Estimating Robustness*

Bálint Szőke

New York University

balint.szoke@nyu.edu

September 24, 2019

Abstract

I estimate and evaluate a model with a representative agent who is concerned that the persistence properties of her baseline model of consumption and inflation are misspecified. Coping with model uncertainty, she discovers a pessimistically biased worst-case model that dictates her behavior. I combine interest rates and aggregate macro series with cross-equation restrictions implied by robust control theory to estimate this worst-case distribution and show that (1) the model's predictions about key features of the yield curve are in line with the data, and (2) the degree of pessimism underlying these findings is plausible. Interpreting the worst-case as the agent's subjective belief, I derive model implied interest rate forecasts and compare them with analogous survey expectations. I find that the model can replicate the dynamics and average level of bias found in the survey.

Keywords— Robustness, relative entropy, asset prices, survey expectations, zero-coupon yields, exponential quadratic stochastic discount factor

*I thank Jarda Borovicka, Monika Piazzesi, Jonathan Payne, Martin Schneider, Jiao Shi, Shenghao Zhu, and especially Lars Peter Hansen for useful criticisms and suggestions. I owe special thanks to Tom Sargent for his guidance and constant support throughout this project.

1 Introduction

Survey expectations are commonly thought to provide valuable information about 'subjective' beliefs.[1] Interestingly, these measures often display large and systematic differences compared to 'objective' forecasts from estimated statistical models. By way of illustration, Piazzesi, Salomao, and Schneider (2015) provide evidence that survey expectations of financial forecasters about future interest rates are formed as if the forecasters believed that the dynamics of yield curve were more persistent than they appear with hindsight. Although findings like this are important first steps to understanding belief formation, they leave unanswered the question: why do beliefs deviate from the observable dynamics?

Building on the robust control model of Hansen et al. (forthcoming), this paper offers a potential answer by presenting a particular mechanism of belief formation and proposing a general method to evaluate its plausibility as an explanation for the seeming inconsistencies between survey expectations and realized time series. The basic idea is that, under mild conditions, robust control theory can supply a model of subjective beliefs, providing testable predictions about how these beliefs deviate from forecasts derived from fitted statistical models. In particular, model implied beliefs distort physical probabilities by overweighting states with adverse utility consequences, so the resulting expectations appear to be *pessimistic* relative to the observable state dynamics.

Importantly, while this theory departs from rational expectations, it provides a set of powerful cross-equation restrictions between the decision maker's preferences, beliefs, and environment. Studying an endowment economy with a robust representative household who faces an endowment stream subject to long-run risk, I use these restrictions along with data on interest rates, consumption, and inflation to estimate the model parameters. In doing so, I switch viewpoint and consider robust control theory as a particular model for the stochastic discount factor. Using the estimated model, I derive implied interest rate expectations and contrast them with analogous forecasts from the *Blue Chip Financial Forecasts* (BCFF) survey. Assuming that these are good proxies for beliefs allows me (1) to test the model's prediction about pessimistic beliefs, and (2) to gain insights into how expectations are formed.

I find that the model implied belief well approximates the average forecast bias of professional forecasters for various maturity and forecast horizons. The model can also reproduce the finding of Piazzesi, Salomao, and Schneider (2015) that, under the subjective (or survey) belief, both the level and the slope of the nominal yield curve appear to be more persistent than what the observed yields suggest. On the other hand, the volatility of model implied expected yield changes falls short of the survey analogues, which indicates that the increased persistence that I estimate from observed prices might be somewhat excessive from the viewpoint of the surveys.

My decision maker acknowledges that her *baseline model* can only approximate her environment, so, rather than using a single model, she considers a *set* of alternatives that are difficult to distinguish from the baseline. Seeking a decision rule that is robust to these alternative models, she ends up discovering a *worst-case model* to which she responds optimally. Being a probability distribution that justifies the robust decision rule as a best response to beliefs, this worst-case model can be interpreted as the decision maker's subjective belief. Hansen and Sargent (2001) represent a set of models as a collection of

[1] For conceptual and methodological issues related to survey expectations, see Pesaran and Weale (2006).

likelihood ratios whose relative entropies with respect to a baseline are bounded by a *single* parameter encoding the decision maker's desired *degree of robustness*. Hansen et al. (forthcoming) refine this setting by introducing a new object, ξ, a nonnegative state dependent function, meant to guarantee that certain parametric models are included in the set. Twisting the set toward particular parametric alternatives embodies *structured uncertainty*, which expresses that the decision maker is concerned about particular misspecifications among many others.

I use a multivariate extension of this model with a quadratic ξ that is well-suited to direct the agent's misspecification concerns to the *persistence* of the baseline dynamics. As a comparison, I also consider the special case of constant ξ, which gives back the original Hansen and Sargent (2001) model with *unstructured uncertainty* that has well-known close connections with recursive utility of Duffie and Epstein (1992). The question then arises as to how important the state dependence of ξ might be. I demonstrate that it is essential to replicate robust features of the interest rate data. In particular, however large I set the risk aversion parameter, recursive utility with a unitary elasticity of substitution cannot generate an upward sloping average nominal yield curve.

In contrast, by allowing for state dependence in ξ, the model not only replicates the observed average nominal yield curve, but it also generates substantial fluctuations in nominal yields with long maturities, features that are strongly backed by the data, but that a constant ξ fails to account for. Fluctuations in long-term yields are implications of state dependent, counter-cyclical market prices of model uncertainty[2] that emerge from the agent's concerns that the persistence of the baseline dynamics is misspecified. The estimated worst-case suggests that she is most worried that inflation is more persistent and that the correlation between consumption growth and lagged inflation is more negative than in her baseline model. This offers a reinterpretation of the 'bad news' channel of Piazzesi and Schneider (2007).

Combining recursive utility with sufficiently high risk aversion and a particular long-run risk feature of their data set, Piazzesi and Schneider (2007) obtain an upward sloping nominal yield curve. Underlying this result is their finding that inflation predicts low future consumption growth. My failure to replicate a positive term premium with recursive utility follows from the fact that the forecasting ability of inflation is more modest in my sample, signaling its *fragility* in the data.[3] Misspecification concerns about that feature therefore seem plausible and indeed, this is what I find by looking at the data through the lens of the estimated worst-case model. One advantage of this interpretation, emphasized by Barillas, Hansen, and Sargent (2009), is that implausibly large risk aversion parameters can be replaced with plausible concerns about robustness to (specific kinds of) misspecifications. Calculating a detection error based measure associated with the estimated worst-case model, I show that the agent's belief is indeed reasonable in the sense that it is not easily rejectable by the data.

The paper contributes to a literature that aims to estimate and empirically evaluate models with robustness, that is, models in which (1) agents exhibit uncertainty aversion, and (2) the set of pri-

[2]That is, the compensation for facing *ambiguity* about the distribution of shocks. Ambiguity can amplify and induce fluctuations in prices of exposures to the *original* shocks without introducing extra shocks (with known *risk*).

[3]The main difference between the two data sets is that theirs ends in 2005:Q4, while mine goes on until 2017:Q2. Using the shorter sample and a scalar ξ (recursive utility), I can replicate the upward sloping yield curve and the high risk aversion that they find. Nonetheless, even in the shorter sample I find evidence for the importance of state-dependence in the tilting function. For more details, see Appendix C.

ors is tightly constrained by statistical discrimination based considerations. Prominent examples are Hansen, Sargent, and Tallarini (1999), Cagetti et al. (2002), Bidder and Smith (2012), and Bhandari, Borovička, and Ho (2019). More broadly, this paper relates to a literature focusing on the quantitative importance of ambiguity aversion in macroeconomics with a prime example being Ilut and Schneider (2014). Although similar in spirit, their paper differs from mine in that it is built on the alternative decision theoretic framework of Epstein and Schneider (2003), which gives rise to significantly more permissive restrictions on the agent's set of priors than those used here.

The closest to my paper is the work by Bhandari, Borovička, and Ho (2019). Similar to the present setting, they extend the model of Hansen and Sargent (2001) in ways that imply time variation in the worst-case drift distortion, and interpret the worst-case as a subjective belief, which enables them to utilize survey expectations in their analysis. However, instead of using surveys to evaluate the model's predictions, they use them to identify exogenous variations in the degree of ambiguity. Another key difference is that they estimate a full-blown general equilibrium model with endogenous consumption. While this allows them to investigate the impact of robustness on aggregate macro dynamics, due to the complexity of the model, they must rely on approximations. In contrast, the endowment economy and my functional form assumptions guarantee a tractable linear quadratic framework, at the cost of limiting my focus to how robustness affects asset prices and beliefs.

An important feature of my paper's decision problem is that the set of models is considered to be a result of careful deliberation, an honest "best guess", that the decision maker does not seek to improve over time through learning. This is because eliminating the kinds of misspecifications that she worries about would require so much data that, due to discounting the future, she simply accepts model misspecification as a permanent state of affairs. Rather than struggling with slow learning and fully embracing one of the 'wrong' models, my decision maker designs a decision rule that works well under a set of models. Hansen and Sargent (2007) and Hansen and Sargent (2010) are attempts to incorporate learning into robust decision problems.

The rest of the paper is structured as follows. Section 2 summarizes the essential features of robust preferences and the particular refinement used in this paper. It also discusses the functional form assumptions and solves the planner's problem to derive formulas for the equilibrium yield curve. Section 3 proposes a two-step maximum likelihood procedure to estimate the model parameters, reports details about the estimation and discusses the results. In Section 4, I compare model implied beliefs with survey expectations using some informative moments. Section 5 concludes. The appendices contain derivations and further results.

2 Robust decision problem

I consider an endowment economy with a representative household who has robust preferences. In the spirit of Lucas (1978), I focus on the consumption side of the economy abstracting from production or storage, the idea being that one can price assets from marginal utilities evaluated at the equilibrium consumption process ignoring the deeper features that made them into equilibrium outcomes. More precisely, I suppose that the observed aggregate consumption series is induced by a *robust* equilibrium decision rule of some dynamic economic model. Instead of specifying the details of this economic model, I presume that its equilibrium exhibits a convenient statistical representation that can be depicted by a fictitious endowment economy.

Let W denote a 2-dimensional Brownian motion on a probability space $(\Omega, \mathcal{F}, Pr)$ and let $\{\mathcal{F}_t\}_{t\geqslant 0}$ denote the completion of the filtration generated by W.[4] A representative household receives a stochastic real endowment stream D and faces an exogenously given price process of consumption goods P. Suppose that both processes are multiplicative functionals[5] of the Markov states X, so the nominal endowment DP is also multiplicative. For simplicity, define the vector $Y \equiv [D, P]'$. The household considers a following baseline model of Y:[6]

$$\begin{aligned} d\log Y_t &= \beta(X_t)dt + \alpha(X_t)dW_t \\ dX_t &= \mu(X_t)dt + \sigma(X_t)dW_t, \end{aligned} \tag{1}$$

where X contains all relevant state variables with given initial X_0. The household thinks this model is a good *approximation* to aggregate output and prices, but is concerned about hard-to-detect, welfare relevant model misspecifications. As a result, she gives us optimality under the single model (1) and instead seeks decision rules that perform *relatively well* under *all* alternative models that she finds reasonable. The heart of the analysis is to characterize the set of models $\mathcal{Z}$ against which the household seeks robustness. This set is captured by parameters that express misspecification fears and its estimation is an important purpose of this paper.

Hansen and Sargent (2001) argue that a useful way of characterizing the set $\mathcal{Z}$ is to include models that are statistically difficult to distinguish from (1) based on a finite history of data, and propose using discounted relative entropy to quantify this difficulty. Following this approach, I treat (1) as the reference point to which all considered models are statistical perturbations and define $\mathcal{Z}$ as a 'ball' around this reference point using an entropy based measure as statistical distance.

2.1 Econometrician's model

Although the theory does not say much about how the baseline model gets determined, common sense suggests that it should not be easily rejectable by the data. It seems minimal to require that the baseline represents observable time series well. This consideration leads to the following assumption:

Assumption 1. *The decision maker's baseline model is the 'best' approximation of the data generating process within a particular class of parametric statistical models that fits the data well.*

As a justification, notice that this assumption places the agent on comparable footing with us, econometricians. While we typically have a bunch of prejudices about the form of well fitting statistical models, i.e. specific β, α, μ and σ functions in (1), and can use *finite* samples to pin down particular

[4] I will adopt the following convention. Let X (without time subscript) denote a stochastic process, X_t the process at time t, and x a realized value of the process.

[5] For an in-depth analysis of the properties of multiplicative functionals, see Hansen and Scheinkman (2009).

[6] In the following, I will use the word *model* to describe joint probability distributions over sequences of random variables, a possible representation of which is the stochastic differential equation (1).

parameter values, we almost never possess large enough datasets to rule out the possibility that the data were actually generated by a model other than our *point* estimate. Assumption 1 basically asserts that the decision maker faces the same difficulties and discovers her baseline in the same way as an econometrician does. As a result, it provides the same economical way of endowing her with a model as rational expectations econometrics does. Motivated by Assumption 1, in what follows, I will use the baseline as a stand-in for the physical distribution, so that whenever I calculate moments under the 'objective' measure (denoted by $\mathbb{E}$), I will use the baseline model.

Specifically, I assume that the household's baseline model takes the following affine form

$$d\log Y_t = (\beta_0 + X_t)\,dt + \alpha dW_t \tag{2}$$

$$dX_t = \quad -\kappa X_t dt + \sigma dW_t \tag{3}$$

where X is a 2-vector. The discrete time representation of (2)-(3) is identical to the state space model used by Piazzesi and Schneider (2007) to capture the dynamics of quarterly consumption growth and inflation. As they pointed out, although this form nests a first-order vector autoregressive process (VAR), it is more flexible due to the presence of the latent variables that allow for first-order moving-average style dynamics, crucial to capture the long-run behavior of the inflation process.[7] In section 3.1, I show that this simple model indeed fits the joint dynamics of consumption growth and inflation well, thereby meeting the requirement of Assumption 1.

2.2 Twisting probabilities

To represent alternative models of Y, I perturb the reference measure Pr by multiplying it with a likelihood ratio process Z^H that follows

$$dZ_t^H = Z_t^H H_t \cdot dW_t \tag{4}$$

with initial condition $Z_0^H = 1$. Here, H is a progressively measurable vector-valued process that satisfies $\int_0^t |H_s|^2 ds < \infty$, with probability one. Each alternative model has a likelihood ratio with respect to the baseline, so that one can index them by their corresponding H processes. Since my goal is to estimate the set $\mathcal{Z}$, it is natural to seek a parsimonious parametrization. In a series of papers,[8] Hansen and Sargent use discounted relative entropy to measure statistical distance between alternative models and construct $\mathcal{Z}$ with only one parameter. In particular, for a given perturbation H and initial $X_0 = x$, discounted relative entropy, with discount rate δ, is

$$\Delta(H \mid x) := \frac{\delta}{2}\int_0^\infty \exp(-\delta t)\mathbb{E}\left[Z_t^H |H_t|^2 \mid X_0 = x\right] dt.$$

[7] Inflation is assumed to be exogenous, which might not be innocuous given that the worst-case model depends on the various channels through which inflation can affect the household's welfare. In the current setting, the only functions of inflation are (1) to determine the real cash-flow of assets and (2) to forecast future consumption growth; consequently, the worst-case model can reflect only these roles. If inflation were endogenous, it could have other welfare relevant impacts that would probably alter the worst-case and thus the agent's subjective beliefs. I ignore these alternative roles.

[8] A selective list is Hansen and Sargent (2001), Anderson, Hansen, and Sargent (2003), Hansen et al. (2006).

The set $\mathcal{Z}$ can then be defined by a scalar upper bound $\zeta \in \mathbb{R}_{++}$, so that the decision maker worries only about those perturbations that satisfy $\Delta(H \mid x) \leqslant \zeta$. With the use of relative entropy, one can include a myriad of parameteric and non-parametric models with non-trivial dynamics. In fact, for some purposes, the set defined by $\Delta(H \mid x)$ is too general in the sense that it can represent only *unstructured* misspecifications. It does not let the decision maker focus on particular parametric misspecifications.

Hansen et al. (forthcoming) refine this specification by endowing the decision maker with a *twisted* set that focuses her specification doubts on particular parametric aspects of the baseline model. Although the extension brings additional free parameters, for my purposes this is a price worth paying. By estimating the *twisted set*, I can 'let the data speak' about the aspects of the baseline model that concern the decision maker most. The extra parameter is a time invariant, nonnegative function of the states $\xi(X_t)$ – call it a *twisting function* – that can be used to define the twisted set as follows:

$$\mathcal{Z}(\xi)[x] := \left\{ Z^H \ : \ \frac{\delta}{2}\int_0^\infty \exp(-\delta t)\mathbb{E}\left[Z_t^H\left[|H_t|^2 - \xi(X_t)\right] \mid X_0 = x\right] dt \leqslant 0 \right\}. \tag{5}$$

State dependence in the function ξ is a device to guarantee that certain parametric alternatives to (1) are included in $\mathcal{Z}(\xi)$. These *worrisome* alternatives twist the entropy ball in particular parametric directions and make sure that the household's valuations and behavior are robust to the kinds of misspecifications that they represent.[9]

An important special case is a scalar $\xi \geqslant 0$. It is straightforward to see that by setting $\xi = 2\zeta$, one obtains the non-twisted set of Hansen and Sargent (2001) parameterized by ζ. Furthermore, the standard robust control model has well-known close connections to recursive utility when the intertemporal elasticity of substitution is one, which raises the question whether the state dependence of ξ is a crucial feature or not. Thanks to the nested nature of this relationship, I can and will investigate this question in section 3.3.1.

