Restrictions on Asset-Price Movements Under Rational Expectations: Theory and Evidence*

Ned Augenblick
UC Berkeley Haas

Eben Lazarus
MIT Sloan

JANUARY 2022

Abstract

How restrictive is the assumption of rational expectations in asset markets? To address this, we derive bounds on admissible asset-price variation under general assumptions. The challenge in this task is that valuations reflect both beliefs and preferences. We gain traction by considering market-implied, or risk-neutral, probabilities of future outcomes, and we provide a mapping between the variation in these probabilities and the minimum risk aversion — or slope of the stochastic discount factor — required under rationality. Implementing these bounds empirically using index options, we find that very high risk aversion is needed to rationalize the behavior of risk-neutral beliefs.

JEL Codes: D84, G14, G41

*Contact: ned@haas.berkeley.edu and elazarus@mit.edu. We are grateful to Laura Blattner, Jarda Borovička, John Campbell, Gabriel Chodorow-Reich, Emmanuel Farhi, Xavier Gabaix, Niels Gormsen, Sam Hanson, Bryan Kelly, David Laibson, Ian Martin, Matthew Rabin, Andrei Shleifer, Juan Sotes-Paladino, David Sraer, Jim Stock, Adrien Verdelhan, Jessica Wachter, and numerous seminar and conference participants for advice, comments, and helpful discussions.

1. Introduction

Asset prices are volatile. Whether they are *too* volatile to reflect rational variation in expected future cash flows, though, is by itself an unresolvable question: the true distribution of future cash-flow streams is unobservable, as are aggregate risk and time preferences. The relevant question is instead what *joint* set of assumptions on (1) beliefs and (2) risk and time preferences — the two elements at the heart of almost all modern theories of asset prices — can be rejected in the data.

Shiller (1981) provides a classic example. He documents excess volatility in equity-index prices relative to a proxy for fundamental value, but this proxy is constructed under the assumption that discount rates over future cash flows are constant over time. Shiller's test therefore provides a *joint* rejection of (1) rational pricing and (2) constant discount rates. But given the ample evidence that discount rates vary over time, it is unclear what to make of this joint rejection.[1] It would seem that changes in unobserved discount rates are capable of explaining arbitrary movements in asset values, so can any meaningful statements be made about how much rational variation we should observe in the data?

In this paper, we show that, even with a meaningful degree of flexibility in preferences and discount rates, there are still certain bounds on asset-price movements that must hold under the assumption of rational expectations (RE). Our theoretical results apply quite generally to valuation processes for any agent with RE over an asset's future payoffs; when we then take these bounds to the data, the "agent" in question can be thought of as the market as a whole. Our results accordingly bear on the efficiency of market valuations, as we show that this null can be jointly tested alongside a much less restrictive assumption than those used in past literature. When we implement our bounds empirically, we find that there is so much variation in the asset prices we consider that it is difficult to rationalize the data with plausible assumptions on time and risk preferences.

As in Shiller's case, we focus our analysis on expectations over the future value of an equity index. But the key feature distinguishing our analysis from previous literature is that we consider the behavior of so-called *risk-neutral (RN) beliefs* over the underlying index's future price, rather than the behavior of the underlying index itself. The risk-neutral belief distribution can be calculated directly using option prices — options allow for bets over the future asset price, and thus the prices of these bets allow us to back out a probability distribution over this future price — so as is standard, we treat risk-neutral beliefs as observable. These risk-neutral beliefs represent the probability distribution that would be equal to a hypothetical risk-neutral agent's true (or *physical*) belief distribution about the future asset price, but risk-neutral beliefs are in general distorted relative to the marginal investor's physical beliefs in the case that the investor is risk-averse. Intuitively, the probability distribution we observe using asset prices will overweight states in which the marginal investor has low wealth (e.g., when the underlying asset has a low return), since the investor will be willing to pay to insure against these high-marginal-utility states.

[1] For an overview of time-varying discount rates, see Cochrane (2011). Fama (1991, p. 1586) provides a succinct summary of the issues in interpreting Shiller's results: "volatility tests are [a] useful way to show that expected returns vary through time," but they "give no help on the central issue of whether the variation in expected returns is rational."

