with the histogram implied by answers to question 2(e) of the survey, which asks for dealers’ percent chance of the rate falling in certain ranges.

The impact of the tilt on the distribution of the short rate at the end of 2015 is illustrated in figure 1. Clearly, incorporating the information from the SPD makes a substantial difference to the distribution as the benchmark implies a very low rate (primarily capturing a late ‘liftoff’ from the ZLB), whereas the dealers envisaged a swifter normalization. Given the simplicity of the benchmark model and the importance of soft information in the historically unprecedented normalization period, it seems natural to incorporate the dealers’ opinions

¹⁰We assume a constant growth rate in assets and loans, as in Hirtle *et al*

¹¹To be sure, the aim of CLASS was not to provide precise *bank specific* projections but, instead, to generate projections that could be reliably aggregated to reveal timely insights into the broader financial system. As noted earlier, in comparison with the projections of income from the richer CCAR models, CLASS performs very well. In addition, we will be working with a simplified version of the model, based on tables 1 and 2 in the appendix of Hirtle, Kovner, Vickery, and Bhanot (2015), but omitting certain judgmental adjustments made within the implementation of the model for more practical applications and also simplifying the treatment of the connection between charge-offs and provisions (we assume net charge-offs are equal to provisions - see the CLASS working paper for more details of the different possible assumptions).

¹²See https://www.newyorkfed.org/markets/primarydealer_survey_questions.html

and override the benchmark in this dimension if one is to use the benchmark for probabilistic modeling of bank vulnerabilities.

The short rate distribution implicit in figure 1 is not the only dimension in which the tilt is manifest. In figure 2 we illustrate how the mean paths over the CLASS scenario variables differ, relative to the benchmark.¹³ Setting aside the higher levels of the short-rate, we also observe that the tilted scenarios are associated with higher growth, lower unemployment and higher inflation. The impact on asset prices is more nuanced. There is little impact on the mean path of the stock market although there appears to be a slight near term boost and credit spreads are lower (as is the term spread). Interestingly, while commercial property prices also are higher than in the benchmark, house prices initially track and then ‘underperform’ the benchmark.

Recalling that the tilt stays as close as possible to the benchmark subject to the moment conditions, what these results seem to suggest is that the benchmark quite strongly associates periods of increasing short rates with expansionary conditions, with some sort of ‘Taylor rule’ association at play in the data. Thus, since the primary dealers envisaged a higher short rate in the near term, it is to be expected that the aggregate ‘macro’ variables should be tilted towards firmer realizations. It is also interesting that two asset classes highly sensitive to interest rate news - the stock market and house prices - seem to be somewhat damped. Apparently, the association of stronger economic conditions with better cashflow news seems to cancel out the discount rate effect for the stock market - though a full structural interpretation of these tilts is somewhat difficult.

Banks typically lend long and borrow short - as reflected in a positive loading of net interest margins on the term spread in the CLASS model equations - so the decline in the term spread, all else equal, is a negative for bank performance under the benchmark. House price returns enter positively (negatively) in PPNR (charge-off) regressions. Thus, there are some offsetting effects of this particular tilt, to be set against the benefits for bank performance of the apparently greater weight placed on a stronger aggregate economic environment under the tilt. Given these offsetting patterns, it is perhaps unsurprising that the mean paths for the tier 1 capital ratio under the tilt are not hugely changed (in fact they are *higher*) relative

¹³Note that we convert the unemployment growth scenarios into the implied path for the level of unemployment and only report the short rate and term spread rather than the ten year yield and the its changes. As mentioned above, we also report inflation, as it is included in the benchmark VAR, despite not being a CLASS scenario variable.

to the benchmark, as shown in figure 3. Thus, it turns out that this tilt is not so much about increasing the ‘stress’ to the banking system but more a question of ‘improving’ the model so that if risk based-analysis is undertaken, such as for a VaR analysis, the model has certain pathological tendencies (the very slow normalization) corrected.¹⁴

5.2. Robustness - Bank specific

We now turn to a systematic search for damaging misspecifications, as captured by the worst case distribution generated by standard robustness analysis.

5.2.1. The degree of robustness

We fix the level of θ to be such that the KL distance between the benchmark and worst case distribution is common across banks, to enhance comparability. This might be desirable in practice if the regulator has a deep preference for how extreme the worst case should be and wishes it to be uniform across banks. The precise degree of robustness in our case is not so important as it can be easily varied according to the desired application. In our case, we set θ such that the mean paths for the unemployment rate reach approximately 6 – 7% at the end of the projection horizon, in comparison with 5% under the benchmark. This is a large enough deviation to illustrate our methods and also comparable to the peak level reached under the 2016 CCAR ‘adverse’ scenario, though the nature of the CCAR projections is however, different from ours.

An alternative to fixing the K-L distance is to use a common θ across banks, thereby penalizing entropic deviations uniformly. To gain intuition for this approach, consider that the distribution over capital ratios is pinned down by the known bank-specific initial conditions, the deterministic mapping from scenario variables to the components of net income that feed through to the capital ratios and the distribution over scenario variables, which is common across banks. Consequently, for any candidate tilted distribution being compared, the value of the penalty term in the minimization problem will be the same across banks. Thus, if the worst case distribution differs across banks it is because of the different exposures each bank has, which in turn are traded off differently against the entropic penalty.

Intuitively, if a bank is relatively heavily exposed to a certain distortion to the benchmark (loosely we can imagine a certain moment of interest being of particular importance)

¹⁴At least, at the time the predictions were made, the slow pace of normalization was apparently thought unlikely.

then the negative contribution to the welfare component of the penalized objective function of marginally distorting the distribution outweighs the positive contribution to the entropic penalty component, where for a different (less exposed) bank, the marginal contributions might have balanced out. Thus, the solution will feature a greater distortion along this dimension until the marginal costs and benefit components of the objective function *do* balance out. Overall, this can lead to the worst case distributions for different banks being different distances from the benchmark in the K-L sense. We do not pursue this approach although it can easily be implemented.¹⁵

5.2.2. Examining the worst case

In figures 4 and 5 we illustrate the initial variation across our example banks in the amounts of different loan types (relative to risk weighted assets) and in the amounts of interest-earning, trading and available-for-sale (AFS) asset holdings (relative to total assets). Combined with variation in the loadings of the components of income and charge-offs (taken from the regressions reported in Hirtle *et al* and illustrated in figures 6 and 7) we see that what would constitute severe misspecification will vary across banks according to the importance of different business lines and their dependence on different scenario variables.

In general, the nature of the distortions to mean paths in figure 8 are reasonably intuitive. Asset prices and spreads come under pressure, growth and unemployment are adversely affected and the short rate remains essentially stuck at its lower bound. The qualitative properties of the worst case are essentially the same across our three banks, but there are interesting differences among them.

IB-B seems to have somewhat less distortion to the ‘macro’ variables - unemployment, inflation and (in the short run at least) GDP - which is perhaps suggestive of a more procyclical tendency among the other CCAR banks. The gap between the unemployment path distortions in particular seems intuitive since, as shown in figure 4, IB-B has relatively little exposure to C&I loans and various types of consumer loans (such as Credit Card loans, for example), the charge-offs for which are estimated to load positively on unemployment. Similarly, at least in the short run, the fact that IB-B features less distortion to property prices likely partly reflects its relatively small (scaling by RWA) loan exposures to residential

¹⁵As Hansen and Sargent (2008) note, θ is not so much the deep parameter in robustness problems as the agent’s view of how much ‘uncertainty’ is thought reasonable. The same level of uncertainty can, in different models (as with different banks, say), be associated with very different θ .

and commercial real estate. The positive distortion to the mean of the ten-year Treasury yield, implied by the short rate and term spread projections, is also greater for IB-Bank. This is perhaps due to IB-B's smaller AFS holdings (shown in figure 5) relative to CCAR-B and CCAR-SB, as the return on AFS securities is estimated to load negatively on the change in the ten-year yield.¹⁶

