FEDERAL RESERVE BANK OF SAN FRANCISCO

WORKING PAPER SERIES

Robust Stress Testing

Rhys Bidder
Federal Reserve Bank of San Francisco

Andrew McKenna
Federal Reserve Board of Governors

September 2015

Working Paper 2015-13
http://www.frbsf.org/economic-research/publications/working-papers/wp2015-13.pdf

Suggested citation:

Bidder, Rhys, Andrew McKenna. 2015. "Robust Stress Testing." Federal Reserve Bank of San Francisco Working Paper 2015-13. http://www.frbsf.org/economic-research/publications/working-papers/wp2015-13.pdf

Robust Stress Testing

Rhys Bidder[*]
Federal Reserve Bank of San Francisco

Andrew McKenna[†]
Federal Reserve Board of Governors

This draft: September 22, 2015

Abstract

In recent years, stress testing has become an important component of financial and macro-prudential regulation. Despite the general consensus that such testing has been useful in many dimensions, the techniques of stress testing are still being honed and debated. This paper contributes to this debate in proposing the use of robust forecasting analysis to identify and construct adverse scenarios that are naturally interpretable as stress tests. These scenarios emerge from a particular pessimistic twist to a benchmark forecasting model, referred to as a 'worst case distribution'. This offers regulators a method of identifying vulnerabilities, even while acknowledging that their models are misspecified in possibly unknown ways.

We first carry out our analysis in the familiar Linear-Quadratic framework of Hansen and Sargent (2008), based on an estimated VAR for the economy and linear regressions of bank performance on the state of the economy. We note, however, that the worst case so constructed features undesirable properties for our purpose in that it distorts moments that we would prefer were left undistorted. In response, we formulate a finite horizon robust forecasting problem in which the worst case distribution is required to respect certain moment conditions. In this framework, we are able to allow for rich nonlinearities in the benchmark process and more general loss functions than in the L-Q setup, thereby bringing our approach closer to applied use.

[*]Economic Research Department, Federal Reserve Bank of San Francisco, 101 Market Street, San Francisco, CA 94115; Email: rhys.bidder@sf.frb.org

[†]Economic Research Department, Federal Reserve Bank of San Francisco, 101 Market Street, San Francisco, CA 94115; Email: andrew.mckenna@frb.gov.

The authors thank Jaroslav Borovicka, Ian Dew-Becker, Fred Furlong, Raffaella Giacomini, Bev Hirtle, Anna Kovner, Jose Lopez, Yulei Luo, Marius Rodriguez, Glenn Rudebusch, Matthew E. Smith, Tom Tallarini Jr. and seminar participants at the Bank of England, HKU SEF, UC Davis, KC Fed and ECB for comments and advice. The authors also thank Ben Pyle for excellent research assistance. 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 Board of Governors or the Federal Reserve System.

Back in August 2007, the Chief Financial Officer of Goldman Sachs, David Viniar, commented to the Financial Times: 'We are seeing things that were 25-standard deviation moves, several days in a row'

. . . A 7.26-sigma daily loss would be expected to occur once every 13.7 billion or so years. That is roughly the estimated age of the universe.

. . . Fortunately, there is a simpler explanation: the model was wrong.

Haldane (2009) - Speech at the Marcus-Evans Conference on Stress-Testing, London

1 Introduction

Stress testing has become an important component of financial and macroprudential regulation. The nature of the financial crisis of 2007-9 prompted regulators and financial institutions to model multidimensional scenarios for macroeconomic and financial variables to assess their impact on firm capital adequacy and, thereby, reveal vulnerabilities in the financial system and suggest a policy response (Schuermann (2013), Furlong (2011) and Hirtle, Schuermann, and Stiroh (2009)). The Federal Reserve's Supervisory Capital Assessment Program (SCAP) and the subsequent Comprehensive Capital Analysis and Review (CCAR) exercises are perhaps the most prominent examples of the stress testing approach and are typically regarded as having contributed significantly to the strengthening of the financial system during and immediately after the recent crisis. Beyond the United States, the stress testing paradigm is also becoming more prominent, notably in the EU-wide stress tests undertaken by the European Banking Authority and the ECB.

Despite the general consensus that such testing has been useful along many dimensions, the techniques of stress testing design and implementation are still evolving. This partly reflects certain misgivings that have been raised over the nature of the stress scenarios as currently applied and debate over exactly how they should be constructed and interpreted.

In this paper, we use a stylized approach to stress testing based on the methods of robust forecasting analysis (see Hansen and Sargent (2008)) to identify and construct adverse forecasting distributions that generate scenarios naturally interpretable as stress tests. We draw on the rapidly expanding literature on model uncertainty and ambiguity to inform regulatory practice. This is a natural approach as, in the world of regulation, model uncertainty is pervasive and the pessimism that emerges from models of choice under ambiguity is a useful way of identifying vulnerabilities.

We also show how to 'focus' ambiguity within a finite horizon robust forecasting problem. We do this by requiring that the adverse forecasting distributions respect certain moment restrictions. In doing this, we retain the unstructured nature of uncertainty that is the hallmark of robustness, while disciplining the agent not to seek robustness against misspecifications that are thought *a priori* irrelevant. We refer to these techniques as 'tilted robustness'.[1] Furthermore, this approach nests more standard conditional forecasting techniques so that the methods we use provide a unified and practical approach to constructing stress scenarios, even in environments without Knightian uncertainty.

1.1 Motivation

We are motivated by ongoing debates over stress testing practice. A particularly important area for discussion is the question of how, first, to characterize the plausibility and severity of stress scenarios and, second, to decide on the appropriate tradeoff between the two. Stress scenarios should be sufficiently severe to be informative about banks' vulnerability but also not so severe as to appear absurd.

If one wishes to obtain a sense of plausibility, acknowledging that the context is stochastic, then there are many statistical tests that could be used. For example, Breuer, Jandacka, Rheinberger, and Summer (2009), Covas, Rump, and Zakrajsek (2013) and White, Kim, and Manganelli (2012), among others, discuss various approaches.[2] Although a consensus has not yet emerged on what measures should be used, a heavy emphasis is typically placed on characterizing tails of distributions, which are typically taken to be 'extreme events'. However, as noted by Haldane (2009) and captured in the quote above, one could have gone (essentially) arbitrarily into the tail of various risk management models prior to the crisis and still not revealed vulnerabilities of the sort that ultimately were exposed.

A problem with the 'tail based' approach to stress testing, therefore, is that the models underpinning them (presumably estimated from historical data) are reasonable approximations of the world but in some dimensions are misspecified in damaging and unknown ways. They are likely to be particularly misspecified in their predictions for the behavior of economies in unfamiliar times of stress. Explicitly accounting for model uncertainty is one of the most important challenges of stress test design and one which robustness is perfectly suited to address.

1.2 Why robustness?

Robust forecasting provides a formal method for confronting model misspecification and how to evaluate randomness in this context. An agent (in our case a regulator) possesses a 'benchmark' model that implies a probability distribution over random variables in the

[1] Glassermna and Xu (2013) and Glasserman and Xu (2014) use a very similar approach in the context of assessing portfolio risk, and many of these results build on the work of Petersen, James, and Dupuis (2000).

[2] See also Christensen, Lopez, and Rudebusch (2013) for similar issues raised in the context of stress testing the Fed's assets and income.

economy. The agent expresses her doubts of her benchmark model by considering alternative distributions that are twisted versions of the distribution implied by the benchmark.

In order to construct a 'robust' forecast (one that puts a particular pessimistic slant on the behavior of the financial system), the agent considers adverse distributions and balances the damage that an implicit misspecification would cause, against the plausibility of that misspecification. This yields a particular joint distribution over sequences that encodes the vulnerabilities implicit in the estimated system - dimensions in which misspecifications would be particularly painful. We use this distribution (typically referred to in the robustness literature as a 'worst case distribution') to generate candidate scenarios and simulations for stress tests.

Under the robustness approach there is a very clear and tightly parameterized tradeoff between severity of scenarios and their plausibility. This tradeoff is expressed in terms of probability 'distributions' rather than 'realizations'. The latter, although easier to plot in diagrams, are arguably difficult to interpret and utilize in a comprehensive risk management framework. The explicit acknowledgement of model misspecification also, to some extent, protects the agent from a false sense of security based on calculating tail probabilities. Historical data informs the process (the benchmark model will presumably be based on it) but the robust regulator allows for other possible data generating processes and concentrates on those that would be damaging.

To the extent that historical experience contains useful information, the methods we propose are very 'informationally efficient' in the sense that the benchmark model identifies suggestive dimensions in which the system is vulnerable. This (very complicated) information is encoded into the worst case distribution via a particular change of measure, based on the forecaster's 'value function'. Thus, it can reveal very subtle and perhaps counterintuitive comovements and correlations that would be damaging but, also, unlikely to be anticipated by a regulator using introspection to identify vulnerabilities (see Schuermann (2013)). Robustness helps identify and confront 'unknown unknowns' by yielding a worst case distribution with statistical properties that can then be interpreted and used to identify economically interpretable weaknesses in the system that could lead to similar statistical properties or scenarios.[3,4]

Another useful feature of the robustness approach is that what constitutes the worst case distribution will depend on what variables the regulator cares about. That is, if the regulator's

[3] See Borio, Drehmann, and Tsatsaronis (2013) and Abdymomunov and Gerlach (2013) for discussions of the difficulty of envisaging vulnerabilities in tranquil times and the difficulty of using historical experience and hypothesizing appropriately.

[4] Breuer and Csiszr (2013) appeal to much of the same intuition that we do and formulate similar finite horizon problems but without the presence of tilting to respect moment conditions.

benchmark model is concerned with exposures of the aggregate system then the scenarios generated from the worst case will be systemically damaging. If, however, the regulator's benchmark connects the performance of a particular bank or group of banks to the state of the economy, the worst case will be different, reflecting the different vulnerabilities. Therefore, if desired, we can tune scenarios, rather than constructing a 'one size fits all' scenario. This might be relevant in cross checking bank specific scenarios volunteered by banks themselves as the scenarios they feel would be most dangerous.[5]

Finally, the robustness framework is very parsimonious and (fairly) easily explainable. It features very few parameters (beyond those in the benchmark model), a clear connection to axiomatized decision theory and a precise sense of plausibility that can be used to defend the final scenarios if they are accused of being inappropriate in the future.[6]

1.3 Application

We begin with a Gaussian Vector Autoregression (VAR) model of macroeconomic and financial variables as our benchmark law of motion for the state of the economy. We estimate linear regressions of measures of aggregate bank performance on the state variables, where our measures of performance is aggregate return on equity. In terms of period payoff, we adopt a quadratic form. Given this Linear-Quadratic setup we can appeal to the results discussed in Hansen and Sargent (2008) and directly compute the worst case distribution, implicit in a 'worst case VAR'.

Although we show that the worst case arising from the simple L-Q model has some interesting properties, it also features some undesirable characteristics. In particular, the long run properties of the economy and, in particular, the unconditional means under the worst case, seem implausible and qualitatively unlike what a regulator might wish for in constructing stress scenarios. For this reason, we undertake a similar (though not recursive) analysis whereby we restrict these moments when constructing the worst case.

At first we adopt an apparently mechanical method of ensuring that certain moments are respected. We first derive the distribution over a finite forecast horizon, implied by the worst case from the basic L-Q approach. We then use exponential tilting to adjust, *ex post*, this distribution to respect certain moment restrictions. This is a way of embedding our

[5]Pritsker (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.

[6]Anticipating greater opposition in to future stress tests, despite initial successes, seems an important concern. Indeed, quoting Hirtle, Schuermann, and Stiroh (2009), 'The positive reaction to [the] release of the SCAP results may not have been [because of] transparency *per se*, but simply because the results were viewed as credible [...] Whether the reception would have been positive if the results and process were not seen in this way or if there had been a negative 'surprise' about a firm or group of firms is open to debate.'

judgement that certain properties of the basic worst case are inappropriate.[7] Within this context we also show how to nest 'direct' (conditioning on adverse behavior of the economy) and 'reverse' (conditioning on adverse behavior of a bank) approaches to stress testing and relate it to conditional forecasting techniques as discussed in Waggoner and Zha (1999).

We go on to show, however, that this apparently mechanical approach can be understood as the outcome of a robustness problem in which the agent only considers worst case distributions that satisfy certain moment conditions. We refer to this approach as 'tilted robustness' and show that it relates to the important and delicate issue of 'focusing' ambiguity.

2 Comprehensive Capital Analysis and Review (CCAR)

We briefly describe the nature of the Comprehensive Capital Analysis and Review (CCAR) program undertaken by the Federal reserve, as it can make concrete some of the issues we hope to address.

CCAR is run annually with the aim of ensuring that bank holding companies' (BHC) capital planning is robust to periods of economic and financial adversity, so that they are able to continue operation during such environments. An important element of the framework is the provision of supervisory stress test scenarios under which the institutions capital adequacy is assessed. These scenarios are not necessarily 'likely' but are regarded as valuable inputs into the regulatory process. Indeed, quoting the CCAR documentation (see of Governors (2012))

> ...the Supervisory Stress Scenario is not a forecast, but rather a hypothetical scenario to be used to assess the strength and resilience of BHC capital in a severely adverse economic environment

Assessment of banks under the stress scenario focuses on the nature of the banks' proposed capital plans and, in particular, whether the institutions are able to maintain capital above certain minimum levels throughout the planning horizon. The scenarios considered are in terms of a variety of macroeconomic and financial data series.[8] Three supervisory scenarios are considered: baseline, adverse and severely adverse. The baseline scenario can be thought of as similar to a reasonable forecast of a likely path of the economy. The other two scenarios capture hypothetical paths of varying severity. In figure 1 we include examples of the baseline, adverse and severely adverse scenarios for real GDP growth, the yield on 10-year Treasuries, the unemployment rate and the yield on 3-month Treasury.

[7] See Robertson, Tallman, and Whiteman (2005), Cogley, Morozov, and Sargent (2005) and Giacomini and Ragusa (2013) for discussions and applications of exponential tilting.

[8] For a complete list of series for the supervisory scenarios and further details see http://www.federalreserve.gov/bankinforeg/stress-tests-capital-planning.htm.

3 Seeking Robustness - The Hansen-Sargent Approach

In this section we lay out an approach to stress testing, based on a robust forecasting problem. We begin with an abstract problem and then specialize to a workhouse linear-quadratic Gaussian (LQG) setup, which is common in the robustness literature (as described in described in Hansen and Sargent (2008)). Although we will ultimately advocate a different, but related approach, much of the intuition of this section will be retained.