We show that statements about the "correct" amount of variation in risk-neutral beliefs under RE require less-restrictive assumptions than statements about variation of the index price itself. Previous analyses focusing on index-price variation require keeping track of some measure of the index's fundamental value; in contrast, we show that one can place restrictions on the intertemporal behavior of RN beliefs without any knowledge of the asset's fundamental value, or knowledge of the marginal investor's underlying physical beliefs. Aside from the maintained assumptions of RE and no arbitrage, our main results require only one general restriction on the stochastic discount factor (SDF), the random variable that determines an asset's ex-ante price by discounting the asset's random future cash flows: we assume that the SDF realization does not depend on the path of unobservable state variables realized between a given trading date and the option expiration date. We refer to this assumption as *conditional transition independence (CTI)*, and this assumption is met in many common macro-finance models. Further, we provide sufficient conditions under which our bounds are robust to violations of the CTI assumption.

To demonstrate the logic of our test, it is useful first to consider the simple case in which we can directly observe a person's subjective belief $\pi_t(\theta = 1)$ over some binary outcome $\theta \in \{0, 1\}$. Suppose we observe that π_t continually moves from 0.10 to 0.90 as t progresses. While it is of course possible to rationalize this movement ex post by constructing a set of signal realizations from a particular signal data-generating process (DGP), this amount of movement appears intuitively "rare" and might bring into question the hypothesis that the agent has RE. Formalizing this intuition, Augenblick and Rabin (2021) note that, when uncertainty is resolved by some period T, the expectation of the sum of squared changes in beliefs across all periods (*belief movement*, $\mathbb{E}\sum_{t=0}^{T-1}(\pi_{t+1} - \pi_t)^2$) must equal the agent's *initial uncertainty* ($\pi_0(1 - \pi_0)$) under RE, *regardless* of the signal DGP. Equivalently, regardless of DGP, the expected *excess movement* — belief movement minus initial uncertainty, which we denote by $\mathbb{E}[X]$ — must always equal zero. Intuitively, under RE, changes in beliefs must on average correspond to the resolution of uncertainty (from its initial level $\pi_0(1 - \pi_0)$ to 0); rational belief movements must mean the person is learning something about the outcome in question.[2] If beliefs are instead continually shifting dramatically relative to initial uncertainty, the RE assumption can be statistically rejected at some confidence level.

The main theoretical contribution of this paper is to show how this logic — that movement in beliefs must correspond on average to reduction in uncertainty — can be used to restrict excess movement in *risk-neutral* beliefs, $\mathbb{E}[X^*]$, in the general case in which only risk-neutral rather than physical beliefs are observable. Our task becomes considerably more difficult in this case: observable risk-neutral beliefs need not follow a martingale under RE given their distortion relative to physical beliefs, and this means that $\mathbb{E}[X^*]$ can be non-zero.[3] But given that the distortion

[2] As a simple example, take a situation in which two coins are flipped sequentially and a person with RE states her beliefs about the likelihood that she will observe two heads. With probability $\frac{1}{4}$, the agent observes two heads, leading to beliefs that move from 25% to 50% to 100%, which has movement of $\frac{5}{16}$; she observes an initial tails with probability $\frac{1}{2}$, leading to belief movement of $\frac{1}{16}$; and she observes heads and then tails with probability $\frac{1}{4}$, leading to movement of $\frac{5}{16}$. As with any signal DGP with a prior of $\frac{1}{4}$, the expectation movement is equal to $\frac{3}{16}$, the initial uncertainty.

[3] Equivalently, some positive average amount of apparent excess movement in risk-neutral beliefs can be rationalized under RE, as long as the person (or market) exhibits some risk aversion over the time-T outcome.

between risk-neutral and physical beliefs is indexed by risk aversion over the terminal states, we show in this case that the admissible $\mathbb{E}[X^*]$ under RE can be bounded as a simple function of this risk-aversion value (or, more generally, the slope of the SDF across states). The broad intuition is the same: if a person's valuation for a binary option that pays \$1 in a given state at time T continually oscillates between \$.10 and \$.90, her unobservable risk aversion renders her precise physical beliefs unidentified, but these extreme valuation changes imply belief movements that must be rare under RE regardless of her exact risk aversion.