However, these interpretations of the source of the distortions should only be regarded as tentative. There are also some apparently counterintuitive patterns in the distortions, such as the fact that the return on trading assets loads negatively on the change in the corporate bond spread (and particularly so if that change is positive, as shown in figure 6) and yet the distortion in the corporate spread is *smaller* for IB-B despite it having a more significant trading book. In addition, as shown in figure 4, IB-B has a relatively large amount of loans classified as 'Other' (though in line with CCAR-B) and since this category's charge-offs are also estimated to load positively on changes in unemployment one might have expected it to enhance the distortion to the unemployment path under the the IB-B worst case distortion.¹⁷

The difficulty in identifying direct reasons for particular properties of the worst case lies in the fact that there are likely offsetting vulnerabilities across the balance sheet which demand certain distortions, *ceteris paribus*. In addition, the web of correlations and interdependencies among the variables mean that because the entropy penalty ensures that the distortions are minimized for a given degree of 'pain', these distortions may be in tension with each other. It may be that a variable is distorted in a 'marginally' counterintuitive way because a variable with which it is correlated (where that correlation is still approximately respected under the worst case) may be a more damaging variable to distort. The manner in which this latter variable is distorted, given the correlation, may overwhelm the natural marginal distortion to the first variable. Such mechanisms mean that it is difficult to draw out precisely the source of the distortions in the worst case. However, to some degree, this is the point of robustness in that it is 'revelatory' and can indicate counterintuitive vulnerabilities.

In terms of the implications of the worst case distributions for the performance of the banks, figure 9 shows the mean paths of the banks' capital ratios in addition to the benchmark paths. Naturally, they are somewhat lower, with the the reductions at the end of the horizon

¹⁶The gap relative to CCAR-SB is reduced when scaling by RWA, rather than total assets as in figure 5

¹⁷It is also possible that IB-B's distortions may be concentrated in higher moments since generally its mean paths appear visually to be somewhat less distorted, despite the fact that the K-L distance from the benchmark has been fixed to be equal across banks.

being approximately 0.38% for CCAR-B, 0.40% for IB-B and 0.48% for CCAR-SB. Given that we have fixed the K-L deviation to be common among the banks we can loosely interpret these differences as suggesting a ranking of vulnerability for a given degree of uncertainty that the regulator is willing to tolerate. Subject to our earlier caveats regarding the drawbacks of taking the *bank-specific* predictions of CLASS too literally, this suggests that, in our simple framework at least, CCAR-SB is somewhat more exposed. However, the differences appear minor.

5.3. Robustness - Banking sector

We also could envisage taking a broader perspective of the state of the banking system and concern ourselves with the distribution of capital ratios across multiple banks. This approach is distinct from constructing a single composite bank to be a weighted average of all banks as it allows for preferences over the distribution of bank characteristics - such as the distribution of capital ratios and so forth. It is clearly the case that a regulator would not be indifferent between two situations in which average capital ratios are a certain number but in the former, all institutions are at that ratio, while in the latter an important bank has a perilously low capital ratio.

For each draw of the scenarios we can run the CLASS regressions for each of the top 200 (by risk weighted assets) and calculate the minimal capital ratio over the forecast horizon. We can then, on sorting the banks by their minimal capital ratios we can obtain a measure, mapping from a given ratio value to the fraction of risk weighted assets held by banks with minimal capital ratios less than or equal to that value. This is done for each draw and the i^{th} draw one then obtains a discrete measure, call it F_i , defined by a collection of cumulative shares and minimal capital ratio nodes. An example of such a measure, taken from an arbitrarily selected draw, is shown in figure 10. We observe several large jumps at certain ratios, reflecting the fact that the distribution of bank size is extremely positively skewed, as illustrated in figure 11.

For each draw, we again wish to map to a measure of welfare and in this case, we need a function that expresses a regulator's preference for a given F_i . For this one could envisage any number of weighting functions - and designing such a function carefully is a very interesting avenue for future research - but for our purposes we will use the function shown in figure 12, which is based on a logistic function that features a somewhat smoothed kink at a 10% minimum capital ratio. Implicitly, the regulator essentially cares about what fraction of the

risk weighted assets are at banks with minimal capital ratios below 10%.

Based on this setup we can again obtain a worst case distribution, and again can compare the mean forecasts of the scenario variables, shown in figure 13. Reflecting substantial homogeneity from the common CLASS regression coefficients we see that the nature of the worst case, at least in terms of mean paths, is qualitatively much as in our earlier bank specific cases.¹⁸ This, however, is likely highly specific to our particular benchmark and the weighting function and welfare functions chosen.

5.4. *Tilted robustness*

Now, suppose a regulator is relatively confident in certain aspects of her benchmark or has extra model information or institutional constraints that require consistency with some other set of forecasts, such as in a conjunctural forecast round common at central banks. In this case, one might wish to restrict the worst case to respect these trusted elements of the benchmark. We illustrate this tilted robustness approach here by requiring that the ‘term structure’ elements of the benchmark model be respected. Specifically, we impose restrictions on the mean path for the short rate, term spread and 10-year yield, such that the worst case mean is equal to that of the benchmark.

In figure 14 we illustrate how imposing these restrictions moderates the worst case distribution. We plot (holding θ fixed in this case) the benchmark mean paths, together with the unrestricted worst case (HS) and the restricted worst case (TR) analogs. One can see that, up to slight sampling variability, the short rate and term spread under TR line up with the benchmark, as intended. While the impact on the short rate, economically, is relatively minor (the difference is approximately one *25bp* rise by the end of 2017), the impact on the term spread is somewhat larger and, given that it induces a steepening in the earlier quarters, it is likely to be associated with simulation draws with somewhat more expansionary characteristics - as reflected also in the more moderate GDP and unemployment deteriorations, relative to the HS worst case.

Associated with the restrictions placed on the worst case, we observe a more moderate deterioration in the path of the capital ratio, as shown in figure 15. Indeed in figure 16 we see the impact of the various tilts on the distribution of CCAR-B’s minimum (over the horizon)

¹⁸We do not align the K-L divergence with that of our earlier bank-specific worst case analysis as the loss function and mappings from realizations to the variable of interest (fraction of assets held by banks with minimum capital ratios below 10% mean that the results are not strictly comparable.

capital ratio, first showing only the benchmark and then with the two tilts layered on top. The HS worst case distribution is almost unimodal and shifted leftwards, while the TR worst case distribution, while shifted left, retains considerable mass in the upper tail, leaving it slightly bi-modal.

As discussed above, tuning θ provides a parsimonious way of controlling how extreme one wishes to make the stressed distributions. Indeed, the CCAR stress scenarios come in two varieties - adverse and severely adverse. We can see from figure 16 that relating θ to how large a left tail one wishes to consider could be a practical method of calibration. Since the severity of the worst case will vary monotonically with θ it is simple for a regulator in practice to identify a calibration that yields the desired degree of stress.

Finally, however, we note some of the restrictions arising from relying only on draws from a fairly well behaved benchmark model. In order fully to explore damaging misspecifications, one requires a proposal density that can generate an adequate number of ‘tail draws’ to begin with as HS and TR tilts can only reweigh existing draws (as can be seen from the common tails in figure 16). In our case, none of the minimal capital ratios generated under the benchmark reach the sort of levels that one might associate with a crisis. Designing more elaborate models to give the robustness approach more to work with is an important avenue for further work.

6. Conclusions

In this paper we have laid out a parsimonious, practical and theoretically grounded toolkit for stress testing and illustrated the methods with an application using the New York FED’s CLASS model. The method revealed what adjustments to a simple forecasting model would imply for bank performance, such as incorporating information from market beliefs about the path of the short rate, and also what misspecifications would be particularly damaging. While the complexity of the financial system makes it difficult to make confident assertions about what vulnerabilities regulators should be worried about, our results tentatively suggested that there were interesting variations among banks depending on their business models, emphasizing the problems with a one-size-fits-all approach in stress testing. The methods also emphasized that introspection may lead regulators astray in that several elements of the worst case distributions we constructed were not *a priori* intuitive. The tools are easy to implement and generalize various techniques that are already familiar to policymakers. Ongoing work

on developing models that allow for severe episodes is an important complementary steps to this technique as, while the tilted models can compensate for certain misspecifications, re-weighting draws from the benchmark model does, to some extent, limit the severity of the tilted distributions.