Concern about long-run risk

The model (2)-(3) incorporates a long-run risk channel, similar (but not identical) to Bansal and Yaron (2004), in the sense that the first entry of X is a predictable, serially correlated component of the consumption growth process. Additionally, the possible interaction between the state variables enables expected inflation (second entry of X) to affect the long-run predictability of consumption growth. To illustrate how the twisting function can be used to express "parameter uncertainty", suppose that the decision maker is concerned that her baseline model might miscalibrate long-run features of consumption growth and as a result, she insists that a particular worrisome model $Z^{\bar{H}}$ is included in her set $\mathcal{Z}(\xi)$. For simplicity, let this worrisome model have the same parametric specification as the

[9] From a somewhat different viewpoint, $\xi(X)$ can also be viewed as a state dependent (additive) adjustment to the relative entropy cost that makes the effective size of the ball $\mathcal{Z}$ vary with the state. This is one way to model the phenomenon that the decision maker's attitude toward misspecification might differ in good and bad times.

baseline (2)-(3), but with different drift parameters $(\bar{\phi}, \bar{\kappa}, \bar{\beta}_0, \bar{\beta}_1)$:

$$d \log Y_t = \left(\bar{\beta}_0 + \bar{\beta}_1 X_t\right) dt + \alpha dW_t^{\bar{H}} \tag{6}$$

$$dX_t = \left(\bar{\phi} - \bar{\kappa} X_t\right) dt + \sigma dW_t^{\bar{H}}. \tag{7}$$

The corresponding drift distortion process $\bar{H}$ satisfies

$$\begin{bmatrix} \alpha \\ \sigma \end{bmatrix} \bar{H}_t = \begin{bmatrix} \bar{\beta}_0 - \beta_0 \\ \bar{\phi} \end{bmatrix} + \begin{bmatrix} \bar{\beta}_0 - I_2 \\ \kappa - \bar{\kappa} \end{bmatrix} X_t.$$

Hansen et al. (forthcoming) show that by specifying the quadratic twisting function:

$$\xi(X_t) = |\bar{H}_t|^2 \tag{8}$$

the implied twisted set $\mathcal{Z}(\xi)$ will include worrisome model (6)-(7), because $\bar{H}$ satisfies the condition of (5) with equality. This is the sense in which the entropy ball is twisted toward $Z^{\bar{H}}$, and the quadratic form and parameter values of ξ can encode the decision maker's parameter uncertainty about long-run risk. Note, however, that while ξ ensures that certain parametric models are included, the set $\mathcal{Z}(\xi)$ contains a myriad of other non-parametric models with much less structure that are nonetheless statistically similar to the worrisome alternatives.[10]

Interestingly, as I will show in section 2.6, the household's struggle to manage doubts about long-run risk can give rise to time varying, counter-cyclical risk premia despite the fact that the baseline model (2)-(3) does not include stochastic volatility. Motivated by this fact, in what follows, I assume that the twisting function is quadratic in X:

$$\xi(X_t) = \xi_0 + X_t'\xi_2 X_t = [1, X_t'] \underbrace{\begin{bmatrix} \bar{\xi}_0 & 0 & 0 \\ 0 & \bar{\xi}_1 & \bar{\xi}_2 \\ 0 & \bar{\xi}_2 & \bar{\xi}_3 \end{bmatrix}}_{\equiv \Xi} \begin{bmatrix} 1 \\ X_t \end{bmatrix} \tag{9}$$

subject to the restriction that the matrix Ξ is positive semidefinite. In this way, I introduce 4 extra parameters $(\bar{\xi}_0, \bar{\xi}_1, \bar{\xi}_2, \bar{\xi}_3)$.

2.3 Worst-case model...

Turning to behavior, that $\mathcal{Z}(\xi)$ is a closed, convex set makes the max-min expected utility theory of Gilboa and Schmeidler (1989) directly applicable. More precisely, one can formulate a two-player, zero-sum game, where the two players are a utility maximizing decision maker and an auxiliary agent whose aim is to minimize the decision maker's expected utility by choosing a distribution from the set $\mathcal{Z}(\xi)$.[11] This formulation is useful because the equilibrium determines the *robust decision rule* and the

[10] Hansen et al. (forthcoming) show that the set $\mathcal{Z}(\xi)$ includes entire relative entropy neighborhoods of the parametric alternatives described with ξ.

[11] In spite of the language, this is still a single-agent decision problem; the second player is just a convenient device that the maximizing agent can use to contemplate the possible adverse consequences of her different policies.

worst-case model as optimal decisions of the two players. Exploiting the martingale formulation, the Lagrangian of the time 0 game can be written as

$$\max_{C \in \mathcal{C}} \min_{Z^H} \int_0^\infty \delta \exp(-\delta t) \mathbb{E}\left[Z_t^H \left(u(C_t) + \frac{\ell}{2}\left[|H_t|^2 - \xi(X_t)\right]\right) \mid \mathcal{F}_0 \right] dt \tag{10}$$

where ℓ is the multiplier associated with the restriction $Z^H \in \mathcal{Z}(\xi)$ and $u(\cdot)$ is the maximizing agent's period utility function. Fleming and Souganidis (1989) and Hansen et al. (2006) show how this game can be made recursive through a Markov perfect game where every instant the players choose their decision rules as functions of the state vector X. The justification is the so called Bellman-Isaacs condition

Assumption 2. ***Bellman-Isaacs:*** *For a fixed ℓ, there exists a value function V such that*

$$\begin{aligned}
\delta V(x) &= \max_c \min_h \; u(c) + \frac{\ell}{2}\left[|h|^2 - \xi(x)\right] + \nabla V_x \cdot \left[\mu(x,c) + \sigma(x,c)\cdot h\right] + \frac{1}{2} tr\left[\sigma(x,c)' V_{xx} \sigma(x,c)\right] \\
&= \min_h \max_c \; u(c) + \frac{\ell}{2}\left[|h|^2 - \xi(x)\right] + \nabla V_x \cdot \left[\mu(x,c) + \sigma(x,c)\cdot h\right] + \frac{1}{2} tr\left[\sigma(x,c)' V_{xx} \sigma(x,c)\right]
\end{aligned}$$

This condition defines the Hamilton-Jacobi-Bellman (HJB) equation and requires the order in which the two players choose to be interchangeable without affecting the optimal decisions. Although the game in which the maximizing player chooses first is closer to the idea of fear from misspecification, the possibility of switching the order opens up another interpretation.

2.4 . . . or subjective belief

Hansen et al. (2006) provide sufficient conditions under which the Bellman-Isaacs condition is satisfied, so in environments that meet these we can study min-max as well as the max-min games. The advantage of this is that if the minimizing player chooses first, his choice will be the distribution with respect to which the maximizing player's (robust) decision rule is a best response. In other words, *ex post* (after the minimizing H^* is determined), one can look at the worst-case model as beliefs that justify the robust decision rule. Hansen et al. (2006) demonstrate that this so called *ex post Bayesian* interpretation leads to an ordinary optimal control problem, where the decision maker has full confidence in a single model that happens to be the worst-case model, and that results in the robust decision rule. In this interpretation, however, the *ex post* label is crucial. First we have to solve for the equilibrium H^* before we make it equal to the decision maker's subjective belief. That is, the worst-case belief embodies not only what the decision maker finds possible *ex ante*, but also what is worth considering to be possible *ex post*. In this sense, subjective beliefs are endogenous and the model is interpretable as a particular (pessimistic) mechanism of belief formation.

2.5 Equilibrium drift distortion

One major benefit of the endowment economy assumption is enabling us to use the planner's problem and the associated shadow values to derive Arrow-Debreu prices that would emerge in a decentralized, competitive environment. With robust preferences, however, a non-trivial question arises: is the proposed price system induces the same consumption rule and worst-case model as those of the planner from which the prices were derived in the first place? Appendix B shows that the answer is affirmative.

The planner shares the household's baseline model and twisting function ξ. Suppose that the period utility function is logarithmic, $u(C) = \log C$. There is no way to move resources over time or states, so $C_t = D_t$ for all t. That being said, the max-min problem becomes a simple minimization that affects valuation not quantities. As before, let ℓ be the multiplier on the constraint $\mathcal{Z}(\xi)$ and define the planner's value function as $V(x, \log d) = \frac{1}{\delta}\log d + v(x)$ that solves the corresponding HJB equation

$$\begin{aligned}\delta V(x, \log d) = \min_h \quad & \log d + \nabla v_x(x)\cdot(-\kappa x + \sigma h) + \frac{1}{\delta}e_1 \cdot (\beta_0 + x + \alpha h) + \\ & + \frac{1}{2}\mathrm{tr}\left[\sigma' v_{xx}(x)\sigma\right] + \frac{\ell}{2}\left(|h|^2 - \xi(x)\right) \end{aligned} \tag{11}$$

where e_j is a selector vector for the jth entry. One can show (see Appendix A) that the function $v(x)$ is quadratic, so the *worst-case drift distortion* becomes affine in the states

$$H_t^* = -\frac{1}{\ell^*}\left((\alpha' e_1)\cdot \underbrace{\nabla V_{\log d}}_{=1/\delta} + \sigma' \underbrace{\nabla V_x}_{=v_1 + v_2 X_t}\right) =: \eta_0 + \eta_1 X_t \tag{12}$$

where ℓ^* is the value of the Lagrange multiplier that makes the constraint $Z^H \in \mathcal{Z}(\xi)$ bind and is a nonlinear function (25) of the value function V. The coefficients η_0 and η_1 are 2×1 and 2×2, respectively, and are both functions of ξ and the baseline parameters $(\kappa, \sigma, \beta_0, \alpha)$. These functions embody cross-equation restrictions across parameters describing the environment, preferences toward robustness and *ex post* beliefs. In my case, for a *fixed* baseline model, these restrictions essentially express the 6 unknowns of (η_0, η_1) with the 4-dimensional parameter vector $(\bar{\xi}_0, \bar{\xi}_1, \bar{\xi}_2, \bar{\xi}_3)$.[12]

The entries of the vector H^* are drift distortions to the respective Brownian shocks and, as (12) suggests, the relative magnitudes of these distortions are determined by $(\nabla V_x, \nabla V_{\log d})$, representing how much the continuation value is exposed to changes in the states, and (σ, α), that is the loadings of the states on the shocks. The distortion is larger in states where ∇V_x is relatively large and in this way the worst-case model overweights states that the continuation value identifies as adverse. The form of H^* is analogous to the standard robust control model of Anderson, Hansen, and Sargent (2003), with the exception that here the (possible) state dependence of ∇V_x is shaped by the twisting function ξ.

A principal advantage of the affine H^* is that it leads to a worst-case model that belongs to the same parametric family as the baseline. In particular, we can write the worst-case model as

$$d\log Y_t = \left[(\beta_0 + \alpha\eta_0) + (I_2 + \alpha\eta_1)X_t\right]dt + \alpha d\widetilde{W}_t =: \left(\widetilde{\beta}_0 + \widetilde{\beta}_1 X_t\right)dt + \alpha d\widetilde{W}_t \tag{13}$$

$$dX_t = \left[\sigma\eta_0 - (\kappa - \sigma\eta_1)X_t\right]dt + \sigma d\widetilde{W}_t =: \left(\widetilde{\phi} - \widetilde{\kappa}X_t\right)dt + \sigma d\widetilde{W}_t \tag{14}$$

where $\widetilde{\cdot}$ denotes objects from the worst-case model. Therefore, while the decision maker considers a

[12] In fact, in Section 3.3 I show that $\bar{\xi}_0$ turns out to be 'redundant' in explaining time series of interest rates.

plethora of non-parametric models with subtle, complicated dynamics against which she seeks robustness, the model that determines her robust decision rule is particularly simple.

2.6 Robust stochastic discount factor and equilibrium yield curve

The stochastic discount factor (SDF) can be constructed by evaluating the representative consumer's marginal rate of substitution at the exogenous consumption process. Given the logarithmic utility function, it follows that *real* and *nominal* marginal rate of subsitutions can be written as:

$$M_t := \exp(-\delta t)\frac{C_0}{C_t} \qquad \text{and} \qquad M_t^{nom} := \exp(-\delta t)\frac{(CP)_0}{(CP)_t}.$$

As a result, the date 0 price of an arbitrary date t nominal Markov payoff $\varphi(X_t)$ is given by

$$p_\varphi(x) = \widetilde{\mathbb{E}}\left[M_t^{nom}\varphi(X_t) \mid X_0 = x\right] = \mathbb{E}\left[Z_t^{H^*} M_t^{nom}\varphi(X_t) \mid X_0 = x\right]. \tag{15}$$

The first equality expresses the price under the investor's subjective belief (hence the $\widetilde{\cdot}$ notation), while the second equality depicts p_φ under the baseline using the equilibrium likelihood ratio, Z^{H^*}, as the change of measure. This likelihood ratio is a multiplicative martingale capturing the impact of pessimistic belief distortions on equilibrium prices. Real and nominal stochastic discount factors are

$$S_t := M_t Z_t^{H^*} \qquad \text{and} \qquad S_t^{nom} := M_t^{nom} Z_t^{H^*}$$

Using Ito's lemma, one can show that the dynamics of S_t^{nom} under the baseline model follow

$$\frac{dS_t^{nom}}{S_t^{nom}} = -\underbrace{\left(\delta + \iota\cdot\widetilde{\beta}_0 + \iota'\widetilde{\beta}_1 X_t - \frac{|\iota'\alpha|^2}{2}\right)}_{\equiv r^{nom}(X_t)} dt - \underbrace{\left(\alpha'\iota - H_t^*\right)}_{\equiv\rho(X_t)} \cdot dW_t \tag{16}$$

with ι denoting a 2-vector of ones. After multiplying the right hand side by minus one, the drift term gives the equilibrium nominal risk-free rate, r^{nom}, while the volatility term, ρ, represents equilibrium prices that are usually interpreted as local risk prices of the Brownian shocks.[13] Motivated by the distinct forces behind its two components, however, Hansen et al. (forthcoming) reinterpret ρ as follows. The first term, $\alpha'\iota$, is the *price of risk* defined as the risk averse investor's compensation for bearing risk, whereas the second term, $-H^*$, is the *uncertainty price* induced by the investor's doubts about the baseline model. When $\eta_1 \neq \mathbf{0}$, the uncertainty prices and the associated uncertainty premia fluctuate over time with X_t. Note that these fluctuations emerge endogenously from a baseline model with homoskedastic shocks as a result of the investor's uncertainty about those shocks.

The fact that the local mean is affine, while the local variance is quadratic in X suggests an exponential-quadratic specification for S_t^{nom} and S_t, hence they fit nicely in the multi factor affine family proposed by Duffie and Kan (1996). Consequently, I can borrow well-known formulas from the asset pricing literature (in particular, see section 5.1. in Borovička et al. (2011)) and express the yield

[13] By replacing ι with e_1 in the above formulas one can readily obtain the real counterparts.

curve recursively as a function of the state vector. I am interested in the zero-coupon yield curve, which is made up of zero-coupon bonds with different maturities. These are assets that promise to deliver one sure dollar in τ periods. Substituting the payoff $\varphi = 1$ into (15), the date t price of a zero-coupon bond with τ-maturity is obtained by evaluating the τ-period conditional expectation of S^{nom}

$$p^{(\tau)}(x) = \mathbb{E}\left[S^{nom}_{\tau} \mid X_0 = x\right] = \exp\Big(a(\tau) + b(\tau)\cdot x\Big)$$

with the coefficients $a(\tau) \in \mathbb{R}$ and $b(\tau) \in \mathbb{R}^2$ solving the system of ordinary differential equations

$$\frac{d}{dt}b(t)' = -\iota'\widetilde{\beta}_1 - b(t)'\widetilde{\kappa} \tag{17}$$

$$\frac{d}{dt}a(t) = -\delta - \iota\cdot\widetilde{\beta}_0 + b(t)'\widetilde{\phi} + \frac{1}{2}\left|\alpha'\iota - \sigma' b(t)\right|^2 \tag{18}$$

with $a(0) = 0$ and $b(0) = 0$. It follows that the zero-coupon yields, $y^{(\tau)}(x)$, are affine in x

$$y^{(\tau)}(x) := -\frac{\log p^{(\tau)}}{\tau} = \widetilde{a}(\tau) + \widetilde{b}(\tau)\cdot x \qquad \text{where} \quad \widetilde{a}(\tau) := -\frac{a(\tau)}{\tau} \quad \text{and} \quad \widetilde{b}(\tau) := -\frac{b(\tau)}{\tau}$$

Notice that the parameters of the above system stem from the worst-case model, so they are altered by η_0 and η_1 relative to the baseline values. As a result, one can infer $(\bar{\xi}_0, \bar{\xi}_1, \bar{\xi}_2, \bar{\xi}_3)$ from observable zero-coupon yields, provided that the baseline parameters and the realized path of X_t are 'known'. In section 3.2, I propose a two-step procedure along these lines to estimate the model parameters.

In spirit, this exercise is similar to the method often used in the arbitrage-free affine term structure literature to estimate market prices of risk and the so called risk-neutral probabilities.[14] Nevertheless, there is a key difference: in my case, subjective beliefs are derived endogenously as part of the minimizing player's decision rule, so the theory imposes tight restrictions on how η_0 and η_1 hinge on the agent's environment $(\kappa, \sigma, \beta_0, \alpha)$ and preferences ξ, inducing a reduction of the number of parameters to estimate. In contrast, while the affine term structure literature exploits the cross-equation restriction implied by no-arbitrage by using a system of ODEs for the yield coefficients similar to (17)-(18), it typically leaves the change of measure, that is the form of η_0 and η_1, unrestricted.

3 Estimation

3.1 Baseline model

Because of the multivariate Ornstein-Uhlenbeck structure of the baseline model (2)-(3), the associated discrete time sampling leads to a time invariant linear state space model with coefficient matrices being

[14] For a nice summary about the objectives and main techniques used in the affine term structure literature, see Piazzesi (2010). An accessible textbook treatment can be found in Chapter 14 of Ljungqvist and Sargent (2012).

explicit functions of the continuous time parameters:[15,16]

$$\begin{bmatrix} \Delta c_{t+1} \\ \pi_{t+1} \end{bmatrix} = \beta_0^D + X_t + \alpha^D \varepsilon_{t+1} \tag{19}$$

$$X_{t+1} = \kappa^D X_t + \sigma^D \varepsilon_{t+1} \tag{20}$$

where $\varepsilon_{t+1} \overset{iid}{\sim} \mathcal{N}(0, I_2)$, while Δc and π denote log consumption growth and inflation respectively. From the agent's perspective, X is assumed to be observable, so there is no need for filtering on her part. In contrast, from an outside observer's point of view, the state vector is latent. I estimate the parameters of this system and the path of the latent state vector X with maximum likelihood using data on quarterly consumption growth and inflation. Given the parametrization, the states can be viewed as (centered) conditional expectations

$$X_t = [X_{\Delta c,t}, X_{\pi,t}]' = \mathbb{E}[\Delta \log Y_{t+1} \mid \mathcal{F}_t] - \mathbb{E}[\Delta \log Y] \qquad \text{and} \qquad \beta_0^D = \mathbb{E}[\Delta \log Y]$$

where $\mathcal{F}_t$ is the information set available to the decision maker at date t containing current and past values of consumption growth and inflation.

Data: The sample consists of quarterly observations over the period 1952:Q2 - 2017:Q2. Real consumption growth and inflation rate are constructed from the National Income and Product Accounts (NIPA) chain-type quantity and price indexes for personal consumption expenditures on non-durable goods and services using the official NIPA methodology.[17] To calculate per capita series, I use quarterly data for population reported in Line 40 from NIPA Table 2.1.