A robust agent 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 by her benchmark model. 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 can be thought of as a 'worst case distribution' that encodes these concerns and allows insight into the fears that inform the agent. We now formalize this intuition.

3.1 General Case

Let us suppose that the robust agent entertains a benchmark model in which the state and innovation sequences are related according to the (possibly nonlinear) vector valued equation

$$x_{t+1} = g(x_t, u_t, \varepsilon_{t+1}) \tag{1}$$

where x_t is the state vector and $\{\varepsilon_t\}$ is a sequence of random innovations. Given a density, $p_\varepsilon(\varepsilon_{t+1}|x_t)$, for ε_{t+1}, equation (1) implies a benchmark transition density $p(x_{t+1}|x_t)$. It is convenient to partition the state, x_t into elements unknown on entering the period, which we identify with ε_t, and those elements that are predetermined, denoted s_t. We capture the dependence of s_t on the state prevailing in the previous period by the function f, such that $s_t = f(x_{t-1})$. With this decomposition we have $p(x_{t+1}|x_t) = p_\varepsilon(\varepsilon_{t+1}|x_t)\delta_{f(x_t)}(s_{t+1})$.[9]

We adopt multiplier preferences as a way of representing the agent's fear of model misspecification (Hansen and Sargent (2008)). In this case, the value function of the agent satisfies

$$V_0 = \min_{\{m_{t+1}\}} \sum_{t=0}^{\infty} E\left[\beta^t M_t \left\{h\left(z_t\right) + \beta\theta E\left(m_{t+1}\log m_{t+1}|\Im_t\right)\right\}|\Im_0\right]$$

[9]Note that the x_t may contain ε_t as an element of the state so that an identity mapping is implicit in g. Note also that $\delta_{f(x_t)}(\cdot)$ takes the value of unity at $f(x_t)$ and zero elsewhere. In the case of our benchmark VAR models below, $x_{t+1} = Ax_t + C\varepsilon_{t+1}$ and, thus, $f(x_t) = Ax_t$.

where $h(\cdot)$ is the robust agent's period payoff function, $z_t = g(x_t)$ is a 'target' variable related to the state by the function $g(\cdot)$ and the problem is subject to equation (1), $M_{t+1} = m_{t+1}M_t$, $E[m_{t+1}|\Im_t] = 1$, $m_{t+1} \geq 0$ and $M_0 = 1$.[10] Thus, $\{m_{t+1}, t \geq 0\}$ is a sequence of martingale increments that recursively define a non-negative martingale $M_t = M_0 \prod_{j=1}^{t} m_j$. The martingale defines Radon-Nikodym derivatives that twist the measures implicit in the benchmark model so as to obtain absolutely continuous measures that represent alternative distributions considered by the agent. Under these twisted measures one can form objects interpretable as expectations taken in the context of a distorted alternative model. This can be seen by defining a distorted conditional expectation operator to be

$$\tilde{E}[b_{t+1}|\Im_t] \equiv E[m_{t+1}b_{t+1}|\Im_t]$$

for some $\Im_{t+1}$ measurable random variable b_{t+1} given $\Im_t$. The conditional relative entropy associated with the twisted conditional distribution is given by the term $E[m_{t+1} \log m_{t+1}|\Im_t]$, which is a measure of how different the distorted measure is from the benchmark. We refer the reader to Strzalecki (2011) and Maccheroni, Marinacci, and Rustichini (2006) for axiomatic foundations of multiplier preferences.

The agent's desire for a robust evaluation of the stochastic process for the target is reflected in the minimization over the sequence of martingale increments that twist the distributions used to evaluate continuation values towards realizations of the state that are painful to the robust agent. 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 our earlier 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. The m that solves the minimization, balancing concerns of 'pain' and 'plausibility', implies a particular distribution, which is typically labeled the 'worst case distribution'.[11]

Typically one would discipline θ with detection error probabilities, which relate to whether or not, with a limited amount of data, an agent could accurately distinguish between the worst case and benchmark distributions using likelihood ratio tests. If the two models have similar stochastic properties, they will be difficult to detect using sample sizes that are typically available for analysis. In this case the detection error probability will be close to 0.5, indicating that the models are almost indistinguishable. Models that have very different characteristics will be easily distinguishable and imply a detection error probability of close to 0. High

[10]We assume that the robust agent's information set, $\Im_t$ contains the entire history of states.

[11]Note that the word 'case' does not, in this context, imply a particular realization - we are working in terms of distributions and, indeed, the sense in which it is 'worst' is also restricted.

detection error probabilities suggest that the competing models are hard to distinguish using the amount of data available and thus represent misspecifications that it is plausible to worry about. We will make use of detection error probabilities in our calibration below.

3.2 Linear Quadratic Gaussian Framework

We now specialize to a standard LQG robust forecasting framework. Although this framework puts strong restrictions on the nature of the worst case, which we will discuss below, it provides a familiar setup that can illustrate the concepts at play. As noted in Bidder and Smith (2012) highly non-linear models can be accommodated within the robustness framework although at increased computational cost.

We posit a linear transition law for the state, x_t, given by

$$\begin{aligned} x_{t+1} &= Ax_t + C\hat{\varepsilon}_{t+1} \\ \hat{\varepsilon}_{t+1} &\sim N(0, I) \end{aligned}$$

where u_t is a vector of controls and $\{\hat{\varepsilon}_t\}$ is an *iid* sequence. The mapping from the state, x_t to the target, z_t will be as follows. A payoff variable c_t and a 'bliss' point variables, b_t are related to the state by

$$\begin{aligned} c_t &= H_c x_t \\ b_t &= H_b x_t \end{aligned}$$

If we let $z_t \equiv c_t - b_t$ and $H \equiv H_c - H_b$, then the period payoff is given by $g(z_t) = z_t' W z_t$, a quadratic form where W captures the weighting scheme. It is useful to note that the period payoff can also be expressed as $x_t' Q x_t$ where $Q \equiv H' W H$.

Given this framework the worst case distribution over sequences can be represented recursively in a particularly tractable form, as shown in appendix 7.2. That is, the worst case transition law is given by a VAR, that is a distorted form of the VAR implied under the benchmark model.

$$\begin{aligned} x' &= \tilde{A}x + \tilde{C}\varepsilon' \\ \tilde{A} &= A + CK \\ K &= \theta^{-1}\left(I - \theta^{-1}C'PC\right)^{-1} C'PA \\ \tilde{C}\tilde{C}' &= C\left(I - \theta^{-1}C'PC\right)^{-1} C' \end{aligned}$$

This transition law, and its repeated application allows us to characterize and draw from the worst case distribution over sequences that emerges from the agent's robust forecasting problem.

It is worth noting here that there are various ways of representing the worst case distribution and it is left unspecified what structural misspecification might implicitly be underpinning it. The joint distribution over sequences under the benchmark is, in our baseline case, easily described by a VAR, so that we can speak naturally of 'processes'. Similarly, the worst case distribution to emerge from the infinite horizon L-Q robustness problem also can be induced, or 'represented', by a worst case VAR with adjusted autoregressive and Cholesky matrices. But this is only a representation. For example, the worst case distribution over sequences could alternatively be expressed in terms of (state dependent) distorted innovation distributions for ε_t. This will be emphasized in section 5.2 where we discuss our tilted robustness approach. In that (finite horizon) case we specify the benchmark in terms of a joint forecasting distribution over a sequence, without reference to transition laws or 'innovations'. Similarly, the worst case is simply expressed as a tilted distribution, rather than a 'process'.

4 A Baseline Stress Testing Framework

We cast a regulatory problem into the robust forecasting framework laid out in section 3. Clearly, one would wish to allow for the regulator to operate a control, but developing such a model in this framework is beyond the scope of the paper. Nevertheless, a useful first step is to imagine the regulator assessing the behavior of the financial system, when left to its own devices, as a first step in evaluating how to frame the regulatory environment.

We lay out the (preliminary) benchmark models that will underpin our analysis.[12] We posit a Gaussian VAR for three standard 'macroeconomic' variables (the unemployment rate, inflation and a short rate), augmented with 'financial' variables (the change in the stock market and a term spread). In addition, we estimate linear loadings of target variables on the state (based on regressions of bank performance on the state). Our measure of performance will be annualized Return on Equity (RoE). We assert that the robust regulator/forecaster has a quadratic period payoff in RoE but with a distant satiation point to capture monotonicity in preferences.

4.1 Data

The macroeconomic series in our 'state' are the civilian unemployment rate (LR), Core PCE Inflation (log difference of JXFE) and the 3 Month Treasury Bill secondary market rate (FTBS3). The financial series that we also include in the state are the log change in the

[12] As noted in Covas, Rump, and Zakrajsek (2013) and Guerrieri and Welch (2012) it can be difficult to obtain reliable and strong predictive relationships in stress testing models based on linear frameworks. This is particularly the case when looking at aggregate variables. In ongoing work we intend to enhance our empirical analysis making use of macroprudential data obtained from filings required after the Dodd-Frank reforms.

quarterly Dow Jones stock market index (SPDJI) and the spread between 10 Year Treasuries and the 3 Month rate (FCM10 - FTBS3), where all series codes are from HAVER. Our intention is to begin with a 'standard macro' VAR and then augment it with financial series, whilst retaining sufficient parsimony to estimate the system. The data sample we use to estimate the VAR is from $1975Q2$ to $2011Q3$. We estimate a first order VAR.[13]

We choose to use return on equity (ROE) as the target variable of interest to the robust forecaster. Although there are various concepts of 'payoff' that we could employ and which might be of interest to a regulator to consider, ROE appears a natural starting point for our analysis, although effects of leverage are a concern. We use data obtained from the New York Fed's 'Quarterly Trends for Consolidated U.S. Banking Organizations' website.[14] We estimate simple OLS regressions of RoE series for the aggregate system and for individual institutions on the contemporaneous values of the VAR state and take the estimated coefficients as defining the loadings of the target variable on the state. Our sample is for $1991Q1 - 2011Q3$.

4.2 Estimated Banking System Exposures

In table 1 we report our regression results for the full sample and for two particular financial institutions, where the explanatory variables were first standardized to have zero mean and unit standard deviation.[15]

We first concentrate on the column of table 1 corresponding to the aggregate banking system. We note that the coefficients typically have economically intuitive signs and, in many cases, are statistically significant. For example, higher unemployment, higher stock-market growth and higher term spreads are associated with higher annualized return on equity. Inflation appears to enter negatively, with varying degrees of significance and the short rate appears not to exhibit a significant statistical relationship *ceteris paribus*.[16]

Turning to the institution-specific regressions, we see broadly similar tendencies although there are differences in magnitude and significance. This implies, as will be shown below,

[13] The Akaike Information Criterion (AIC) favors two lags although since this model is intended primarily as an expositional device and additional lags bring very little of interest to our qualitative story, we simplify to a single lag specification. We are actively working on developing a more convincing benchmark although, as section 5 suggests, this will likely not be a Gaussian VAR.

[14] See the New York Fed's quarterly trends for data and associated documentation. Out concept of equity is Tier 1 common equity = tier 1 capital - perpetual preferred stock and related surplus + non-qualifying perpetual preferred stock - qualifying Class A non-controlling (minority) interests in consolidated subsidiaries - qualifying restricted core capital elements (other than cumulative perpetual preferred stock) - qualifying mandatory convertible preferred securities of internationally active bank holding companies.

[15] All the data we use is public but we choose not to identify the two institutions used as the analysis is preliminary in its current form and meant only to be illustrative.

[16] Regardless of significance, we will take the point estimates as given in our following analysis.

that the predictions of the robust forecasts will vary according to whether the forecaster is assumed to be concerned with the aggregate or a particular institution. These varying exposures, which we will take as variations in the maintained benchmark, will ultimately lead to a different worst case distribution, since they hint at different dimensions in which misspecification in the benchmark model would be damaging. That is, our stress scenarios will be tuned according to the regulator's desired emphasis or mandate.

We will defer our discussion of the properties of our estimated VAR models until section 4.4 since it is most natural to characterize the implied moments, when comparing and contrasting them to those of the worst case distributions we derive.

4.3 Calibrating Preference Parameters

We calibrate the satiation point of the forecaster's quadratic period payoff to imply a coefficient of relative risk aversion of unity, when RoE is at its ergodic mean under the benchmark model. It is not immediately obvious that this is the appropriate approach. Return on equity, being a scaled (by equity) version of net income may not be an object over which one can plausibly define a utility function that is then tuned to yield a particular value of a concept of aversion to risk. Nevertheless, this approach yields a satiation point that is distant (in terms of standard deviations under the benchmark) from the average level of RoE and thus ensures that, despite using quadratic preferences, the model typically operates within a region where preferences are essentially monotonic, which seems plausible.[17]

To calibrate the degree of the forecaster's aversion to model uncertainty, we employ detection error probabilities to assess the plausibility of our calibration. We will use the worst cases derived under a pair of detection error probabilities (DEP) of 0.2 and 0.1, with a sample size of 100 observations. The former calibration can be regarded as implying a fairly plausible degree of aversion to model uncertainty (see Hansen and Sargent (2008) for a further discussion).[18] The latter calibration is more extreme (though not ludicrous) and is intended to help illustrate the qualitative nature of the worst case distortions. More practically, we follow the CCAR approach in providing scenarios of differing degrees of severity.

[17]Fixing the satiation point in this way also ensures that we obtain intuitive properties of the worst case in that low RoE is revealed as a 'bad'. Initially we fixed the satiation point to be two standard deviations above the average RoE but this seemed somewhat arbitrary. Nevertheless, when we target a risk aversion of 1 under the benchmark for the aggregate banking system, estimated with the full sample, it happens to imply a similar satiation point.

[18]However, we initialize our simulations from a point common to both the worst case and benchmark simulations.

4.4 Results

In this section we discuss our results from our baseline L-Q framework. We emphasize the differences between the benchmark and worst case models since these are suggestive of the nature of the implicit misspecifications that particularly concern the robust forecaster. The worst case VARs can be used to derive various moments of interest. We will briefly lay out some unconditional moments, but our main aim is to generate stress scenarios.