References

- BORIO, C., M. DREHMANN, AND K. TSATSARONIS (2013): “Stress-testing macro stress testing: Does it live up to expectations?,” *Journal of Financial Stability*, (0), –.
- BREUER, T., AND I. CSISZR (2013): “Systematic stress tests with entropic plausibility constraints,” *Journal of Banking and Finance*, 37(5), 1552 – 1559.
- BREUER, T., M. JANDACKA, K. RHEINBERGER, AND M. SUMMER (2008): “Macroeconomic Stress and Worst Case Analysis of Loan Portfolios,” .
- (2009): “How to find plausible, severe and useful stress scenarios,” *International Journal of Central Banking*.
- CHRISTENSEN, J. H., J. A. LOPEZ, AND G. D. RUDEBUSCH (2015): “A probability-based stress test of Federal Reserve assets and income,” *Journal of Monetary Economics*, 73, 26 – 43, Carnegie-Rochester-NYU Conference: Monetary Policy: An Unprecedented Predicament, November 14-15, 2014.
- COGLEY, T., S. MOROZOV, AND T. J. SARGENT (2005): “Bayesian fan charts for U.K. inflation: Forecasting and sources of uncertainty in an evolving monetary system,” *Journal of Economic Dynamics and Control*, 29(11), 1893–1925.
- COVAS, F. B., B. RUMP, AND E. ZAKRAJSEK (2013): “Stress-Testing U.S. Bank Holding Companies: A Dynamic Panel Quantile Regression Approach,” FEDS Working Paper 2013-55, Federal Reserve Board of Governors.
- DANIELSSON, J., K. R. JAMES, M. VALENZUELA, AND I. ZER (2016): “Model risk of risk models,” *Journal of Financial Stability*, 23(C), 79–91.
- FRAME, S. F., K. GERARDI, AND P. S. WILLEN (2013): “Supervisory stress tests, model risk and model disclosure: Lessons from OFHEO,” .
- GIACOMINI, R., AND G. RAGUSA (2013): “Theory-coherent Forecasting,” *Journal of Econometrics (in press)*.

- GLASSERMAN, P., AND X. XU (2013): “Robust Portfolio Control with Stochastic Factor Dynamics,” *Operations research*, 61(4), 874–893.
- GLASSERMAN, P., AND X. XU (2014): “Robust risk measurement and model risk,” *Quantitative Finance*, 14(1), 29–58.
- GUERRIERI, L., AND M. WELCH (2012): “Can Macro Variables Used in Stress Testing Forecast the Performance of Banks?,” FEDS Working Paper 2012-49, Federal Reserve Board of Governors.
- HANSEN, L. P., AND T. J. SARGENT (2008): *Robustness*. Princeton University Press, Princeton, NJ.
- HANSEN, L. P., AND T. J. SARGENT (2012): “Three types of ambiguity,” *Journal of Monetary Economics*, 59(5), 422–445.
- HIRTLE, B., A. KOVNER, J. VICKERY, AND M. BHANOT (2015): “Assessing Financial Stability: The Capital and Loss Assessment under Stress Scenarios (CLASS) Model,” FRBNY Staff Report 663, Federal Reserve Bank of New York.
- HIRTLE, B., T. SCHUERMAN, AND K. STIROH (2009): “Macroprudential supervision of financial institutions: lessons from the SCAP,” Discussion paper.
- KWON, H., AND J. MIAO (2013): “Three types of robust Ramsey problem in a Linear Quadratic Framework,” .
- LEHNERT, A., AND B. HIRTLE (2015): “Supervisory Stress Tests,” *Annual Review of Financial Economics*, 7(1), 339–355.
- PETERSEN, I., M. JAMES, AND P. DUPUIS (2000): “Minimax optimal control of stochastic uncertain systems with relative entropy constraints,” *Automatic Control, IEEE Transactions on*, 45(3), 398–412.
- PRITSKER, M. (2013): “Knightian uncertainty and interbank lending,” *Journal of Financial Intermediation*, 22(1), 85 – 105, [jce:title;Research on the Financial Crisis;ce:title;.](#)
- PRITSKER, M. G. (2012): “Enhanced Stress Testing and Financial Stability,” Federal Reserve Bank of Boston.

ROBERTSON, J. C., E. W. TALLMAN, AND C. H. WHITEMAN (2005): “Forecasting Using Relative Entropy,” *Journal of Money, Credit and Banking*, 37(3), 383–401.

WU, J. C., AND F. D. XIA (2016): “Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound,” *Journal of Money, Credit and Banking*, 48(2-3), 253–291.

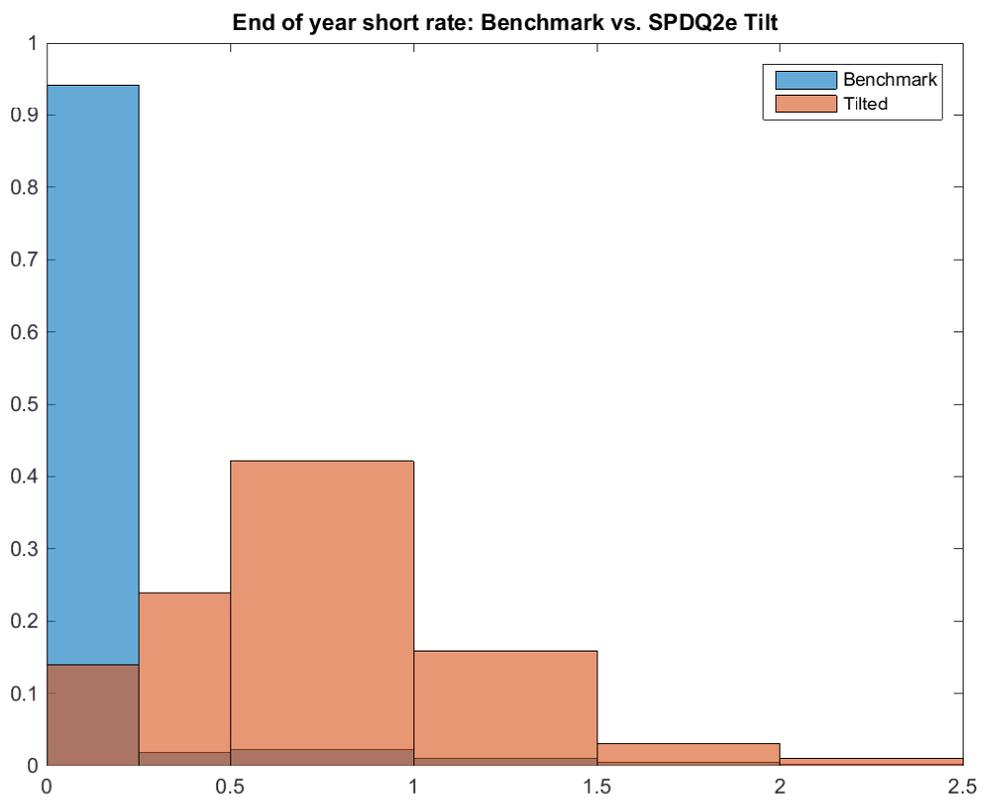


Figure 1: Comparing benchmark and tilted distribution of the short rate at the end of 2015 (tilt is to align with the answers from Q2e of the 2015Q1 Survey of Primary Dealers.)