κ^D		σ^D		β_0^D	α^D	
0.703	-0.026	0.154	-0.020	0.472	0.417	0.000
(0.084)	(0.034)	(0.026)	(0.025)	-	(0.018)	-
0.137	0.947	-0.001	0.178	0.847	-0.081	0.351
(0.072)	(0.028)	(0.024)	(0.026)	-	(0.022)	(0.015)

Table 1: Maximum Likelihood estimates and asymptotic standard errors (in parentheses) for the baseline model (19)-(20). The likelihood is initialized at the stationary distribution of X. The column for β_0^D shows the sample averages. The matrix α^D is normalized to be lower triangular.

Estimating the baseline linear state space model (19)-(20) with maximum likelihood gives rise to the point estimates and asymptotic standard errors in Table 1. To get these numbers, I follow Piazzesi and Schneider (2007) in estimating β_0^D, the unconditional expectations for the observables, separately with the respective sample averages. Note that the exposure matrix α is assumed to be lower-triangular,

[15] See the Online Appendix for how the vector $(\kappa^D, \sigma^D, \beta_0^D, \alpha^D)$ relates to the continuous time parameters $(\kappa, \sigma, \beta_0, \alpha)$.

[16] Given that the number of shocks equals to the number of observables, this form might look like a (time invariant) *innovations representation* derived from some other state space model built up from 'more primitive' shocks (observed by the agent). Importantly, I assume no such underlying structure.

[17] The used series are from Line $2, 5, 6$ of Table 1.1.3, 1.1.4 and 1.1.5. For real growth rates and the associated price changes NIPA uses the so called Fisher formula. See, for example, BEA (2016).

which is essentially a normalization given that the worst-case model does not depend on the particular decomposition of the covariance matrix $\alpha\alpha'$.[18] Using the Kalman smoother, I derive the estimated state paths associated with the point estimates; these are represented by the thick solid lines of the bottom panel of Figure 2. Properties of the SDF, discussed in section 2.6, suggest that if the model is correctly specified, the zero-coupon yields of the top panel should be well approximated by affine functions of these filtered state variables.

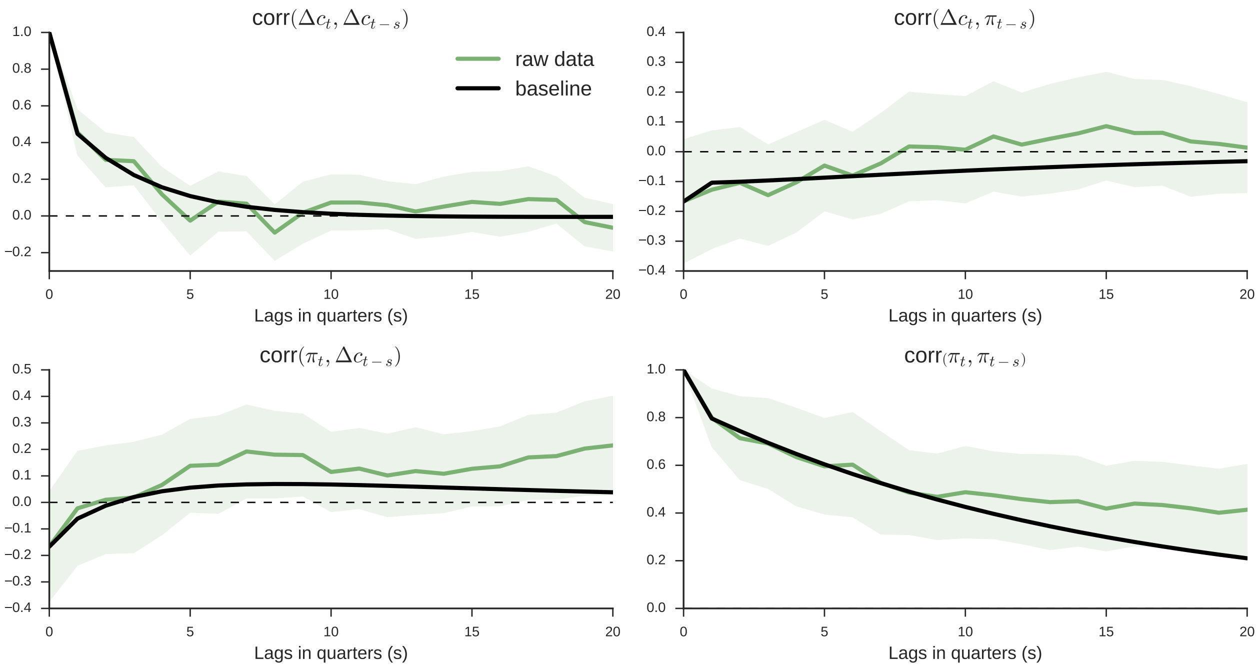

Figure 1: Autocorrelation functions computed from raw data (green solid lines) and from the baseline model (black solid lines). Shaded area represents 2×GMM standard error bounds computed with the Newey-West estimator including 4 quarter lags.

In order to get a better sense of the estimated dynamics, Figure 1 reports autocorrelation functions computed from the data and from the baseline model. The estimated state space model provides a relatively good approximation of the data, which fulfills the main requirement of Assumption 1. In fact, the White information matrix test cannot reject the null that the model is correctly specified. This property of the linear Gaussian model, just as well as the shape of the auto- and cross-correlations between consumption growth and inflation more or less replicate the results of Piazzesi and Schneider (2007). One key difference is that the presence of their main driving force, the role of inflation as bad news about future consumption growth, is not as strong in my sample as in the one that they use. In particular, based on the upper right panel of Figure 1, the relationship between lagged inflation and consumption growth is statistically insignificant, contrary to Piazzesi and Schneider (2007), where a similar figure shows a significant negative contemporaneous and lagged correlation. Note, however, that the standard errors are relatively large in all three cases where inflation is involved, signaling the fact that it is hard to estimate the inflation dynamics and especially the cross-correlations between

[18] One can see this, for example, by combining (12) with the worst-case dynamics (13)-(14).

consumption growth and inflation accurately. This gives ample cause for the decision maker to be concerned about the nature of the relationship between the two variables. Based on this observation, in section 3.5 I offer a reinterpretation of the 'bad news' channel of Piazzesi and Schneider (2007) in terms of model uncertainty.

3.2 Estimation of the worst-case model

An appealing feature of robust control theory is that it lets us deviate from rational expectations, but still preserves a set of powerful cross-equation restrictions on the decision maker's beliefs, like that of (12), provided that the ex post Bayesian interpretation is applicable. Consequently, estimation can proceed essentially as with rational expectations econometrics (Hansen and Sargent, 1980). The main difference is that now restrictions through which we interpret the data emanate from the decision maker's best response to the worst-case model instead of the econometrician's model.

Building on this idea, I proceed to estimate parameters of the twisting function ξ with two-stage maximum likelihood using data on real consumption growth, inflation and interest rates. For simplicity, define the composite parameter vectors $\psi_1 := (\kappa, \sigma, \beta_0, \alpha)$, $\psi_2 := (\bar{\xi}_0, \bar{\xi}_1, \bar{\xi}_2, \bar{\xi}_3)$ and $\psi := (\psi_1, \psi_2)$. In case of the baseline ψ_1, one can also define the discrete time analog, ψ_1^D. Recall that this vector is a one-to-one function of ψ_1, so having a point estimate for $\widehat{\psi}_1^D$ implies an estimate for $\widehat{\psi}_1$.

Two-step estimator for the vector ψ

1. Estimate ψ_1^D using the discrete time version of the baseline model (19)-(20) with *maximum likelihood*. This step produces $\widehat{\psi}_1^D$ and an estimated path for the states $\left\{\widehat{X}_t\right\}_{t=0}^T$ that I derive by running the Kalman smoother.

2. Derive and fix the continuous time baseline parameters $\widehat{\psi}_1$ along with $\left\{\widehat{X}_t\right\}_{t=0}^T$ and construct model implied measurement error ridden zero-coupon yields as functions of ψ_2

$$y^{(\tau)}(\psi_2;\ \widehat{\psi}_1, \widehat{X}_t) = a^{(\tau)}(\psi_2;\ \widehat{\psi}_1) + b^{(\tau)}(\psi_2;\ \widehat{\psi}_1) \cdot \widehat{X}_t + \sigma_m^{(\tau)} \varepsilon_t$$

   where $\varepsilon_{t+1} \overset{iid}{\sim} \mathcal{N}(0, I_2)$, whereas $a^{(\tau)}$ and $b^{(\tau)}$ denote the solution of the system of ODEs parametrized by $(\psi_2;\ \widehat{\psi}_1)$. Estimate ψ_2 along with σ_m by maximizing the conditional likelihood—subject to the constraint that ψ_2 induces a positive semidefinite Ξ matrix defined in (9)—for yields with 1-, 5-, and 15-years of maturity.

A natural question to ask: why two steps? Given the quasi-analytic form of zero-coupon yields, in principle, I could simply estimate the parameters of ξ along with the baseline model in one step by maximum likelihood, thus obtaining a potentially more efficient estimator. I sacrifice efficiency in order to address concerns raised by Hansen (2007) and Chen, Dou, and Kogan (2017) about 'overusing' cross-equation restrictions by including asset price series as a way to improve inference about (demonstrably hard to estimate) models that assign a special role to beliefs.[19] In addition, estimating the

[19] These papers describe examples in which inferences about the behavior of quantities, say consumption, become

vector X without using price data allows me to interpret them as macro risk factors and investigate how these factors accompanied with a SDF can predict prices.

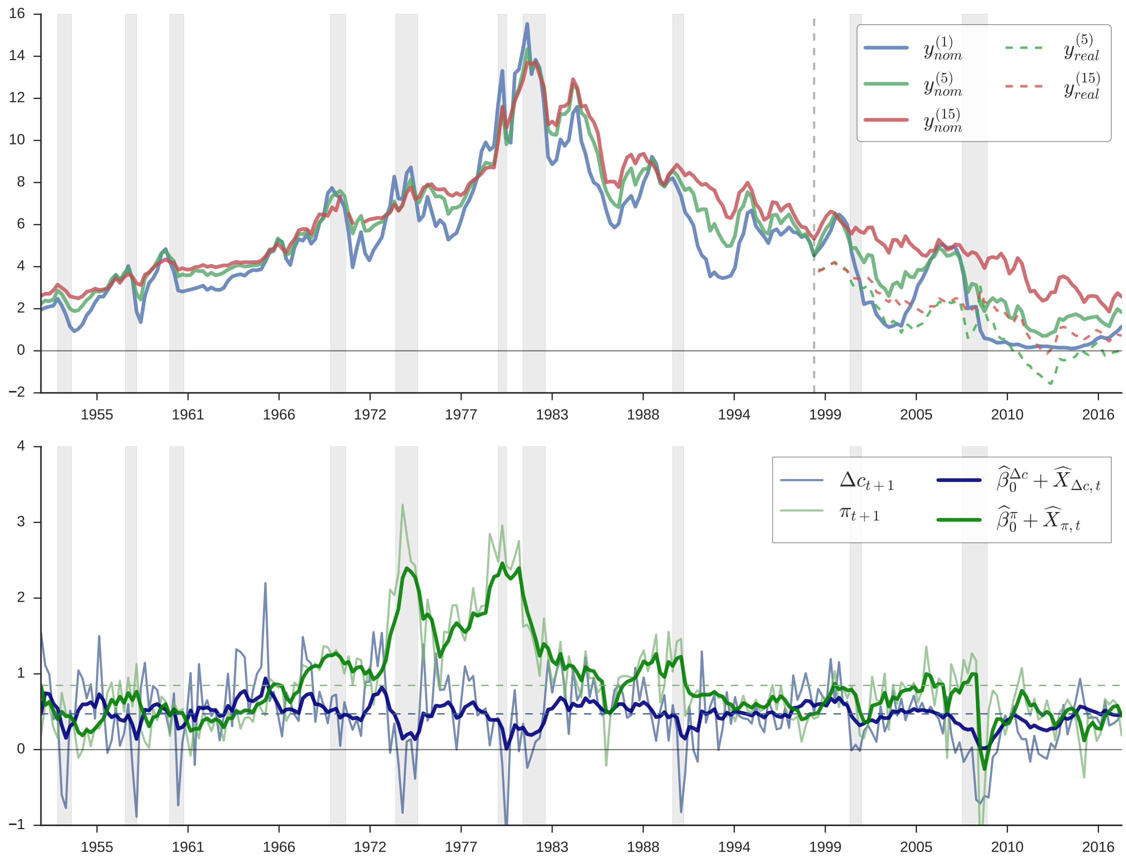

Figure 2: The top panel displays nominal and real zeros-coupon yields with different maturities. Superscripts show maturities in years. The dashed vertical lines identify dates from which additional data are available. The bottom panel shows time-series of real consumption growth and inflation (thin lines) and the corresponding estimated latent factors (thick lines) shifted by the unconditional means. NBER recessions are shaded.

Data: Data on nominal zero-coupon yields with maturities one to 15 years are from combining the zero-coupon estimates of McCulloch and Kwon (1993) covering the period 1952:Q2 - 1971:Q4 with the Gurkaynak, Sack, and Wright (2007) data, which is available since 1971:Q4. I also use *real* zero-coupon yields using the estimated TIPS yield curve in Gurkaynak, Sack, and Wright (2010) starting from 1999.[20]

The top panel of Figure 2 displays a set of nominal and real zero-coupon yields. Over most of the sample period, the nominal yield curve is upward-sloping: yields of bonds with longer maturities

more precise once one includes asset prices and utilizes the associated extensive set of cross-equation restrictions, relative to when inferences are made from the quantity data alone. The question then arises “about how the investors who are supposedly putting those cross-equation restrictions into returns came to know those quantity processes before *they* observed returns.” Hansen and Sargent (2019).

[20]Both Gurkaynak, Sack, and Wright (2007, 2010) data sets are continually updated. The latest versions can be conveniently downloaded through the Quandl Financial and Economic Database using the keys `FED/SVENY` and `FED/TIPSY`. These are daily series that I turn into quarterly data by averaging.

tend to be higher than those of bonds with shorter maturities. Another salient feature of the data is the recurring narrowing (and subsequent widening) of yield spreads, i.e., the difference between long-term and short-term interest rates, prior to U.S. recessions. This is a well-documented fact usually summarized as the slope of the yield curve being a reliable predictor of future economic activity (Hamilton and Kim, 2002). These are features of the data that the model is expected to replicate.

3.3 Twisting function

Table 2 reports the point estimates for the parameters of ξ resulting from the two-stage MLE described above. Interestingly, the constant term ξ_0 turns out to be practically zero and the constraint $\xi_0 = 0$ does not change the estimates significantly. This indicates a critical role for the twisting function's state dependence. As we have seen in section 2, it is exactly this state dependence, emerging from $\xi_2 \neq \mathbf{0}$, in which the twisted set $\mathcal{Z}(\xi)$ deviates from the standard framework of Hansen and Sargent (2001).

	$\bar{\xi}_0$	$\bar{\xi}_1$	$\bar{\xi}_2$	$\bar{\xi}_3$
Unrestricted case	≈ 0.0	0.120	-0.141	0.086
	(0.02)	(0.62)	(0.07)	(0.01)
Only state dependence	0.0	0.170	-0.123	0.088
	-	(0.42)	(0.07)	(0.01)

Table 2: Second stage QMLE estimates and asymptotic standard errors (in the parentheses) for the parameters of ξ. The standard errors take into account the two-step nature of the estimator. The rows labeled as "only state dependence" report estimates for the case when $\bar{\xi}_0 = 0$ is imposed. One should take the standard error of the intercept term $\bar{\xi}_0$ cautiously given that the estimate is very close to the boundary of the constraint.

If I impose no state dependence with the constraint $\xi_2 = \mathbf{0}$, the estimated ξ_0 becomes exactly zero (at the boundary of the constraint). This suggests that logarithmic utility with rational expectations, $\xi = 0$, does a better job of explaining zero-coupon yields than the Hansen and Sargent (2001) model, at least when one uses the proposed two-step procedure. This finding is remarkable, especially in lights of the Hansen and Sargent (2001) model's close connections with recursive utility (Skiadas, 2003).

3.3.1 Recursive utility with unit elasticity of intertemporal substitution

For the sake of comparison, consider an investor with preferences represented by a utility function of the type proposed by Kreps and Porteus (1978), Epstein and Zin (1989) and Duffie and Epstein (1992) with unitary elasticity of intertemporal substitution. Suppose that the investor's belief is given by the baseline model of section 3.1 evaluated at the point estimates. Borovička et al. (2011) show that the associated (nominal) stochastic discount factor S_t^{RU} can be written as

$$S_t^{RU} = \exp(-\delta t)\frac{(PC)_0}{(PC)_t}\widetilde{M}_t = M_t^{nom}\widetilde{M}_t$$

where $\widetilde{M}_t$ is a multiplicative martingale following

$$d \log \widetilde{M}_t = -\frac{|\widetilde{\alpha}|^2}{2} dt + \widetilde{\alpha} \cdot dW_t \qquad \text{with} \quad \widetilde{\alpha} := (1-\gamma) \left[\frac{1}{\delta} \alpha' e_1 + \sigma' \nabla V_x^{RU} \right]$$

In this formulation, γ represents the risk aversion coefficient and ∇V_x^{RU} is the shadow value associated with the value function, which turns out to be linear given the lognormal (3)-(2). The main difference between S^{RU} and the robust SDF is the multiplicative martingale terms, $\widetilde{M}_t$ and $Z_t^{H^*}$, respectively. While the exposure of $\widetilde{M}$ is a constant $\widetilde{\alpha}$, the likelihood ratio Z^{H^*} has a possibly state dependent exposure H^*. It turns out, however, that the elimination of the state dependence renders the robust value function linear and the two preferences observationally equivalent. In particular, one can establish the following relationship between the risk aversion parameter γ and the constant term of ξ:

$$(\gamma - 1) = \frac{1}{\ell^*(\xi_0)} = \sqrt{\frac{\xi_0 \delta^2}{\left| \alpha' e_1 + \sigma' \left(\delta I_2 + \kappa' \right)^{-1} e_1 \right|^2}}$$

under which the recursive utility investor with unitary elasticity of substitution becomes observationally equivalent with the robust decision maker provided that $\xi(x) = \xi_0 \geqslant 0$.

Using the relationship between $\widetilde{\alpha}$ and H^*, one can reinterpret the effect of recursive utility as a pessimistic adjustment to the means of $\log Y$ and X through ϕ and β_0. On top of this, by activating state dependence in ξ, the structured uncertainty model might induce further adjustments in the persistence of the baseline dynamics.[21]. Indeed, the finding that the estimated ξ_0 is zero (or equivalently, $\gamma = 1$) suggests that it is the persistence of the baseline that needs to be altered in order to explain the observed zero-coupon yields.