4.4.1 Unconditional Moments

In table 2 we depict unconditional means and standard deviations of aggregate RoE and the state variables under the benchmark and under the worst case.[19] We observe that the average return on equity is markedly lower under the $DEP = 0.2$ case and even lower with $DEP = 0.1$, to such an extent that it is substantially below zero. In addition, we observe pessimistic upward distortions of the unconditional standard deviation of RoE under the worst cases. Both these patterns are to be expected: the forecaster fears distortions to his model that would induce lower and more volatile payoff (recall we have set the satiation point so that the agent essentially has monotonic preferences, at least within a range of realizations that are likely under the benchmark or nearby models).

In addition, we observe that the volatilities of the state variables are also inflated, although the volatility of stock market growth is only slightly distorted. With regard to the means, the patterns of distortions are not entirely intuitive given the signs of the estimated exposure coefficients in table 1. The somewhat counterintuitive signs likely reflect patterns of unconditional correlation among the states that render the nature of the worst case somewhat hard to predict.

In table 3 we observe the unconditional correlations among the states under the benchmark, the worst case based on aggregate RoE regressions and the worst case based on the first institution's RoE regression. Both worst cases are calculated for $DEP = 0.1$. It is perhaps illustrative to concentrate on the $\{2, 1\}$ and $\{3, 1\}$ elements (the correlations between unemployment and inflation and between unemployment and the short rate). We see that under the benchmark the $\{2, 1\}$ term is slightly positive (0.14) whereas under the worst cases the positive correlation is exaggerated, to 0.31 and 0.28, respectively. This suggests that a world in which unemployment and inflation are more positively correlated than in the data would be damaging for the health of the banking sector. This might be interpreted as suggesting a

[19]Note that the state variables were standardized to be mean 0, standard deviation 1 before estimation of the VAR and we did not restrict the intercepts in the VAR equations to be zero. Our VAR(2) specification performs better in terms of capturing the standardizations in the estimation, but we continue with the VAR(1) setup as we intend the framework to be illustrative.

fear that the benchmark model underestimates the importance of supply shocks - although this particular structural interpretation is not strictly implied by the worst case.

With regard to the $\{3,1\}$ term we see the moment is barely changed in the 'aggregate' worst case (from -0.49 to -0.53), but moderated in the 'institution 1' worst case (from -0.49 to -0.37). In the latter case it appears that misspecifications representable by a weaker negative correlation between unemployment and the short rate would be damaging for the financial system. An example of a structural phenomenon that could give rise to this, and perhaps also relates to the distortions to the $\{2,1\}$ term, might be a world where 'stagflation' is a problem, necessitating the raising of rates in environments where, despite high unemployment, inflation is also high. This interpretation is perhaps more than our preliminary empirical analysis can bear, but it illustrates the sort of suggestive and diagnostic role the worst case can play.

4.4.2 Stress Scenarios

In this section we consider the evolution of the economy under benchmark and worst cases. We will identify stress scenarios with conditional mean paths, from particular initial conditions. Clearly, though, the worst case distribution can be used to generate quantile-based paths and many other moments.

In figure 2 we depict evolutions from a given initial condition. The initial condition is common to all the paths and is based on what would prevail after an orthogonalized unemployment shock under the benchmark, given the economy was initially at steady state. The worst cases are calculated taking aggregate RoE as the target variable. This diagram is a stylized version of figure 1 (taken from the 2013 CCAR stress scenarios) - these are our 'robust stress test' scenarios.

We think of evolution under the benchmark as being an expression of the 'baseline' scenario of CCAR and the evolutions under the worst case as an expression of the adverse and severely adverse scenarios. The degree of severity is identified with detection error probabilities of 20% and 10%. We also depict the associated paths for aggregate RoE in figure 3.

In addition, as aforementioned, one can 'tune' the worst case according to the variables with which the regulator is concerned. If she wishes to focus on a particular institution or class of institutions, this is easily handled in the robustness framework. In figure 4 we depict benchmark and worst case scenarios ($DEP = 0.2$) based on aggregate RoE and the two specific institutions discussed above. We observe that the scenarios differ substantially, indicating that the worst case distributions differ and, implicitly, that the vulnerabilities the regulator should worry about are different. One could perhaps also envisage this analysis being used in sub-samples to identify how vulnerabilities vary over time.

Furthermore, a regulator could choose to 'tweak' her VAR or regression loadings on the basis of extra-model information or a hunch about a structural break within the financial system. It may be unclear intuitively how this suspected structural break may affect the vulnerabilities of the system in a complicated, multivariate model (CCAR has more than 20 dimensions to its scenarios).[20] However, to the extent that the regulator's judgement and the benchmark model are *reasonable* descriptions of the world (even if they are flawed), the process of calculating the value function and associated worst case change in measure can be used as a powerful dimensionality reduction tool to reveal weaknesses that would have been difficult or impossible to intuit.

However, one could argue that the scenarios (and indeed other moments of the worst case distributions) exhibit certain undesirable properties. First, from a mechanical perspective, they do not seem to share some of the qualitative properties of the CCAR scenarios we depict in figure 1. In figure 1 we observe a tendency for reversion that gives the impression of the stress scenarios featuring transitory events, from which the economy will ultimately recover. Second, and related, we might believe that the unconditional 'average' behavior of the economy is (reasonably) well captured by our models, even if there is concern that the economy is vulnerable to dramatic stresses.[21] Thus we may be suspicious that, although our analysis is disciplined by DEPs, the worst case means that we report in table 2 may be *conceptually* implausible.

The scenarios we depict under the worst case, in fact, largely reflect transitions to the differing stochastic steady states detailed in table 2.[22] Given the unstructured nature of uncertainty that is implicit in robustness analysis, all moments are regarded as possibly misspecified by the forecaster. A particularly damaging misspecification would be one representable by a negative distortion to the unconditional mean of RoE, due to the presence of a the satiation point in our analysis. Hence, the worst case captures this. In section 5 we will lay out methods to discipline the worst case and the dimensions in which we allow for misspecification.

[20]Schuermann (2013) addresses this issue: 'Finding coherent outcomes in such a high dimensional space, short of resorting to historical realizations is daunting indeed... Compounding this problem is the challenge of finding a scenario where the real and the financial factors are jointly coherent.'

[21]We are using these scenarios as a vehicle for our discussion and, again, for official positions on the nature of these objects we refer the reader to http://www.federalreserve.gov/bankinforeg/stress-tests-capital-planning.htm.

[22]There are *some* distortions to transitory dynamics also, as indicated by the impulse response diagrams shown in figure 6. In addition, one might argue that we do or should doubt the unconditional properties of our models - perhaps reflecting a concern that what we perceive as a sustainable pace of financial innovation or a 'technology miracle' is actually wishful thinking.

5 Tilted Robustness

In this section we examine a different, yet closely related approach to obtaining a robust forecast in which we restrict the worst case distribution beyond simply penalizing deviations from the benchmark in terms of relative entropy. We will begin by taking an apparently *ad hoc* method to impose that the distributions generating our scenarios respect certain moment conditions, using the techniques of exponential tilting. In doing so we illustrate connections with conditional forecasting and show how to nest 'direct' and 'reverse' stress testing within our framework. However, we go on to show that these apparently mechanical methods can be interpreted from a particular decision theoretic framework, which we refer to as 'tilted robustness'.[23]

5.1 Exponential Tilting

Suppose we judge that certain moment restrictions should be respected by a model. Typically we will be concerned with situations where the restrictions we wish to impose are not initially respected. We therefore must adjust the distribution. However, we wish to do this in a way that incorporates the 'least' extra-model information, thus minimizing the manipulation of the distribution and retaining parsimony.

5.1.1 Theory

Let X denote a generic random variable with p.d.f, π. Let us express the desired moment restriction as

$$E\left[g\left(X\right)\right]=0$$

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

$$\min_{\pi^*}\int\log\frac{\pi^*\left(X\right)}{\pi\left(X\right)}\pi^*\left(X\right)dX$$

such that

$$\int g\left(X\right)\pi^*\left(X\right)dX=0$$

[23] For a more thorough treatment of 'tilted robustness' with an application to interest rate risk, see Bidder and Giacomini (2015).

which yields a solution, π^*, where

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

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

As we show in appendix 7.4, in the Gaussian case where only first and second moments are restricted by g, the exponential tilt can be derived analytically and returns a tilted distribution that remains Gaussian. 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

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

and set

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

5.1.2 Application

We can use exponential tilting to twist the worst case forecasting distributions obtained in section 4.4. For example, in figure 8 we illustrate the conditional mean paths for RoE under the distributions implied by the benchmark and worst case VARs (based on aggregate RoE regressions and with $DEP = 0.2$), along with the path obtained by tilting the distribution implied by the worst case VAR but to respect the moment condition that at the end of the forecast horizon, it should have the same mean as under the benchmark.[24] In this linear Gaussian case with a restriction in terms of the first moment, we can calculate the twisted Gaussian distribution analytically and its associated conditional mean path. We see the red line bend back up to strike the benchmark at the final horizon. Thus, we have 'fixed' one of the undesirable properties of the L-Q worst case distribution. We also show, in figure 9, the associated conditional mean paths for the states.

[24] We actually also restrict the standard deviation at the final horizon to be the same as under the benchmark and we have added a small amount of measurement noise to the system (an additive shock to the connection between RoE and the states), for mechanical coding reasons.

There are several points to note from these diagrams. Firstly, although our restriction is only in terms of the final forecast horizon, we do not see the tilted worst case aligned with the untilted until the final horizon and then leap discontinuously up (in terms of conditional means). The reason for this is that the distortion to the distribution is derived in the most efficient manner, and it happens to be the case that the K-L divergence minimizing way of imposing the desired moment, is to distort other moments also. This is obvious in the shape of the red path for RoE and the differing shapes of the paths for the states. Note also, that in the final horizon, the states do not return to their benchmark means, despite the fact that RoE is restricted to do so.

Although we have not done so here, we could require that the divergence between the distributions generating the blue (benchmark) and green (untilted W.C.) lines and between the blue and red (tilted) worst case, are equal. Intuitively, one might expect that, once one has removed the final horizon mean from the set of moments that can be distorted (which was the dominant feature of our untilted worst case), there is greater scope for more interesting conditional moment and co-movement distortions among the variables. This appears to be an important avenue which we are actively pursuing. Essentially, we free up the relative entropy 'budget' faced by the robust agent to be used on more 'relevant' distortions.

5.1.3 Direct and Reverse Stress Testing

Finally, using these techniques we can nest 'reverse' and 'direct' stress testing, which is essentially a question of conditioning either on adverse scenarios for the state or for RoE. This is comparable to the sort of approach discussed in Waggoner and Zha (1999) but is more general, since it operates in terms of moments in (theoretically) arbitrarily nonlinear models. In figure 10 we depict conditional mean paths obtained under a version of direct and reverse stress testing.

We first define a tightly parameterized expression of an 'adverse' path. We pick a finite horizon and a sequence of scale parameters, one for each horizon, that will be used to scale the standard deviations at each horizon, which will then be added (or subtracted) from the untilted unemployment (or RoE) conditional mean paths. We assert that the scale sequence attains a maximum at horizon $\tau = 6$ and must be zero at the beginning and end of the forecast period. Between these points we linearly interpolate. Clearly there are much more sophisticated methods we could use to define 'adverse' but this parsimonious approach allows us to tell a simple story.

Looking at the red lines in the diagrams, we see that if we restrict the unemployment path to respect certain moment conditions, we can back out a particular path for RoE and, unsurprisingly, it is pessimistically twisted downwards. Similarly, if we impose that the tilted distribution respect an adverse sequence of conditional means for RoE, we can back out an implied adverse sequence of means for unemployment (green lines). This example is very simple and conceptually not enormously removed from conditional forecasting approaches that are already commonly employed by researchers and policy-makers. However, as we now show, we can provide a decision theoretic perspective on some of these approaches.

5.2 A 'Tilted Robustness' Problem

We will introduce some slightly different notation in this section, recognizing that we are switching to a finite horizon case, while retaining the essence of the basic robustness problem discussed in section 3.[25]

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 possesses a model describing the behavior of X, characterized by a 'benchmark' distribution, π. Were the agent to fully trust this model, she would evaluate her welfare on the basis of expected utility, $\int v(X)\pi(X)\,dX$. 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. In addition, however, we posit a constraint requiring that the twisted distribution respects moment conditions captured by the function, g.

$$
\begin{aligned}
W &= \min_{m} \int m(X)\,v(X)\,\pi(X) + \theta m(X)\log(m(X))\,\pi(X)\,dX && (2)\\
&\quad s.t. \\
1 &= \int m(X)\,\pi(X)\,dX \\
0 &= \int m(X)\,\pi(X)\,g(X)\,dX && (3)
\end{aligned}
$$

Whereas in section 3 we operated within a setting that easily allowed a recursive representation of the problem, here we will simply posit that the agent is facing a (finite horizon) sequence problem. Thus, as alluded to at the end of section 3.2, although our benchmark model can be naturally spoken of in terms of a tightly parameterized 'process' (a VAR), we here focus on conditional forecasting distributions and attack the distributions over sequences directly. We will also not be expressing our worst case in terms of a VAR. This makes explicit the unstructured nature of uncertainty (despite the restriction on moments). The agent is

[25] There is also slight abuse and reuse of notation - we are running out of letters.

concerned that 'something' is misspecified in her benchmark that would induce some other distribution over sequences. That 'something' is left implicit.

Solving problem 2 yields a particular minimizing likelihood ratio

$$m^{*}(X)=\frac{\exp\left\{-\frac{v(X)-\varphi^{*}g(X)}{\theta}\right\}}{E\left[\exp\left\{-\frac{v(X)-\varphi^{*}g(X)}{\theta}\right\}\right]} \tag{4}$$

where φ^{*} is the Lagrange multiplier on the constraint 3 at the solution. We refer the reader to appendix 7.3 for derivations

The minimizing likelihood ratio, $m^{*}(X)$, is similar to the familiar exponential tilt obtained under the approach to robust control advocated by Hansen and Sargent, $m^{*}_{HS}(X)\propto \exp\left\{-\frac{v(X)}{\theta}\right\}$, but differs due to the presence of the moment restrictions that the twisted distribution must satisfy. Similar expressions are obtained in Kwon and Miao (2013), Hansen and Sargent (2012) and also in Petersen, James, and Dupuis (2000) and Glasserman and Xu (2014). The two former works remain within a recursive LQG framework and introduce moment restrictions in the sense of characterizing a particular type of robustness in which a policymaker forms a robust policy, informed by a distribution that respects the intertemporal optimality conditions of the private sector.