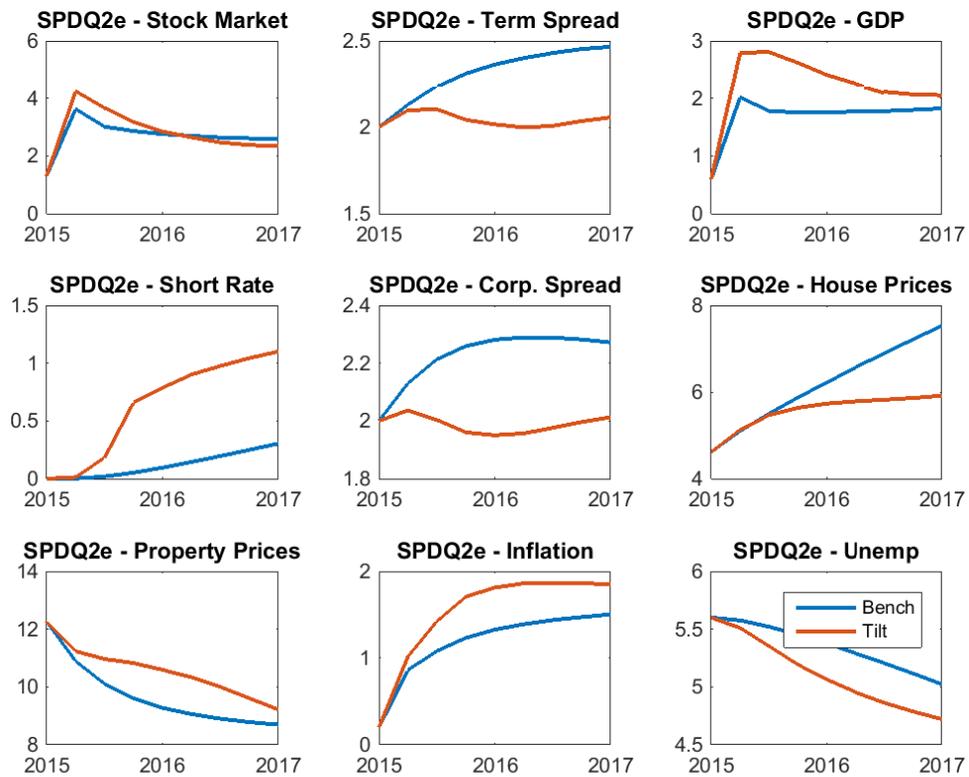


Figure 2: Comparing benchmark and tilted means of the CLASS scenario variables (tilt is to align the distribution of the short rate at the end of 2015 with the answers from Q2e of the 2015Q1 Survey of Primary Dealers.)

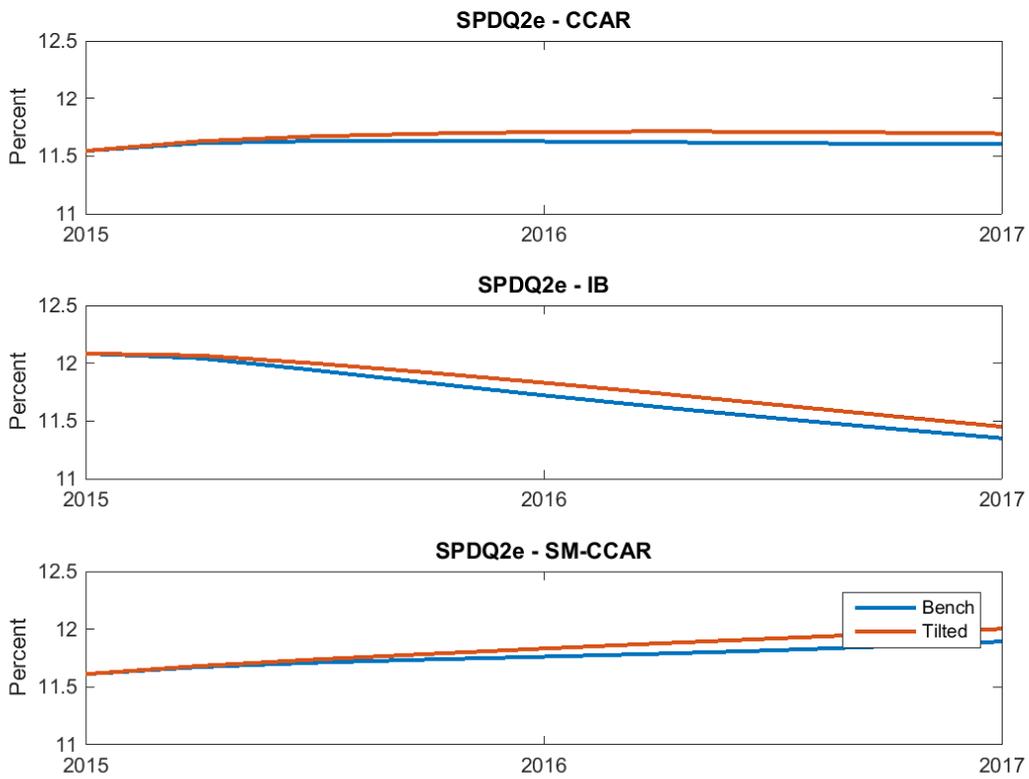


Figure 3: Comparing benchmark and tilted means of capital ratios over the forecast horizon (tilt is to align the distribution of the short rate at the end of 2015 with the answers from Q2e of the 2015Q1 Survey of Primary Dealers.)