This result might be surprising in light of Piazzesi and Schneider (2007), who employ a stochastic discount factor identical to S^{RU}. Using similar consumption and inflation data as I do, with a sample ending in 2005, they choose γ (and δ) so that the model implied average yield curve matches both the 1-year and the 5-year average zero-coupon yields. In fact, if I restrict my sample to their shorter period, the two-step procedure gives $\gamma = 32$ (or $\xi_0 = 0.07$), which is close to the value that they find. Extending the sample with the last ten years, however, changes the estimated dynamics of the baseline model: most importantly, inflation becomes much less persistent. As a result, there is no $\gamma > 1$ or $\xi_0 > 0$, which would make the average nominal yield curve upward sloping and the best fit is reached by logarithmic expected utility. In other words, recursive utility and the argument of Piazzesi and Schneider (2007) alone are insufficient for an upward sloping nominal yield curve.

3.4 Nominal yield curve

Figure 3 compares the observed yield curve dynamics with those implied by the model with quadratic ξ. Although the resulting fit is far from perfect, the robust control model with the state dependent twisting function can capture important features of the data. In particular, both high short rates and

[21]Importantly, ξ_2 affects every term in the value function (see Appendix A), so both the level and the persistence parameters change. Indeed, even if $\xi_0 = 0$, ξ_2 alone can influence all parameters ϕ, κ, β_0 and β_1 (see Table 3)

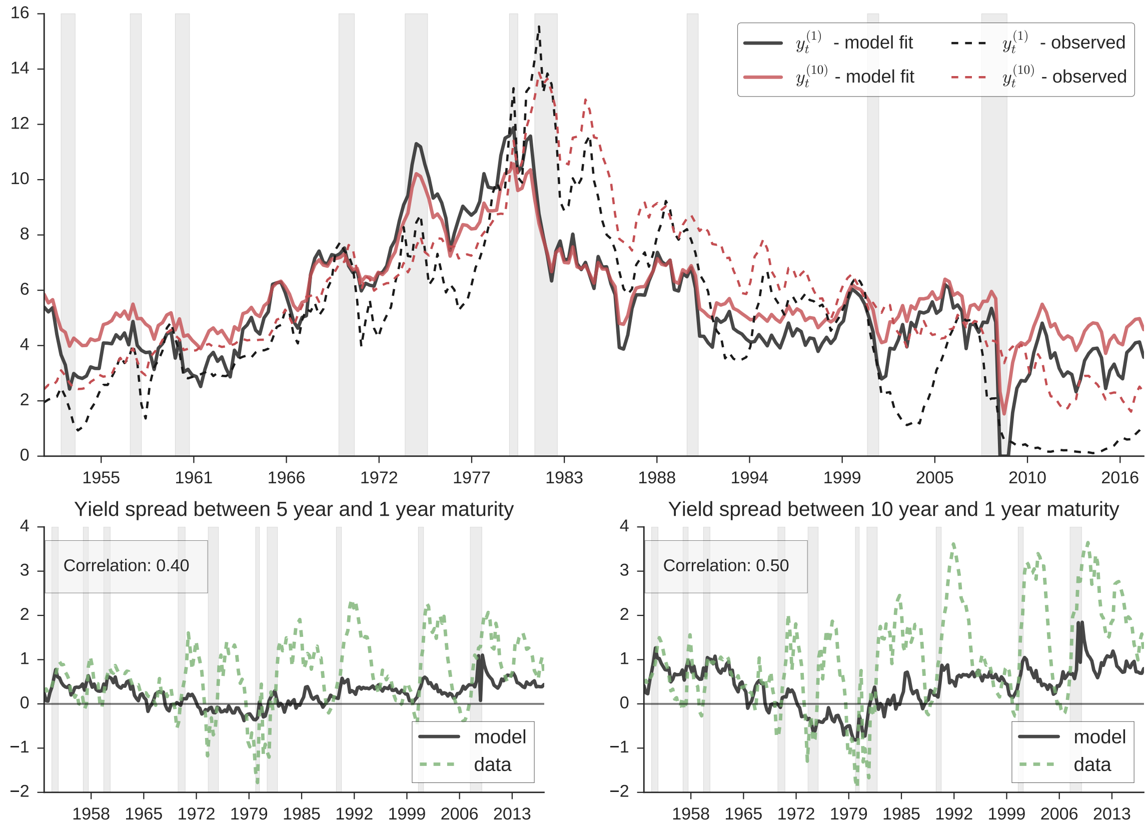

Figure 3: Comparison of the estimated yields and yield spreads with their observed counterparts. Top panel shows zeros-coupon yields with different maturities (superscripts denote maturities in years). The bottom panels display a medium- (left) and a long-term yield spread (right) and the correlation coefficients between the model implied and realized series. NBER recessions are shaded.

low yield spreads seem to predict low consumption growth, and they move in opposite directions during recessions – short rates fall while yield spreads run up (see Figure 2). The model can also replicate the changing sign of the slope of the yield curve in the late 1970s and early 1980s. On the other hand, the average level of yields in the periods of the 1980s and post-2007 look puzzling from the model's point of view. Also, the sensitivity of the yield spreads to the states, determined by the endogenous vector $b(\tau)$, seems to be somewhat low relative to the data.

As for unconditional moments, Figure 4 shows the stationary distribution of the model implied yield curve along with sample analogues of the nominal yield curve's mean and standard deviation. The left panel represents the distribution that arises from the estimated quadratic ξ, while, as a comparison, the right panel shows what recursive utility with unitary elasticity of substitution would result in. Interestingly, contrary to the rational expectations scenario, the fitted robust control model induces an upward sloping mean nominal yield curve following closely the sample averages, and it also appears to generate a substantial amount of volatility for both short and long maturities. Although the model implied volatilities still fall short of the standard deviations computed from the sample, this is clearly an improvement relative to the rational expectations model that heavily suffers from the so called

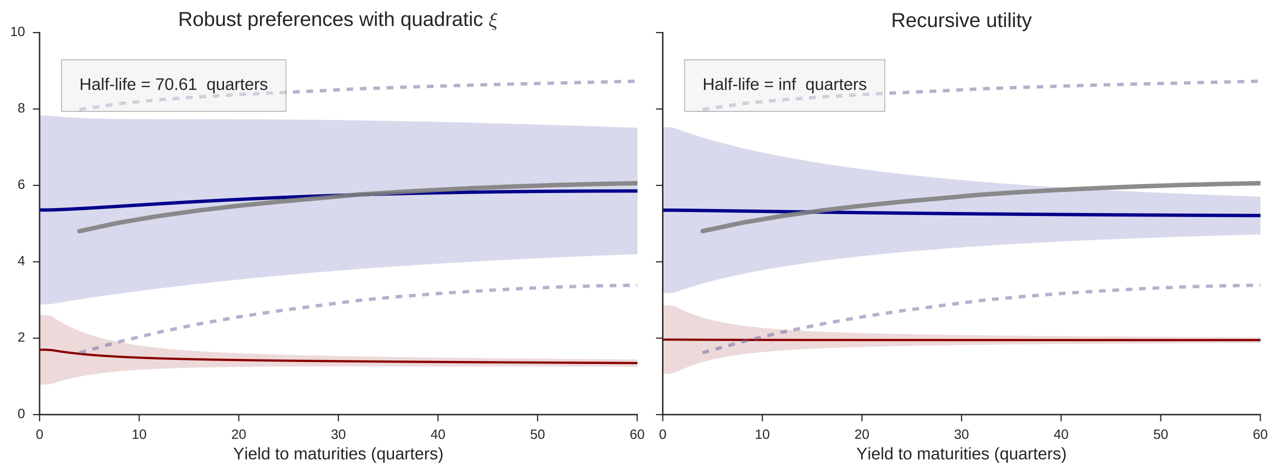

Figure 4: Model implied stationary distributions of the nominal (blue) and real (red) yield curves and the corresponding sample moments. Shaded areas represent one standard deviation bands around the means (solid lines). Grey solid lines show the sample average, the dashed lines are one standard deviation bands using the sample standard deviations. The left panel is for the benchmark model with quadratic ξ. The right panel shows the case with $\xi_2 = \mathbf{0}$. The boxes contain associated half-lives.

excess volatility puzzle (Shiller, 1979).[22]

The reason for the relatively large volatility of long yields is the presence of state dependent *uncertainty prices*. Because a risk premium is the product of $\rho(x)$ and the asset's shock exposure, this property effectively breaks the expectation hypothesis that requires constant risk premium (Cochrane, 2005). It is the extra fluctuation in risk premia that helps to reconcile the relative volatility of long and short yields. Notice that this fluctuation arises from the agent's misspecification concerns as opposed to other forces commonly employed in the asset pricing literature to obtain variable (and counter-cyclical) risk premia.[23]

Regarding *real* interest rates, the model produces a downward sloping average yield curve. In fact, real yields with long maturities are predicted to be lower than in the case without robustness.[24] The unconditional variance of yields appears to decay sharply with maturities, suggesting a relatively modest state dependence in the uncertainty price of the shocks to consumption growth. In other words, misspecification concerns regarding the persistence of consumption growth process alone do not seem to be significant (see also Figure 5).

Notwithstanding its shortcomings, the model does a decent job in explaining the unconditional

[22]The puzzle can be summarized as follows: long-term interest rates derived from the expectations hypothesis are not volatile enough to be aligned with the data. Models that exhibit constant risk premium satisfy the expectations hypothesis, so they are exposed to this type of puzzle. The recursive utility model with homoskedastic consumption is one example. Cochrane (2005) argues compellingly that excess volatility is exactly the same phenomenon as return predictability.

[23]Notable among these are the long-run risk model with stochastic volatility by Bansal and Yaron (2004), which assumes recursive utility and introduces a new shock, and the external habit model of Campbell and Cochrane (1999), which hinges on a particular history-dependent utility function that implies time varying local risk aversion.

[24]This follows from the downward (pessimistic) adjustment in the expected consumption growth rate, $\beta_0^{\Delta c}$, caused by the worst-case model (see Table 3). This suggests that as long as the baseline dynamics – in this case log utility with rational expectations – does not induce an upward sloping real yield curve, robust preferences cannot 'fix' this. As a similar adjustment appears in models with recursive utility, the same implication applies as the findings of Piazzesi and Schneider (2007) nicely illustrate.

moments of the nominal yield curve as well as the dynamics of its slope. This is remarkable given that it can be regarded as a parsimonious two factor model of the yield curve with the factors being estimated from macro aggregates without any reference to prices.

3.5 Subjective beliefs

Table 2 shows that the key role of ξ in shaping the household's stochastic discount factor is to make the size of $\mathcal{Z}$ sensitive to X, the deviation of expected inflation and consumption growth from their long run means. Recall that the larger is the value of $\xi(X_t)$, the more relaxed the minimizing agent's constraint, implying a more concerned decision maker. For example, the household appears to be more worried in times when the baseline inflation expectation deviates more from its long term mean in either direction, that is, when the green thick line of the bottom panel of Figure 2 lies far above (in the 1970s) or far below (early 1960s and post 2007) the dashed horizontal line.

Baseline parameters

ϕ^D	κ^D		β_0^D	β_1^D	
0.00	0.70	-0.03	0.47	1.00	0.00
0.00	0.14	0.95	0.85	0.00	1.00

Worst-case parameters

$\widetilde{\phi}^D$	$\widetilde{\kappa}^D$		$\widetilde{\beta}_0^D$	$\widetilde{\beta}_1^D$	
-0.03	0.71	-0.02	0.37	1.03	0.03
0.03	0.16	1.00	0.94	0.05	1.13

Table 3: Parameter values of the state-space model under the Baseline and the worst-case distribution

The household copes with model uncertainty by choosing a worst-case model that is interpretable as her subjective belief. Because of the chosen functional forms, this belief can be represented with a model from the same parametric family as the baseline. The resulting worst-case parameters are reported in Table 3. Clearly, the constant terms $\widetilde{\phi}$ and $\widetilde{\beta}_0$ change pessimistically relative to the baseline. This kind of pessimistic adjustment of unconditional expectations is typical in standard robust control models with unstructured uncertainty or in recursive utility models with unitary elasticity of substitution (see section 3.3.1) and it remains present with the state dependent ξ as well. On the other hand, a novel feature is that the worst-case model also alters the correlation structure of the baseline. In particular, the state variables, especially the conditional inflation expectation, become more persistent and more strongly correlated with each other.

Displaying autocorrelation functions that emerge from the estimated worst-case model (red lines), Figure 5 illustrates how the robust household perceives the dynamics of fundamentals, i.e., the belief that supports the observed yield curve dynamics. Evidently, inflation is much more persistent and the negative correlation between inflation and consumption growth is more pronounced than what the baseline model suggests. In other words, the household's subjective belief overweights the long-run risk component of inflation and, indirectly, of consumption. Similar conclusion can be drawn from Figure 6, plotting impulse responses, i.e., changes in conditional forecasts (entries of X) following a one percent shock ε_{t+1}. Given the lower-triangular nature of α, the two entries of ε_{t+1} can be viewed as unanticipated changes in consumption growth and inflation, respectively. As before, the black lines represent the 'objective' forecasts (baseline model), while the red lines show forecasts under the worst-case model. Evidently, worst-case beliefs predict a much more pronounced and more persistent increase in inflation after a consumption growth surprise (bottom left panel). Similarly, while a one-percent inflation surprise lowers growth forecasts under both models, the effect is rather short-lived under the baseline model. In contrast, under the worst-case beliefs, this effect is permanently negative suggesting that inflation is a much worse news under the worst-case model than under the baseline.

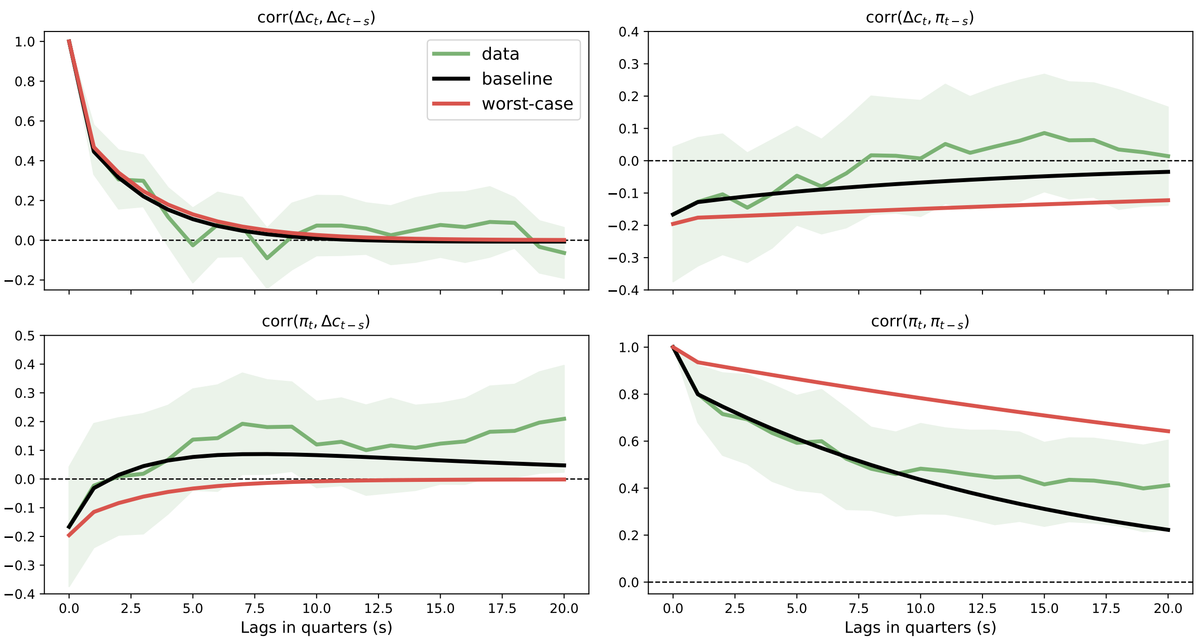

Figure 5: Autocorrelation functions computed from the estimated baseline (black lines) and worst-case model (red lines).

These findings offer a reinterpretation of the key driving force in Piazzesi and Schneider (2007). They use recursive utility to make the representative agent averse to persistence and endow her with (rational) expectations derived from a state space model similar to that of Table 1, except that they estimate significant negative comovement between lagged inflation and consumption growth. Recursive utility along with the investor's beliefs that inflation carries bad news about future consumption growth (long-run risk) induces a premium for assets like nominal bonds that pay off little when inflation is high. Because inflation is persistent, it affects long bonds worse than short bonds, giving rise to a term premium. One weakness of this argument, however, is its dependence on a particular correlation that is a relatively fragile feature of the data. The robust control model with structured uncertainty offers a possible remedy: while it is hard to measure the forecasting ability of inflation accurately, the implied welfare consequences are significant, so this is exactly the kind of misspecification that a robust agent would be worried about. Indeed, as we have seen above, the estimated subjective belief credits inflation with strong forecasting ability about future consumption growth and very high persistence, indicating that the market perceives the economy to be more prone to long-run risk than what the observed data as captured by the baseline model would imply.

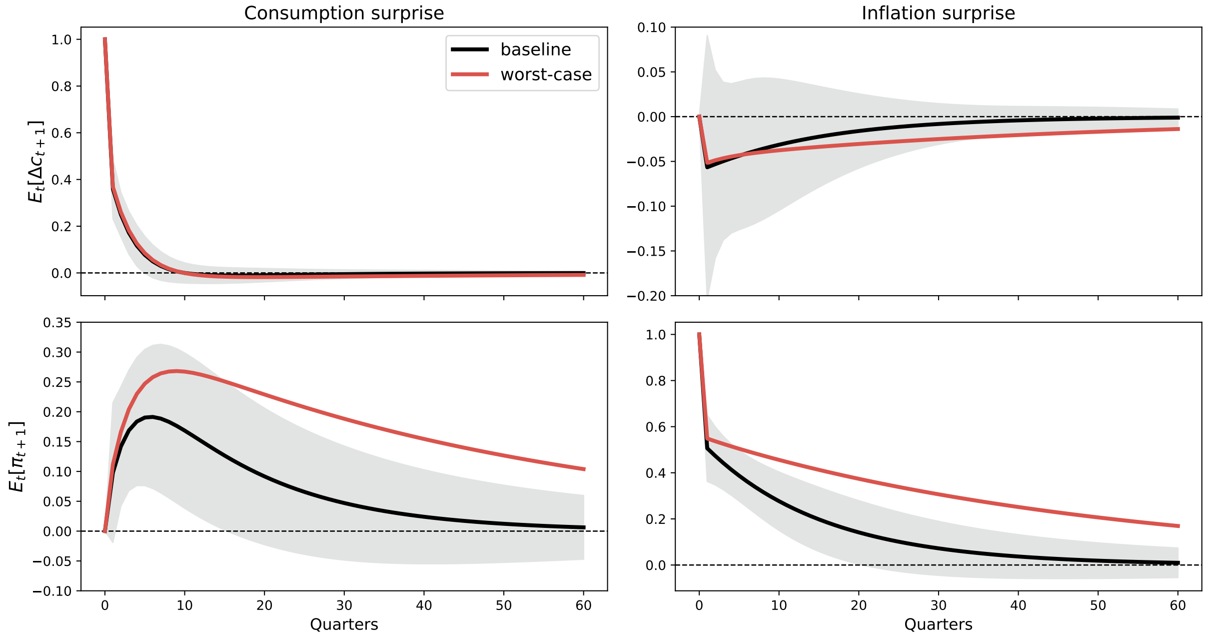

Figure 6: Impulse responses to one-percentage point surprises in consumption growth (left column) and inflation (right column) under the baseline model (black lines) and under the estimated worst-case model (red lines). The first row show the effect on conditional consumption growth forecast, $X_{\Delta c,t}$, the second row is for conditional inflation forecasts, $X_{\pi,t}$. The responses are measured in percent. Shaded areas are $2\times$ standard error bounds based on the maximum likelihood estimates.

