

Robust Stress Testing

Rhys Bidder*

Federal Reserve Bank of San Francisco

Andrew McKenna†

Federal Reserve Bank of San Francisco

This draft: February 14, 2014

Preliminary and incomplete. Please do not quote or distribute without the authors' permission.

Abstract

In recent years, stress testing has become an important component of financial and macroprudential 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 control 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’.

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 make a contribution to robustness theory within economics by formulating 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 bring 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@sf.frb.org. The authors thank Jaroslav Borovicka, Ian Dew-Becker, Fred Furlong, Bev Hirtle, Anna Kovner, Jose Lopez, Yulei Luo, Marius Rodriguez, Glenn Rudebusch, Matthew E. Smith and seminar participants at the Bank of England and the HKU School of Economics and Finance for helpful 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 aforementioned people, the Federal Reserve Bank of San Francisco, the Federal Reserve Board of Governors or the Federal Reserve System.

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 (2013b), 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 (see Board of Governors (2013), Borio, Drehmann, and Tsatsaronis (2013) and Hirtle, Schuermann, and Stiroh (2009)). 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 propose a stylized approach to stress testing based on the methods of robust control 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.

Aside from introducing a new perspective on stress testing, we also propose a significant theoretical contribution to robustness theory in economics by showing 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.

1.1 Motivation

We are motivated by ongoing debates over stress testing practice. One particular area for discussion is the question of how to trade-off the plausibility and severity of the scenarios. As noted in Breuer, Jandacka, Rheinberger, and Summer (2009) 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, the work of Breuer, Jandacka, Rheinberger, and Summer (2009), Covas, Rump, and Zakrajsek (2013) and White, Kim, and Manganelli (2012), among others, proposes various measures of distance and ways of using moments estimated from data. 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) 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.

Alternatively put, the models were misspecified so that what, from a shocks perspective were apparently highly implausible scenarios under the maintained distributions, were, in fact, eminently plausible and more worthy of concern. A problem with the 'tail based' approach to stress testing 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 ways. They are likely to be particularly misspecified in their predictions for the behavior of economies in unfamiliar times of stress. It is therefore desirable to derive extreme scenarios and a precise measure of distance that are informed by historical data, but not shackled to them as the implied distributions may be wrong (we face 'unknown unknowns'). This appears to be one of the most important challenges of stress test design and one which robustness is perfectly suited to address.

1.2 Why robustness?

Our proposed methodology addresses all the above issues explicitly. Briefly, robust control 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 economy. The agent expresses his doubts of his 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 also 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 forecaster (or regulator) allows for other possible data generating processes and concentrates on those that would be damaging. One can also incorporate different degrees of uncertainty about different dimensions of the benchmark model, such as in rarely visited regions of the state space.

To the extent that the 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 very 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. 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.¹

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 benchmark model is for exposures of the aggregate system then the scenarios generated from the worst case will be systemically damaging. If, however, the regulator’s

¹See Borio, Drehmann, and Tsatsaronis (2013) for a discussion of the difficulty of envisaging vulnerabilities in tranquil times and the difficulty of using historical experience and hypothesizing appropriately. See also concerns raised in Schuermann (2013a).

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.²

1.3 Application

We begin with a 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 a 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. Since there is little intuition for a satiation point in the target variables we are using, we set the satiation point to be distant from the values that the target is estimated to take with high probability under the benchmark. Given this Linear-Quadratic setup we can appeal to the results discussed in Hansen and Sargent (2008) and directly compute the worst case distribution.

We show that the worst case arising from this simple model has some interesting properties, but also some undesirable characteristics. In particular, the long run properties of the economy (the unconditional mean of the worst case model’s variables) 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. In the finite horizon framework that we adopt for this part of the analysis, we show that a broader class of benchmark data generating processes can be used, allowing for non-linearities, estimation uncertainty and other properties that would be beyond the scope of standard Linear-Quadratic robustness and difficult to handle using global methods in a recursive framework. It is this more general setup that apparently holds the most promise for incorporating robust control theory into macro-prudential practice.

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.

²Pritsker (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.

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 a supervisory stress test scenario 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.³ 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.

3 A Robust Specification of Preferences

In this section we lay out the theoretical basis of our approach to stress testing. We begin with an abstract robust forecasting problem.⁴ We then specialize to a a workhouse linear-quadratic Gaussian (LQG) setup which will form the basis of our approach. Finally, we include a preliminary illustrative example of our method for disciplining the scenarios obtained from robustness analysis by restricting the worst case distribution to respect certain moments.

3.1 Seeking Robustness - the Hansen-Sargent approach

A robust agent is endowed with a ‘benchmark’ model but fears that it is misspecified. He is concerned that the world is actually described by a model that is similar to the benchmark

³For a complete list of series for the supervisory scenarios and further details see <http://www.federalreserve.gov/bankinfo/stress-tests/supervisory-baseline-adverse-and-severely-adverse-scenarios.htm> and <http://www.federalreserve.gov/bankinfo/stress-tests-capital-planning.htm>

⁴We suppress any reference to a ‘control’ in these derivations. Other than that the discussion follows closely the treatment in Bidder and Smith (2012).

but distorted in some way. The agent expresses his doubts of his model by considering alternative distributions that are distorted versions of the distribution implied by his benchmark model. In order to explore the fragility of his model the agent considers adverse distributions and balances the damage that an implicit misspecification would cause him, 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 in the context of an abstract model, following the methodology described in Hansen and Sargent (2008).

3.1.1 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

$$x_{t+1} = g(x_t, u_t, \epsilon_{t+1}) \quad (1)$$

where x_t is the state vector and $\{\epsilon_t\}$ is a sequence of random variates. Given a density, $p_\epsilon(\epsilon_{t+1}|x_t)$, 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 $p(x_{t+1}|x_t) = p_\epsilon(\epsilon_{t+1}|x_t)\delta_{f(x_t)}(s_{t+1})$.⁵

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

$$V_0 = \min_{\{m_{t+1}\}} \sum_{t=0}^{\infty} E [\beta^t M_t \{h(z_t) + \beta \theta E(m_{t+1} \log m_{t+1} | \mathfrak{S}_t)\} | \mathfrak{S}_0] \quad (2)$$

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), $M_{t+1} = m_{t+1}M_t$, $E[m_{t+1}|\mathfrak{S}_t] = 1$, $m_{t+1} \geq 0$ and $M_0 = 1$.⁶ 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

⁵Note 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, $x_{t+1} = Ax_t + C\epsilon_{t+1}$ and, thus, $f(x_t) = Ax_t$.

⁶We assume that the robust agent’s information set, $\mathfrak{S}_t$ contains the entire history of states.