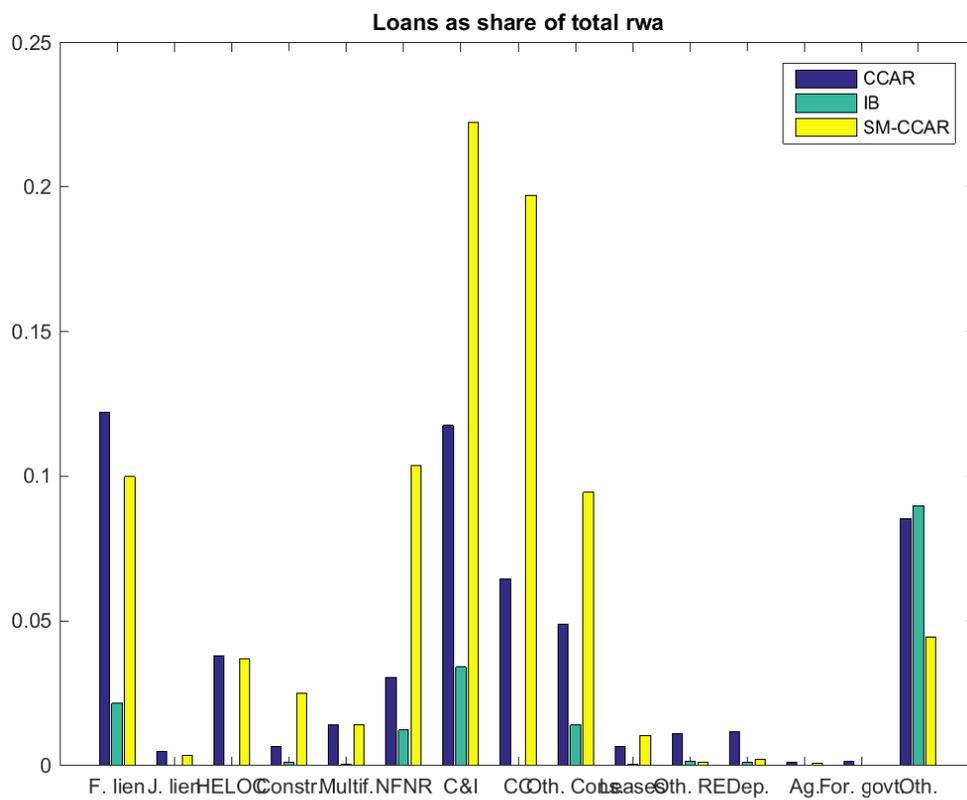


Figure 4: Variation in loan shares across example banks

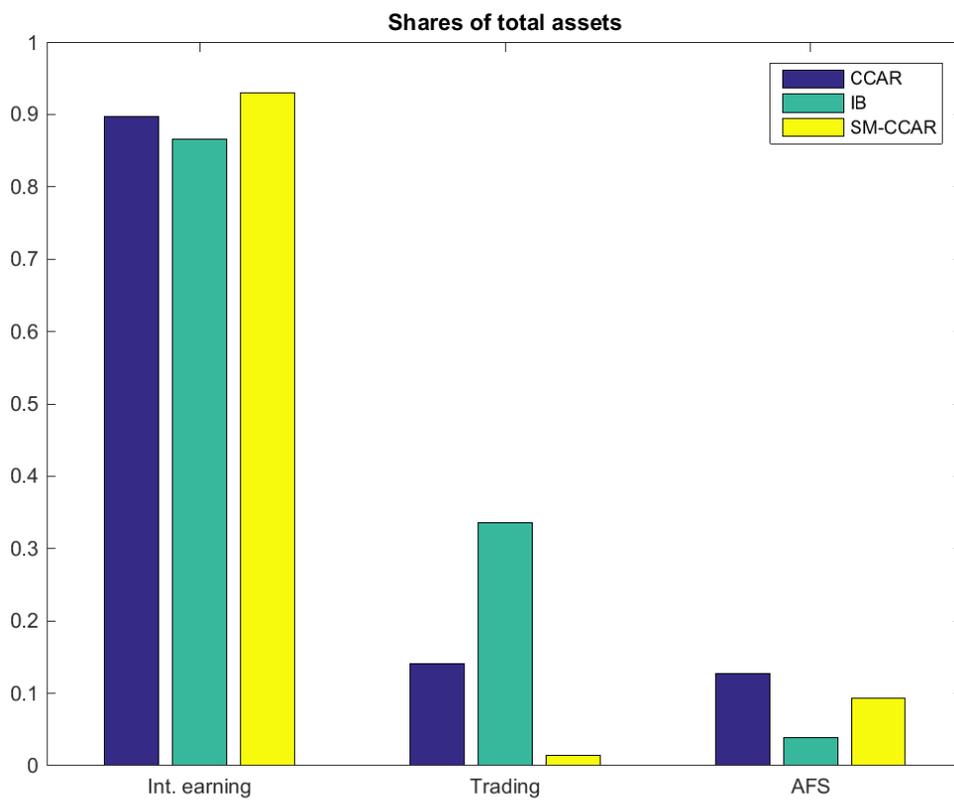


Figure 5: Variation in types of assets (as proportion of total assets) across example banks

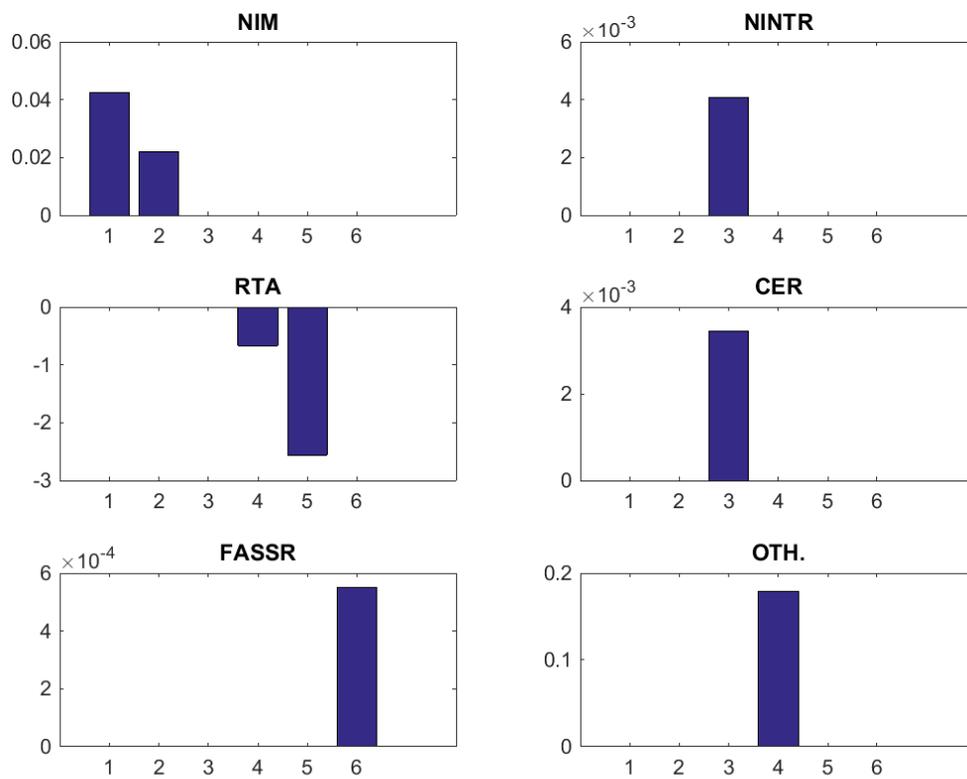


Figure 6: Variation in regression coefficients on CLASS scenario variables across components of PPNR (key for scenario variables: 1) Term Spread, 2) Short Rate, 3) Stock Returns, 4) Δ Corporate Spread, 5) Δ Corporate Spread (if positive), 6) GDP Growth)

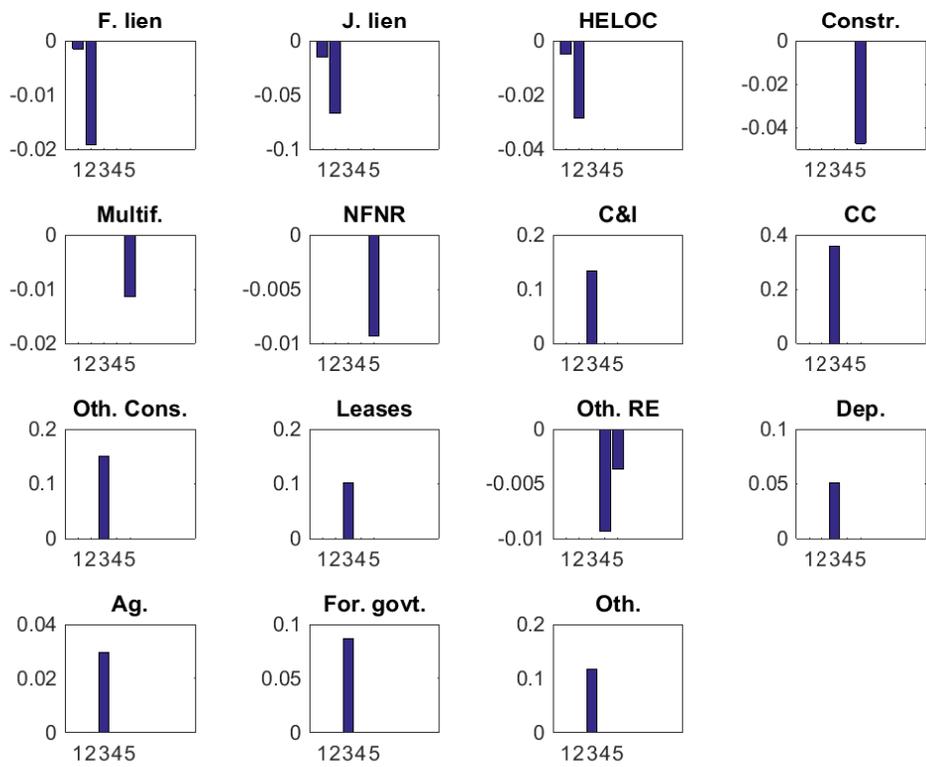


Figure 7: Variation in regression coefficients on CLASS scenario variables across components of Net charge-offs (key for scenario variables: 1) House Price Growth, 2) House Price Growth (if negative), 3) Change in Unemployment, 4) Commercial Real Estate Price Growth, 5) Commercial Real Estate Price Growth (if negative))