With the subjective belief at hand it is possible to study whether the robust control model can replicate the empirical findings of Piazzesi, Salomao, and Schneider (2015) about the expectations of professional forecasters. As mentioned before, contrasting survey based yield forecasts with predictions of a simple statistical model, they find that survey expectations are formed as if both the level and slope of the yield curve were more persistent than they appear in hindsight. Figure 7 displays the autocorrelations of $y_t^{(1)}$ (level) and $y_t^{(5)}-y_t^{(1)}$ (slope) computed from the baseline that approximates the physical measure and the model implied subjective belief, as well as the corresponding sample statistics. Apparently, the agent in the model perceives the yield curve dynamics overly persistent, which is qualitatively consistent with the pattern found by Piazzesi, Salomao, and Schneider (2015). Since zero-coupon yields are affine functions of the states, the discrepancies between the model implied objective and subjective autocorrelations are implications of the different state dynamics depicted by the two measures. It is interesting, however, that the subjective autocorrelation functions exceed even the sample analogues. Given that I do not use any price information in the estimation of the baseline model, the closeness of the black lines to the data are quite remarkable.

3.6 Size of the ball

These pessimistic beliefs are inferred from bond prices seen through the lens of the robust control model of Hansen et al. (forthcoming). It is not clear, however, how 'reasonable' these beliefs are. In section 2, I called those models reasonable that the decision maker cannot differentiate easily from the

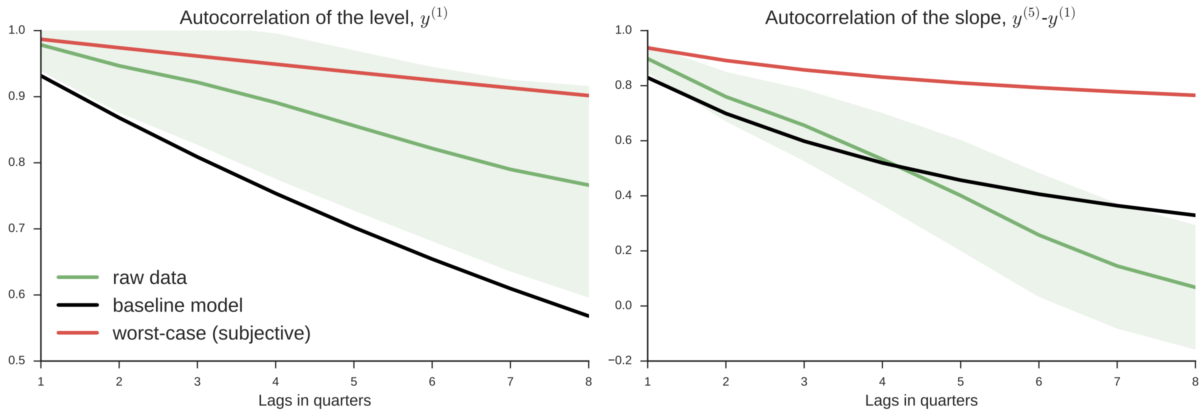

Figure 7: Autocorrelation functions of zero-coupon yield with 1 year maturity (left) and the yield spread between the 5- and 1-year maturity zero coupon bonds (right) computed from raw data (green solid lines), from the estimated baseline model (black solid lines) and from the worst-case model (red solid lines). Shaded area represents 2×GMM standard error bounds computed with the Newey-West estimator including 4 quarter lags.

baseline when she can only use a finite sample. How easy is it to distinguish the estimated worst-case model from the baseline? To address this question, I use Chernoff entropy to measure the statistical distance between two models. Chernoff entropy is proposed by Anderson, Hansen, and Sargent (2003) and Hansen and Sargent (2019) as a tool to calibrate the plausible amount of robustness. Exploiting Chernoff entropy's subtle connection to statistical decision theory, they calibrate the worst-case model, so that it gives a prechosen level of a *detection error probability*. The detection error probability for a given sample size T is defined as the average[25] probability of choosing the 'wrong' model associated with a likelihood ratio test between the worst-case and the baseline. Chernoff entropy, χ, is then defined as the upper bound for the asymptotic decay rate of this mistake probability as T goes to infinity. High values of χ suggest a fast decay rate indicating that the two models in question are relatively easy to differentiate from each other.

To facilitate the interpretation of χ, Hansen and Sargent (2019) introduce the notion of half-life. This is based on the *hypothetical* scenario in which the decay rate of mistake probability is constant $\bar{\chi}$ over time. If this were the case, the half-life, i.e., the increase in sample size $T_2 - T_1 > 0$ necessary to halve the mistake probability, would be

$$\frac{\text{mistake probability after } T_2}{\text{mistake probability after } T_1} = \frac{\exp(-T_2\bar{\chi})}{\exp(-T_1\bar{\chi})} = \frac{1}{2} \qquad \Rightarrow \qquad T_2 - T_1 = \frac{\log 2}{\bar{\chi}}$$

Estimating both models from the data, I can invert the calibration exercise of Hansen and Sargent (2019) and use the implied half-life as a sanity check for the plausibility of the estimated belief. In my case, the estimated ξ gives rise to a half-life of 70 quarters, which means that the worst-case model is so close (statistically) to the baseline that if the decision maker kept running likelihood ratio tests each period between the two models, it would take more than 15 years for her to halve the mistake

[25]By average I mean that the conditional error probabilities, say probability of mistakenly choosing model B while model A generates the data, are weighted by prior probabilities associated with the two models. Anderson, Hansen, and Sargent (2003) assign .5 to each model, whereas Hansen and Sargent (2019) put probability one on the baseline.

probability. This seems reasonable, especially given that the constant $\bar{\chi}$ is more of an upper bound than an accurate measure of the (finite sample) decay rate. Furthermore, discriminating among a set of different models, the actual task faced by the robust decision maker, is significantly more complicated than the pair-wise comparison suggested by the above thought experiment.

4 Survey expectations

As we saw in section 2.4, Assumption 2 suggests a reinterpretation of the worst-case model as the robust decision maker's subjective belief. Interestingly, this feature of robust control theory can be utilized to test a particular prediction of the model with an independent data source. Notice that if the equilibrium drift-distortion H^* happens to be nonzero, our theory predicts that the decision maker's subjective belief is 'wrong', in the sense that it systematically differs from the econometrician's model. As this discrepancy follows a specific pattern, this prediction is testable provided that one has a good proxy for beliefs.

A possible approach is to study surveys of professional forecasters about their expectations of future aggregate variables. In the analysis that follows, I use the *Blue Chip Financial Forecasts* survey, which I take to represent 'prevailing opinion' about the future level of U.S. interest rates. In each month, this survey asks approximately 40-50 professional forecasters about their h-quarter-ahead forecasts of variable R, so one can consider their median answer as a rough approximation to the market's period-t subjective expectation about R_{t+h}. In order to be able to use such survey forecasts to test the predictions of robust control theory, I need the following assumption:

Assumption 3. *Professional forecasters understand the survey question as an inquiry about the expectation that justifies their behavior, asking for forecasts that they would act upon.*

Assumption 3 ensures that the answers of professional forecasters represent the same object that the model of this paper is about: a robust decision maker's worst-case belief.[26]

To assess the accuracy of survey forecasts, a popular approach in the literature is to contrast time series of survey expectations with (objective) conditional expectations coming from statistical models that one fits to the realized R series. As a prominent example, Piazzesi, Salomao, and Schneider (2015) utilize survey data of interest rate expectations and study their joint dynamics with realized interest rates. Following the tradition of the affine term structure literature, they impose little structure on how two probability measures differ other than forcing the stochastic discount factor to belong to the same affine parametric family under both measures. Thanks to the particular functional form assumptions in section 2, this paper presents a very similar setting, where the models of interest exhibit an affine

[26] It might be tempting to extend this logic to higher moments as well and use the cross-sectional distribution of survey answers as a proxy for the market's view about the conditional distribution of future R. Cross-sectional dispersion of survey answers, for example, tends to be substantial, which one might interpret as the 'degree of confidence' (Ilut and Schneider, 2014) or the 'size of the worst-case drift distortion' (Ulrich, 2013). Nevertheless, the robust control model of this paper offers no direct link to cross-sectional dispersion in survey forecasts. In general, the variance of the conditional expectation of R does not equal to the conditional variance of R. Instead, I consider a representative agent economy that does not feature heterogeneity in baseline models or attitudes toward robustness. The representative agent serves as a stand-in for the market and it allows me to focus the analysis on the effect of pessimistic belief distortions, which is separate from the notion of belief heterogeneity.

structure. Nevertheless, in contrast to the affine term structure literature, the change of measure H^* is itself an equilibrium object, restricted by cross-equation relationships among the model parameters. As we saw in section 2.5, the central feature of these restrictions is that the worst-case measure overweights adverse states where the degree of adversity is determined by the derivative of the value function.

Like Piazzesi, Salomao, and Schneider (2015), I use survey expectations of interest rates, which have the advantage that prices of traded assets already contain valuable information about H^* through the market's stochastic discount factor. The central idea of my testing strategy is that I can make inferences about the same object, the equilibrium drift distortion H^*, from two independent sources. If the model is correct, the two sources should lead to similar outcomes, a testable prediction.

4.1 Interest rate expectations

The task of forecasting subsumes a 'new role' of subjective belief besides determining the coefficients $a(\tau)$ and $b(\tau)$ for the conditional expectation of the stochastic discount factor: it encodes the agent's view about future values of the *state vector*. If $Z^{H^*} \neq 1$, these views are systematically, pessimistically biased relative to the baseline state dynamics. Being affine functions of the state vector, model implied yield forecasts are propelled by these altered state dynamics, so by utilizing surveys, I can effectively test whether the subjective belief elicited from prices is consistent with the state dynamics encoded in survey expectations.

To investigate the validity of this statement, I compare model predictions, evaluated at the estimated $\widehat{\psi}$, about subjective belief with the survey data along some informative conditional and unconditional moments. The idea is that if the model is correctly specified, the twisting function should give rise to a worst-case that can simultaneously explain the realized yields and the observed biases of survey expectations. By comparing moments, I hope to identify dimensions along which the model performs well and ones along which it fails. That said, let $E_t^\star y_h^{(\tau)}$ denote the *median* of survey yield forecasts in period t, where h identifies the forecast horizon (in quarters) and τ stands for the maturity of the bond (in years). Similarly, let $E_t^\star sp_h^m$ and $E_t^\star sp_h^l$ denote the h-quarter ahead median survey forecasts of medium- (5 year minus 1 year maturity) and long-term (10 year minus 1 year maturity) yield spreads, respectively. Consider the following variables[27]

(a) conditional level: h-period ahead expectation of zero-coupon yield with maturity τ:

$$\widetilde{y}_h^{(\tau)}(X_t) := \widetilde{\mathbb{E}}\left[y^{(\tau)}(X_{t+h})|X_t\right] = a^{(\tau)}(\psi) + b^{(\tau)}(\psi) \cdot \widetilde{\mathbb{E}}\left[X_{t+h}|X_t\right]$$

This is the direct model counterpart of survey expectations $E_t^\star y_h^{(\tau)}$.

(b) conditional yield change: expected h-period change of yield with maturity τ:[28]

$$\Delta\widetilde{y}_h^{(\tau)}(X_t) := \widetilde{\mathbb{E}}\left[y^{(\tau)}(X_{t+h+1}) - y^{(\tau)}(X_{t+1})|X_t\right] = b^{(\tau)}(\psi) \cdot \widetilde{\mathbb{E}}\left[X_{t+h+1} - X_{t+1}|X_t\right]$$

[27] By conditional I mean conditioning on the agent's information set, which is X_t in the model.

[28] I shift the starting date one period ahead in order to get a variable that hinges only on the survey data. The aim is to eliminate the risk of comparing the model with a variable that was in part used in the estimation step. Replacing $E_t^\star y_1^{(\tau)}$ with $y_t^{(\tau)}$ does not affect any of my results significantly.

The sample counterpart is constructed by $\Delta E_t^\star y_h^{(\tau)} := E_t^\star y_{h+1}^{(\tau)} - E_t^\star y_1^{(\tau)}$.

(c) conditional change in spread: expected h-period change in yield spreads

$$\Delta\widetilde{sp}_h^i(X_t) := \widetilde{\mathbb{E}}\left[sp^i(X_{t+h+1}) - sp^i(X_{t+1})|X_t\right] = \Delta^i b(\psi)\cdot\widetilde{\mathbb{E}}\left[X_{t+h+1} - X_{t+1}|X_t\right] \qquad i = m, l$$

The sample counterpart is constructed by $\Delta E_t^\star sp_h^i := E_t^\star sp_{h+1}^i - E_t^\star sp_1^i$ for $i = m, l$.

(d) subjective forecast error and bias: difference between the conditional expectation under the subjective measure and the actual realization

$$\begin{aligned}\mathbf{error}_{t,h}^{(\tau)} &= \widetilde{\mathbb{E}}\left[y^{(\tau)}(X_{t+h})|X_t\right] - y^{(\tau)}(X_{t+h}) = \\ &= \underbrace{\widetilde{\mathbb{E}}\left[y^{(\tau)}(X_{t+h})|X_t\right] - \mathbb{E}\left[y^{(\tau)}(X_{t+h})|X_t\right]}_{=\mathbf{bias}_{t,h}^{(\tau)}} + \underbrace{\mathbb{E}\left[y^{(\tau)}(X_{t+h})|X_t\right] - y^{(\tau)}(X_{t+h})}_{=\mathbf{objective\ forecast\ error}_{t,h}^{(\tau)}} = \\ &= b^{(\tau)}(\psi)\cdot\Big(\widetilde{\mathbb{E}}\left[X_{t+h}|X_t\right] - \mathbb{E}\left[X_{t+h}|X_t\right] + \mathbb{E}\left[X_{t+h}|X_t\right] - X_{t+h}\Big)\end{aligned}$$

The natural sample analogue is $E_t^\star y_h^{(\tau)} - y_{t+h}^{(\tau)}$. This is a crude measure of the experts' forecast accuracy that is expected to be zero on average if the used forecasting model is 'correct'. However, a more informative measure would compare the subjective expectations with statistical forecasts. To that end, I decompose subjective forecast error into a *bias* and an *objective forecast error* terms, the former being defined as the difference between the subjective and objective conditional expectations of a particular zero-coupon yield.[29] In the model, the bias can be readily computed by using yield forecasts from the worst-case and baseline models. In order to get a sense of the size of bias in the sample, I construct a measure for 'objective' conditional forecast, $\widehat{E}_t y_h^{(\tau)}$, based on a simple unrestricted VAR meant to approximate the interest rate dynamics well.[30]

Within the model, using the stationary distribution of the states *under the baseline*, one can compute the unconditional moments of these variables, which are directly comparable with the sample averages and standard deviations of the data analogues. In case of the subjective forecast error this implies

$$\mathbb{E}\left[\mathbf{error}_{t,h}^{(\tau)}\right] = \mathbb{E}\left[\mathbf{bias}_{t,h}^{(\tau)}\right] + \mathbb{E}\left[\mathbf{objective\ forecast\ error}_{t,h}^{(\tau)}\right] = \mathbb{E}\left[\mathbf{bias}_{t,h}^{(\tau)}\right]$$

that is, according to the model, the unconditional mean of the subjective forecast error is equal to the unconditional mean bias. Of course, if we used the 'true' data generating process to compute objective forecasts $\widehat{E}_t y_h^{(\tau)}$, a similar law of iterated expectations type argument would hold in the sample as well. Using a VAR approximation, however, I am obviously less ambitious and will report sample averages for the objective forecast error.

[29] Piazzesi, Salomao, and Schneider (2015) use the bias term to construct a measure of *subjective risk premia* from common *statistical* measures of risk premia.

[30] For macro aggregates, Bhandari, Borovička, and Ho (2019) use the difference in survey expectations between the Michigan Survey of household expectations and the Survey of Professional Forecasters as a proxy for $\mathbf{bias}_{t,h}^{(\tau)}$, the idea being that the professionals use the 'true' model. As for interest rates, due to lack of data, a similar proxy is infeasible.

4.2 Survey data

As an observable proxy for the market's subjective expectations of future yields, I use the Blue Chip Financial Forecasts (BCFF) survey over the period 1982:Q4 through 2017:Q2. This is a monthly survey, where the information is typically collected over a two-day period at the end of each month. Forecasts are averages over calendar quarters and cover horizons up to five quarters ahead. To obtain quarterly series, I use answers gathered at the end of each quarter.[31] The survey forecasts are for U.S. Treasury *par yields* on coupon bonds with maturities of 1, 2, 5, 10, and 30 years, hence they are not directly comparable to model implied zero coupon yields. The main difficulty is that, unlike zero-coupon yields, par yields are non-linear functions of the states. In order to avoid this complication, I use the bootstrap technique proposed by Fama and Bliss (1987) to interpolate the forecasts of par yields and obtain approximate forecasts of zero-coupon yields.