Our framework, provided that one can evaluate the moment condition and draw from the benchmark model, allows for a richer class of non-linear dynamics and moment conditions. In addition, our motivation is more related to questions of the 'plausibility' of distortions to certain moments. Our restriction to finite horizon, although something we hope to relax, does not seem inappropriate for our (important) policy application as invariably, regulatory oversight is expressed in terms of fixed horizons. Petersen, James, and Dupuis (2000) lay out a very general framework and then specializes to the LQG case. Glasserman and Xu (2014) apply these methods in the context of portfolio analysis. We bring these powerful techniques to bear in our stress testing application.

A similar approach is taken by Breuer and Csiszr (2013) in the sense that they construct worst case distortions to finite horizon distributions that are derived from possibly non-linear and non-Gaussian models. They do not employ the tilting methods we advocate and apparently also work with linear and quadratic approximations to the worst case distribution, rather than fully exploiting the scope of importance sampling to obtain draws (approximately) from the underlying worst case.[26]

The restriction to finite horizon brings significant benefits in terms of enhancing tractability, such that we can go well beyond the Gaussian framework. It is not necessary for our

[26] Of course, in many applications the approximations they use to obtain analytical results may be advisable, due to the sampling error that might arise from Monte-Carlo methods.

analysis to be able to evaluate the pdf of the benchmark distribution, provided that we can simulate from it. In that case we would work with an equally weighted set of draws that represents an empirical approximation to the distribution. Our methods would then simply involve changes in discrete measures implied by importance weights. This seems important in practice where one might be working with large, complicated models built up from various 'satellite' sub-models separately maintained and drawn from, as perhaps is common in regulatory environments. In appendix 7.5 we illustrate the use of the exclusively importance sampling approach in a (very) stylized example that allows for non-linearity and non-Gaussianity over a finite multi-period horizon.

5.2.1 A Simple Example

For expositional purposes we set aside our estimated VAR example and here deal with a simpler case, to illustrate the tilted robustness method. We take a Gaussian AR(1) as a primitive for the evolution of the state

$$\begin{aligned} x_{t+1} &= \rho x_t + \varepsilon_{t+1} \\ \varepsilon_t &\sim N\left(0, \sigma^2\right) \end{aligned}$$

Let us envisage a situation in which we are concerned with realizations of the state over a horizon from $t+1$ to $t+\tau$. From the perspective of time t, one can think of a stacked vector of the state in the above system as a multivariate Normal random variable that has a particular structure on its distribution. Thus, in the notation of the earlier analysis π is the stacked Normal distribution and x is $x^\tau \equiv (x_{t+1}, \ldots, x_{t+\tau})$. We use $v(x) = \sum_{j=1}^{\tau} \beta^j u(x_{t+j})$. Then, if we wanted to ensure that the worst case distribution respected the same expectation for $x_{t+\tau}$ as the benchmark, we would set $g(x) = x_{t+\tau} - \rho^\tau x_t$. We could envisage u as $u(x_{t+j}) = u_1 \circ u_2(x_{t+j})$ where u_2 maps the state realization in a period into 'consumption' and then u_1 is a period utility function. We take u_2 to be an identity mapping in the example below and use a CRRA utility function for u_1. Although in this particular example the tilted worst case can be derived analytically (much as in section 5.1.2), we will use the more general importance sampling techniques mentioned above, based on reweighting draws from the benchmark.

In figure 11 we depict the conditional mean paths (jagged due to multinomial resampling variability) under the benchmark distribution, 'Hansen-Sargent' worst case distribution and the 'Tilted Robustness' worst case distribution.[27] We observe that a pessimistic twist is at

[27]We have picked θ in the two twisted cases to yield the same equally weighted (of both divergence directions) K-L divergences, although that is not obvious from the diagram, where, at least the conditional mean path under the tilted robustness case seems 'closer' to the benchmark. Perhaps other moments not

play in the two worst case paths but, in the tilted robustness case, the mean restriction (to be equal to the benchmark) at the final horizon seems to be respected. In figure 12 we also show that both worst cases feature elevated standard deviations, by horizon.

5.2.2 Interpretation of Tilted Robustness: Focusing Ambiguity

In the basic Hansen-Sargent case, we allow the agent to distort all moments of the benchmark distribution, subject to the relative entropy penalty and a requirement of absolute continuity. There is only one parameter, θ, that releases the agent from the requirement that she fully trusts her model.[28] In a sense, then, this parameter is the only degree of freedom for determining the worst case and it is 'used up' in deviating from Expected Utility. So it is unsurprising that it can have pathological properties that may not be consistent with our ideas of what elements of a model are most worth doubting. This captures the intuition that the agent is facing completely unstructured uncertainty that conceivably could render all moments of the benchmark distribution misspecified. Here, by restricting a subset of moments under the worst case we retain much of the unstructured nature of the uncertainty faced by an agent. But we implicitly assert that the agent trusts certain dimensions of her benchmark model, even if she does not trust it entirely. That is we 'focus' ambiguity.

Now, there are other ways of 'focusing' ambiguity in the uncertainty literature. However, these methods typically entail violation of the 'intuition of ambiguity'. By this we mean that the modeler tends to specify very particular dimensions in which the agent is ambiguous and often does so while introducing free parameters picked by the analyst to assert certain 'known unknowns'. This approach arguably is unsuited to truly ambiguous situations. Frequently, it involves positing an interval around some parameter or scalar object within a model and asserting that the agent behaves as if the value taken is at the 'adverse' end of the interval. Although this approach may make sense in certain situations, it often seems to be an inappropriate transfer of some of the intuition from the famous 'Ellsberg' examples to very different situations (see Ellsberg (1961)).

In the Ellsberg case (how many black balls in urns full of black and red balls etc.) a 'model' is essentially a question of a relative frequency - that is, a real scalar. In this case, when one speaks of ambiguity and constructing a set of priors (over which we ultimately minimize) it is utterly natural to end up with 'intervals' capturing the multiple models that an ambiguous agent may be concerned could be generating the data (it surely cannot be plausible that

illustrated here are being distorted more under the tilted robustness case, such as, implicitly, moment related to serial dependence. In our finite horizon case, it is not entirely obvious how to calibrate this divergence as the standard approach to calculating DEPs seems to have a slightly different interpretation.

[28] Or two if one counts the specification of the robustness problem (and particularly the use of relative entropy in the penalty) as a 'generalized' parameter.

there would be ‘holes’ in the set of priors). But, our intuition does not begin with intervals, it begins with a desire to construct a plausible set of priors that convey ambiguity. In the Ellsberg case, it happens to make complete sense that we should be working with intervals. But it may not always make sense to try to manhandle priors into the form of intervals on arbitrarily selected objects in a given model. Through tilting to respect moment conditions, we focus ambiguity by asserting *confidence* in a limited set of moments, rather than asserting a *lack of confidence* in a, perhaps arbitrarily, selected set of moments. Thus our approach appears less obtrusive on the part of the modeler and more true to the intuition of ambiguity.

the methods used here nest fairly standard conditional forecasting methods that can be used even if the theory of robustness is thought uncompelling: simply switch off the value function part of the minimizing likelihood ratio and do exponential tilting (much as in Cogley, Morozov, and Sargent (2005)). But our methods also allow the regulator to acknowledge the Knightian model uncertainty that they invariably face in a disciplined and theoretically grounded way - this is an ambiguity paper and the world of financial regulation is an ambiguous place. Our emphasis on finite horizons, though undesirable in some respects, also buys the regulator much more generality in terms of the models that can be applied numerically when carrying out the testing. This hopefully renders our approach not only theoretically interesting but also practically useful.

6 Conclusion

We have used a stylized approach to constructing stress test scenarios, based on the tools of robust forecasting. We take a simple model of the economy and banks’ exposures and twist the probabilities implied by this model in a particular pessimistic manner to identify dimensions in which the system is vulnerable. The tools are easily implemented and can yield a set of moments that can be used to derive distributions over objects of interest (in this case bank performance) in a way that emphasizes possible model misspecifications in dimensions in which the system is vulnerable.

We initially operate in a simple, but illustrative, linear-quadratic framework, but then generalize our analysis (in a finite horizon context) to allow for nonlinearities in the model of the economy and a broader class of uncertainties. In doing so, we articulate a way in which ambiguity can be ‘focused’. Beside its theoretical interest, this approach also renders our analysis closer to what might ultimately be implementable in the real world.

In ongoing work, we are attempting to solidify the, as yet, rather preliminary benchmark model we use, exploiting the rich data set obtained from regulatory filings required since the Dodd-Frank reforms. Allowing for regime switching, parameter or estimation uncertainty, latency and non-standard shock distributions are important avenues. These models seem to hold more promise in our stress testing application than the L-Q framework and hopefully can help provide a theoretical and practically useful set of techniques. Importantly, these methods also seem well suited to addressing other areas of policy, beyond financial regulation.

7 Appendix

In this section we include additional details to help with understanding the results in the paper.

7.1 Recursive representation of the worst case distribution (general case)

We seek a recursive expression of the problem and, invoking results in Hansen and Sargent (2008), obtain a value function of the following form

$$\begin{array}{rcl} V(\varepsilon_t, s_t) & = & \min_{m(\varepsilon_{t+1}, s_{t+1})} h(z_t) \\ & + & \beta \int m(\varepsilon_{t+1}, s_{t+1}) V(\varepsilon_{t+1}, s_{t+1}) p_\varepsilon(\varepsilon_{t+1}|x_t) \\ & + & \theta m(\varepsilon_{t+1}, s_{t+1}) \log m(\varepsilon_{t+1}, s_{t+1}) p_\varepsilon(\varepsilon_{t+1}|x_t) d\varepsilon_{t+1} \end{array} \tag{5}$$

subject to $\int m(\varepsilon_{t+1}, s_{t+1}) p(\varepsilon_{t+1}|x_t) d\varepsilon_{t+1} = 1$ for all values of s_{t+1}. If one solves the inner minimization problem (interpretable as that of the 'evil' agent) one obtains the minimizing martingale increment, which has the form

$$m(\varepsilon_{t+1}, s_{t+1}) = \frac{e^{-\frac{V(\varepsilon_{t+1}, s_{t+1})}{\theta}}}{E\left[e^{-\frac{V(\varepsilon_{t+1}, s_{t+1})}{\theta}} |\varepsilon_t, s_t\right]} \tag{6}$$

If one substitutes this solution into the original problem, then we obtain the following expression for the Bellman equation (with slight abuse of notation), where we note that the expectation in equation (7) is with respect to the benchmark transition density.

$$V(x_t) = h(z_t) - \beta\theta \log E\left[\exp\left(-\frac{V(x_{t+1})}{\theta}\right) |x_t\right] \tag{7}$$

The martingale M_t from the solution of the agent's problem is a ratio of joint densities, $\frac{\tilde{p}(x_{1:t})}{p(x_{1:t})}$, where $\tilde{p}$ denotes the density implied by the worst case model while p denotes the benchmark model's density. The martingale increment, $m(x_{t+1})$, is a ratio of conditional densities, $\frac{\tilde{p}(x_{t+1}|x_t)}{p(x_{t+1}|x_t)}$. Thus we have $\tilde{p}(x_{t+1}|x_t) = m(x_{t+1}) p(x_{t+1}|x_t)$. While $\tilde{p}$ is not directly interpretable as the conditional 'beliefs' of the agent, the fact that it differs from p emphasizes that more than one distribution plays a role in this problem, in contrast to the case where the agent fully trusts his model.

7.2 LQG Robustness

We posit a linear transition law for the state, x_t, given by

$$\begin{array}{rcl} x_{t+1} & = & Ax_t + C\hat{\varepsilon}_{t+1} \\ \hat{\varepsilon}_{t+1} & \sim & N(0, I) \end{array}$$

where u_t is a vector of controls and $\{\hat{\varepsilon}_t\}$ is an *iid* sequence. To represent misspecification in this case we first consider distorted models represented by allowing the mean of the Gaussian innovation in $t+1$ to depend, possibly in a nonlinear fashion, on the history of the state up to and including t. Thus, alternative models, capturing some unknown misspecification, are represented by the distorted transition law

$$\begin{aligned} x_{t+1} &= Ax_t + C\left(\varepsilon_{t+1} + w_{t+1}\right) \\ \varepsilon_{t+1} &\sim N\left(0, I\right) \\ w_{t+1} &= g_t\left(x_t, x_{t-1,\ldots}\right) \end{aligned}$$

We will ultimately show that the agent will envisage a particular distortion featuring a twist to the innovation covariance matrix. However, we will defer that discussion because it turns out that the solution of the robust problem, in terms of the distortion to the mean, does not depend on this twist.

Given $\theta \in (\theta_{bd}, +\infty]$ the multiplier problem considered is[29]

$$\min_{\{w_{t+1}\}_{t=0}^{\infty}} E_0 \sum_{t=0}^{\infty} \beta^t \left\{h\left(z_t\right) + \beta\theta w'_{t+1} w_{t+1}\right\}$$

subject to $z_t = g\left(x_t\right)$ and the distorted law of motion

$$\begin{aligned} x_{t+1} &= Ax_t + C\left(\varepsilon_{t+1} + w_{t+1}\right) \\ \varepsilon_{t+1} &\sim N\left(0, I\right) \\ w_{t+1} &= g_t\left(x_t, x_{t-1,\ldots}\right) \end{aligned}$$

The mapping from the state, x_t to the target, z_t will be as follows. A payoff variable c_t and a 'bliss' point variables, b_t are related to the state by

$$\begin{aligned} c_t &= H_c x_t \\ b_t &= H_b x_t \end{aligned}$$

If we let $z_t \equiv c_t - b_t$ and $H \equiv H_c - H_b$, then the period payoff is given by $g_(z_t) = z_t' W z_t$, a quadratic form where W captures the weighting scheme. It is useful to note that the period payoff can also be expressed as $x_t' Q x_t$ where $Q \equiv H' W H$.