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

$$\tilde{E}[b_{t+1}|\mathfrak{S}_t] \equiv E[m_{t+1}b_{t+1}|\mathfrak{S}_t] \quad (3)$$

for some $\mathfrak{S}_{t+1}$ measurable random variable b_{t+1} given $\mathfrak{S}_t$. The conditional relative entropy associated with the twisted conditional distribution is given by the term $E[m_{t+1} \log m_{t+1}|\mathfrak{S}_t]$, which is a measure of how different the distorted measure is from the benchmark.

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.

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 approximating models 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 distinguishable characteristics will be easily identifiable and imply a detection error probability of close to 0. High 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.

3.1.2 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 illuminate 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.

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

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.

3.2 Seeking Robustness - tilting the worst case

In this section we examine a different, yet closely related approach to obtaining a robust forecast, but in which we restrict the worst case distribution beyond simply penalizing deviations from the benchmark in terms of relative entropy. We also restrict it to respect certain moment conditions. This method will be used in section 5.2 where we attempt to remove certain ‘undesirable’ properties exhibited by the worst case. However, beside these practical issues there are deeper reasons why one might wish to explore this way of ‘disciplining’ the worst case distribution.

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.⁷ This captures the intuition that the agent is facing completely unstructured

⁷There is only one parameter (or two if you count the relative entropy specification penalty as a ‘generalized’ parameter) that releases the agent from the requirement that he fully trusts his model. In a sense, then, this

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, implicitly assert that the agent trusts certain dimensions of his benchmark model, even if he does not trust it entirely. That is we ‘focus’ ambiguity.

Now, there are other ways of ‘focusing’ ambiguity but these methods tend to work by violating 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 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.

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 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 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.

3.2.1 Statement of problem

We will introduce some slightly different notation in this section.⁸ 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 . Clearly, this distribution can easily be decomposed into randomness arising from shock innovations, data generating process,

parameter is the only degree of freedom for determining the worst case. 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.

⁸There is also slight abuse of and reuse of notation - we are running out of letters.

parameterization and initial state (we are actively developing these applications).⁹

Were the agent to fully trust this model, he would evaluate his welfare on the basis of expected utility, $\int v(x) \pi(x) dx$. However, the agent distrusts his 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 his 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.

At this level of generality we could equivalently be referring to problem 2. However, whereas before 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 and possesses a distribution over sequences, which he will twist in a particular way. Specifically, we posit a constraint requiring that the twisted distribution respects moment conditions captured by the function, g .

$$W = \min_m \int m(x) v(x) \pi(x) + \theta m(x) \log(m(x)) \pi(x) dx \quad (4)$$

$$s.t. \quad (5)$$

$$1 = \int m(x) \pi(x) dx \quad (6)$$

$$0 = \int m(x) \pi(x) g(x) dx \quad (7)$$

Solving this problem (as shown in appendix 7.3) yields a particular minimizing likelihood ratio

$$m^*(x) = \frac{\exp \left\{ -\frac{v(x) - \varphi^* g(x)}{\theta} \right\}}{E \left[\exp \left\{ -\frac{v(x) - \varphi^* g(x)}{\theta} \right\} \right]} \quad (8)$$

where φ^* is the Lagrange multiplier on the constraint 7 at the solution.

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) and Petersen, James, and Dupuis (2000). The former work remains within a recursive LQG framework and introduces 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

⁹Markov switching and allowing for more general forms of nonlinearity has been identified by, among others, Borio, Drehmann, and Tsatsaronis (2013) as important ways in which one can generate adverse scenarios without resorting to implausibly large ‘shocks’.

can evaluate the moment condition and the pdf associated with the benchmark model, allows for a richer class of non-linear dynamics and moment conditions.¹⁰ The latter work lays out a very general framework and then specializes to the LQG case. We bring these powerful methods to bear in our economically motivated application.

We note that, unlike in the methods of section 3.1, here we will end up with a worst case distribution, rather than a worst case ‘process’. For our purposes this is not a problem and, in fact, provided we can simulate from the benchmark, the restriction to finite horizon brings significant benefits in terms of enhancing tractability, such that we can go well beyond the Gaussian framework.¹¹ We are actively working on allowing for parameter uncertainty, regime switching, latency and other nonlinearities within the benchmark. In addition, we are able use very general forms of period loss function since we are not undertaking difficult problems of solving for a fixed point in a functional space when constructing the value function - we are not working in a recursive framework.

3.2.2 Gaussian Example

Suppose we take as a primitive a Gaussian VAR for the evolution of the state

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

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

$$\begin{aligned}x^\tau &\sim N\left(\begin{pmatrix} A \\ A^2 \\ \vdots \\ A^\tau \end{pmatrix} x_t, \begin{pmatrix} \Omega_{11} & \Omega_{12} & \dots & \Omega_{1\tau} \\ \Omega'_{12} & \Omega_{22} & & \\ \vdots & & \ddots & \\ \Omega'_{1\tau} & & & \Omega'_{\tau\tau} \end{pmatrix}\right) \\ \Omega_{ij} &= \begin{matrix} A^{|j-i|}\Omega_{jj} & \text{if } i \geq j \\ \Omega_{ii}(A^{|j-i|})' & o/w \end{matrix} \\ \Omega_{ii} &= \sum_{k=1}^i A^{k-1}\Omega(A^{k-1})'\end{aligned}\tag{9}$$

¹⁰The restriction to finite horizon, although something we are hoping to relax, does not seem inappropriate for our (important) policy application as invariably, regulatory oversight is expressed in terms of fixed horizons.

¹¹It is not necessary for our 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.

Thus, in the notation of the earlier analysis π is the Normal distribution given in (9), x is x^τ and we could 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 $g(x) = x_{t+\tau} - A^\tau x_t$. 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. This particular example will be used in our stress testing framework below.¹²

4 Robust Stress Testing

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 various benchmark models that will underpin our analysis. Specifically, we posit a linear evolution for the state (a Gaussian homoscedastic vector autoregression (VAR)), linear loadings of target variables on the state (based on estimated linear regression of bank performance on the state) and a quadratic period payoff (with a distant satiation point). We will take annualized return on equity as our target variable and consider a VAR in 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).

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 quarterly Dow Jones stock market index (SPDJI) and the spread between 10 Year Treasuries and the 3 Month rate (FCM10 - FTBS3).¹³ We include the change in the stock market in all our specifications, but also consider adding the bond and term spreads individually. The data sample we use to estimate the VAR is from 1975Q2 to 2011Q3.

We allow the lag length of our VARs to be determined by the Akaike Information Criterion (AIC) and the three specifications we consider are for $x_t = (x_t^M, x_t^F)$ where x_t^M comprises the unemployment rate, inflation and the short rate, while x_t^F comprises the change in the stock market and the term spread. Our intention is to begin with a ‘standard macro’ VAR

¹²In fact, in this case the tilted worst case can be derived analytically using well known expressions for deriving conditional Normals.