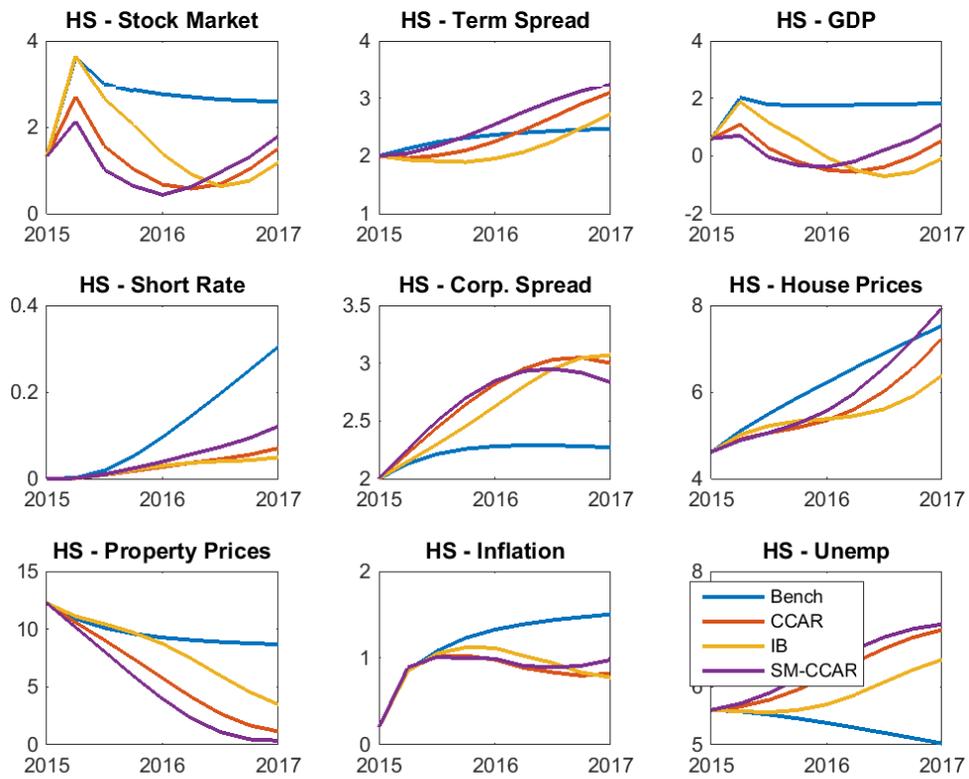


Figure 8: Comparing benchmark and HS worst case means of the CLASS scenario variables over the forecast horizon - common K-L distance

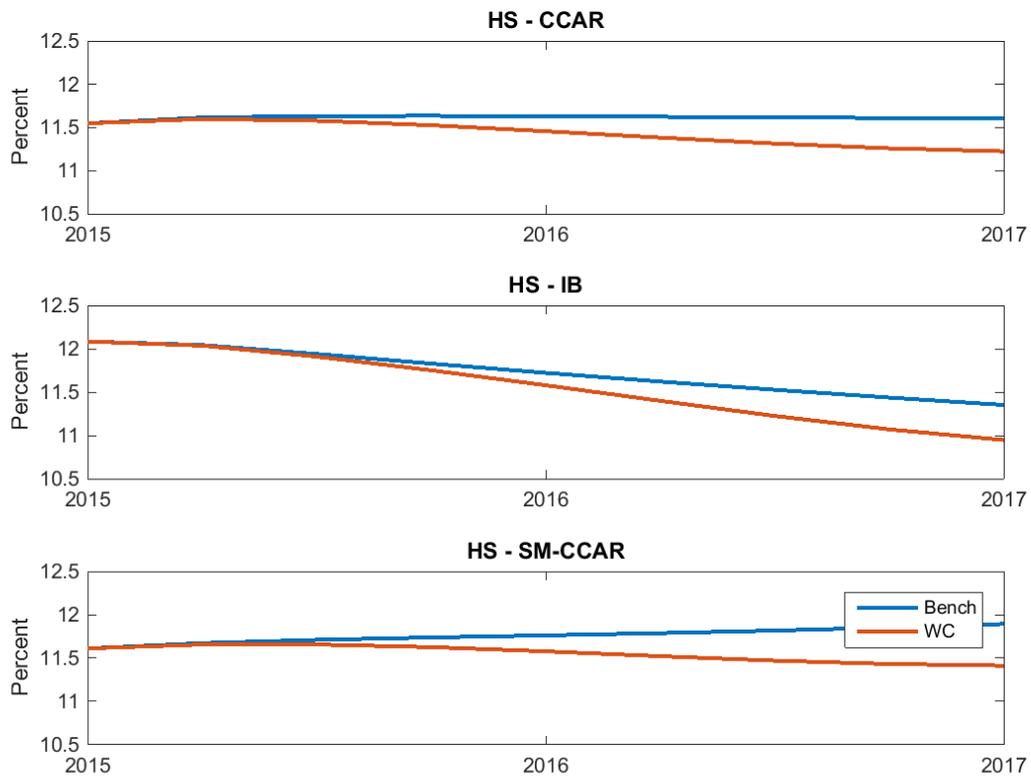


Figure 9: Comparing benchmark and HS worst case means of capital ratios over the forecast horizon

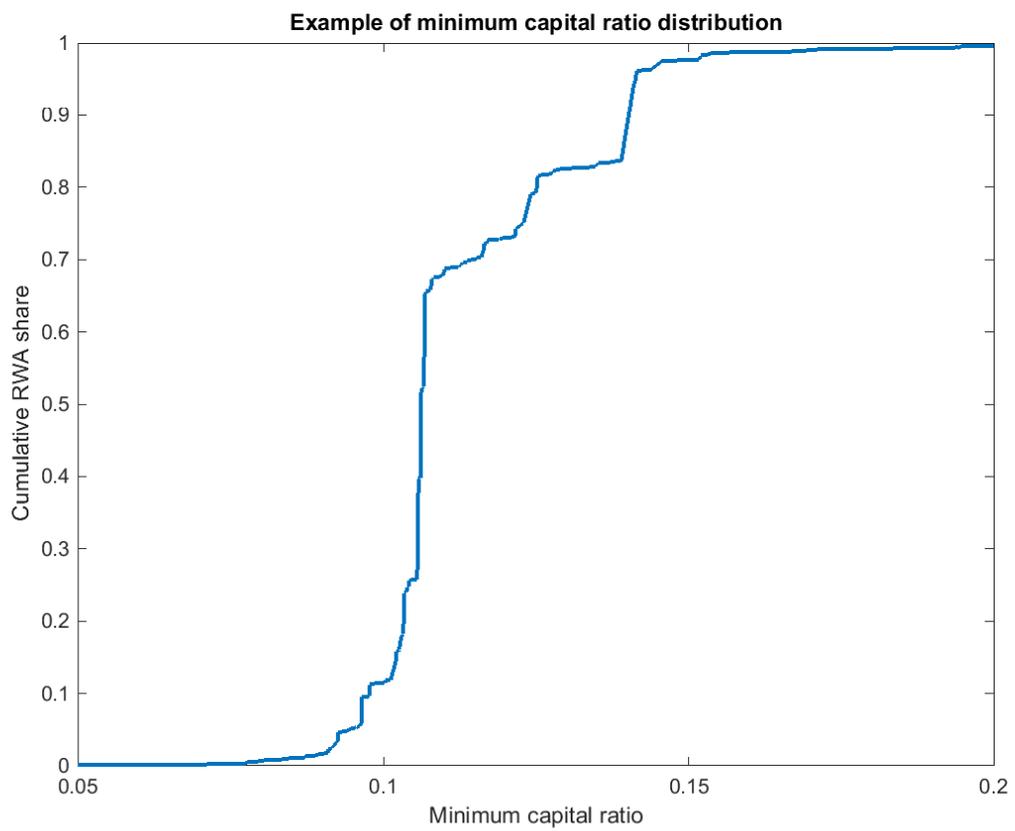


Figure 10: Example (for a given draw) CDF of minimal capital ratios (top 200 banks)