As discussed in Hansen and Sargent (2008) the solution to this problem implies a stationary rule relating the distorted conditional mean of the $t+1$ innovation to the state in t, $w_{t+1} = Kx_t$. Letting $-x_0' P x_0 - p$ be the value of the problem and $h\left(z\right) = z' W z \left(= x' Q x\right)$, then we have the following Bellman equation

$$-x'Px - p = \min_{w} E\left\{h\left(z\right) + \theta\beta w' w - \beta x^{*\prime} P x^* - \beta p\right\}$$

[29]The lower bound or 'breakdown' point considered for θ, θ_{bd}, ensures that the problem remains well posed.

subject to

$$x^* = Ax + C(\varepsilon + w) \tag{8}$$

$$\varepsilon \sim N(0, I) \tag{9}$$

Now, P can be recovered from solving an associated certainty equivalent problem in which $\varepsilon_{t+1} \equiv 0$. This allows us to omit ε_{t+1} from the problem and abstract from p. Based on this insight, we can solve for many of the objects of interest by solving a deterministic robust linear forecasting problem:

$$\min_{\{w_{t+1}\}} \sum_{t=0}^{\infty} \beta^t \left\{ h(z_t) + \theta \beta w'_{t+1} w_{t+1} \right\}$$

given x_0 and subject to equation (9). If $-x'_0 P x_0$ is the value of the sequence problem, then the value of the agent's problem can be expressed recursively according to the Bellman equation

$$-x'Px = \min_{w} \{ h(z) + \theta \beta w'w - \beta x^{*\prime} P x^* \}$$

subject to

$$x^* = Ax + Cw$$

If one considers the inner minimization problem we observe that it induces a pessimistic twist to the continuation value, captured by the application of an operator $D(P)$, defined as follows

$$-x^{*\prime} D(P) x^* = -x' A' D(P) Ax = \min_{w} \{ \theta w'w - x^{*\prime} P x^* \}$$

where the minimization is subject to the dynamics of the distorted model

$$x^* = Ax + Cw$$

Thus we have that

$$D(P) = P + \theta^{-1} PC \left(I - \theta^{-1} C'PC \right)^{-1} C'P$$

It transpires that after allowing for the solution of the inner minimization problem and the pessimistic twist to the continuation value that it implies, one can represent the Bellman equation as

$$-x'Px = h(z) - \beta x^* D(P) x^*$$

subject to the approximating model[30]

$$x^* = Ax$$

[30] Recall we are working with deterministic cases due to the aforementioned augmented certainty equivalence result.

The maximization in the Bellman equation implies a particular operator that maps from a given 'continuation P' to the P that captures the value of the agent's problem in the current period, given his robust control. This operator, $T\left(\tilde{P}\right)$ is given by

$$T\left(\tilde{P}\right) = Q + \beta A'\tilde{P}A$$

Therefore, the P associated with the solution of the robust problem is the fixed point of the composite operator $T \circ D$. Associated with this P is the distorted mean law, $w = Kx$ where

$$K = \theta^{-1}\left(I - \theta^{-1}C'PC\right)^{-1}C'PA$$

Applying these laws to the evolution equation {equation} yields dynamics under the deterministic worst case given by

$$\begin{aligned} x' &= (A + CK)\,x \\ &= \tilde{A}x \end{aligned}$$

which can be contrasted with the dynamics that emerge under the benchmark, but allowing for the agent's robust control law, given by

$$x' = Ax$$

Allowing for randomness, but still restricting ourselves only to consider distortions to means, implies that the evolution of the state under the worst case is characterized by

$$x' = \tilde{A}x + C\varepsilon'$$

and, under the benchmark,

$$x' = Ax + C\varepsilon'$$

However, when one allows for more general distortions in this framework than simply those representable by a state dependent distortion to the mean of innovations, the worst case also features a distortion to the covariance matrix of the innovations. That is, the worst case transition law is given by

$$\begin{aligned} x' &= \tilde{A}x + \tilde{C}\varepsilon' \\ \tilde{C}\tilde{C}' &= C\left(I - \theta^{-1}C'PC\right)^{-1}C' \end{aligned}$$

This transition law, and its implicit repeated application allows us to characterize and draw from the worst case distribution over sequences that emerges from the agent's robust forecasting problem.

7.3 Tilted Robustness Derivations

The first order conditions of the minimization problem 2 (leaving aside the constraints, for now) are

$$v(x) + \theta(\log(m(x)) + 1) = \lambda + \varphi g(x)$$

where λ and φ are, respectively, the Lagrange multipliers on the constraints that the twisted distribution integrates to one and that the moment conditions are satisfied. Exploiting the first constraint (that the twisted measure integrates to 1) we obtain

$$\begin{aligned} 1 &= \exp\left\{-1+\frac{\lambda}{\theta}\right\}\int \exp\left\{-\frac{v(x)-\varphi g(x)}{\theta}\right\}\pi(x)\,dx \\ &= \exp\left\{-1+\frac{\lambda}{\theta}\right\}E\left[\exp\left\{-\frac{v(x)-\varphi g(x)}{\theta}\right\}\right] \end{aligned}$$

which implies the minimizing likelihood ratio is given by[31]

$$m^*(x;\varphi^*) = \frac{\exp\left\{-\frac{v(x)-\varphi^* g(x)}{\theta}\right\}}{E\left[\exp\left\{-\frac{v(x)-\varphi^* g(x)}{\theta}\right\}\right]} \tag{10}$$

We can obtain φ^* by defining

$$m(x;\varphi) = \frac{\exp\left\{-\frac{v(x)-\varphi g(x)}{\theta}\right\}}{E\left[\exp\left\{-\frac{v(x)-\varphi g(x)}{\theta}\right\}\right]}$$

substituting into the moment condition constraint

$$\int m(x;\varphi)\,\pi(x)\,g(x)\,dx = 0 \tag{11}$$

and then exploiting a root-finding subroutine (and numerical integration) to find the φ that satisfies this equation, call it φ^* . We then define $m^*(x) \equiv m^*(x,\varphi^*) \equiv m(x;\varphi^*)$.

7.4 Tilting in a Linear Gaussian Framework

We will consider a special case of the tilting framework discussed in section 5.1, that of a linear Gaussian state space where only restrictions on first and second moments are imposed. It is therefore useful first to examine how tilting in a multivariate Normal context is implemented. So let us consider a random variable $y \sim N(\zeta, \Sigma)$. Suppose we wish to impose that $E[y_2] = \mu_2$

[31]We make explicit the dependence on the multiplier φ as we have not yet discussed how it is obtained.

and $Var(y_2) = \Omega_{22}$ where we partition $y = (y_1', y_2')'$.[32] Then we have

$$\begin{aligned} \mu_1 &= \zeta_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mu_2 - \zeta_2) \\ \Omega_{12} &= \Sigma_{12}\Sigma_{22}^{-1}\Omega_{22} \\ \Omega_{11} &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} + \Omega_{12}\Sigma_{22}^{-1}\Sigma_{21} \end{aligned}$$

Now, in our linear Gaussian model, the state space we consider takes the form

$$\begin{aligned} y_t &= Gx_t + w_t \\ x_{t+1} &= Ax_t + \varepsilon_t \\ \begin{pmatrix} w_t \\ \varepsilon_t \end{pmatrix} &\sim N(0, \Omega) \end{aligned}$$

where

$$\Omega \equiv \begin{pmatrix} \Omega_w & \Omega_{w\varepsilon} \\ \Omega_{w\varepsilon}' & \Omega_\varepsilon \end{pmatrix}$$

We will find it convenient to stack the variables to create a VAR system for $\tilde{x}_t \equiv (y_t', x_t')'$ which takes the form

$$\begin{aligned} \tilde{x}_{t+1} &= \tilde{A}\tilde{x}_t + \tilde{\varepsilon}_t \\ \tilde{\varepsilon}_t &\sim N\left(0, \tilde{\Omega}\right) \end{aligned}$$

where

$$\begin{aligned} \tilde{A} &\equiv A_0^{-1}A \\ \tilde{\Omega} &\equiv A_0^{-1}\Omega \\ A_0 &\equiv \begin{pmatrix} I_{n_y} & -G \\ 0 & I_{n_x} \end{pmatrix} \\ A_1 &= \begin{pmatrix} 0 & 0 \\ 0 & A \end{pmatrix} \end{aligned}$$

Clearly, from the perspective of time t, one can think of a stacked vector of the variables in the above system at different horizons as a multivariate normal random variable that has a particular structure on its distribution, arising from the underlying system being a VAR.

[32] Ordering variables so that the partitions are contiguous is clearly without loss of generality.

That is, denoting $\tilde{x}_t^\tau \equiv \left(\tilde{x}_{t+1}', \tilde{x}_{t+1}', ..., \tilde{x}_{t+\tau}'\right)'$, we have that, conditional on $\tilde{x}_t$

$$
\begin{aligned}
\tilde{x}_t^\tau &\sim N\left(\begin{pmatrix} A \\ A^2 \\ ... \\ A^\tau \end{pmatrix} x_t, \begin{pmatrix} \Omega_{11} & \Omega_{12} & ... & \Omega_1 \\ \Omega_{12}' & \Omega_{22} & ... & \vdots \\ \vdots & ... & \ddots & \vdots \\ \Omega_{1\tau}' & ... & ... & \Omega_{\tau\tau} \end{pmatrix}\right) \\
\Omega_{ij} &= \begin{cases} A^{|j-i|}\Omega_{jj} & if\ i \geq j \\ \Omega_{ii}\left(A^{|j-i|}\right)' & o/w \end{cases} \\
\Omega_{ii} &= \sum_{k=1}^{i} A^{k-1}\tilde{\Omega}\left(A^{k-1}\right)'
\end{aligned}
$$

Consequently, we can invoke the analytic expressions for the tilted moments in the multivariate Normal case laid out above.

7.5 Robustness Using Importance Sampling in a Non-Linear and Non-Gaussian Framework

In this example we construct a very stylized example to illustrate some of the properties of the finite horizon approach to robustness described in the main body of the paper, without employing additional tilting.[33] Thus, this is akin to some of the examples in Breuer and Csiszr (2013) (based on the framework of Jimnez and Menca (2009)) although extended to convey some stylized points of interest. In particular, we wish to emphasize the importance of non-linearity, parameter uncertainty and the unfamiliarity of 'crisis' regimes and carry out the analysis entirely with the use of importance sampling, rather than analytical approximations. The calibration is only qualitatively illustrative.

The framework is that of a Markov Switching VAR(1) process for a bivariate state, $x_t = (y_t, u_t)$ where $y_t = \log\left(\left(1 - pd\right)/pd_t\right)$ is a transformation of a 'default rate', pd_t, and u_t is a 'macro factor' (see Fong and Wong (2008) for an empirical application of a MSVAR setup to a non-robust regulatory problem). The two variables feed back on each other such that a higher (worse) realization for u_t implies higher default probabilities (lower y_t). Similarly, higher default probabilities imply higher realizations of the macro factor, *ceteris paribus.* The dynamic relationships among the variables are described by

$$
\begin{aligned}
x_t &= \mu_t + A_t x_{t-1} + C_t \varepsilon_t \\
\varepsilon_t &\backsim \mathcal{N}(0, I)
\end{aligned}
$$

where μ_t, A_t and C_t are governed by a discrete state, s_t which follows a first order Markov

[33]This section is extremely preliminary.

chain:

$$\begin{aligned}
\mu_t &= \mu(s_t) \\
A_t &= A(s_t) \\
C_t &= C(s_t) \\
s_t | s_{t-1} &\backsim T
\end{aligned}$$

In our application, we will assume that $s_t \in \{1, 2\}$ and that the first state is both 'better' and more 'familiar' than the second, 'crisis' state. We set

$$\begin{aligned}
\mu_1 &= (I - A(1)) \cdot \begin{pmatrix} \bar{y}_1 \\ 1 \end{pmatrix} \\
\mu_1 &= (I - A(2)) \cdot \begin{pmatrix} \bar{y}_2 \\ 1.3 \end{pmatrix} \\
\bar{y}_1 &= \log\left(\frac{1 - 0.05}{0.05}\right) \\
\bar{y}_2 &= \log\left(\frac{1 - 0.25}{0.25}\right) \\
A(1) &= A(2) = \begin{pmatrix} 0.5 & -0.4 \\ -0.4 & 0.5 \end{pmatrix} \\
C(1) &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\
C(2) &= \sqrt{1.2}C(1) \\
T &= \begin{pmatrix} 0.75 & 0.25 \\ 0.8 & 0.2 \end{pmatrix}
\end{aligned}$$

Thus, the the first state will be more frequently occupied under the ergodic distribution of the chain (more familiar) and is associated with lower volatilities of innovations and VAR intercepts (better). In addition, we assert that there is a distribution over loss given default (LGD). in the two regimes. We assert that the fixed (but unknown) LGD are distributed as beta random variables

$$\begin{aligned}
LGD_{s_t=1} &\backsim \mathcal{B}(24.9, 58.1) \\
LGD_{s_t=2} &\backsim \mathcal{B}(7.9, 5.3)
\end{aligned}$$

implying means and variances of $\{0.3, 0.6\}$ and $\{0.05^2, 0.13^2\}$ respectively. Thus, the best estimate of the LGD under the bad regime is higher than that for the good regime, but the uncertainty around it is also greater, capturing the intuition that our information regarding unusual regimes (which are perhaps the most relevant for stress testing) is inferior.

In terms of the loss function, we calculate the expected payoff in t as

$$EAD * (1 - pd_t * LGD_t)$$

where EAD is a notional 'exposure at default'. We do not discount these expected payoffs over the forecast horizon.

We initialize the economy with $s_0 = 1$ and with the x_t at its regime-1 'unconditional mean' but with the macro factor shocked by a negative one standard deviation innovation. We simulate over a horizon of 10 periods, drawing $N = 300000$ times and each time also drawing LGD for the two possible regimes. Note that the LGD draws are fixed over each simulation of x_t over the forecast horizon. Thus we obtain equally weighted draws from the benchmark. We then re-weight and redraw according to m_{HS} using $\beta = 1$ and the period loss function described above.

In figure 13 we plot the point-wise medians at each horizon for pd_t and u_t, together with the relative frequency of the 'good' regime ($s_t = 1$) under the benchmark and worst case. We observe that the default rate and macro factors are elevated under the worst case, relative to the benchmark. Consistent with this, the economy spends less time in the good regime under the worst case at each horizon. Indeed, if one fits a first order Markov chain to the state transitions under the worst case simulations one obtains an estimated transition matrix

$$T_{wc} = \begin{pmatrix} 0.69 & 0.31 \\ 0.73 & 0.27 \end{pmatrix}$$

Furthermore, if one examines the simulated ε_t in figure 14 we also observe adverse shifts in the distributions for the innovations. The differences between the worst case and benchmark are small but they can be seen more clearly for regime 2. This is likely largely because of the same mechanisms outlined in Barillas, Hansen, and Sargent (2009) and Bidder and Smith (2012) where the size of distortions to innovations is increasing in the level of volatility and $C(2) > C(1)$ in our case. It is also possible that the unfamiliarity of the second regime induces greater distortion in that regime for a given relative entropy between the joint distribution over regimes, innovations and LGD under the benchmark and worst case.