¹³All series codes are from HAVER.

and then augment it with financial series, whilst retaining sufficient parsimony to estimate the system. Our lag selection criterion favors two lags.

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. We use data for the ‘aggregate’ banking system, as obtained from the New York Fed’s ‘Quarterly Trends for Consolidated U.S. Banking Organizations’ website.¹⁴ We estimate a simple OLS regression of this series on the contemporaneous values of the VAR state and take the estimated coefficients as defining the loadings of the target variable on the state.¹⁵ We estimate for 1991Q1 – 2011Q3 and also consider subsamples obtained by splitting at 2000Q4 in order to show how variation in loadings may affect what constitutes a worst case at different times.¹⁶ In addition, we

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.¹⁷

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*.

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

¹⁴See http://www.newyorkfed.org/research/banking_research/quarterly_trends.html for data and associated documentation. In ongoing work we intend to undertake similar analysis at the individual bank level. In addition, allowing for the effects of leverage seems an important avenue to pursue. We also are investigating the effects of (various) yield curve changes on net interest income.

¹⁵In future analysis, we will allow for the residual in the relationship also to enter the the robust forecasting problem, but for now we take the estimated coefficients as exclusively defining the exposure of a ‘representative’ bank to the elements of the state. In addition, this analysis remains primarily, not least because we choose to omit a lagged dependent variable in our regressions which, in other studies, has been found to be an important element in modeling the relationship between bank performance and the wider economy (see Guerreri and Welch (2012)). One avenue would be to include the measure of aggregate performance in the VAR, comparable to the analysis undertaken by Hoggarth, Sorensen, and Zicchino (2005).

¹⁶Clearly, the lengths of the subsamples are rather short and the associated results should only be regarded as illustrative.

¹⁷All the data we use is public but we choose not to identify the two institutions used as the analysis is extremely empirically simplistic in its current form and meant only to be illustrative.

that the predictions of the robust forecasts will vary according to whether the forecaster is implicitly assumed to be concerned with the aggregate or a particular institution. For example, we note that the second institution does not exhibit a significant loading on unemployment and that the point estimate is of a lower magnitude than those obtained in the aggregate regression or for the first 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.

We will defer our discussion of the properties of our estimated VAR models until section 5 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 is somewhat plausible.¹⁸

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 0.1 and 0.01, 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). The latter calibration is rather extreme but is intended to help illustrate the qualitative nature of the worst case distortions, to the extent that the distortions do indeed retain the same qualitative properties as the class of models the agent regards as plausible increases. More practically, we follow the CCAR approach in providing scenarios of differing degrees of severity. Furthermore, it is perhaps arguable that from a regulatory perspective

¹⁸Fixing 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.

it is worth characterizing the qualitative nature of damaging misspecifications without being as concerned with DEPs as one would be when dealing with the behavior of a robust agent in, say, an endowment economy.

5 Results

In this section we discuss our results. 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. We will first apply standard, recursive techniques as discussed in section 3.1.2. The worst case VARs derived in this context yield interesting insights and in some dimensions are quite revealing. However, we will also show that they exhibit a property, specifically a distortion to unconditional mean bank performance, that may not be intuitively desirable and which certainly renders our stress scenarios qualitatively different from those in CCAR. In response to these concerns, we then apply the methods discussed in section 3.2.

5.1 Basic Robustness

We use worst case VARs to derive unconditional moments and then, also, conditional moments that are more easily comparable to the sort of objects that feature in stress tests.

5.1.1 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 $DEP = 0.1$ case and even lower under the $DEP = 0.01$ case, 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 somewhat surprisingly the volatility of the financial variables (stock market growth and term spread) are 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

¹⁹Note 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.

somewhat counterintuitive signs likely reflect patterns of unconditional autocorrelation among the states that render the nature of the worst case somewhat hard to predict. For example, if one observes that the banking system loads positively on a particular state and also loads positively on another, more ‘important’ state, which happens to be negatively correlated with the first state, then it is possible that damaging misspecifications that imply low RoE could be represented with a positive distortion of the first state since it will be associated with a more negative value of the more influential state. So the worst case respects the negative correlation and thus the intuitive ‘marginal’ distortion of the first state is overwhelmed.

In table 3 we observe the unconditional correlations among the states, under the benchmark and worst cases. 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 (0.17) whereas under the worst cases the positive correlation is exaggerated, to 0.5 in the more extreme case. 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 fear that the benchmark model underestimates the importance of supply shocks - although this structural interpretation is not strictly implied by the worst case.

With regard to the $\{3,1\}$ term we see a sign flip under the worst case, relative to the benchmark. In this case it appears that misspecifications representable by a positive 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.²⁰

5.1.2 Impulse Responses and Transition Paths

In this section we consider the evolution of the economy under the benchmark and worst cases. Thus, whereas in section 5.1.1 we discussed the unconditional properties of the various ‘models’, here we will discuss certain conditional moments. In addition, although before, we

²⁰Although we omit tables and plots for reasons of space, we also find that the worst case features greater power at low frequencies. This implies that the forecaster fears misspecifications representable by greater persistence in the processes (see Bidder and Smith (2013) and Bidder and Dew-Becker (2014) for related discussions). Finally, and perhaps surprisingly, the mean square errors of h-step forecasts are not much different between the benchmark and worst cases. Essentially, this reflects the fact that the distortion to the covariance matrix of innovations is small under the worst case. The distortion to the mean path taken by the economy is the primary way in which the agent mechanically represents his doubts through the worst case.

concentrated on the results for the aggregate system, here we will exploit our institution-specific regressions. Our empirical analysis is still rather preliminary.

Let us first consider impulse responses obtained under the benchmark and worst case distributions. In figure 2 we plot Choleski orthogonalized impulse responses corresponding to a shock to the inflation equation in the VAR.²¹ There are several stylized points to note about this figure. Firstly, the responses of some of the variables to the shocks, particularly the financial variables in the bottom two panels, are very similar between the benchmarks and the worst cases. However, overall, we observe that there are differences and, in some cases, the differences are fairly substantial.

Ultimately, we are interested in our measure of performance, RoE, and we observe considerable differences in the responses of this variable, especially after the first few periods. Perhaps unsurprisingly, that the diagrams suggest that the agent fears misspecifications that can be represented with a more severe and persistent response of RoE to a shock. A regulator might thus wish to use these alternative responses to elicit from a bank what their assessment of the impact of such a shock would be and induce them to ask the bank to consider, not just an adverse shock, but one whose effects reverberate for many periods. Indeed, it is conceivable (in a much richer model) that this sort of analysis could help a regulator direct a bank towards considering a shock that they themselves had not realized would be damaging for them.