Our upper bound for excess RN belief movement $\mathbb{E}[X^*]$ is tight in the sense that it is possible to construct a (somewhat perverse) signal DGP that produces movement that is arbitrarily close to the bound. But this bound is, by construction, very conservative, and so we also provide a tighter bound that holds under somewhat more restrictive assumptions. In both cases, the bound depends on a risk-aversion parameter ϕ that corresponds to the relative marginal value of a dollar across time-T states. For example, if the person is forming beliefs over two possible terminal consumption states $C_T \in \{C_{low}, C_{high}\}$ (so that the option payoff is, e.g., $Y_T = 1$ if $C_T = C_{low}$), then $\phi = \frac{U'(C_{low})}{U'(C_{high})}$. In the empirically relevant case where we observe market-implied RN beliefs over the index return, then $\phi = \frac{\mathbb{E}[M_{t,T} \mid R_T = R_{low}]}{\mathbb{E}[M_{t,T} \mid R_T = R_{high}]}$, where $M_{t,T}$ is the stochastic discount factor and R_{low}, R_{high} are arbitrary index return values. Since our bound for $\mathbb{E}[X^*]$ depends on this parameter ϕ, any observed average excess movement in the data is informative as to the minimal ϕ required in order for the bound to be satisfied. We are effectively inverting the logic of the joint hypothesis problem: under the assumption of RE, the data is informative as to the minimal ϕ needed to rationalize the observed amount of excess movement in RN beliefs. We are, in other words, asking how restrictive it is — in particular, how much utility curvature one must assume — to maintain the hypothesis of RE.

Our main assumption throughout — the assumption of conditional transition independence (CTI) introduced above — is that ϕ is constant over time for a given belief stream, which generally lasts weeks to months in our empirical exercise. This assumption holds in many leading economic models: given a stable utility function, the marginal value of \$1 in some time-T state generally does not change relative to its marginal value in some other state.[4] Nonetheless, we provide additional results suggesting that reasonable variation in ϕ is unlikely to meaningfully affect our conclusions. This might seem counterintuitive: for example, one might imagine that if ϕ oscillates between 2.9 and 3.1 over and over, the movement of RN beliefs can be unbounded over time even given no movement in physical beliefs, strongly violating our results. This logic, however, ignores the fact that the person must also have rational expectations over ϕ, making such persistent oscillations rare under RE. To demonstrate this principle, we first theoretically prove that if expectations about ϕ evolve as a martingale or supermartingale, our bounds continue to hold. We then numerically simulate hundreds of DGPs in which both the underlying state and ϕ are uncertain, so that ϕ varies within a stream. We find that, across all DGPs, larger initial uncertainty in ϕ (leading to larger average movement in ϕ across the stream) has little to no impact on the average and maximum

[4] In asset-pricing terms, while permanent shocks to the SDF may alter $\mathbb{E}_t[M_{t,T}]$ for all states, they generally do not affect the expected *relative* SDF realizations across possible states.

$\mathbb{E}[X^*]$ statistics. Finally, we simulate a calibrated version of the habit-formation model of Campbell and Cochrane (1999), which implies a changing ϕ. We find that our bounds continue to hold and continue to be conservative.

We then take our bounds to the data using S&P 500 index option prices obtained from OptionMetrics. We use standard methods to infer the distribution of the market's RN beliefs over index returns for each option expiration date in the sample. In order to map to our two-state theoretical setting, we then translate each full distribution into a set of binary RN beliefs $\pi_t^*(R_T = \theta_j \mid R_T \in \{\theta_j, \theta_{j+1}\})$; these correspond to the RN probability that the index return will be equal to (or in a range close to) θ_j, conditional on being either θ_j or θ_{j+1}. (We set our return states to correspond to five-percentage-point bins for the S&P return.) We then implement our theoretical bounds, which allow us to infer the minimal risk-aversion value ϕ (at each point in the return distribution) needed to rationalize the observed variation in RN beliefs over the index return.