Finally, we consider the distributions for LGD. We observe in figure 15 that these also are pessimistically twisted under the worst case with the twist particularly obvious in the lower panel, corresponding to $LGD_{s_t=2}$. The reasons for this are likely the fact that (especially under worst case) the probability of default is higher in this regime. The value function encodes this fact and emphasizes distortions to the benchmark distribution that reveal this vulnerability (if defaults are higher then it will be particularly damaging to have higher losses given default). In addition to this, however, is the additional uncertainty around LGD in the second regime under the benchmark, which renders a given distortion less expensive in terms of its contribution to the relative entropy penalty.

8 Tables

Table 1: OLS regressions of RoE on the VAR state, in the aggregate and for two different institutions.

	(1) Aggregate	(2) Inst. 1	(3) Inst. 2
Unemployment	-3.918***	-3.238***	-1.643
Inflation	-4.713**	-3.909**	-3.745
3-M Tbill	2.226	-1.068	7.496
ΔD.J.	3.875***	1.655***	3.097***
Term Spread	2.493**	1.440	3.312**
Const	8.757***	9.678***	12.44***
N	83	58	43
R^2	0.443	0.460	0.440

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 2: Unconditional moments of aggregate RoE and states under the benchmark and worst cases based on aggregate RoE regressions and for DEPs of 10% and 20%

	Benchmark		W.C. (0.2)		W.C. (0.1)	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
RoE	8.21	9.41	0.97	11.06	-4.09	11.88
Unemployment	0.31	2.36	2.28	2.75	3.65	2.95
Inflation	-0.29	0.91	-0.12	0.97	0.01	0.99
3-M Tbill	-0.61	1.27	-1.33	1.31	-1.84	1.33
ΔD.J.	0.00	0.98	-0.01	0.99	-0.03	0.99
Term Spread	0.26	1.55	1.45	1.73	2.28	1.83

Table 3: Unconditional correlations of states under the benchmark and worst cases based on aggregate RoE regressions and institution 1 (for DEP of 10%)

(a) Benchmark

$$\begin{Bmatrix} 1.00 & \cdot & \cdot & \cdot & \cdot \\ 0.14 & 1.00 & \cdot & \cdot & \cdot \\ -0.49 & 0.59 & 1.00 & \cdot & \cdot \\ 0.16 & 0.02 & -0.08 & 1.00 & \cdot \\ 0.77 & -0.15 & -0.72 & 0.14 & 1.00 \end{Bmatrix}$$

(b) Aggregate - W.C.(0.1)

$$\begin{Bmatrix} 1.00 & \cdot & \cdot & \cdot & \cdot \\ 0.31 & 1.00 & \cdot & \cdot & \cdot \\ -0.53 & 0.44 & 1.00 & \cdot & \cdot \\ 0.13 & 0.01 & -0.08 & 1.00 & \cdot \\ 0.82 & 0.04 & -0.73 & 0.14 & 1.00 \end{Bmatrix}$$

(c) Institution 1 - W.C.(0.1)

$$\begin{Bmatrix} 1.00 & \cdot & \cdot & \cdot & \cdot \\ 0.28 & 1.00 & \cdot & \cdot & \cdot \\ -0.37 & 0.62 & 1.00 & \cdot & \cdot \\ 0.15 & 0.01 & -0.09 & 1.00 & \cdot \\ 0.77 & -0.05 & -0.65 & 0.15 & 1.00 \end{Bmatrix}$$

9 Figures

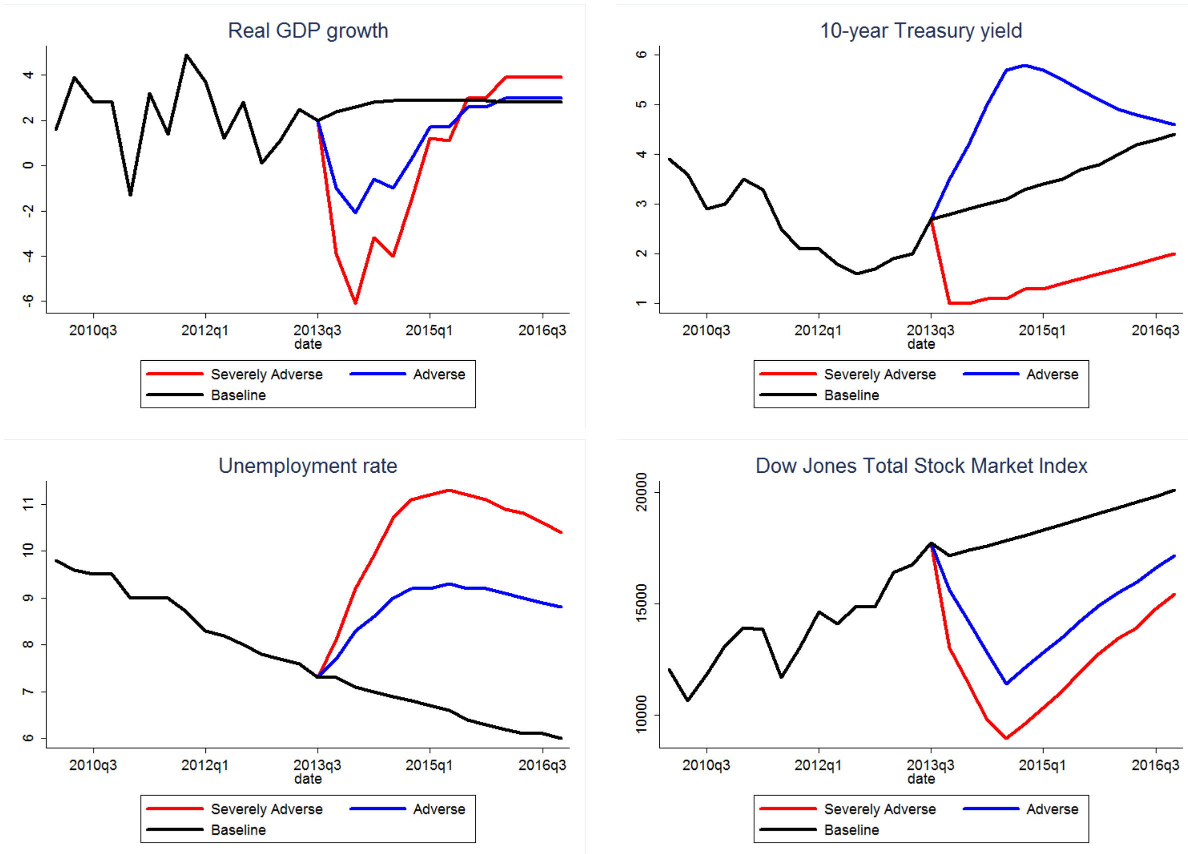

Figure 1: A subset of CCAR scenarios (2013).

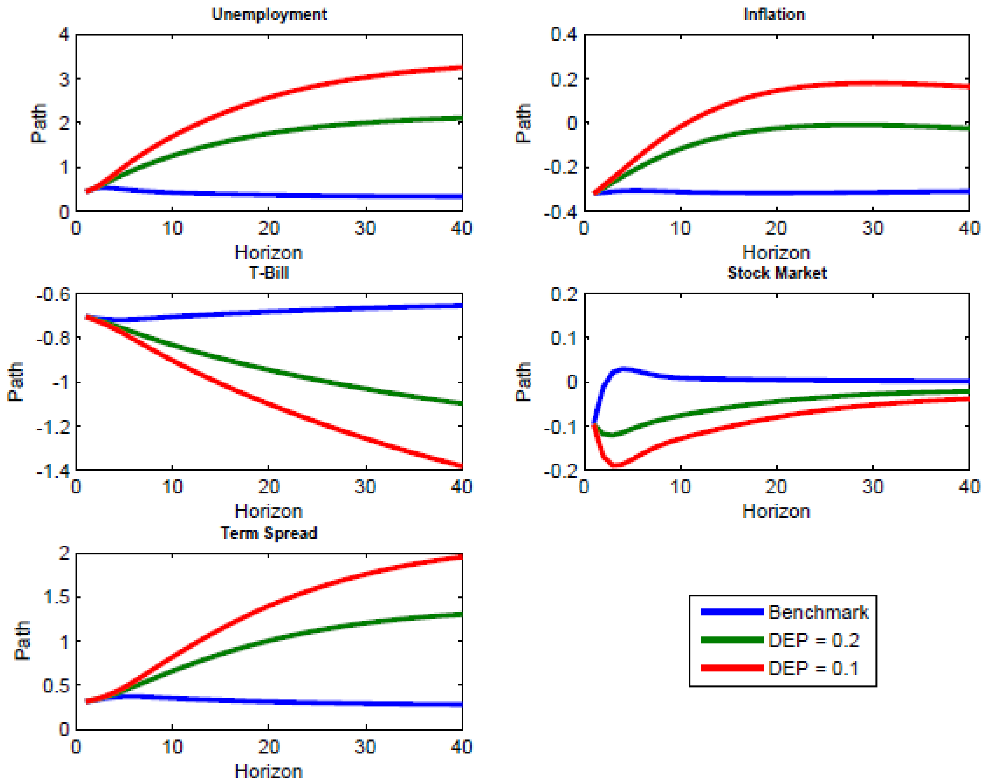

Figure 2: Conditional mean evolution for the states, generated under the benchmark and worst case VARs where the worst case is based on the aggregate RoE regressions and the DEPs considered are 20% and 10%. Initialization is based on an orthogonalized unemployment shock.

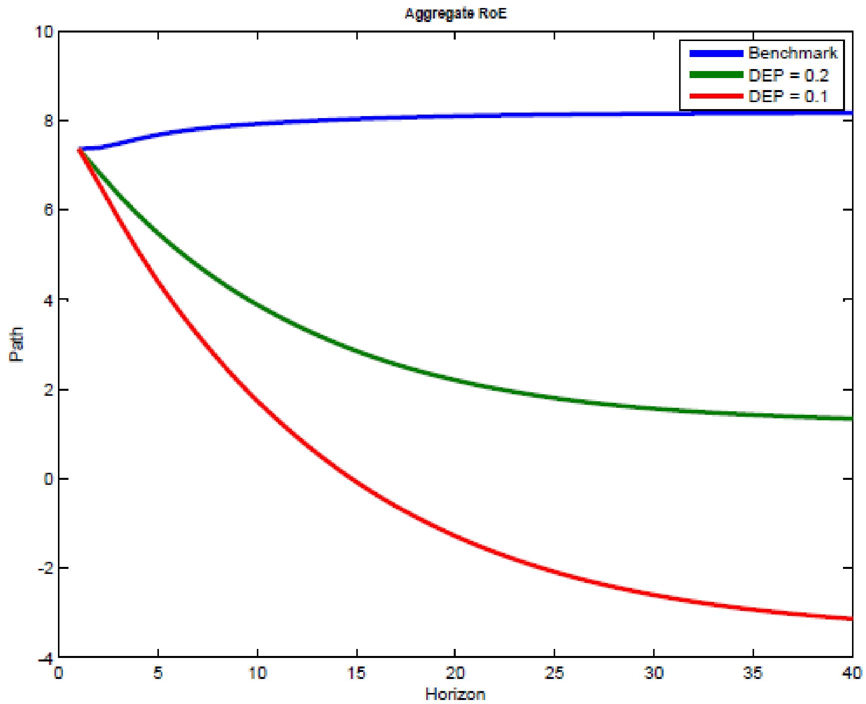

Figure 3: Conditional mean evolution for aggregate RoE, generated under the benchmark and worst case VARs where the worst case is based on the aggregate RoE regressions and the DEPs considered are 20% and 10%. Initialization is based on an orthogonalized unemployment shock.

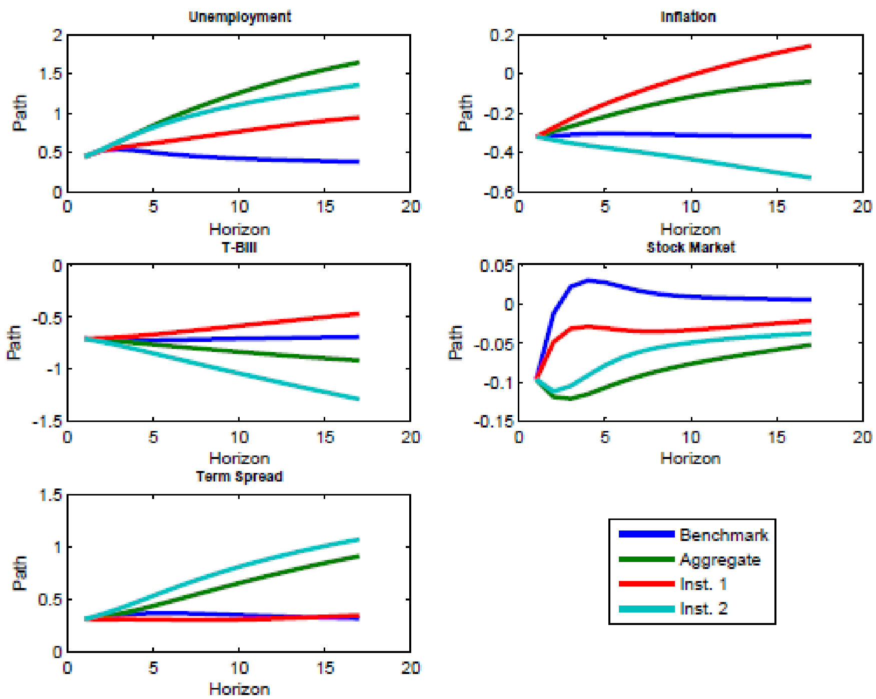

Figure 4: Conditional mean evolution for states under the benchmark and worst cases (with DEP of 20%) based on the aggregate and institution-specific RoE regressions. Initialization is based on an orthogonalized unemployment shock.