Although the impulse responses we have shown are interesting in providing a sense of how the dynamics of the benchmark and worst case models differ, and the unconditional moments in section 5.1.1 are interesting in terms of the ergodic behavior of the economy, we now present a combination of the two: depictions of how the economy evolves from a given initial state. These scenarios are the objects that are perhaps most easily comparable to the scenarios of the CCAR stress tests, some of which were depicted in figure 1.

In figure 3 we depict the economy’s evolution (under the benchmark and worst cases) from a particular initial state. In this case, the initial state happens to be that which would prevail after an orthogonalized shock to inflation under the benchmark.²² We simply depict the path the economy takes if one evolves the state according to the conditional mean law of motion under the relevant VAR.

We observe that, depending on the DEP considered, the path of RoE in the top left

²¹The ordering of variables is as in table 1, excluding RoE. The orthogonalization should not be taken too seriously in this case as the main issue is to examine how a particular (in this case quite complicated) moment changes between the benchmark and worst case, depending on the fear of misspecification and the estimated exposures. In future work we hope to investigate identification strategies further and also consider whether applying the strategy to the worst case VAR has an interesting interpretation.

²²So in some sense these are impulse responses but with the same absolute initial condition representing a different distance from the eventual steady state, which differs between models.

panel is pessimistically distorted relative to the benchmark. The path reflects the estimated loadings combined with the state evolutions depicted in the remaining panels. To further emphasize this point we also depict in figure 4 conditional mean paths for the states under the benchmark and the various worst cases implied by the institution-specific RoE regressions, as well as those for the aggregate. The estimated loadings differ and hence, so do the worst case paths.²³ As the horizon extends the gaps between the paths become more extreme as they tend towards the different average levels listed in table 2.

These scenarios are tailored to the state of the economy and the banks' exposures in a formal and easily explainable way. Through the logic of detection error probabilities one can refer to an intuitive measure of plausibility when defending the scenarios' validity. Furthermore, to the extent that the benchmark model is a 'good' description of the data generating process, one hopes that it can reveal scenarios that one might not ever have intuited and yet capture important vulnerabilities in the system and our presumably imperfect modeling of it.

5.2 Tilted Robustness

[THIS SECTION IS PRELIMINARY AND INCOMPLETE] We now show an example of restricting worst case distributions obtained under the standard Hansen-Sargent methodology to respect particular moment conditions, as discussed in section 3.2. To operationalize this we fix a finite horizon and take the benchmark model to be the multivariate normal distribution over the stacked forecasts implied by an AR(1).

The moment restriction we consider is that the conditional mean at the end of the forecast horizon is the same as under the benchmark, given that we start from an initial state. In figure 5, we show the conditional mean under the benchmark, the H-S worst case (which does not respect the mean restriction) and then the path under 'Tilted Robustness'.²⁴ We have adjusted θ to ensure the same distance measure between the HS worst case and the benchmark and the TR worst case and the benchmark.²⁵ This is not obvious in the means diagram (where the TR, at least according to this moment, looks closer) or, indeed in the standard deviations in figure 6.²⁶ This is perhaps indicative of other distortions, not shown

²³The detection error probability here used is 10%

²⁴The AR(1) and its calibration are simply illustrative (hence this section being labeled incomplete), but clearly similar analysis could be undertaken based on the VARs discussed earlier and more general non-linear and non-Gaussian benchmarks. The period utility function is CRRA. Also, note that these diagrams only really show us properties of marginal forecast distributions at each horizon. Presumably there may be interesting distortions of moments related to serial dependence.

²⁵The distance measure is $0.5 \cdot \text{KL}(\text{weights1}, \text{weights2}) + 0.5 \cdot \text{KL}(\text{weights2}, \text{weights1})$, where KL is the K-L divergence.

²⁶Please also note that in this Gaussian case with a mean restriction this can actually be done analytically (playing around with formulae for conditioning Normals).

in this diagram, perhaps implicitly in time dependence.

We can see the effect of constraining the mean at $\tau = 20$ to be the same as under the benchmark. The red line bends back up and (subject to multinomial sampling variability) hits the desired moment. The HS path does not. However, note that we can see that even the HS path starts bending up towards the benchmark. This relates to the analysis of Hansen and Sargent (1995) on the discounting of entropy in the penalty function. Without discounting in problem (4) of the contribution to the entropy penalty of distortions in future periods, the effects of robustness wear off at far horizons, as perceived from the ‘time-0’ perspective (which would yield implicitly non-stationary policy rules over time). By discounting entropy in the sequence problem, it allows for more distortion in the future, which offsets you caring less about distortions in the future. Here, we just twist the whole distribution over the sequence and have discounting of payoffs. But we do not discount entropy.

6 Conclusion

We have proposed a stylized approach to constructing stress test scenarios, based on the tools of robust control analysis. 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. The methodology also addresses many of the main areas of debate over stress testing theory and provides a response to some of the main concerns over how stress testing is currently implemented.

We initially operate in a simple, but insightful, 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 contribute to robustness theory in economics by articulating 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. Allowing for regime switching, parameter or estimation uncertainty, latency and non-standard shock distributions are promising avenues.

7 Appendix

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

7.1 Appendix 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

$$\begin{aligned} V(\epsilon_t, s_t) &= \min_{m(\epsilon_{t+1}, s_{t+1})} h(z_t) \\ &+ \beta \int m(\epsilon_{t+1}, s_{t+1}) V(\epsilon_{t+1}, s_{t+1}) p_\epsilon(\epsilon_{t+1}|x_t) \\ &+ \theta m(\epsilon_{t+1}, s_{t+1}) \log m(\epsilon_{t+1}, s_{t+1}) p_\epsilon(\epsilon_{t+1}|x_t) d\epsilon_{t+1} \end{aligned} \quad (10)$$

subject to $\int m(\epsilon_{t+1}, s_{t+1}) p(\epsilon_{t+1}|x_t) d\epsilon_{t+1} = 1$ 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

$$m(\epsilon_{t+1}, s_{t+1}) = \frac{e^{-\frac{V(\epsilon_{t+1}, s_{t+1})}{\theta}}}{E \left[e^{-\frac{V(\epsilon_{t+1}, s_{t+1})}{\theta}} | \epsilon_t, s_t \right]} \quad (11)$$

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 (12) is with respect to the benchmark transition density.

$$V(x_t) = h(z_t) - \beta \theta \log E \left[\exp \left(-\frac{V(x_{t+1})}{\theta} \right) | x_t \right] \quad (12)$$

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 Appendix 2 - LQG Robustness