We find that extremely high risk aversion is needed to rationalize the observed excess movement: in many cases, there is in fact no value of ϕ under which the tight version of the bound is met, and the conservative version of the bound generally implies implausibly large values for ϕ. It is thus quite costly (in the sense that a high ϕ is required) to maintain the assumption of RE in light of the degree of excess movement in risk-neutral beliefs over the index return. This suggests that many leading rational frameworks capable of explaining medium-to-low-frequency variation in asset prices have difficulty rationalizing medium-to-high-frequency variation in beliefs.

Given that we conduct our estimation using variation in index options prices, we must also account for the effect of possible market microstructure noise. In order to do so, we derive an additional theoretical result that describes how microstructure noise biases our estimates through its effect on observed excess movement X^*. This bias depends on the variance of the noise component of observed RN beliefs. We accordingly proceed to estimate this noise variance in the data, by turning to a set of high-frequency index option prices (which we obtain directly from the CBOE) for this purpose. Using these intraday data, we implement the microstructure noise variance estimator proposed by Li and Linton (2021), which is particularly well suited for our purposes. We can then construct an empirical noise correction, removing the effect of noise from X^* before we conduct our estimation. All of our results are noise-corrected in this way, which help to ensure that our findings do not depend on idiosyncrasies specific to the options market.

We then briefly explore possible explanations for our findings, with the goal of providing some positive directions for alternative models. Conducting regressions of our excess-movement measure on a range of macroeconomic statistics, we find that X^* has a strong positive relationship with measures of macroeconomic uncertainty and no relationship with measures of liquidity or limits to arbitrage in asset markets. These results provide further evidence against the possibility that factors specific to the option market are the main drivers of our results, but this regression evidence is only suggestive and reduced-form.

Relation to previous literature and interpretation of results. In addition to Shiller (1981), we follow, among others, LeRoy and Porter (1981), De Bondt and Thaler (1985), Campbell and

Shiller (1987), and Stein (1989) in testing for excess volatility in asset prices relative to RE. Kleidon (1986) and Marsh and Merton (1986) emphasize non-stationarity in accounting for apparent excess volatility; much of the literature since then has emphasized time variation in discount rates (Cochrane, 2011). We show that even without imposing any restrictions on the structure of the data-generating process, and imposing only mild restrictions on the variation in discount rates, RE nonetheless restricts the admissible variation in option prices in an empirically testable way.

There are two costs associated with our additional generality. First, we consider derivative prices rather than the behavior of the underlying index directly. In this way, our work is complementary to that of Giglio and Kelly (2018), who document excess volatility in long-maturity claims on equity and currency volatility, inflation swaps, commodity futures, and credit default swaps. Their framework differs from ours in that they achieve identification by parameterizing the DGP for cash flows on the underlying, whereas we restrict the evolution of the SDF. Their parameterization — a low-dimensional affine model under the RN measure — applies well to the term-structure-like claims they consider, but not to claims on the equity index itself, to which our framework does apply. The two frameworks thus provide independent and complementary evidence for excess volatility, and both do so in a manner that allows for discount-rate variation.

Second, rather than allowing for fully binary (rejection vs. non-rejection) empirical tests of RE models, our general framework instead allows for a mapping between the observed asset-price variation and the risk aversion required to rationalize the data. Our results may thus appear similar in spirit to those of Mehra and Prescott (1985), and more generally Hansen and Jagannathan (1991), who find that the SDF must be highly volatile to rationalize the observed excess returns for risky assets. Our results differ from theirs in two respects. First, we obtain our mapping using the second moment of observed returns, while they use the first moment of returns.[5] More importantly, the Hansen–Jagannathan results may in principle be explained by features of the data-generating process for consumption or returns rather than high risk aversion per se; for example, models of rare disasters (e.g., Barro, 2006; Gabaix, 2012) can generate sufficient SDF volatility to rationalize the observed equity premium without requiring high risk aversion. But this is not the case for our results, as we obtain a relationship between local changes in the RN belief distribution and local risk aversion (or the slope of the SDF) at those points of the distribution. If we observe highly variable RN beliefs over the event that the S&P's 90-day return will be between 5% and 10%, we know that this cannot be attributable to disasters that affect the left tail of the return distribution; instead, we conclude either that risk aversion is very high or that there is a departure from RE.