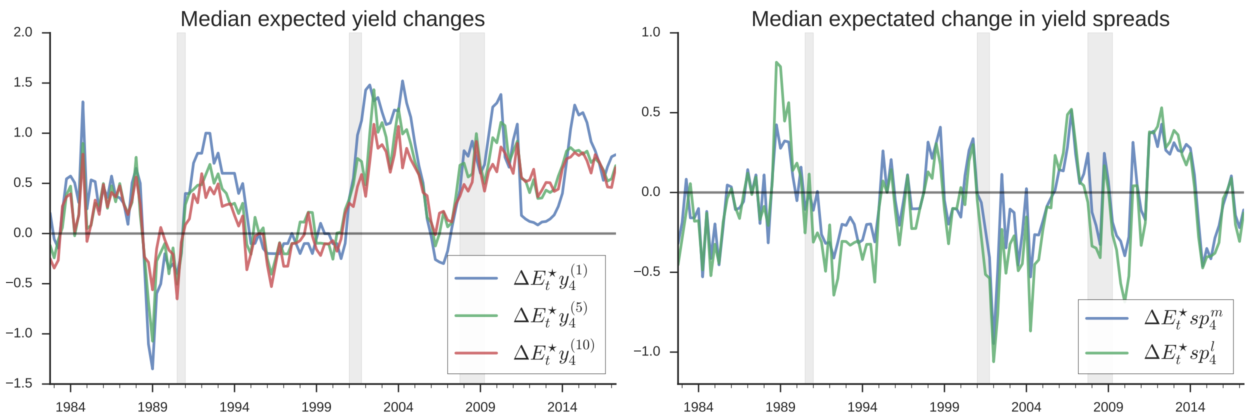

Figure 8: Summary of subjective one-year-ahead zero-coupon yield expectations constructed from the BCFF survey. The left panel shows survey analogues of the *conditional yield change* variable, that is the expected one year percentage-point change (between $t+5$ and $t+1$) in yields of different maturities. The right panel displays the expected change in the medium-term and long-term spreads.

Figure 8 reports sample analogues of the *conditional yield change* and *conditional change in spread* variables defined above for one year horizon. Looking at these graphs, one can see that survey forecasts in period t are more than just simple copies of the period t realized yields. Expected yield and spread changes exhibit large and systematic fluctuations. As the survey statistics in Table 6 also demonstrate, the series exhibit substantial volatility, and on average, market participants anticipate yields to rise. Furthermore, after all recessions included in the sample, the market prognosticated significant surge in yields for all maturities in a manner suggesting a forthcoming flattening in the yield curve. These features already picture the market as being somewhat pessimistic, however, in principle these pessimistic forecasts might be actually correct.

The sample statistics for *subjective forecast error* contained in Table 4 for different maturities and forecast horizons, show that this is not the case. For all combinations, the sample averages are positive and increasing in the forecast horizon, that is, the farther the professionals must look into the future,

[31] For example, answers given at the end of December are viewed as the observation for the fourth quarter of the respective year. The one-quarter ahead forecast is then equal to the expected average yield between January and March.

the more biased (pessimistically) their average views are about interest rates. At horizons up to 2 quarters, the estimates are close to zero, but the statistical significance improves as the forecasting horizon is getting longer. By design, this finding is at odds with rational expectations. On the other hand, as we have seen in the earlier sections, the robust control model with its endogenous subjective belief might be a good candidate to explain these patterns.

	Mean					Standard deviation				
Survey	1-q	2-q	3-q	4-q	5-q	1-q	2-q	3-q	4-q	5-q
$\mathbf{error}_{t,h}^{(1)}$	0.02	0.16	0.34	0.52	0.71	0.60	0.88	1.13	1.36	1.54
	(0.06)	(0.11)	(0.16)	(0.20)	(0.24)	(0.08)	(0.11)	(0.14)	(0.16)	(0.16)
$\mathbf{error}_{t,h}^{(5)}$	0.06	0.21	0.38	0.55	0.71	0.65	0.90	1.07	1.23	1.33
	(0.06)	(0.10)	(0.14)	(0.17)	(0.19)	(0.07)	(0.11)	(0.13)	(0.15)	(0.15)
$\mathbf{error}_{t,h}^{(10)}$	0.05	0.19	0.34	0.50	0.64	0.60	0.83	0.99	1.12	1.20
	(0.06)	(0.10)	(0.13)	(0.16)	(0.17)	(0.07)	(0.10)	(0.12)	(0.13)	(0.14)

Table 4: Survey based means and standard deviations of*subjective forecast error* for different maturities and forecast horizons. GMM standard errors computed with the Newey-West estimator with 4 quarter lags are in parentheses.

4.3 Subjective bias

To evaluate this possibility, Figure 9 and 10 report features of the model implied *subjective forecast error* and *bias* variables, and compare them with their sample counterparts, when the model is evaluated at the worst-case parameters of Table 3. Evidently, subjective forecast error in the model moves closely together with the survey based series. Notice, however, that this measure includes the objective forecast error, which captures more of the difficulty of predicting zero-coupon yields than the systematic bias that the experts make while forecasting future yields. Because I use information about these yield dynamics to estimate the worst-case model, the impressive comovement between the survey- and model-based subjective forecast errors might only come from matching its objective part well.

To control for this, I subtract a proxy for the objective forecast error from the subjective counterpart, where the objective conditional forecasts are computed from running a simple unrestricted VAR for a vector including yields with 1-, 2-, 3-, 4-, 5-, 7-, 10- and 15-year maturities.[32] As anticipated, the resulting measure of survey bias, depicted on the right panel of Figure 9, is much smaller than the subjective forecast error, but it is still substantial, indicating again that models with rational expectations would have a hard time matching the BCFF survey. In contrast, the robust control model of this paper can produce a bias series that comoves with the survey based bias surprisingly well; at least until 2008: just like before with the yield curve, model predictions start deviating from their sample analogues in the post-Lehman period.

Figure 9 shows results for a particular yield and forecast horizon for illustrative purposes, but the observed patterns of comovement are representative for most asset-horizon combinations. In fact, as

[32]More precisely, to get an h-quarter-ahead objective forecast, I project the vector Y_{t+h} directly on Y_t and take $\widehat{\alpha}+\widehat{\beta}Y_t$.

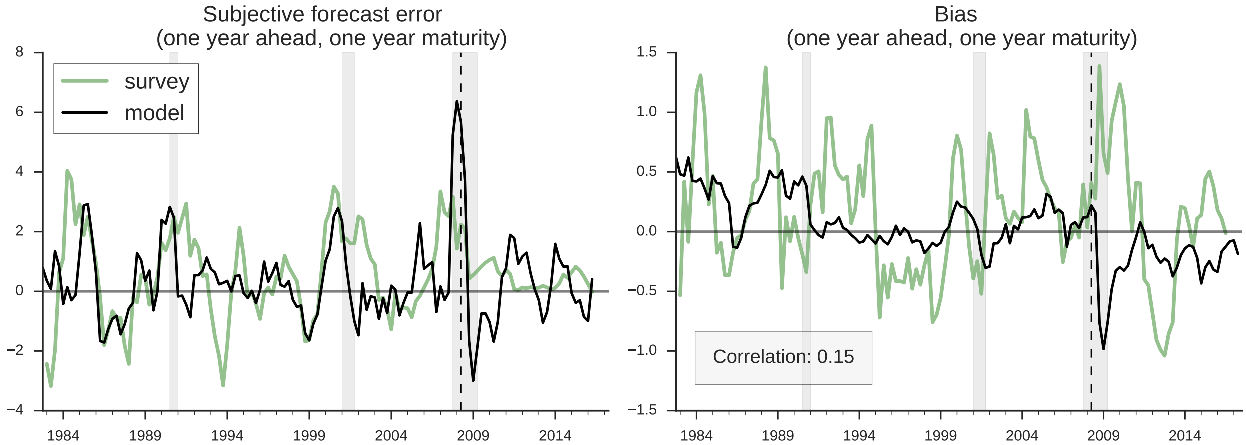

Figure 9: Comparison of the dynamics of model implied *subjective forecast error* (left) and *bias* (right) variables with the sample analogues. To approximate the objective conditional forecast needed to compute survey bias, I used an unrestricted VAR including a wide range of zero-coupon yields with different maturities. The vertical dashed lines indicate the collapse of Lehman Brothers. NBER recessions are shaded.

Table 5 illustrates, the correlation coefficient of 0.15 for the 4-quarter-ahead forecasts of the 1-year zero-coupon yield is at the lower end of the spectrum.

	Correlation (model vs survey bias)				
	1-q	2-q	3-q	4-q	5-q
$y^{(1)}$	0.07	0.16	0.16	0.15	0.15
	(0.15)	(0.16)	(0.17)	(0.17)	(0.17)
$y^{(3)}$	0.38	0.36	0.32	0.25	0.25
	(0.10)	(0.12)	(0.15)	(0.16)	(0.16)
$y^{(10)}$	0.34	0.35	0.30	0.23	0.23
	(0.15)	(0.13)	(0.13)	(0.14)	(0.14)

Table 5: Correlation coefficients between the model implied path for the *bias* measure and its survey based counterpart for different maturities and forecast horizons. GMM standard errors computed with the Newey-West estimator with 4 quarter lags are in parentheses.

The unconditional moments of the bias term are depicted in Figure 10 for various maturities and forecast horizons. Recall that in principle, the mean bias in yield forecasts should approximate the average subjective forecast error well. Although the black lines in Figure 10 clearly fall short of the point estimates in Table 4, they are reasonably close. More strikingly, these model implied mean biases can quite closely match the first two moments of the survey bias estimates in Figure 10, coming from the difference between survey expectations and VAR forecasts. Absent robustness, the predicted mean *bias* is zero. In light of this, it is remarkable that the model can replicate key features of the survey data: the mean bias is *positive* and increasing in the forecast horizon, whereas it is decreasing (or slightly hump-shaped) in the maturity dimension.

As for the average fluctuation, the model implied standard deviations of the *bias* variable are again close to the point estimates. Of course, as the survey based bias statistics depend on the particular

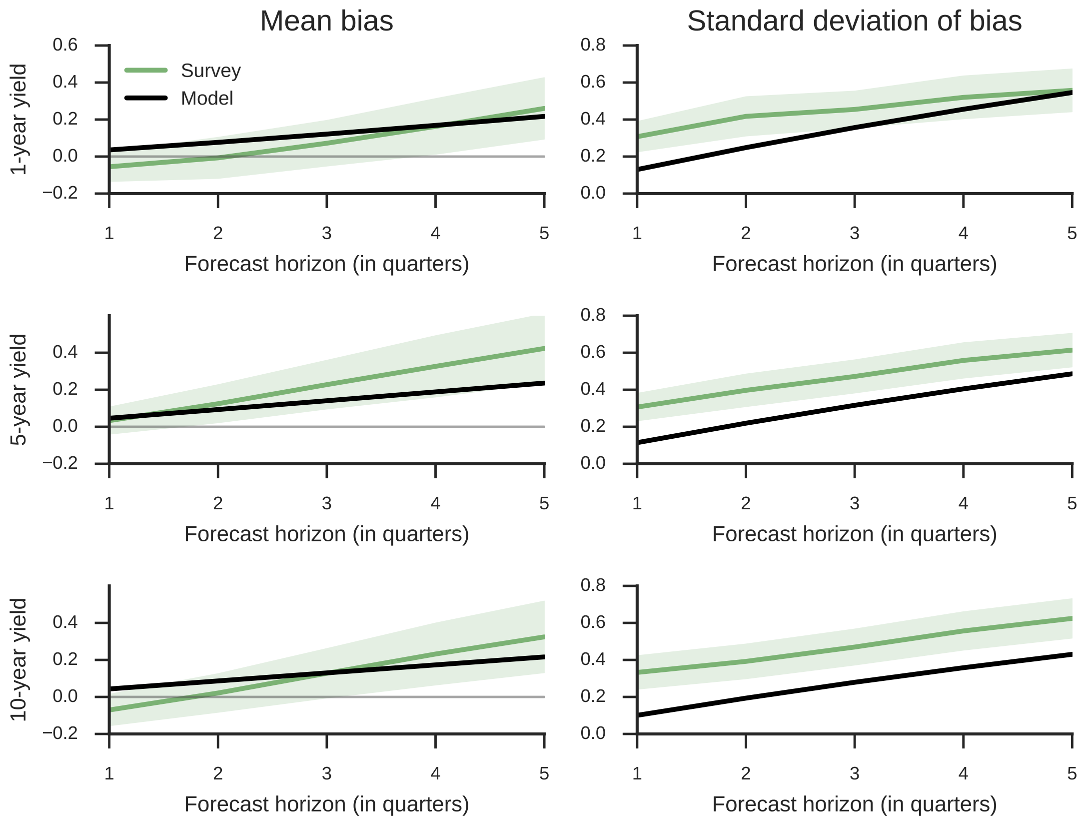

Figure 10: Comparison of the model implied unconditional moments of the bias component with the sample analogues for various maturities and forecast horizons. Model implied means and standard deviations are calculated under the stationary baseline dynamics, when the model is evaluated at the parameters of Table 3. To approximate the objective conditional forecast needed to compute survey bias, I used an unrestricted VAR including a wide range of zero-coupon yields with different maturities.

VAR specification, one should take these results with a grain of salt. Nonetheless, it is encouraging that both the model implied bias and subjective error variance (not shown in the table) are in line with the data both qualitatively and quantitatively. As an illustration, in the model, the unconditional standard deviations of the four quarter ahead subjective forecast error are 1.71, 1.46 and 1.30 for the maturities 1, 5 and 10 years respectively that are comparable with the estimates in Table 4.

4.4 Other moments

To get a more detailed picture about the performance of the elicited subjective belief, Table 6 and Figure 11 compare model predictions with the survey data along the three conditional variables for the horizon of one year. Regarding *conditional level*, the model implied moments are comparable to survey expectations, although there are systematic differences especially for short maturities: the unconditional means in the model are higher, while the standard deviations are somewhat lower than the sample statistics. The main reason behind these differences is the historically low yield expectations of the post-2007 period that the model fails to match. Otherwise, the elicited belief performs reasonably given that the estimator is restricted to be an affine function of two macro factors. It can replicate the main movements and (most of) the downward trend over the sample period. Moreover, relative to the (estimated) rational expectation model with recursive utility, depicted in the bottom two lines of Table 6, the numbers indicate clear improvement.

		Level			**Yield change**			**Spread change**	
		$\widetilde{y}_4^{(1)}$	$\widetilde{y}_4^{(5)}$	$\widetilde{y}_4^{(10)}$	$\Delta\widetilde{y}_4^{(1)}$	$\Delta\widetilde{y}_4^{(5)}$	$\Delta\widetilde{y}_4^{(10)}$	$\Delta\widetilde{sp}_4^{m}$	$\Delta\widetilde{sp}_4^{l}$
Survey	mean	4.51	5.54	6.06	0.41	0.36	0.30	-0.05	-0.11
		(0.47)	(0.45)	(0.41)	(0.09)	(0.07)	(0.06)	(0.03)	(0.05)
	std	2.80	2.70	2.46	0.55	0.43	0.37	0.25	0.32
		(0.47)	(0.80)	(1.11)	(0.05)	(0.04)	(0.03)	(0.03)	(0.04)
Robust model	mean	5.51	5.85	6.10	0.20	0.23	0.21	0.01	-0.01
	std	2.33	2.11	1.89	0.22	0.12	0.10	0.10	0.12
RE model	mean	5.35	5.28	5.23	0.00	0.00	0.00	0.00	0.00
	std	1.51	0.91	0.56	0.61	0.51	0.45	0.12	0.18

Table 6: Model implied and survey based statistics of the 1-year-ahead *conditional level*, *conditional yield change*, and *conditional change in spread* variables for different maturities. The top panel shows sample averages and GMM standard errors (computed with the Newey-West estimator including 4 quarter lags). The middle panel displays the stationary means and standard deviations (evaluated under the baseline) induced by the estimated model with the quadratic ξ. The bottom panel shows moments from the model estimated with a scalar ξ, which corresponds to recursive utility and rational expectations (RE).

A more serious challenge for the model, however, is to match the expected yield and spread changes. Unlike the *subjective forecast error*, the *conditional change* variables hinge only on the subjective state dynamics in the sense that fluctuations in $\Delta\widetilde{y}_n^{(\tau)}$ are induced by fluctuations in the *expected movements* of X. Indeed, Figure 11 reveal that the estimated model (evaluated at $\widehat{\psi}$) cannot produce enough fluctuation in $\Delta\widetilde{y}_h^{(\tau)}$ and $\Delta\widetilde{sp}_h^{(\tau)}$ to match the surveys and their performance are getting worse at longer maturities. Table 6 shows that the model implied unconditional standard deviations of the *conditional yield change* variable for the 1, 5 and 10-year bonds are, respectively, 1/3, 1/4 and 1/4 of the sample statistics. The prime reason for this low volatility is the same feature of the worst-case model that helped to explain realized interest rates: a highly persistent inflation factor X_π. Naturally, since X_π is close to a random walk, the expected change $\Delta X_{\pi,t}$ is almost zero for all t, implying low variability in the expected yield changes, which is at odds with the substantial variation of the survey analogue.