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

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

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

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

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 is²⁷

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

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

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

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 , $w_{t+1} = K x_t$. Letting $-x'_0 P x_0 - p$ be the value of the problem and $h(z) = z' W z (= x' Q x)$, then we have the following Bellman equation

$$-x' P x - p = \min_w E \{h(z) + \theta \beta w' w - \beta x^{*'} P x^* - \beta p\}$$

²⁷The lower bound or ‘breakdown’ point considered for θ , θ_{bd} , ensures that the problem remains well posed.

subject to

$$x^* = Ax + C(\varepsilon + w) \quad (13)$$

$$\varepsilon \sim N(0, I) \quad (14)$$

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:

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

given x_0 and subject to equation (14). 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

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

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

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

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

$$x^* = Ax + Cw$$

Thus we have that

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

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

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

subject to the approximating model²⁸

$$x^* = Ax$$

²⁸ 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, $T(\tilde{P})$ is given by

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

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

$$K = \theta^{-1} (I - \theta^{-1} C' P C)^{-1} C' P A$$

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

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

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 Appendix 3 - Tilted Robustness

The first order conditions of the minimization problem 4 (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 by²⁹

$$m^*(x; \varphi^*) = \frac{\exp \left\{ -\frac{v(x) - \varphi^* g(x)}{\theta} \right\}}{E \left[\exp \left\{ -\frac{v(x) - \varphi^* g(x)}{\theta} \right\} \right]} \quad (15)$$

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

$$\int m(x; \varphi) \pi(x) g(x) dx = 0 \quad (16)$$

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

²⁹We make explicit the dependence on the multiplier φ as we have not yet discussed how it is obtained.

8 Tables

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

	(1)	(2)	(3)
	Aggregate	Inst. 1	Inst. 2
Unemp.	-3.918***	-3.238***	-1.643
PCEPI	-4.713**	-3.909**	-3.745
3M Tbill	2.226	-1.068	7.496
DEquities	3.875***	1.655***	3.097***
Term Spr.	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: Means and standard deviations under benchmark and worst cases for different DEPs - based on full sample estimates.

	Benchmark		W.C. (0.1)		W.C. (0.01)	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
RoE	9.03	6.60	1.06	7.90	-7.80	9.04
Unemp.	-0.01	1.01	1.00	1.14	2.12	1.27
PCPEPI	-0.2	0.93	1.11	1.12	2.57	1.29
3M Tbill	-0.33	1.02	0.64	1.11	1.71	1.21
DEquities	-0.04	0.99	-0.07	1.00	-0.10	1.01
Term Spr.	0.07	1.02	0.12	1.03	0.17	1.03

Table 3: Unconditional correlations among the states under the benchmark and worst cases.

$$\begin{array}{c} \text{(a) Benchmark} \\ \left\{ \begin{array}{ccccc} 1.00 & \cdot & \cdot & \cdot & \cdot \\ 0.17 & 1.00 & \cdot & \cdot & \cdot \\ -0.14 & 0.71 & 1.00 & \cdot & \cdot \\ 0.08 & -0.05 & 0.02 & 1.00 & \cdot \\ 0.56 & -0.31 & -0.57 & -0.03 & 1.00 \end{array} \right\} \end{array}$$

$$\begin{array}{c} \text{(b) W.C. (0.1)} \\ \left\{ \begin{array}{ccccc} 1.00 & \cdot & \cdot & \cdot & \cdot \\ 0.37 & 1.00 & \cdot & \cdot & \cdot \\ 0.05 & 0.74 & 1.00 & \cdot & \cdot \\ 0.05 & -0.07 & 0.03 & 1.00 & \cdot \\ 0.51 & -0.24 & -0.50 & -0.02 & 1.00 \end{array} \right\} \end{array}$$

$$\begin{array}{c} \text{(c) W.C. (0.01)} \\ \left\{ \begin{array}{ccccc} 1.00 & \cdot & \cdot & \cdot & \cdot \\ 0.50 & 1.00 & \cdot & \cdot & \cdot \\ 0.19 & 0.76 & 1.00 & \cdot & \cdot \\ 0.03 & -0.08 & 0.03 & 1.00 & \cdot \\ 0.48 & -0.20 & -0.44 & -0.01 & 1.00 \end{array} \right\} \end{array}$$

9 Figures

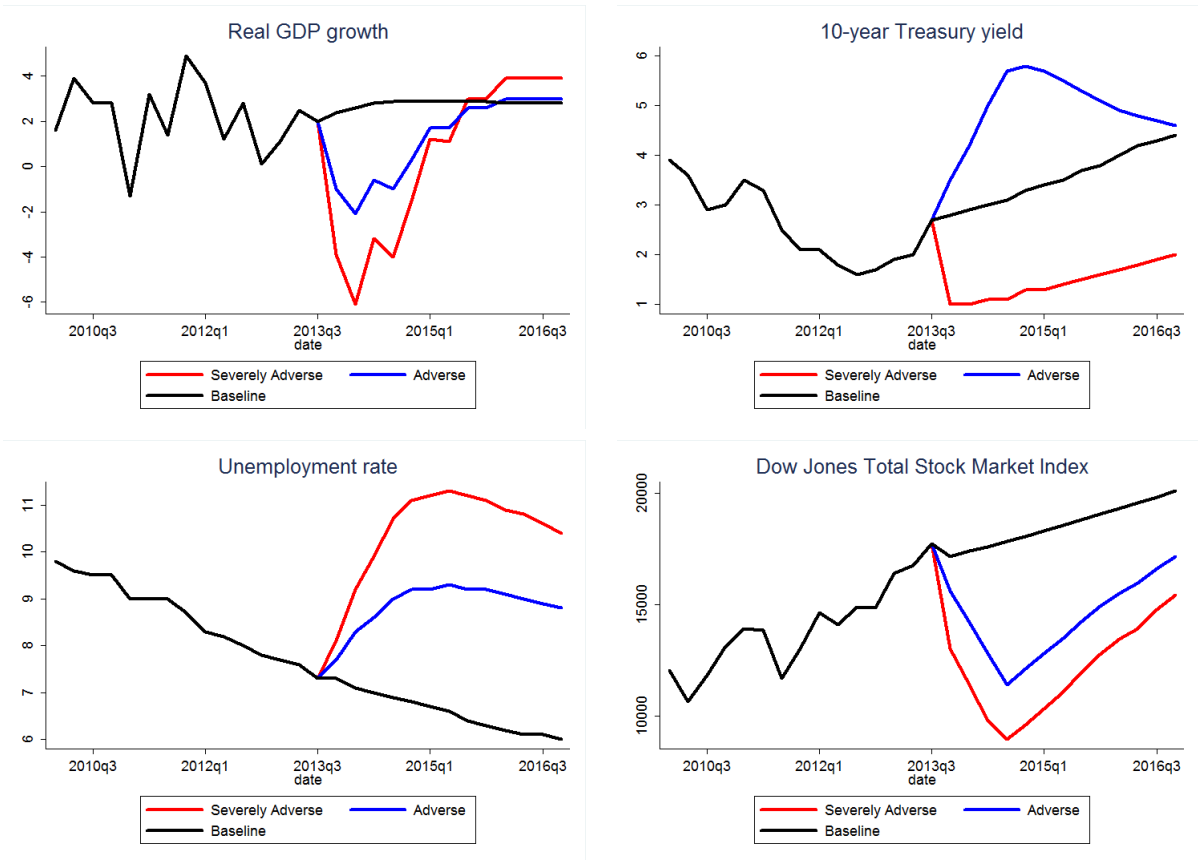


Figure 1: A subset of CCAR scenarios.