Finally, our results complement evidence on beliefs obtained from survey data, as, for example, in Greenwood and Shleifer (2014), Bordalo et al. (2020), and De la O and Myers (2021), as well as the results of Augenblick and Rabin (2021) for settings with directly observable beliefs. Another set of related literature endeavors to measure physical beliefs indirectly from options data, for the purpose of examining either expectations or preferences; see Garcia, Ghysels, and Renault (2010)

[5] One could instead map between variation in returns and the volatility of SDF volatility (or the heteroskedasticity of the SDF), but we consider the results from our mapping to be somewhat more intuitive than this alternative.

for a survey. As a recent example, Ross (2015) assumes a Markov process for transitions between return states and a transition-independence assumption on the SDF more restrictive than the one we use (see Borovička, Hansen, and Scheinkman, 2016, for a discussion), with which one can back out a distribution of physical beliefs (see also D'Arienzo, 2020). Our approach differs from this work in that we need not measure physical beliefs at all or know the true data-generating process for returns to conduct our tests, so we accordingly require less structure. We also complement the findings of van Binsbergen, Brandt, and Koijen (2012), who, like us, provide evidence of excess volatility in short-maturity equity derivatives claims.

Organization. Section 2 introduces our theoretical framework in a simple two-state setting, which allows for clear derivations and intuition for our main results. Section 3 then extends these results to a more general asset-pricing setting, and Section 4 provides additional results speaking to the robustness and economic interpretation of our bounds. We then implement our bounds empirically in Section 5, which describes our data and presents our results. Section 6 concludes. While we provide derivations for many of our theoretical results in the text, our online appendix contains detailed proofs (Appendix A) as well as additional technical material (Appendix B).

2. Theoretical Results in a Simple Setting

We first examine risk-neutral (RN) belief movement in a simplified environment with a single individual and two terminal consumption states. This simple setting allows for transparent derivations of our main results and discussion of their economic intuition, before turning to the more general asset-pricing framework considered in Section 3. We begin by describing our setup and reviewing previous results concerning physical belief movement (Section 2.1), then discuss RN beliefs and the complications introduced by risk aversion (Section 2.2), and finally provide simple versions of our main theoretical results on RN belief movement (Section 2.3).

2.1 Bayesian Belief Movement: Setup and Results

Before considering asset prices and RN beliefs, we first consider the simplified case in which a person's beliefs are directly observable. The setup and results in this subsection build on Augenblick and Rabin (AR, 2021).

Time is discrete and indexed by $t \in \{0, 1, 2, \ldots, T\}$. At the beginning of each period, a person observes a signal $s_t \in S$ regarding two mutually exclusive and exhaustive events, which we call *states* $\theta \in \{0, 1\}$. The data-generating process (DGP) is general: signals are drawn from the discrete signal distribution $DGP(s_t \mid \theta, H_{t-1})$, where H_t represents the history of signal realizations through t. Define $\mathbb{P}(H_T)$ to be the probability of observing history H_T induced by the DGP, and write $\mathbb{E}[\cdot] \equiv \mathbb{E}^{\mathbb{P}}[\cdot]$ for the expectation under $\mathbb{P}$. The person's (*physical* or *subjective*) *belief* in state 1 (vs. state 0) at time t given the DGP and history H_t is denoted by $\pi_t(H_t)$. The *belief stream* $\boldsymbol{\pi}(H_t) = [\pi_0, \pi_1(s_1), \pi_2(\{s_1, s_2\}), \ldots]$ is the collection of beliefs given history H_t. We will commonly suppress the dependence of these objects on H_t to simplify notation. Given our empirical

setting, we focus entirely on *resolving* streams in which the person achieves certainty about the true state by period T with probability 1: $\pi_T(H_T) = \theta \in \{0, 1\}$ for all H_T.[6]