One can see this by comparing the red lines with the black ones in Figure 11, which represent hypothetical objective conditional forecasts. They are constructed by keeping the parameters of the zero-coupon yields fixed, while using the baseline model, instead of the worst-case, to forecast future values of X. Apparently, these objective forecasts, steming from a less persistent X_π variable, are substantially more volatile than the subjective model forecasts. Nevertheless, regarding long-term averages, the model performs well (and better than the objective forecast): the predicted unconditional means fall into the two standard error bands of the survey statistics. Also, there is a noticable positive comovement between the model and survey based measures. The correlation coefficients are 0.34, 0.44

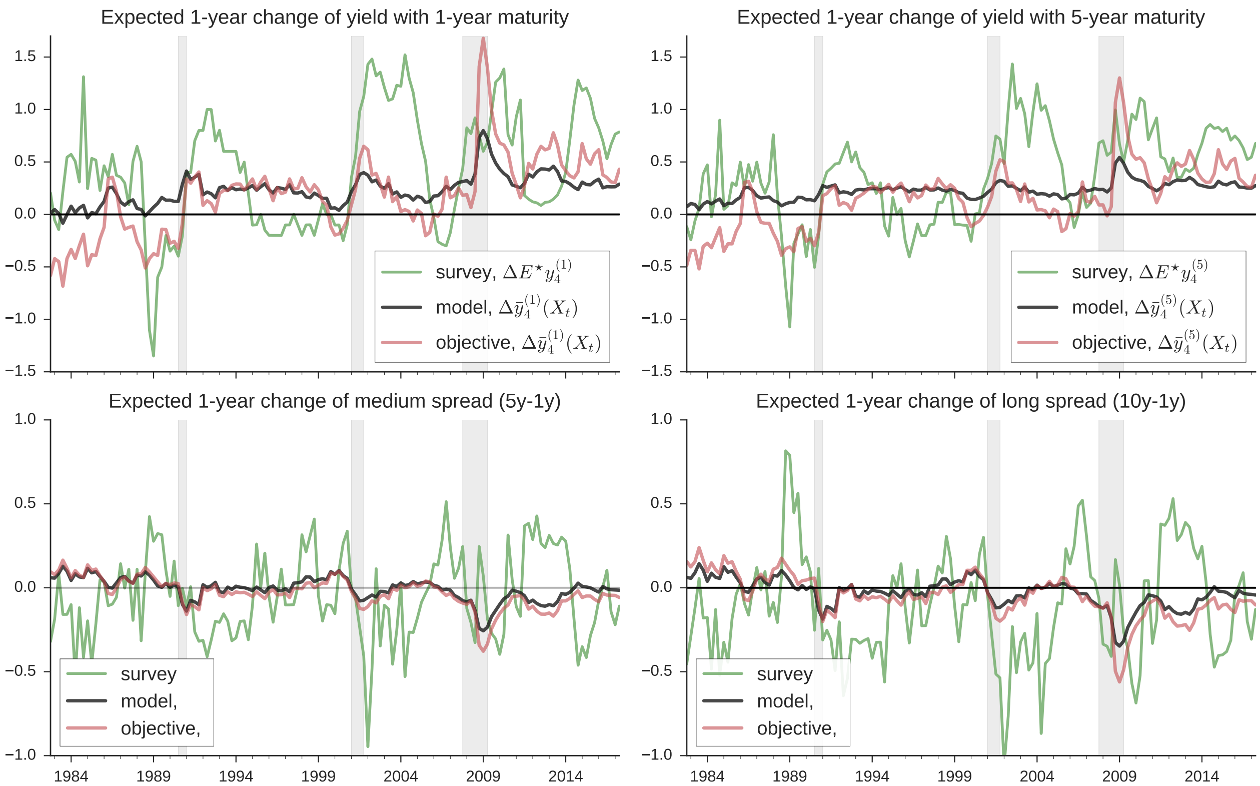

Figure 11: Comparison of the model implied *conditional yield change* (top row) and *conditional change in spread* variables (bottom row) with the BCFF survey. The forecast horizon is 1-year. The model is evaluated at the values of Table 3. Red lines display 'objective' forecasts, which fix the parameters $a(\tau)$ and $b(\tau)$ of the yields, but use the baseline model to forecast future values of the states. NBER recessions are shaded.

and 0.41 for the maturities of 1, 5 and 10 years, respectively.

Summing up, on the one hand the belief that I elicit from macro variables and asset price data by means of the robust control model, can reasonably well approximate the dynamics and average pattern of long-term forecast biases encoded in the BCFF survey. Underlying this result is the constant term $\widetilde{\phi}$ of the state process and the fact that the worst-case model adjusts it in ways to induce a positive long-term bias in yield expectations. On the other hand, the large variability of the expected yield changes in the survey suggests that the persistence of the worst-case state dynamics, inferred from realized yields, is 'too high' to explain the survey data.

5 Concluding Remarks

In this paper, I describe a general framework to estimate and evaluate robust control models using three types of data sets: macro aggregates, asset prices and survey expectations. Although I focus on bond prices and yield expectations, in principle, the idea applies to other assets and more general environments as well. For the former, testing jointly the implications of the model (as a consumption based asset pricing model) for bond and equity returns seems to be a natural next step. In addition, similar to the analysis of this paper, one might use survey expectations on stock returns to investigate whether the pessimism found in the BCFF survey holds for stock returns as well and whether these survey expectations are in line with predictions of the model (as a model for subjective beliefs).

Regarding the environment, I assume exogenous consumption and inflation, by which I seriously limit the possible effects of robustness on the macroeconomy. The reason for doing this is that I want to focus attention on how the assumed dynamics – with the associated value function – gives rise to pessimistic beliefs. Nevertheless, understanding the opposite direction, that is, how pessimistic beliefs affect the aggregate dynamics is just as important. This would necessitate a serious modeling of consumption and inflation and involve 'adding back' the max operator to the minimization problem. Naturally, this direction would complicate the model significantly, so I leave it for future research.

Appendix

A Solution of the planner's problem

This appendix provides details about the derivation of the planner's problem in section 2.5. As in the text, let the state vector be $X \in \mathbb{R}^n$ and collect the multiplicative functionals of interest in the m-vector $Y \in \mathbb{R}^m_+$, so that its first entry is consumption C. For simplicity, define the stacked vector $S := [1, X', \log Y']'$ with $1 + n + m$ elements and let $\mathbb{E}^H$ denote the expectation operator under the alternative model, which can be represented by the m-dimensional drift distrotion H relative to the baseline model. The Lagrangian of the planner's minimization problem, with $\ell \geqslant 0$ being the multiplier on the constraint $Z^H \in \mathcal{Z}(\xi)$, can be written as[33]

$$
\begin{aligned}
V(x, y) &:= \max_{\ell \geqslant 0} \min_{H} \int_0^\infty \exp(-\delta t)\mathbb{E}^H \left[\log C_t + \frac{\ell}{2}\left[H_t \cdot H_t - S_t' \xi S_t\right] \mid S_0 = s\right] dt = \\
&= \max_{\ell \geqslant 0} \min_{H} \int_0^\infty \exp(-\delta t)\mathbb{E}^H \left[S_t' Q S_t + H_t' R H_t \mid S_0 = s\right] dt
\end{aligned}
$$

subject to

$$
dS_t = \begin{bmatrix} 0 & \mathbf{0}_n & 0 & \mathbf{0}'_{m-1} \\ \phi & -\kappa & 0 & \mathbf{0}'_{m-1} \\ \beta_0 & \beta_1 & 0 & \mathbf{0}'_{m-1} \end{bmatrix} S_t dt + \begin{bmatrix} \mathbf{0}'_m \\ \sigma \\ \alpha \end{bmatrix} H_t dt + \begin{bmatrix} \mathbf{0}'_m \\ \sigma \\ \alpha \end{bmatrix} dW_t =: A S_t dt + B H_t dt + \Sigma dW_t
$$

where $H_0 = 0$ and W is an m-dimensional Brownian Motion and

$$
Q = -\frac{1}{2}\begin{bmatrix} \ell\xi_0 & \ell\xi_1' & -1 & \mathbf{0}'_{m-1} \\ \ell\xi_1 & \ell\xi_2 & \mathbf{0}_n & \mathbf{0}'_{n\times(m-1)} \\ -1 & \mathbf{0}'_n & 0 & \mathbf{0}'_{m-1} \\ \mathbf{0}_{m-1} & \mathbf{0}_{(m-1)\times n} & \mathbf{0}_{m-1} & \mathbf{0}'_{(m-1)\times(m-1)} \end{bmatrix} \qquad R = \frac{\ell}{2} I_m
$$

where $\mathbf{0}_n$ denotes an n-vector of zeros, while I_m is an identity matrix of size m.

Evidently, the above problem is a linear-quadratic optimal control problem. The corresponding value function is of the form $V = S_t' P S_t$, where P solves the (discounted) continuous time algebraic Riccati equation (CARE)

$$
A'P + PA - PBR^{-1}B'P + Q = \delta P
$$

Guess the form $V(x, y) = [1, x']v(\ell)[1, x']' + \frac{1}{\delta}\log c$, with $v(\ell)$ being a positive semi-definite square matrix of size $1 + n$ and verify it through the HJB equation

$$
\delta V(x, y) = \min_h \log c + \frac{\ell}{2}\left(|h|^2 - \xi(x)\right) + \nabla V_x' (\phi - \kappa x + \sigma h) + \frac{1}{\delta} e_1' (\beta_0 + \beta_1 x + \alpha h) + \frac{1}{2}\mathrm{tr}\left[\sigma' V_{xx} \sigma\right]
$$

[33] Small capitals x and y denote realizations of random variables X and Y.

This guess translates into the following restricted form of P

$$P = \begin{bmatrix} \upsilon_0(\ell) & \upsilon_1'(\ell) & 1/(2\delta) & \mathbf{0}'_{m-1} \\ \upsilon_1(\ell) & \upsilon_2(\ell) & \mathbf{0}_n & \mathbf{0}'_{n\times(m-1)} \\ 1/(2\delta) & \mathbf{0}'_n & 0 & \mathbf{0}'_{m-1} \\ \mathbf{0}_{m-1} & \mathbf{0}_{(m-1)\times n} & \mathbf{0}_{m-1} & \mathbf{0}'_{(m-1)\times(m-1)} \end{bmatrix}$$

with the feedback control

$$H_t^* = -R^{-1}B'PS_t = -\frac{2}{\ell}\left[\frac{1}{2\delta}\alpha' e_1 + \sigma'\upsilon_1(\ell) + \sigma'\upsilon_2(\ell)X_t\right] \equiv \eta_0 + \eta_1 X_t \tag{21}$$

where η_0 is an $m \times 1$ vector, while η_1 is an $m \times n$ matrix.

The particular pattern of zeros in P leads to an independent CARE for $\upsilon_2(\ell)$

$$\left((-\kappa) - \frac{\delta}{2}\right)'\upsilon_2(\ell) + \upsilon_2(\ell)\left((-\kappa) - \frac{\delta}{2}\right) - \frac{2}{\ell}\upsilon_2(\ell)\sigma\sigma'\upsilon_2(\ell) - \frac{\ell}{2}\xi_2 = 0$$

Notice that this equation is homogeneous in ℓ, i.e. $\upsilon_2(\ell) = \bar{\upsilon}_2\ell$, where

$$\left((-\kappa) - \frac{\delta}{2}\right)'\bar{\upsilon}_2 + \bar{\upsilon}_2\left((-\kappa) - \frac{\delta}{2}\right) - 2\bar{\upsilon}_2\sigma\sigma'\bar{\upsilon}_2 - \frac{1}{2}\xi_2 = 0 \tag{22}$$

Evaluate the HJB with the guess at the optimum

$$\begin{aligned} 0 = &-\delta\left[\upsilon_0(\ell) + 2\upsilon_1(\ell)\cdot x + x'\ell\bar{\upsilon}_2 x\right] + 2\left[x'\bar{\upsilon}_2\ell + \upsilon_1(\ell)'\right](\phi - \kappa x) + \frac{1}{\delta}e_1'(\beta_0 + \beta_1 x) + \\ &-\frac{1}{2\ell}\left|\frac{1}{\delta}\alpha' e_1 + 2\sigma'\left[\upsilon_1(\ell) + \ell\bar{\upsilon}_2 x\right]\right|^2 + \frac{1}{2}\mathrm{tr}\left[\sigma'\bar{\upsilon}_2\sigma\right]\ell - \frac{\ell}{2}\left(\xi_0 + 2\xi_1\cdot x + x'\xi_2 x\right) \end{aligned}$$

Collecting the linear terms yields

$$\begin{aligned} \upsilon_1(\ell) &= \left(\delta I_N - (-\kappa)' + 2\bar{\upsilon}_2\sigma\sigma'\right)^{-1}\left[\ell\left(\bar{\upsilon}_2\phi - \frac{\xi_1}{2}\right) + \frac{1}{\delta}\left(\frac{1}{2}\beta_1' - \bar{\upsilon}_2\sigma\alpha'\right)e_1\right] \\ &\equiv \ell\bar{\upsilon}_{1,1} + \bar{\upsilon}_{1,0} \end{aligned} \tag{23}$$

Matching coefficients for the constant gives

$$\begin{aligned} v_0(\ell) = &\frac{1}{\delta}\left[2\phi'\bar{\upsilon}_{1,0} + \frac{1}{\delta}e_1'[\beta_0 - 2\alpha\sigma'\bar{\upsilon}_{1,1}] - 4\bar{\upsilon}_{1,0}\sigma\sigma'\bar{\upsilon}_{1,1}\right] + \\ &+\frac{1}{\delta}\left(2\phi'\bar{\upsilon}_{1,1} + \frac{1}{2}\mathrm{tr}\left[\sigma'\bar{\upsilon}_2\sigma\right] - \frac{\xi_0}{2} - 2\bar{\upsilon}_{1,1}'\sigma\sigma'\bar{\upsilon}_{1,1}\right)\ell + \\ &-\frac{1}{\delta}\left(\frac{e_1'\alpha\alpha' e_1}{2\delta^2} + \frac{2}{\delta}e_1'\alpha\sigma'\bar{\upsilon}_{1,0} + 2\bar{\upsilon}_{1,0}'\sigma\sigma'\bar{\upsilon}_{1,0}\right)\ell^{-1} \equiv \bar{\upsilon}_{0,1}\ell + \bar{\upsilon}_{0,0} + \bar{\upsilon}_{0,-1}\ell^{-1} \end{aligned} \tag{24}$$

Given the form of the value function for a fixed ℓ, we can maximize over $\ell > 0$ to obtain

$$\ell^* = \sqrt{\frac{\bar{\upsilon}_{0,-1}}{\bar{\upsilon}_{0,1} + 2\bar{\upsilon}_{1,1}\cdot x + x'\bar{\upsilon}_2 x}} \tag{25}$$

The equations (22), (23), (24) and (25) with the guessed (and verified) form of the value function provide the solution of the planner's problem.

B Decentralization of the planner's problem

In this appendix I show that the planner's solution derived in Appendix A can be decentralized by scaled shadow prices coming from the planner's value function

$$V(x, y) = \upsilon_0(\ell) + 2\upsilon_1'(\ell)x + x'\upsilon_2(\ell)x + \log d$$

The first element of the vector Y_t is the exogenous endowment/consumption $D_t = C_t$. The separable log utility and the fact that there is not way to move resources over states and time imply that the planner's shadow price process (without discounting and weighting by probabilities) is given by $u'(D_t) = \frac{1}{D_t} = \frac{1}{C_t}$ with the process D_t being evaluated under the worst-case model.[34] It is useful to express this process under the baseline model and define $p_t \equiv \frac{1}{C_t} Z_t^{H*}$ following

$$\frac{dp_t}{p_t} = -\left(e_1\cdot(\beta_0 + \alpha\eta_0) + e_1'\left(\beta_1 + \alpha\eta_1\right)X_t - \frac{|e_1'\alpha|^2}{2}\right)dt - \big(\alpha' e_1 - \eta(X_t)\big)\cdot dW_t \tag{26}$$

where η_0 and η_1 are the coefficients of the planner's worst-case drift distortion $\eta(X_t) \equiv H_t^*$ defined in (21). Following the strategy of Hansen and Sargent (2008), I show now that if we confront the price-taking household with the price system (26), its worst-case model will be in line with that derived in (21) and the robust decision rule will dictate to consume D_t.

The household faces the exogenous price process p_t, which depends on the planner's worst-case drift distrotion (21), however, in principle she might choose a potentially different drift distortion $\bar{H}$. In order to ensure that the household behaves as a price-taker, introduce a composite state vector $(\bar{X}, X)$, where the exogenous X follows (14), while the conformable $\bar{X}$ denotes the minimizing player's endogenous state, the dynamics of which is driven by $\bar{H}$. Despite the possibly different dynamics, the known initial condition implies that $\bar{X}_0 = X_0$.

The time zero budget constaint of the household is

$$\int_0^\infty \exp(-\delta t)\mathbb{E}^{\bar{H}}\left[p_t(\bar{D}_t - \bar{C}_t)\right]dt \geqslant 0$$

with an associated Lagrange multiplier $\mu \geqslant 0$. Let $\ell \geqslant 0$ be the Lagrange multiplier on $\mathcal{Z}(\xi)$ and guess

[34] By definition, shadow prices are derived from the planner's value function evaluated at the optimum, so the process D_t must be evaluated *under the worst-case* model.

the form

$$W(\bar{x},x,\bar{y}) = \log \bar{d} + \begin{bmatrix} 1 & \bar{x}' & x' \end{bmatrix} \begin{bmatrix} w_0 & w_{\bar{x}}' & w_x' \\ w_{\bar{x}} & w_{\bar{x}\bar{x}} & w_{\bar{x}x} \\ w_x & w_{\bar{x}x}' & w_{xx} \end{bmatrix} \begin{bmatrix} 1 \\ \bar{x} \\ x \end{bmatrix}$$

for the value function of the two-player zero-sum game. The associated HJB equation is

$$\begin{aligned} \delta W(\bar{x},x,\bar{y}) = \max_{\bar{c}} \ \min_{\bar{h}} \ & \delta \log \bar{c} + \frac{\ell}{2}\left[|\bar{h}|^2 - \xi(\bar{x})\right] + \mu\left[p(\bar{d}-\bar{c})\right] + \nabla W'_{(\bar{x},x)} \begin{bmatrix} \phi - \kappa\bar{x} + \sigma\bar{h} \\ \phi - \kappa x + \sigma\bar{h} \end{bmatrix} + \\ & + e_1'\left(\beta_0 + \beta_1\bar{x} + \alpha\bar{h}\right) + \frac{1}{2}\text{trace}\left(\sigma'\nabla^2 W_{(\bar{x},x)}\sigma\right) \end{aligned} \tag{27}$$

Because this problem satisfies the Bellman-Isaacs condition, the order of the max-min operators is interchangeable. Taking the first order condition w.r.t. $\bar{c}$ yields $\bar{c}^* = \frac{\delta}{\mu p}$. Plugging this form into (27) leads to

$$\delta \log \bar{c}^* + \mu p(\bar{d} - \bar{c}^*) = \delta \log \delta - \delta \log \mu - \delta \log p + \mu p \bar{d} - \delta$$

Taking the minimum of this w.r.t. $\mu \geqslant 0$ implies $\mu^* = \frac{\delta}{p\bar{d}}$ and

$$\delta \log \delta - \delta \log \mu^* - \delta \log p + \mu^* p \bar{d} - \delta = \delta \log \bar{d}$$

Regarding the mininization w.r.t. $\bar{h}$, the FOC implies

$$\bar{h}^* = -\frac{1}{\ell}\left[\alpha' e_1 + \begin{bmatrix} \sigma' & 0 \\ 0 & \sigma' \end{bmatrix} \nabla W_{(\bar{x},x)}\right]$$

Plugging all these into (27) leaves us with

$$\begin{aligned} \delta\left(W(\bar{x},x,\bar{y}) - \log \bar{d}\right) = \ & \nabla W'_{(\bar{x},x)}\left(\begin{bmatrix} \phi \\ \phi \end{bmatrix} + \begin{bmatrix} -\kappa & 0 \\ 0 & -\kappa \end{bmatrix}\begin{bmatrix} \bar{x} \\ x \end{bmatrix}\right) + \begin{bmatrix} \bar{x}' & x' \end{bmatrix}\begin{bmatrix} \xi_2 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \bar{x} \\ x \end{bmatrix} \\ & - \frac{1}{2\ell}\left(e_1'\alpha\alpha' e_1 + 2e_1'\alpha \begin{bmatrix} \sigma' & 0 \\ 0 & \sigma' \end{bmatrix} \nabla W_{(\bar{x},x)} + \nabla W'_{(\bar{x},x)} \begin{bmatrix} \sigma\sigma' & 0 \\ 0 & \sigma\sigma' \end{bmatrix} \nabla W_{(\bar{x},x)}\right) + \\ & + \begin{bmatrix} e_1'\beta_1 - \frac{\ell}{2}\xi_1' & 0 \end{bmatrix}\begin{bmatrix} \bar{x} \\ x \end{bmatrix} + e_1 \cdot \beta_0 - \frac{\ell}{2}\xi_0 + \frac{1}{2}\text{trace}\left(\sigma'\nabla^2 W_{(\bar{x},x)}\sigma\right) \end{aligned}$$

where

$$\nabla W_{(\bar{x},x)} = 2\begin{bmatrix} w_{\bar{x}} & w_{\bar{x}\bar{x}} & w_{\bar{x}x} \\ w_x & w_{\bar{x}x}' & w_{xx} \end{bmatrix}\begin{bmatrix} 1 \\ \bar{x} \\ x \end{bmatrix} = 2\begin{bmatrix} w_{\bar{x}} \\ w_x \end{bmatrix} + 2\begin{bmatrix} w_{\bar{x}\bar{x}} & w_{\bar{x}x} \\ w_{\bar{x}x}' & w_{xx} \end{bmatrix}\begin{bmatrix} \bar{x} \\ x \end{bmatrix}$$

Matching coefficients while considering the equations blockwise, it is striaghtforward to see that

$$w_{\bar{x}\bar{x}} = \bar{\upsilon}_2 \qquad \text{and} \qquad w_{\bar{x}x} = w_{xx} = 0$$

where $\bar{\upsilon}_2$ is the solution of the planner's Riccatti equation in (22). This results in the linear terms

$$w_{\bar{x}} = \upsilon_1(\ell) \qquad \text{and} \qquad w_x = 0$$

where $\upsilon_1(\ell)$ comes from (23). This immediately implies that $w_0 = \upsilon_0(\ell)$ from (24) and $\ell = \ell^*$ from (25) proving that $\bar{H} = H^*$ and $\bar{C} = D$.