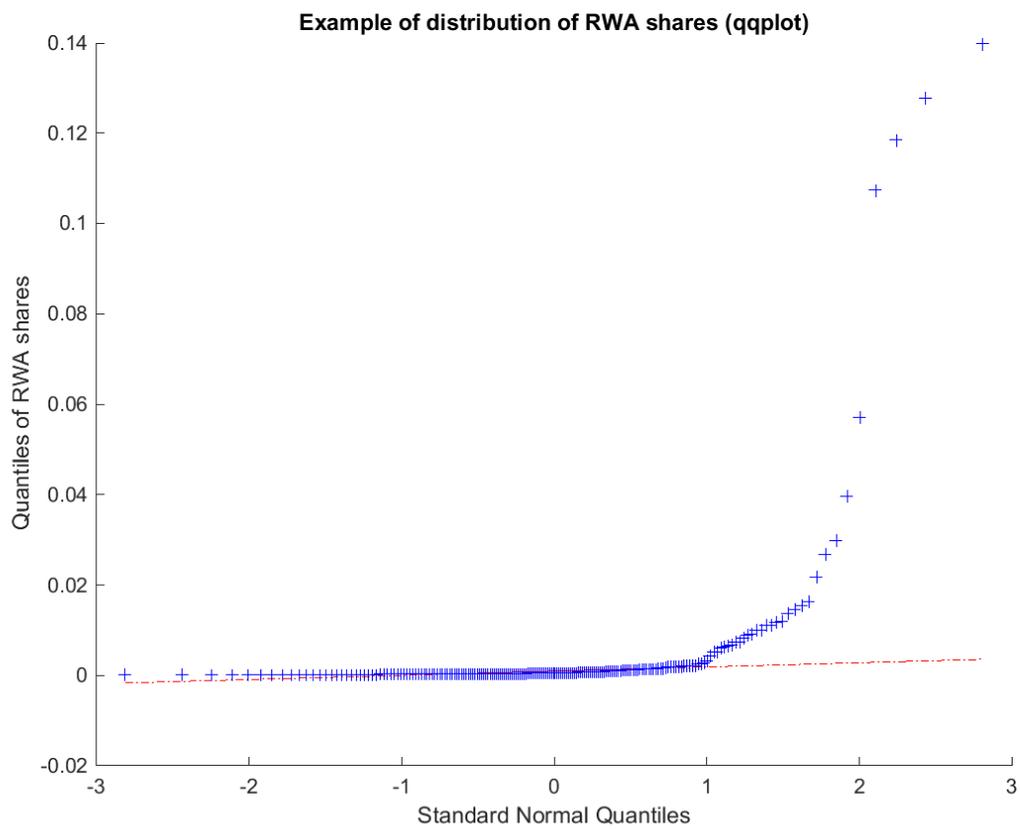


Figure 11: Distribution of risk weighted asset shares (top 200 banks)

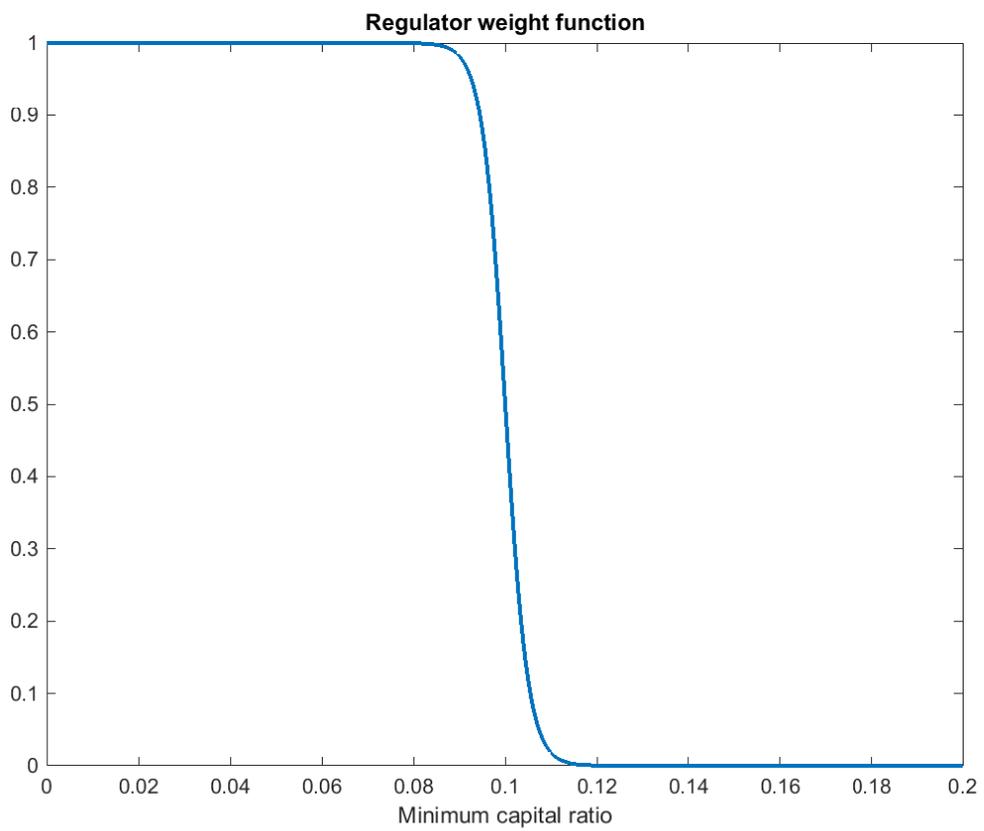


Figure 12: Regulator weight function to rank minimum capital ratio distributions

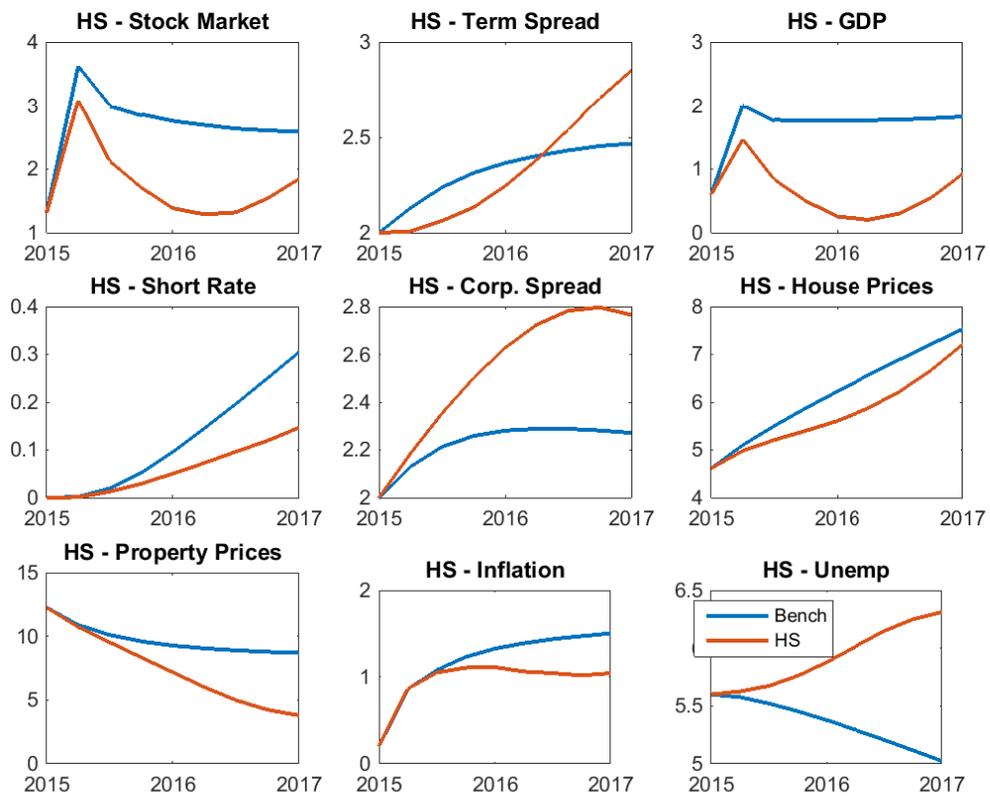


Figure 13: Comparing benchmark and HS worst case means of the CLASS scenario variables over the forecast horizon - based on minimal capital ratios of all banks.