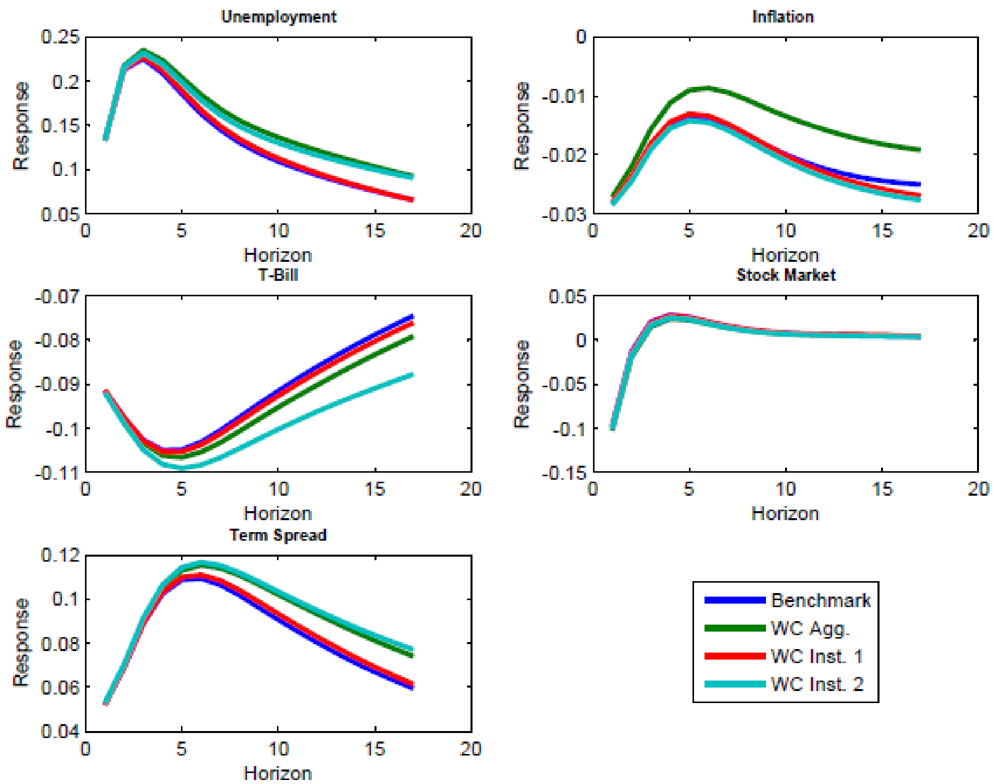

Figure 5: IRFs for states under the benchmark and worst cases (with DEP of 20%) based on the aggregate and institution-specific RoE regressions. Response is to an orthogonalized unemployment shock.

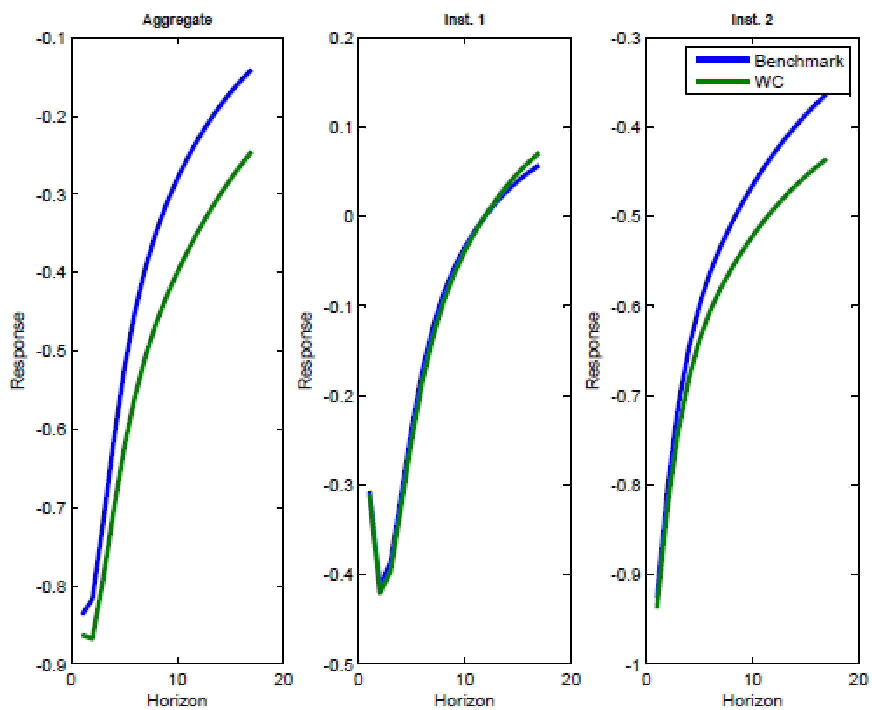

Figure 6: IRFs for RoE under the benchmark and worst cases (with DEP of 20%) based on the aggregate and institution-specific RoE regressions. Response is to an orthogonalized unemployment shock.

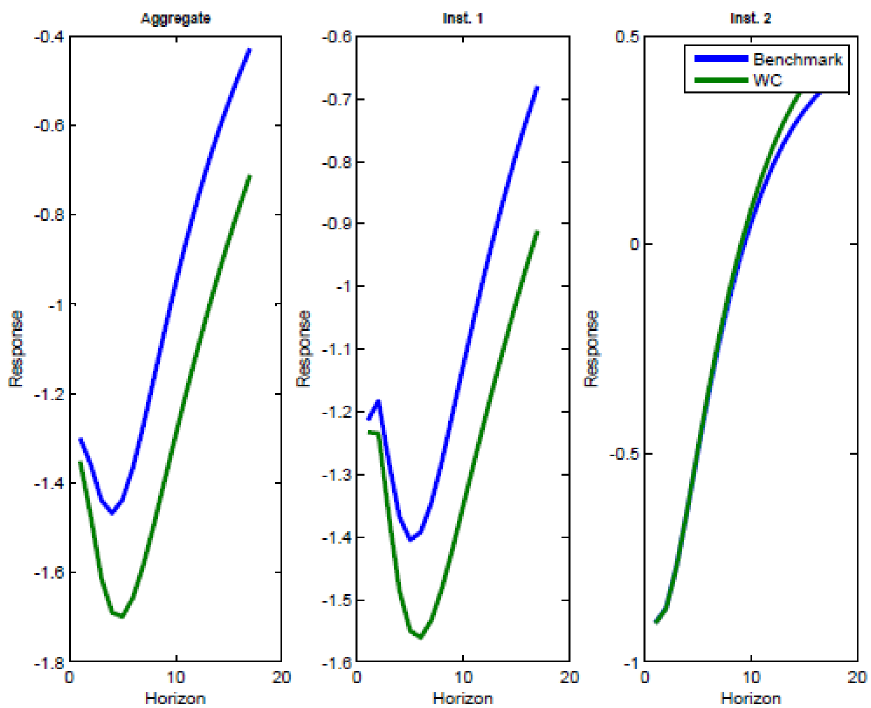

Figure 7: IRFs for RoE under the benchmark and worst cases (with DEP of 20%) based on the aggregate and institution-specific RoE regressions. Response is to an orthogonalized inflation shock.

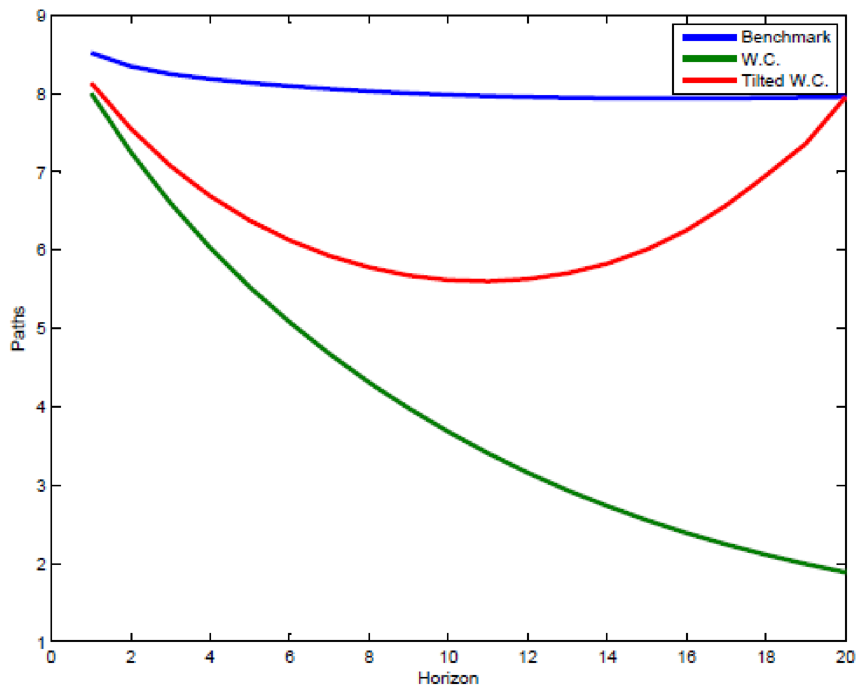

Figure 8: Conditional mean evolution of aggregate RoE under the benchmark, untilted worst case VAR (with DEP of 20%) and the tilted distribution based on the worst case VAR but respecting the same expectation as the benchmark at the end of the forecast horizon.

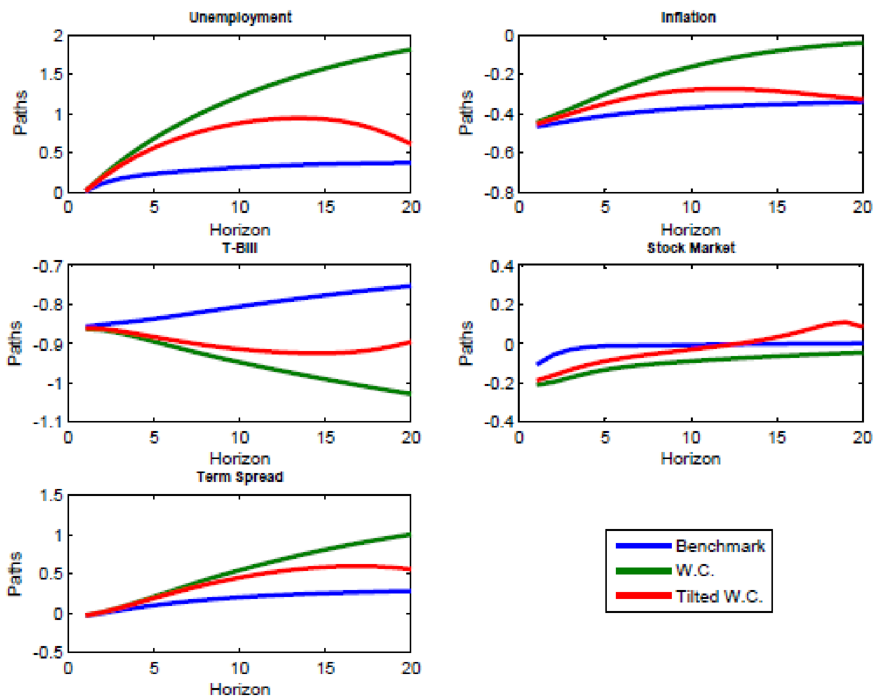

Figure 9: Conditional mean evolution of states under the benchmark, untilted worst case VAR (with WC based on aggregate RoE regressions and DEP of 20%) and the tilted distribution based on the worst case VAR but respecting the same expectation for RoE as the benchmark at the end of the forecast horizon.

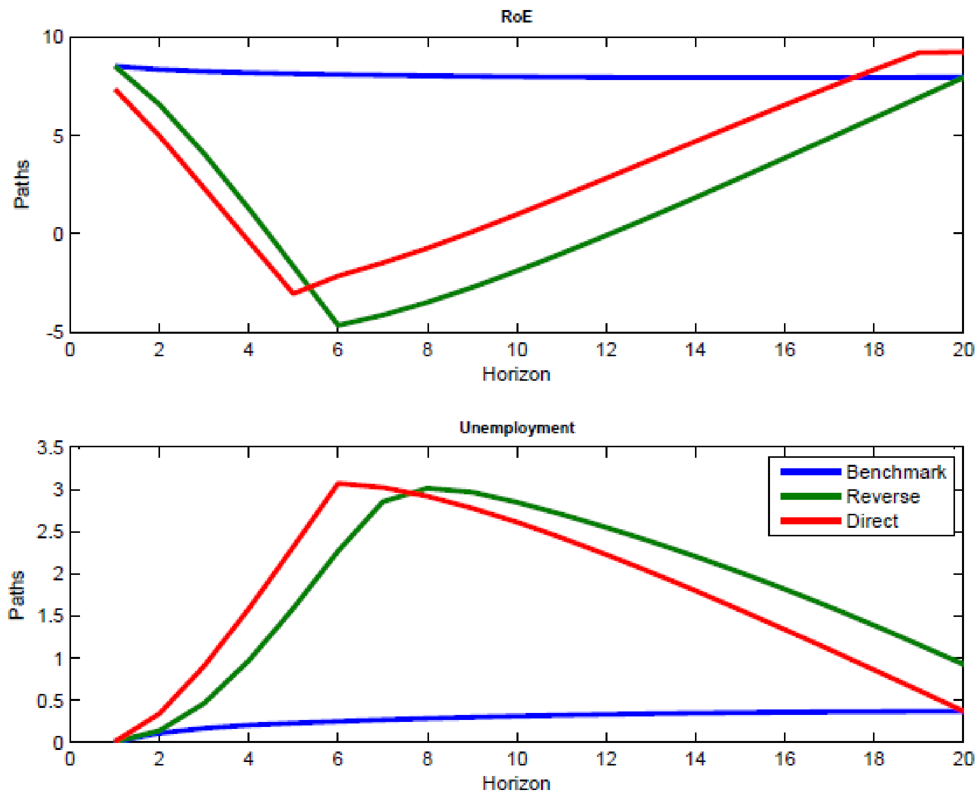

Figure 10: Direct and Reverse Stress Testing: Comparing conditional mean evolutions of (aggregate) RoE and Unemployment. Top panel - RoE. Bottom panel - Unemployment. 'Direct' paths are from conditioning on a particular (adverse) path for the conditional means of unemployment at each horizon and 'reverse' paths are from conditioning on a particular (adverse) path for the conditional means of RoE at each horizon.