Total *belief movement* of $\boldsymbol{\pi}$ is defined as the sum of squared changes in beliefs across all periods:

$$\mathsf{m}(\boldsymbol{\pi}) \equiv \sum_{t=0}^{T-1} (\pi_{t+1} - \pi_t)^2. \tag{1}$$

Initial uncertainty of $\boldsymbol{\pi}$ is defined as the time-0 variance of the Bernoulli random variable $\mathbb{1}\{\theta = 1\}$:

$$\mathsf{u}_0(\boldsymbol{\pi}) \equiv (1 - \pi_0)\pi_0. \tag{2}$$

Given that we focus on resolving belief streams for which $\pi_T \in \{0, 1\}$, final uncertainty is always zero ($\mathsf{u}_T = (1 - \pi_T)\pi_T = 0$ a.s.). Initial uncertainty u_0 is therefore equal to the total amount of uncertainty reduction for $\boldsymbol{\pi}$, $\mathsf{u}_0 - \mathsf{u}_T$, which is helpful in interpreting some of the following results.

Our main variable of interest is the difference between movement and initial uncertainty, which — for reasons that will become clear — we call *excess movement*:

$$X(\boldsymbol{\pi}) \equiv \mathsf{m}(\boldsymbol{\pi}) - \mathsf{u}_0(\boldsymbol{\pi}). \tag{3}$$

Throughout the paper, we maintain the assumption that the person's beliefs are a martingale with respect to the true DGP:

ASSUMPTION 1 (*Martingale Beliefs*)**.** Beliefs satisfy $\pi_t(H_t) = \mathbb{E}[\pi_{t+1} \mid H_t]$ for any H_t.

A person with rational expectations (RE) over θ — equivalently, someone with a correct prior who updates using Bayes' rule according to the true DGP — will satisfy this assumption. Given our focus on resolving streams for which $\pi_T = \theta$, this assumption is in fact equivalent to RE, as $\pi_t = \mathbb{E}[\pi_T \mid H_t] = \mathbb{E}[\theta \mid H_t] = \mathbb{P}(\theta = 1 \mid H_t)$.[7] We thus often refer to Assumption 1 as RE.

To clarify our setting, Table 1 lists some examples of possible signals, belief streams, likelihoods of those belief streams, movement, and uncertainty resolution given three simple binary-signal DGPs that we will return to in the subsequent sections. (Only the first five columns of the table are relevant for now, as the remaining three refer to risk-neutral beliefs as defined and discussed below.) For example, in the first DGP, the person has a rational prior of $\pi_0 = .25$; at $t = 1$, she observes $s_1 \in \{l, h\}$ with symmetric relative likelihood ratios (or *precision*) for each of the two signals, $L_h \equiv \frac{\mathbb{P}[s_t = h \mid \theta = 1]}{\mathbb{P}[s_t = h \mid \theta = 0]} = 3$ and $L_l \equiv \frac{\mathbb{P}[s_t = l \mid \theta = 0]}{\mathbb{P}[s_t = l \mid \theta = 1]} = 3$; and then at $t = 2$, $s_2 \in \{L, H\}$ fully reveals the state (with signal "H" revealing state 1 and vice versa). In the first row, if the person first observes $s_1 = l$ and then $s_2 = L$, she will first update to $\pi_1 = .1$ and then $\pi_2 = 0$. The ex ante likelihood of this signal realization for this DGP is $\frac{9}{16} = .5625$ and the resultant belief stream $[.25, .1, 0]$ has movement of $(.25 - .1)^2 + (.1 - 0)^2 = .0325$ and initial uncertainty of $\pi_0(1 - \pi_0) = .1875$. The

[6] As discussed in AR, a number of statements can also be made about non-resolving (sub)streams. In our setting, these statements would often be excessively conservative and add little benefit to our empirical analysis.

[7] While one can construct cases for which beliefs violate RE but are a martingale under $\mathbb{P}$, such cases are degenerate (e.g., $\pi_t = 1$ for all H_t) and ruled out by $\pi_T = \theta$. But for now we assume RE only over θ, which is weaker than a full rationality assumption, as beliefs over individual histories H_T could still differ from $\mathbb{P}(H_T)$.

second example uses an asymmetric DGP in which the person either sees a high signal and learns with certainty that state 1 is true or sees a low signal and learns that state 0 is more likely. As a result, movements towards a belief of 1 are much larger than movements toward 0. The third example is the opposite with a resolving low signal and noisy high signal.