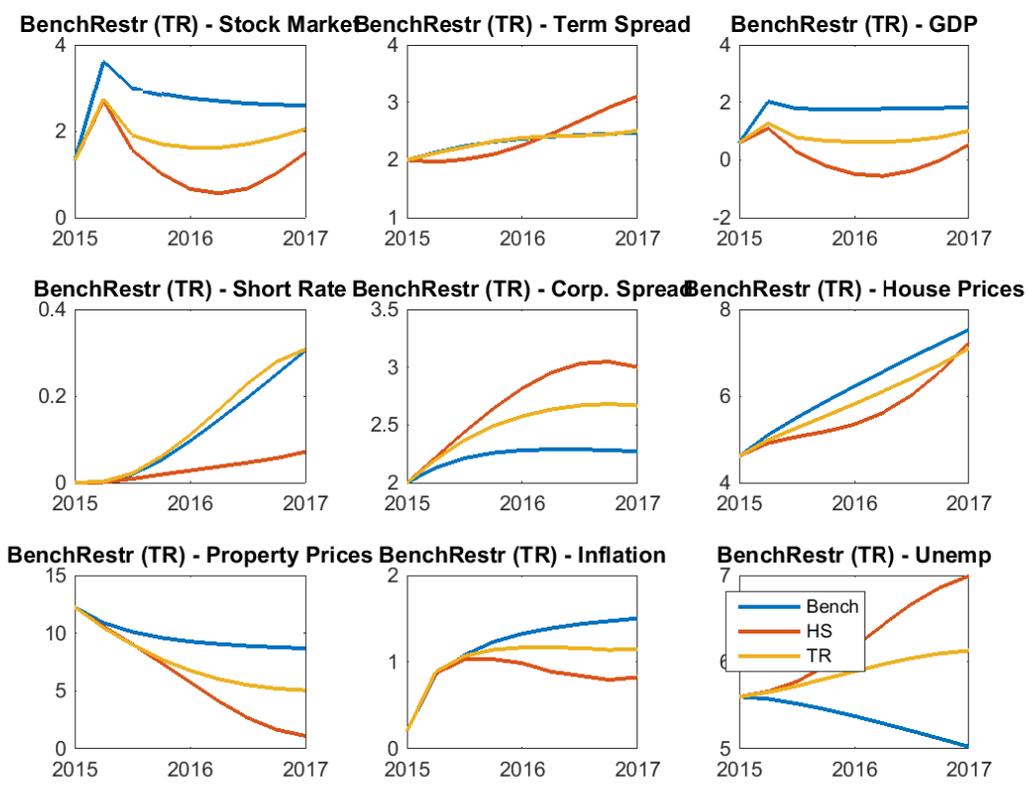


Figure 14: Comparing benchmark, HS worst case and TR worst case means of the CLASS scenario variables over the forecast horizon (tilt is to align means of the short rate, term spread and change in the 10-year Treasury yield with the benchmark model equivalents and using ‘CCAR’ bank as the example institution.)

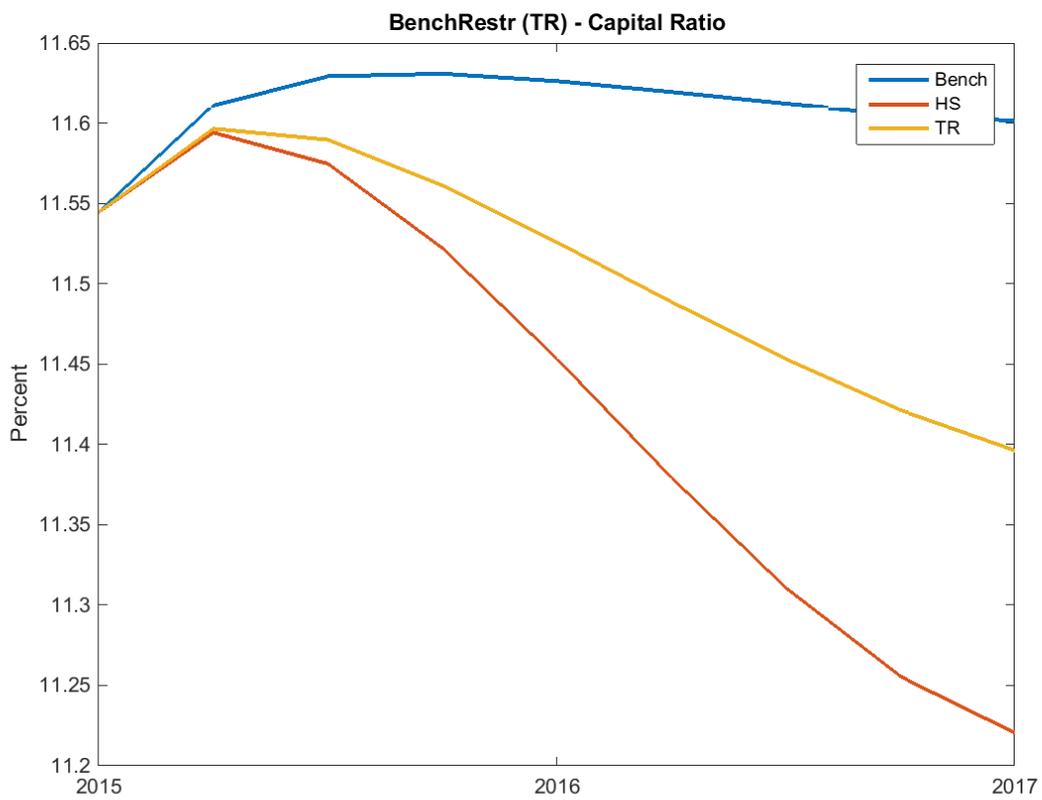


Figure 15: Comparing benchmark, HS worst case and TR worst case means of the capital ratio over the forecast horizon (tilt is to align means of the short rate, term spread and change in the 10-year Treasury yield with the benchmark model equivalents and using 'CCAR' bank as the example institution.)

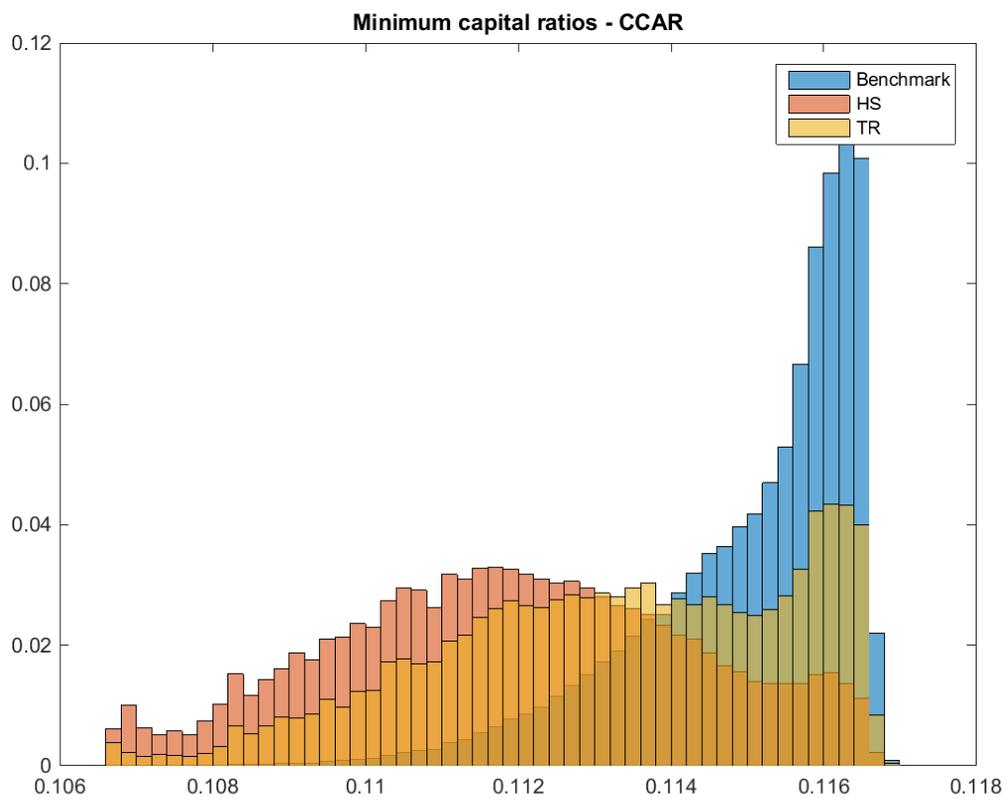


Figure 16: Comparing benchmark, HS worst case and TR worst case distributions over the capital ratio (tilt is to align means of the short rate, term spread and change in the 10-year Treasury yield with the benchmark model equivalents and using ‘CCAR’ bank as the example institution.)