C Results for the restricted sample

This appendix presents results for the period 1952:Q2-2005:Q4, which coincides with the sample period used by Piazzesi and Schneider (2007). To facilite comparison with their analysis, I also include estimates for the case when the tilting function ξ is restricted to be constant, which is interpretable as recursive utility or the robust control model of Hansen and Sargent (2001) with unstructured uncertainty (see Section 3.3.1).

κ^D		σ^D		β_0^D	α^D	
0.568	-0.066	0.129	-0.038	0.533	0.434	0.000
(0.139)	(0.043)	(0.030)	(0.029)	-	(0.021)	-
0.225	0.987	-0.002	0.167	0.920	-0.089	0.300
(0.104)	(0.030)	(0.024)	(0.021)	-	(0.021)	(0.016)

Table 7: Maximum Likelihood estimates and asymptotic standard errors (in parentheses) for the baseline model (19)-(20) when the sample is restricted to the period 1952:Q2-2005:Q4. The likelihood is initialized at the stationary distribution of X. The column for β_0^D shows the sample averages. The matrix α^D is normalized to be lower triangular.

Table 7 contains maximum likelihood estimates of the baseline model parameters. These values are more or less in line with the findings of Piazzesi and Schneider (2007). Apparently, in the restricted sample consumption growth appears to be less persistent, while inflation is even more persistent than in the extended sample used in the main text. Figure 12 displays the corresponding autrocorrelation functions along with the sample analogues calculated using the restricted sample. As hinted before, the main difference between the two samples can be spotted on the top right panel: for the restricted sample, high inflation seems to be a good predictor of low future consumption growth, at least on a 1-2 year horizon.

As for the parameters of the tilting function, Table 8 contains the non-linear least squares estimates. Assuming the quadratic form (9) leads to parameter estimates significantly different from the ones in Table 2. The diagonal elements of the Ξ matrix are an order of magnitude larger and the cross term $\bar{\xi}_2$ even changes sign. This suggests that the properties of the baseline model, especially the estimated persistence of consumption growth and inflation can significantly affect the particular values of the tilting function parameters. Nevertheless, these drastic changes in the parameter values do not

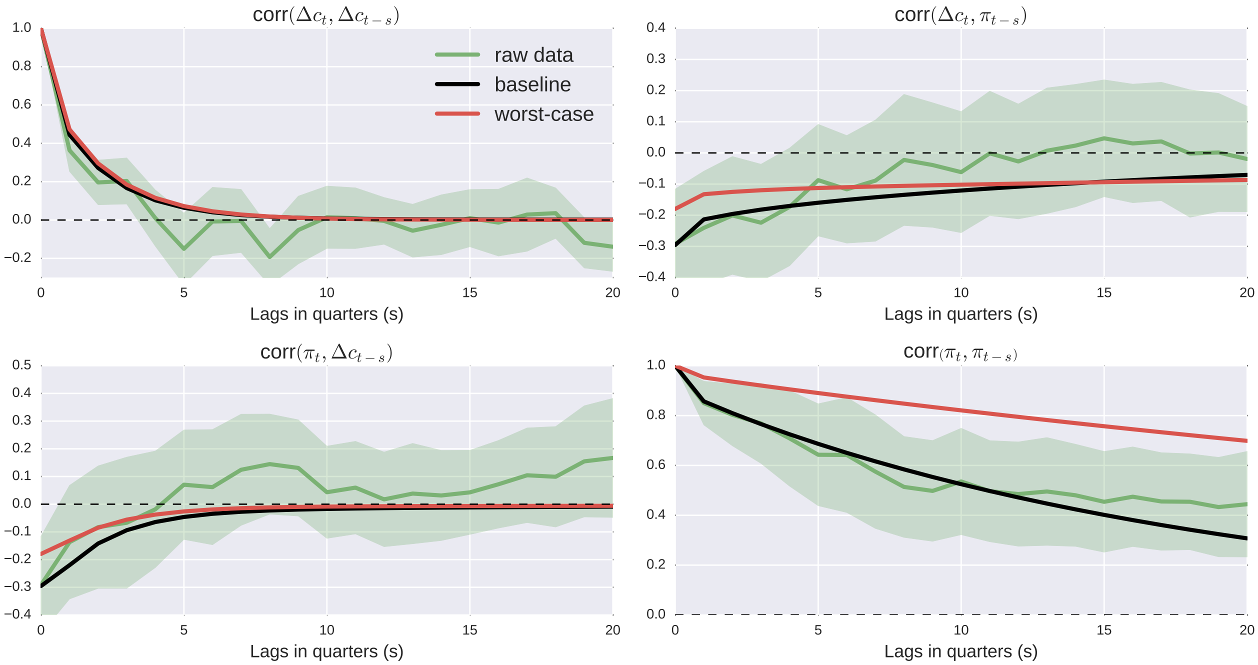

Figure 12: Autocorrelation functions computed from raw data (green solid lines) and from the baseline model (black solid lines) when the sample is restricted to the period 1952:Q2-2005:Q4. Shaded area represents 2×GMM standard error bounds computed with the Newey-West estimator including 4 quarter lags.

alter the model's overall prediction about the worst-case distribution. As the red lines on Figure 12 demonstrate, the worst-case model indicates more persistence than the baseline.

The relationship between the red and black lines on the top right panel provides explanations for the changing sign of $\bar{\xi}_2$. The estimated baseline model assigns such a strong forecasting ability to inflation that the worst-case model wants to counteract it, at least on the short horizon (up to 3-year). Because this forecasting channel is shaped mainly by the parameter $\bar{\xi}_2$, the estimated sign is flipped.

	$\bar{\xi}_0$		$\bar{\xi}_1$	$\bar{\xi}_2$	$\bar{\xi}_3$
Quadratic tilting	0.0		11,829	4,537	1,740
	(0.02)		(8,125)	(1,031)	(201)
Recursive utility	0.077	$\Rightarrow$	$\gamma = 32$		
	(0.013)				

Table 8: Non-linear least squares estimates and asymptotic GMM standard errors (in the parentheses) for the parameters of ξ when the sample is restricted to the period 1952:Q2-2005:Q4. The standard errors take into account the two-step nature of the estimator. γ denotes the risk aversion parameter of recursive utility and is calculated from $\bar{\xi}_0$ using the formula in Section 3.3.1.

As for the special case of recursive utility, the bottom panel of Table 8 shows that in the restricted sample, recursive utility does have a role. The best fit is reached when the risk aversion parameter is around 30, which is in the ballpark of the estimate of Piazzesi and Schneider (2007). Nevertheless, the right panel of Figure 13 reveals that even though recursive utility is able to generate an upward sloping average nominal yield curve in the restricted sample, the predicted variation is quickly decreasing with the horizon, which is at odds with the data. Again, we need state dependence in ξ in order to break the expectations hypothesis and induce time-variation in risk premia.

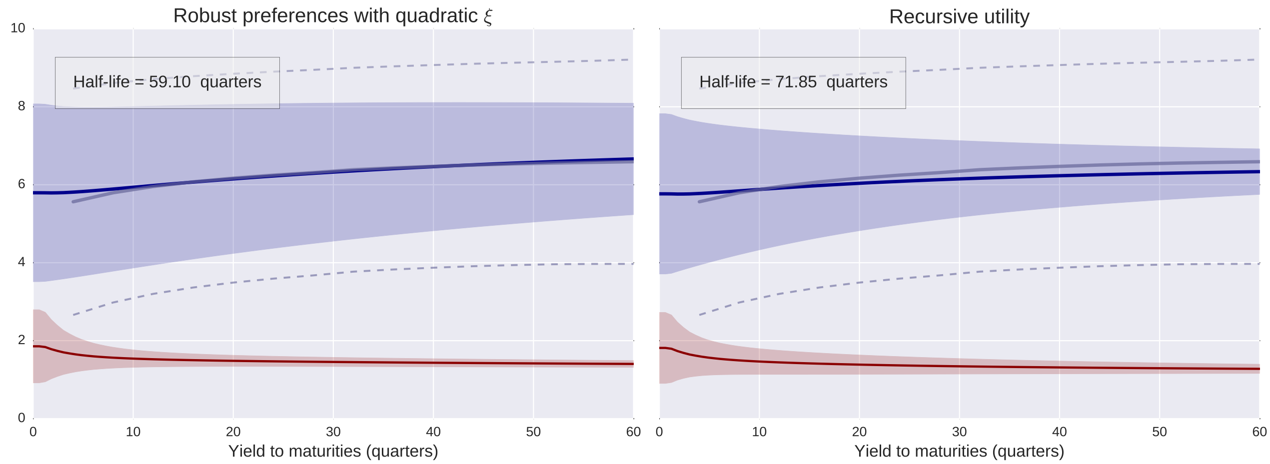

Figure 13: Model implied stationary distributions of the nominal (blue) and real (red) yield curves and the corresponding sample moments when the sample is restricted to the period 1952:Q2-2005:Q4. Shaded areas represent one standard deviation bands around the means (solid lines). Grey solid lines show the sample average, the dashed lines are one standard deviation bands using the sample standard deviations. The left panel is for the model with a constant ξ (recursive utility). The right panel shows the case for expected utiliy with logarithmic preferences. The boxes contain associated half-lives.

References

Anderson, Evan W., Lars Peter Hansen, and Thomas J. Sargent. 2003. "A Quartet of Semigroups for Model Specification, Robustness, Prices of Risk, and Model Detection." *Journal of the European Economic Association* 1 (1):68–123.

Bansal, Ravi and Amir Yaron. 2004. "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles." *The Journal of Finance* 59 (4):1481–1509.

Barillas, Francisco, Lars Peter Hansen, and Thomas J. Sargent. 2009. "Doubts or variability?" *Journal of Economic Theory* 144 (6):2388 – 2418.

BEA. 2016. "NIPA Handbook: Concepts and Methods of the U.S. National Income and Product Accounts." Tech. rep., U.S. Bureau of Economic Analysis. URL http://bea.gov/national/pdf/all-chapters.pdf.

Bhandari, Anmol, Jaroslav Borovička, and Paul Ho. 2019. "Survey Data and Subjective Beliefs in Business Cycle Models." University of Minnesota, New York University and Federal Reserve Bank of Richmond.

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

Borovička, Jaroslav, Lars Peter Hansen, Mark Hendricks, and José A. Scheinkman. 2011. "Risk-Price Dynamics." *Journal of Financial Econometrics* 9 (1):3–65.

Cagetti, Marco, Lars Peter Hansen, Thomas J. Sargent, and Noah Williams. 2002. "Robustness and Pricing with Uncertain Growth." *Review of Financial Studies* 15 (2):363–404.

Campbell, John Y. and John H. Cochrane. 1999. "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior." *Journal of Political Economy* 107 (2):205–251.

Chen, Hui, Winston Wei Dou, and Leonid Kogan. 2017. "Measuring the "Dark Matter" in Asset Pricing Models." Tech. rep., MIT Sloan and Wharton. URL http://www.mit.edu/~huichen/darkmatter.pdf.

Cochrane, John H. 2005. *Asset Pricing (Revised Edition)*. Princeton and Oxford. Princeton University Press.

Duffie, Darrell and Larry G. Epstein. 1992. "Asset Pricing with Stochastic Differential Utility." *Review of Financial Studies* 5 (3):411–436.

Duffie, Darrell and Rui Kan. 1996. "A Yield-factor model of interest rates." *Mathematical Finance* 6 (4):379–406.

Epstein, Larry G. and Martin Schneider. 2003. "Recursive multiple-priors." *Journal of Economic Theory* 113 (1):1–31.

Epstein, Larry G. and Stanley E. Zin. 1989. “Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework.” *Econometrica* 57 (4):937–969.

Fama, Eugene F. and Robert R. Bliss. 1987. “The Information in Long-Maturity Forward Rates.” *The American Economic Review* 77 (4):680–692.

Fleming, W. H. and E. P. Souganidis. 1989. “On the existence of value functions of two player, zero-sum stochastic differential games.” *Indiana University Mathematics Journal* 38 (2):293–314.

Gilboa, Itzhak and David Schmeidler. 1989. “Maxmin expected utility with non-unique prior.” *Journal of Mathematical Economics* 18 (2):141–153.

Gurkaynak, Refet S., Brian Sack, and Jonathan Wright. 2007. “The U.S. Treasury Yield Curve: 1961 to the Present.” *Journal of Monetary Economics* 8 (54):2291 – 2304.

———. 2010. “The TIPS Yield Curve and Inflation Compensation.” *American Economic Journal: Macroeconomics* 1 (2):70 – 92.

Hamilton, James D and Dong Heon Kim. 2002. “A Reexamination of the Predictability of Economic Activity Using the Yield Spread.” *Journal of Money, Credit and Banking* 34 (2):340–60.

Hansen, Lars Peter. 2007. “Beliefs, Doubts and Learning: Valuing Macroeconomic Risk.” *American Economic Review* 97 (2):1–30.

Hansen, Lars Peter and Thomas J. Sargent. 1980. “Formulating and estimating dynamic linear rational expectations models.” *Journal of Economic Dynamics and Control* 2:7 – 46.

———. 2001. “Robust Control and Model Uncertainty.” *American Economic Review* 91 (2):60–66.

———. 2007. “Recursive robust estimation and control without commitment.” *Journal of Economic Theory* 136 (1):1 – 27.

———. 2008. *Robustness*. Princeton University Press.

———. 2010. “Fragile beliefs and the price of uncertainty.” *Quantitative Economics* 1 (1):129–162.

———. 2019. “Macroeconomic Uncertainty Prices when Beliefs are Tenuous.” Tech. rep., University of Chicago and New York University. URL http://www.tomsargent.com/research/tenuous.pdf.

Hansen, Lars Peter, Thomas J. Sargent, and Thomas D. Tallarini. 1999. “Robust Permanent Income and Pricing.” *Review of Economic Studies* 66 (4):873–907.

Hansen, Lars Peter, Thomas J. Sargent, Gauhar Turmuhambetova, and Noah Williams. 2006. “Robust control and model misspecification.” *Journal of Economic Theory* 128 (1):45–90.

Hansen, Lars Peter and José A. Scheinkman. 2009. “Long-Term Risk: An Operator Approach.” *Econometrica* 77 (1):177–234.

Hansen, Lars Peter, Bálint Szőke, Lloyd Han, and Thomas J. Sargent. forthcoming. "Twisted Probabilities, Uncertainty, and Prices." *The Journal of Econometrics* .

Ilut, Cosmin L. and Martin Schneider. 2014. "Ambiguous Business Cycles." *American Economic Review* 104 (8):2368–99.

Kreps, David and Evan L Porteus. 1978. "Temporal Resolution of Uncertainty and Dynamic Choice Theory." *Econometrica* 46 (1):185–200.

Ljungqvist, Lars and Thomas J. Sargent. 2012. *Recursive Macroeconomic Theory, Third Edition*, *MIT Press Books*, vol. 1. The MIT Press, 3 ed.

Lucas, Robert E. 1978. "Asset Prices in an Exchange Economy." *Econometrica* 46 (6):1429–1445.

McCulloch, J.Huston and Heon-Chul Kwon. 1993. "U.S. Term Structure Data, 1947-1991." Tech. Rep. Working Paper 93-6., Ohio State University. URL http://econ.ohio-state.edu/jhm/ts/mcckwon/mccull.htm.

Pesaran, M. Hashem and Martin Weale. 2006. "Survey Expectations." In *Handbook of Economic Forecasting*, vol. 1, edited by G. Elliott, C. Granger, and A. Timmermann, chap. 14. Elsevier, 715–776.

Piazzesi, Monika. 2010. "Affine term structure models." In *Handbook of Financial Econometrics*, edited by Yacine Aït-Sahalia and Lars Peter Hansen, Elsevier Handbooks. Amsterdam: North-Holland, 691–766.

Piazzesi, Monika, Juliana Salomao, and Martin Schneider. 2015. "Trend and Cycle in Bond Premia." Tech. rep., Stanford University. URL http://web.stanford.edu/~piazzesi/trendcycle.pdf.

Piazzesi, Monika and Martin Schneider. 2007. "Equilibrium Yield Curves." In *NBER Macroeconomics Annual 2006, Volume 21*, NBER Chapters. National Bureau of Economic Research, Inc, 389–472.

Shiller, Robert J. 1979. "The Volatility of Long-Term Interest Rates and Expectations Models of the Term Structure." *Journal of Political Economy* 87 (6):1190–1219.

Skiadas, Costis. 2003. "Robust control and recursive utility." *Finance and Stochastics* 7:475–489.

Ulrich, Maxim. 2013. "Inflation ambiguity and the term structure of U.S. Government bonds." *Journal of Monetary Economics* 60 (2):295–309.