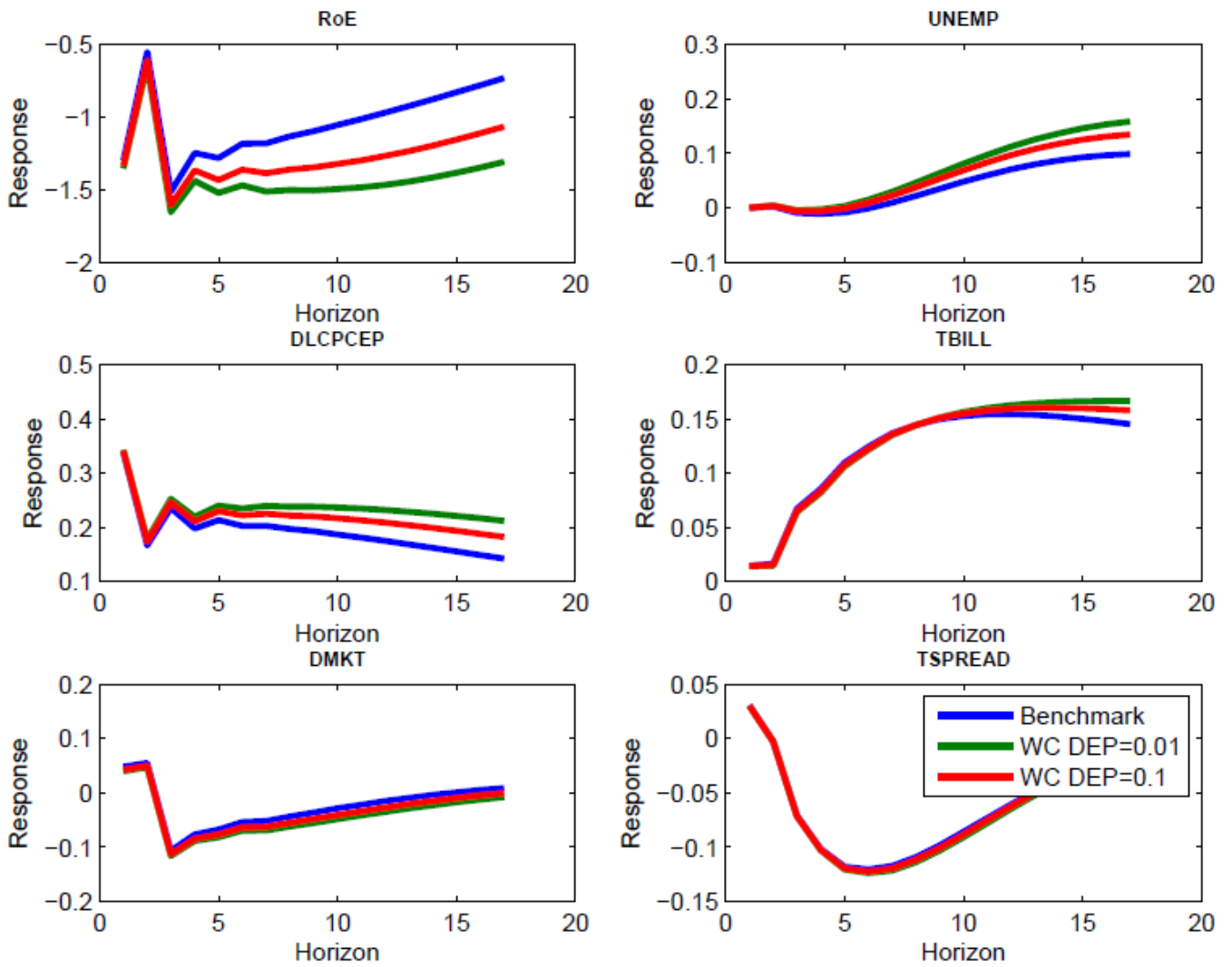


Figure 2: Orthogonalized impulse response to inflation. Based on full sample estimated bank loadings.