Under the RE assumption, belief movement and initial uncertainty are related according to the following proposition, which restates a main result in AR.

PROPOSITION 1 (Augenblick and Rabin, 2021)**.** *Under Assumption 1, for any DGP, expected total belief movement must equal initial uncertainty. Expected excess movement in beliefs must therefore be zero:*

$$\mathbb{E}[X] = 0. \tag{4}$$

As discussed in AR, Proposition 1 is a straightforward implication of the assumption of martingale beliefs.[8] It motivates referring to X as *excess movement*. As an example, consider the first DGP in Table 1: the four signal realizations $[l, L]$, $[l, H]$, $[h, L]$, $[h, H]$ have likelihood .5625, .0625, .1875, .1875, respectively, with respective movement statistics .0325, .8325, .3125, and .3125. Therefore, expected movement is .1875, which equals initial uncertainty, so expected excess movement is zero. Note that the other two DGPs in the table also start with a prior of .25 and initial uncertainty of .1875. Proposition 1 therefore implies that, although these DGPs produce different streams with different movement statistics and likelihoods, the expected movement will be .1875 and expected excess movement will be zero.

AR then use this relationship to create simple tests of RE if one can directly observe beliefs. Given a set of observed belief streams, one can straightforwardly calculate the sample average of the empirical excess movement statistic and statistically test if it differs from zero. The restriction reflects the intuition that if the person's beliefs are moving, this movement must on average correspond to learning about the true terminal state (in the sense that uncertainty is resolved from its initial value to 0). Rewriting $\mathbb{E}[X] = 0$ as $\mathbb{E}[\sum_{t=0}^{T-1}(1-2\pi_t)(\pi_{t+1}-\pi_t)] = 0$ (see footnote 8), it is apparent that expected belief movements toward 0.5 (the point of highest uncertainty) lead to a positive $\mathbb{E}[X]$ statistic, and vice versa. So it could be the case that $\mathbb{E}[\pi_{t+1} - \pi_t] = 0$ unconditionally, but a test based on Proposition 1 would still reject the null of RE if, for instance, low values of π_t (i.e., $1 - 2\pi_t > 0$, or $\pi_t < 0.5$) tend to be revised upward ($\pi_{t+1} - \pi_t > 0$), and high values tend to be revised downward.

2.2 Risk-Neutral Beliefs: Setup

Proposition 1 shows that one can make statements about excess belief movement under RE when beliefs are observable. We now consider the situation in which beliefs are not directly observable, but it is possible to observe the stream of the person's willingness to pay for an Arrow-Debreu

[8] To see this, rewrite X as $\sum_{t=0}^{T-1}(2\pi_t - 1)(\pi_t - \pi_{t+1})$. Using the law of iterated expectations on each term in the sum, $\mathbb{E}[(2\pi_t - 1)(\pi_t - \pi_{t+1})] = \mathbb{E}[(2\pi_t - 1)(\pi_t - \mathbb{E}[\pi_{t+1}|\pi_t])]$, which must be zero under the martingale assumption. This result has appeared in other forms in past literature; for one example, see Barndorff-Nielsen and Shephard (2001).

security that pays \$1 (1 unit of the numeraire consumption good) in period T if state θ is realized. We denote this valuation by $q_t(\theta|H_t)$ for each $\theta \in \{0,1\}$. An object analogous to $q_t(\theta|H_t)$ will be observable using options data for suitably defined states, but we postpone this additional formalism to Section 3 so as to maintain focus on the main economic intuition for now. Our main contribution will be to develop a semi-parametric test of RE using only data on valuation streams.

To start, consider the simple case in which the person values consumption at all periods and in all states equally (i.e., she is risk-neutral and does not discount future consumption). In this case, $q_t(1|H_t) = \pi(H_t)$ and $q_t(0|H_t) = 1 - \pi(H_t)$. Beliefs are thus directly observable from asset values, and we can apply Proposition 1 as a test of the rationality of these beliefs.