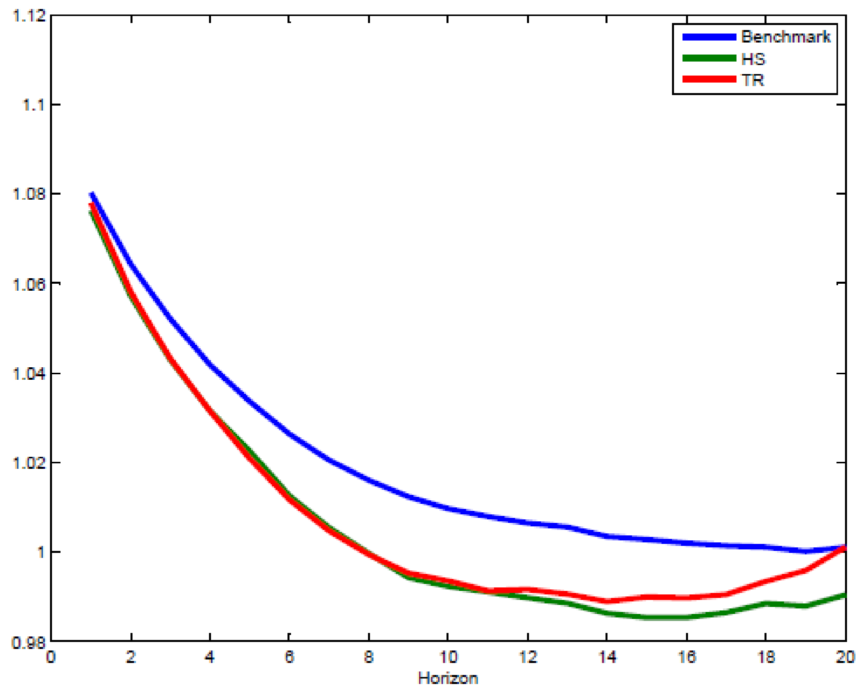

Figure 11: Tilted Robustness: A simple AR example. We plot conditional means by horizon under the benchmark, HS worst case and TR worst case, where the restriction imposed upon the worst case is that it respect the same mean at the end of the forecast horizon as the benchmark. Plots obtained from using importance sampling with the benchmark as proposal.

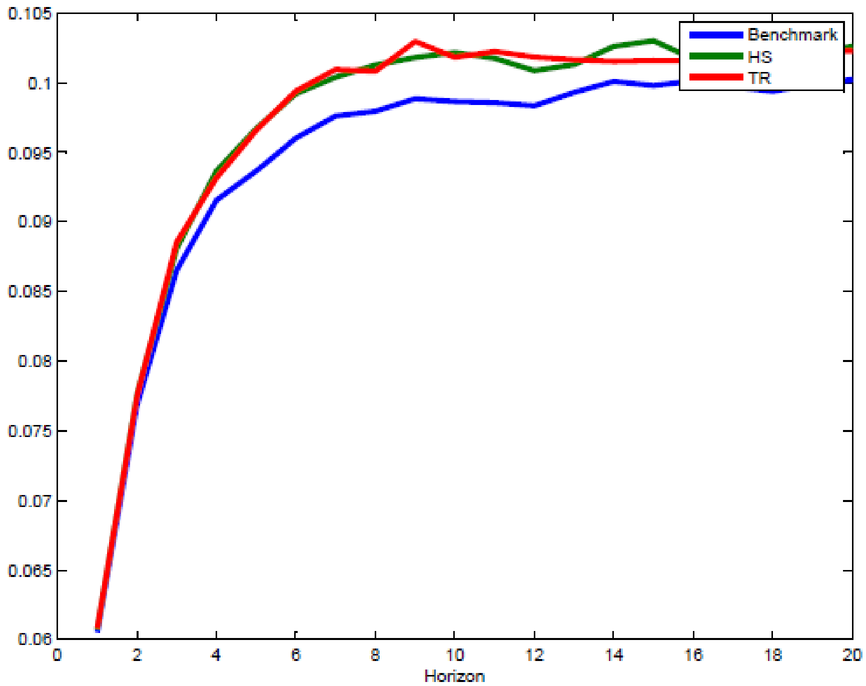

Figure 12: Tilted Robustness: A simple AR example. We plot conditional standard deviations by horizon under the benchmark, HS worst case and TR worst case, where the restriction imposed upon the worst case is that it respect the same mean at the end of the forecast horizon as the benchmark. Plots obtained from using importance sampling with the benchmark as proposal.

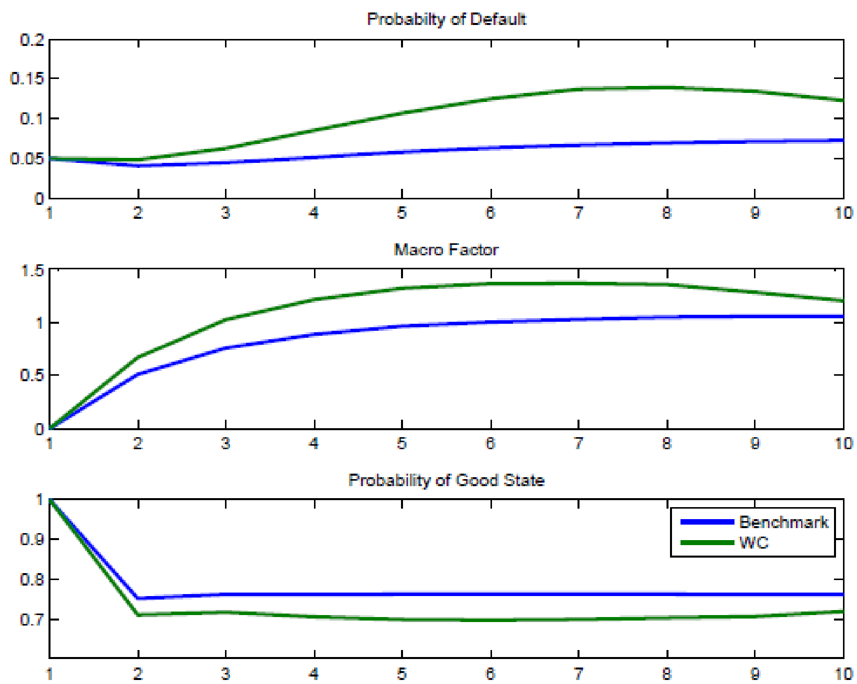

Figure 13: Robustness - non-linear and non-Gaussian example. We plot the point-wise medians of the probability of default and the macro factor over a horizon of 10 periods from an initial condition, and also the probability of being in the 'good' regime, in each period. Objects are plotted under the benchmark and worst case distribution.

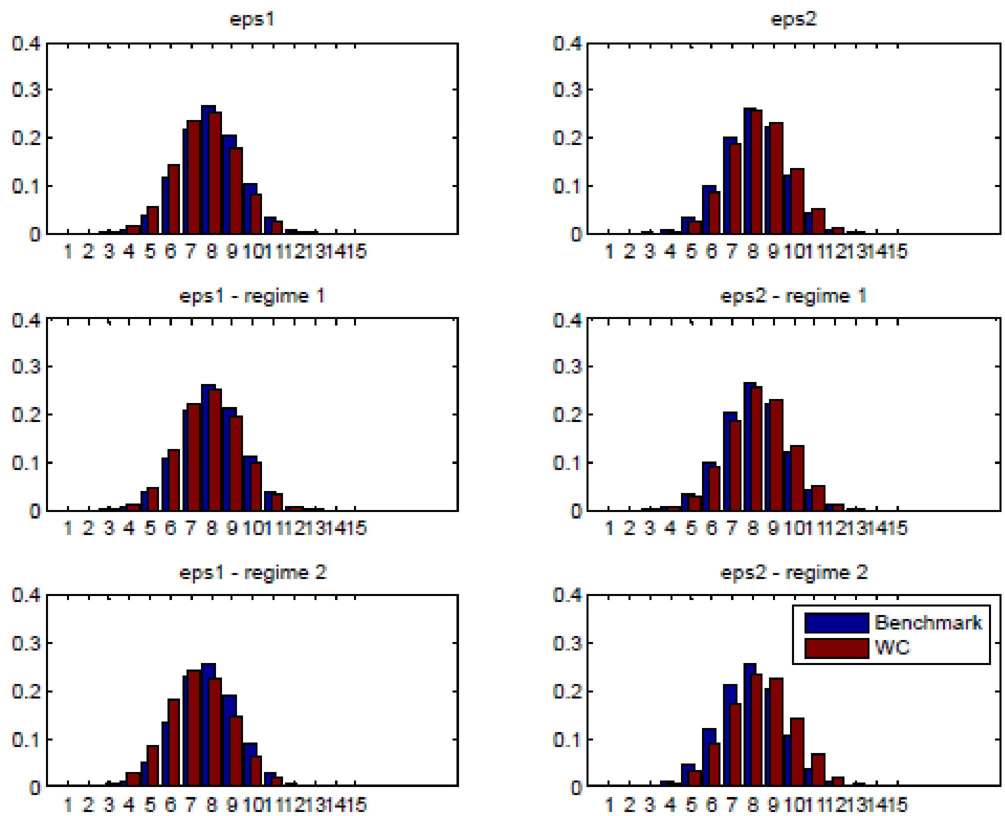

Figure 14: Robustness - non-linear and non-Gaussian example. We plot empirical relative frequencies of the VAR innovations, pooling regimes and then conditioning on regimes. Objects are plotted under the benchmark and worst case distribution.

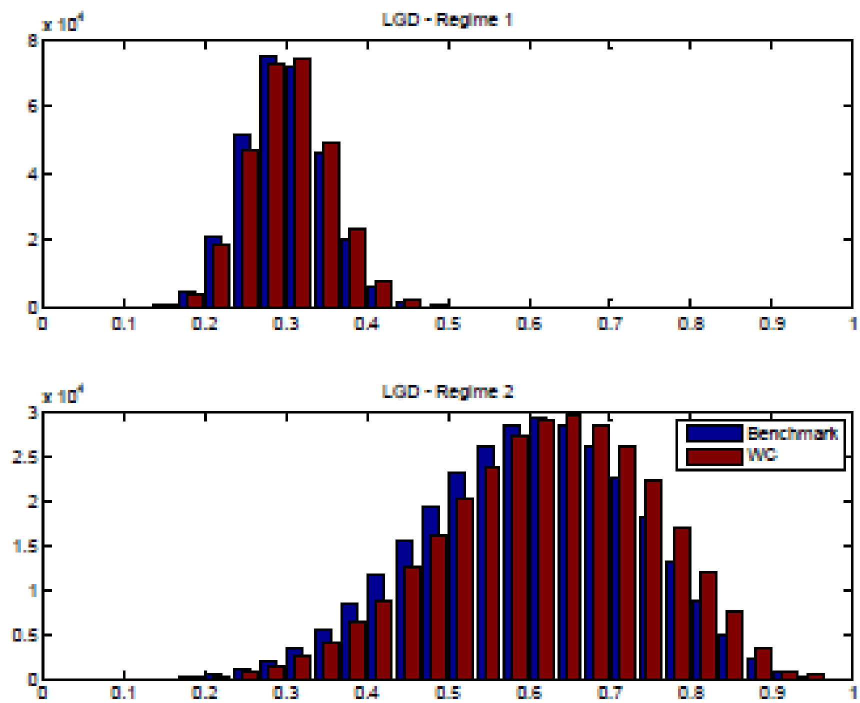

Figure 15: Robustness - non-linear and non-Gaussian example. We plot empirical distributions of LGD under the two regimes. Objects are plotted under the benchmark and worst case distribution.

References

Abdymomunov, A., and J. Gerlach (2013): "Stress Testing Interest Rate Risk Exposures," Working paper series, Federal Reserve Bank of Richmond.

Barillas, F., L. P. Hansen, and T. J. Sargent (2009): "Doubts or variability?," *Journal of Economic Theory*, 144(6), 2388–2418.

Bidder, R., and R. Giacomini (2015): "Stress Testing and Interest Rate Risk," Unpublished paper.

Bidder, R., and M. Smith (2012): "Robust animal spirits," *Journal of Monetary Economics*, 59(8), 738–750.

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 (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 (2013): "A probability-based stress test of Federal Reserve assets and income," Discussion paper.

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.

Ellsberg, D. (1961): "Risk, Ambiguity and the Savage Axioms," *Quarterly Journal of Economics*, 75(4), 643–669.

Fong, T. P.-W., and C.-S. Wong (2008): "Stress testing banks' credit risk using mixture autoregressive models," Working paper, Hong Kong Monetary Authority.

Furlong, F. (2011): "Stress testing and bank capital supervision," *FRBSF Economic Letter*, (June 27).

Giacomini, R., and G. Ragusa (2013): "Theory-coherent Forecasting," *Journal of Econometrics (in press).*

Glasserman, P., and X. Xu (2014): "Robust risk measurement and model risk," *Quantitative Finance*, 14(1), 29–58.

Glassermna, P., and Z. Xu (2013): "Robust Portfolio Control with Stochastic Factor Dynamics," *Operations reserach*, 61(4), 874–893.

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.

Haldane, A. (2009): "Why banks failed the stress test.," Speech at the Marcus-Evans Conference on Stress-Testing, London.

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., T. Schuermann, and K. Stiroh (2009): "Macroprudential supervision of financial institutions: lessons from the SCAP," Discussion paper.

Jimnez, G., and J. Menca (2009): "Modelling the distribution of credit losses with observable and latent factors," *Journal of Empirical Finance*, 16(2), 235 – 253.

Kwon, H., and J. Miao (2013): "Three types of robust Ramsey problem in a Linear Quadratic Framework," .

Maccheroni, F., M. Marinacci, and A. Rustichini (2006): "Ambiguity Aversion, Robustness, and the Variational Representation of Preferences," *Econometrica*, 74(6), 1447–1498.

of Governors, F. R. B. (2012): "Comprehensive Capital Analysis and Review 2012: Methodology for Stress Scenario Projections," Discussion paper.

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, ¡ce:title¿Research on the Financial Crisis¡/ce:title¿.

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.

Schuermann, T. (2013): “Stress Testing Banks,” Oliver Wyman, Wharton Financial Institutions Center.

Strzalecki, T. (2011): “Axiomatic Foundations of Multiplier Preferences,” *Econometrica*, 79(1), 47–73.

Waggoner, D. F., and T. Zha (1999): “Conditional Forecasts In Dynamic Multivariate Models,” *The Review of Economics and Statistics*, 81(4), 639–651.

White, H., T.-H. Kim, and S. Manganelli (2012): “VAR for VaR: Measuring Tail Dependence Using Multivariate Regression Quantiles,” Working papers 2012rwp-45, Yonsei University, Yonsei Economics Research Institute.