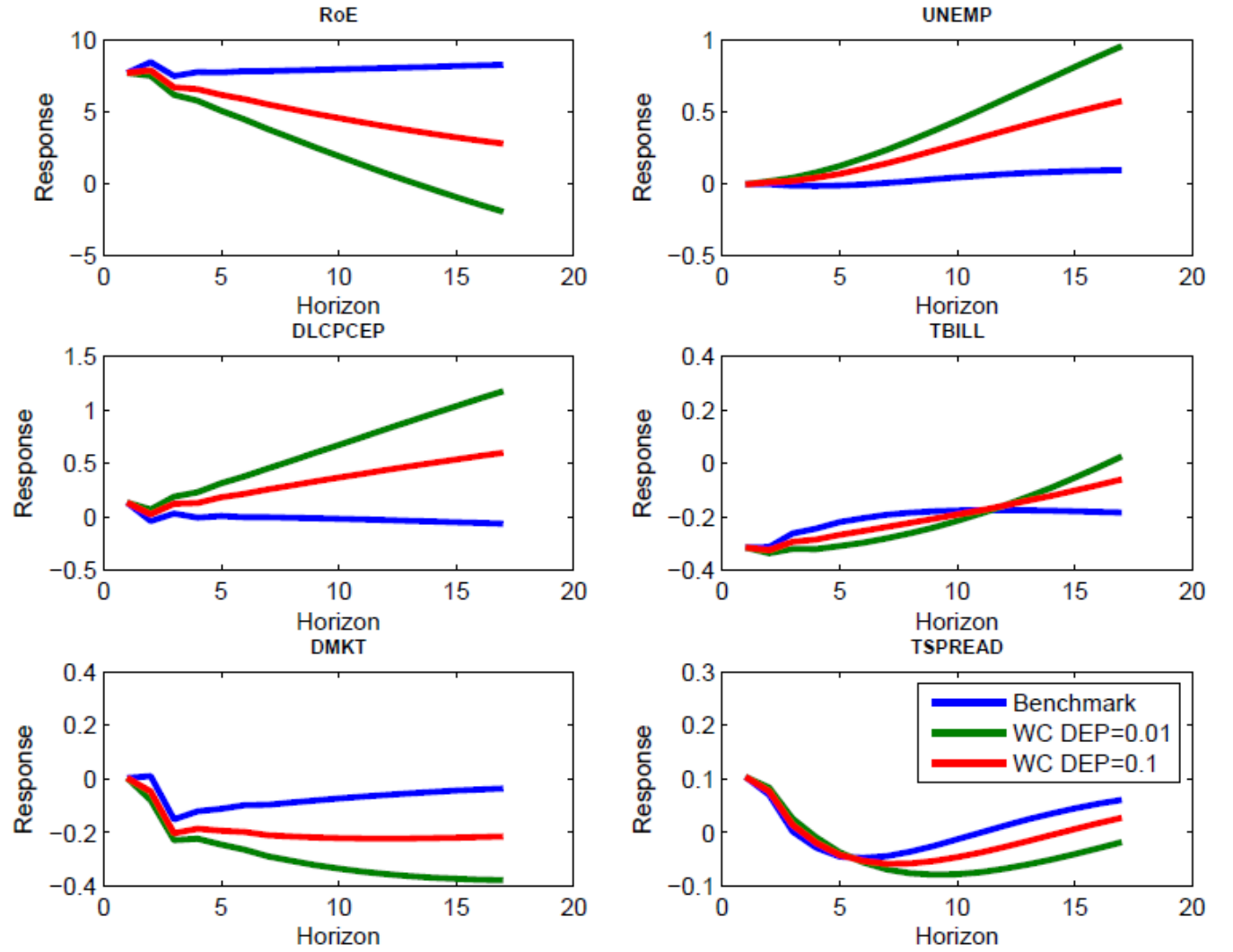


Figure 3: Evolution of economy from a situation where there has been an orthogonalized shock to inflation. Based on full sample estimated bank loadings.

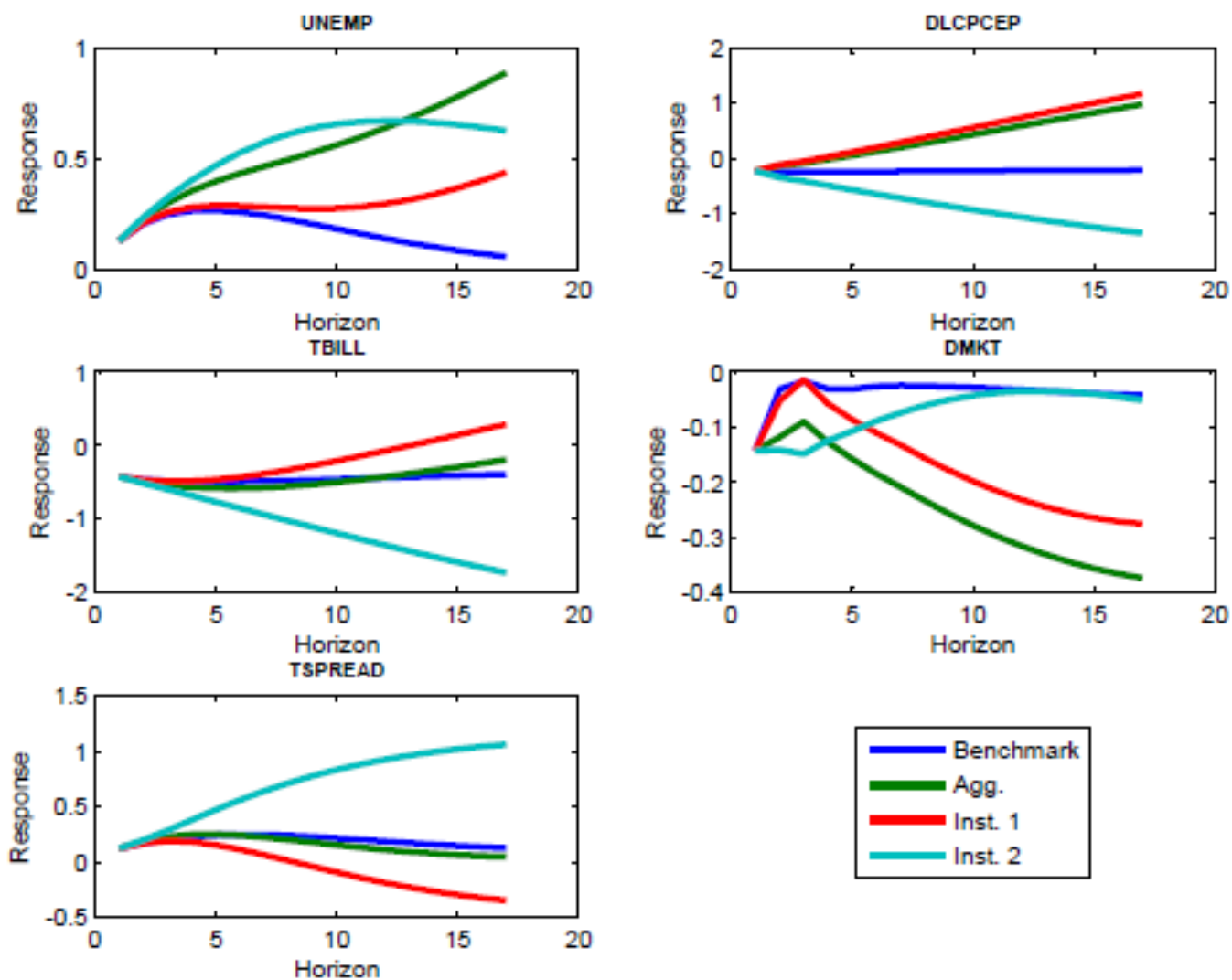


Figure 4: Evolution of economy from a situation where there has been an orthogonalized shock to unemployment. Comparing benchmark and worst cases based on aggregate and institution-specific regressions.

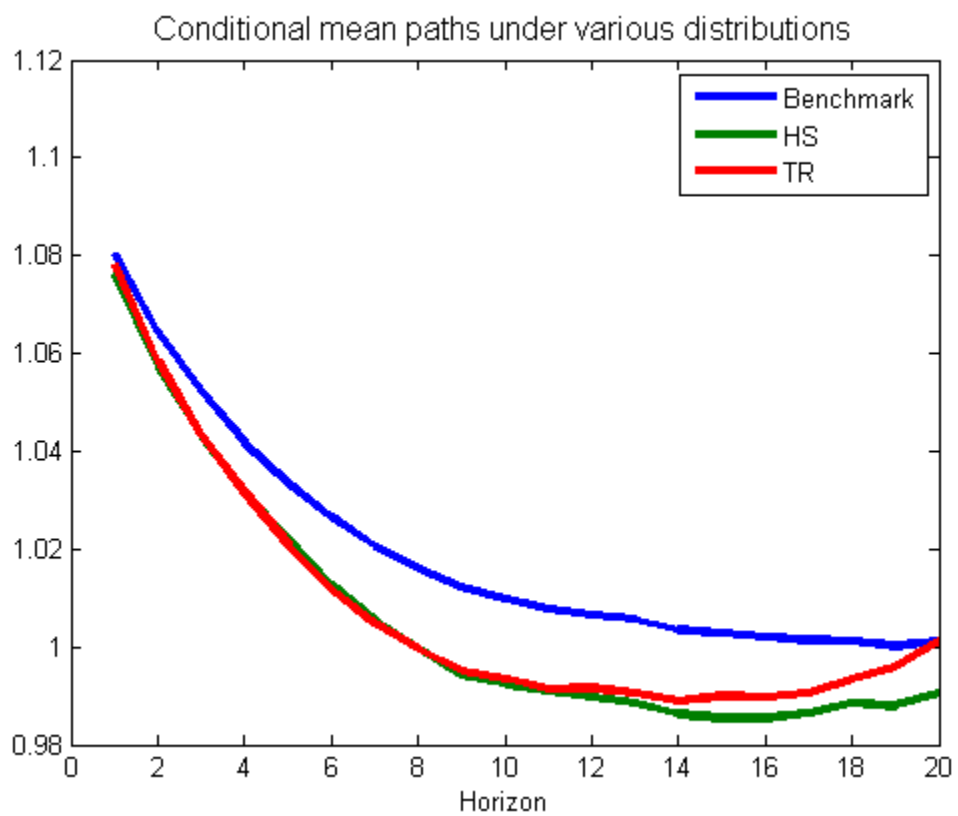


Figure 5: Marginal means of forecasts at each horizon.

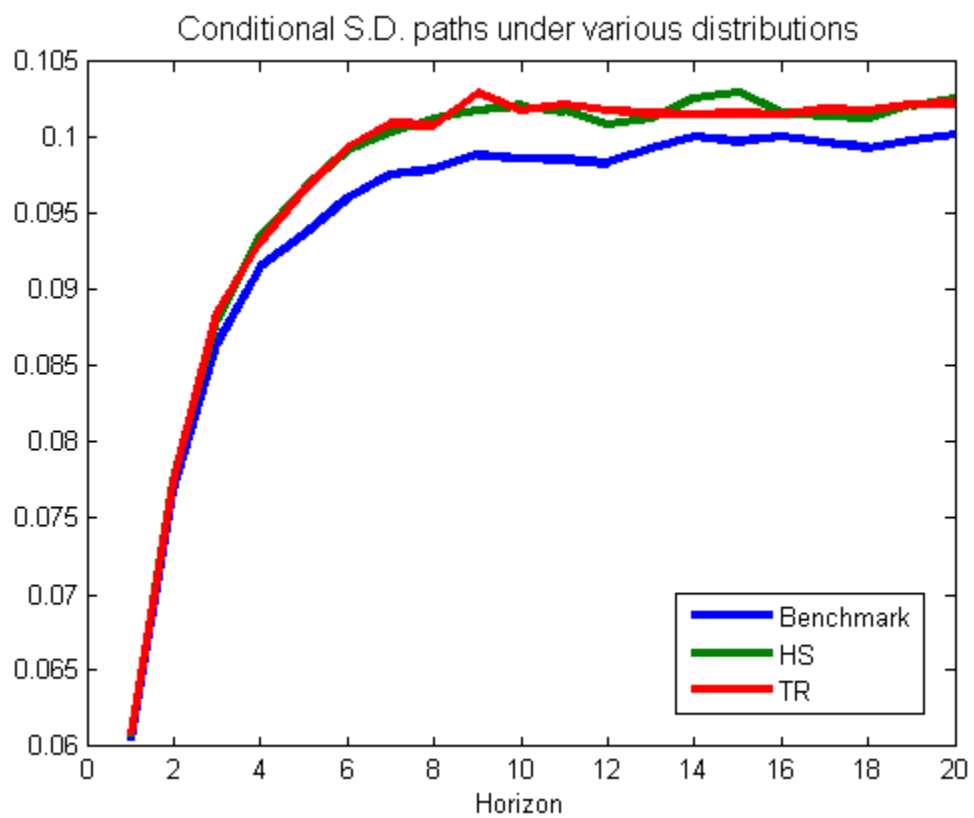


Figure 6: Marginal standard deviations of forecasts at each horizon.

References

- BIDDER, R., AND I. DEW-BECKER (2014): “Long-run risk is the worst-case: ambiguity aversion and non-parametric estimation of the endowment process,” .
- BIDDER, R., AND M. SMITH (2012): “Robust animal spirits,” *Journal of Monetary Economics*, 59(8), 738–750.
- BIDDER, R., AND M. E. SMITH (2013): “Doubts and Variability: A Robust Perspective on Exotic Consumption Series,” Federal Reserve Bank of San Francisco, Working Paper No. 2013-28.
- BOARD OF GOVERNORS, F. R. (2013): “Comprehensive Capital Analysis and Review 2013: Assessment Framework and Results,” Technical report.
- 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., M. JANDACKA, K. RHEINBERGER, AND M. SUMMER (2009): “How to find plausible, severe and useful stress scenarios,” *International Journal of Central Banking*.
- 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.
- FURLONG, F. (2011): “Stress testing and bank capital supervision,” *FRBSF Economic Letter*, (June 27).
- GUERRERI, 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 (1995): “Discounted Linear Exponential Quadratic Gaussian Control,” *IEEE Transactions on Automatic Control*, 40(5), 968–971.
- (2008): *Robustness*. Princeton University Press, Princeton, NJ.
- HIRTLE, B., T. SCHUERMANN, AND K. STIROH (2009): “Macroprudential supervision of financial institutions: lessons from the SCAP,” Discussion paper.

- HOGGARTH, G., S. SORENSEN, AND L. ZICCHINO (2005): “Stress tests of UK banks using a VAR approach,” Bank of England working papers 282, Bank of England.
- KWON, H., AND J. MIAO (2013): “Three types of robust Ramsey problem in a Linear Quadratic Framework,” .
- 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, [jce:titlejResearch on the Financial Crisisj/ce:titlej](#).
- SCHUERMANN, T. (2013a): “The Fed’s Stress Tests Add Risk to the Financial System,” .
- (2013b): “Stress Testing Banks,” Oliver Wyman, Wharton Financial Institutions Center.
- 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.