The problem of identifying excess movement in asset valuations becomes significantly more complicated when the person does not weight consumption equally across time and states, as valuations no longer correspond directly to beliefs. To see this, assume now that the agent has time-separable utility, with concave period utility function $U(C_t)$, and exponentially discounts future consumption with discount factor β. Assume that θ determines period-T consumption $C_{T,\theta} \equiv C_T(\theta)$. At an interior optimum, valuations are

$$q_t(1|H_t) = \frac{\beta^{T-t}U'(C_{T,1})}{U'(C_t)}\pi_t, \qquad q_t(0|H_t) = \frac{\beta^{T-t}U'(C_{T,0})}{U'(C_t)}(1-\pi_t). \tag{5}$$

Valuations thus no longer reveal beliefs.

Following asset-pricing convention, define the *stochastic discount factor* (SDF) $M_{t,T}(\theta)$ as the ratio of valuation to probability for state θ:

$$M_{t,T}(1) \equiv \frac{q_t(1|H_t)}{\pi_t(H_t)}, \qquad M_{t,T}(0) \equiv \frac{q_t(0|H_t)}{1-\pi_t(H_t)}.$$

Continuing from (5), $M_{t,T}(\theta) = \frac{\beta^{T-t}U'(C_{T,\theta})}{U'(C_t)}$ encodes the relative valuation of a marginal dollar in state θ in period T versus one in period t. In our environment, the *SDF ratio* ϕ_t will be particularly important:

$$\phi_t \equiv \frac{M_{t,T}(1)}{M_{t,T}(0)}, \tag{6}$$

which encodes the slope of the stochastic discount factor across the two states. In consumption terms, continuing the above example, we have

$$\phi_t = \frac{\frac{\beta^{T-t}U'(C_{T,1})}{U'(C_t)}}{\frac{\beta^{T-t}U'(C_{T,0})}{U'(C_t)}} = \frac{U'(C_{T,1})}{U'(C_{T,0})}, \tag{7}$$

so ϕ_t can be thought of as the marginal rate of substitution across the states. Assume that $\theta = 1$ corresponds to the low-consumption state relative to $\theta = 0$, or $C_{T,1} < C_{T,0}$; this is without loss of generality, as the states can be relabeled arbitrarily. Given this labeling, with concave utility, we have $U'(C_{T,1}) \geqslant U'(C_{T,0})$ and thus $\phi_t \geqslant 1$. We will maintain the convention of labeling state 1

(corresponding to the belief π_t) as the "bad" state, so we state it as an assumption:

ASSUMPTION 2 (*Labeling of States*)**.** $\phi_t \geqslant 1$ with probability 1 for all t.

In (7), given that θ determines period-T consumption, ϕ_t is in fact constant across t. The person's beliefs may be changing about the relative likelihood of the two states, but her beliefs about the relative marginal utility outcomes conditional on each of the two states being realized are not. This assumption is central to our results:

ASSUMPTION 3 (*Constant SDF Ratio*)**.** $\phi_t = \phi$ is constant with probability 1 for all t.

We discuss this assumption in more detail, for general settings, in Section 3. Intuitively, what is required is that the "severity" of the bad state relative to the good state is constant, so that the relative valuation of a marginal dollar in the two states is constant. This follows naturally in the current example from the assumption of stable time-separable utility and fixed state-contingent consumption values. This illustrates the manner in which restrictions on risk-neutral belief variation require weaker assumptions than restrictions on the underlying asset price: an Arrow-Debreu state price (and associated RN belief) depends on marginal utility in a single state, whereas the price of a consumption claim depends on the probability-weighted sum of marginal utilities over *all* states. Assuming constant discount rates allows for identification in the latter context (e.g., Shiller, 1981), but we need not make this assumption when working with RN beliefs. RN beliefs also eliminate dependence on time-t marginal utility by means of a convenient normalization. Appendix B.1 (in our online appendix) discusses the relationship between RN beliefs and discount rates in greater detail. We also later consider how our bounds are affected when ϕ_t is time-varying; this robustness discussion is postponed to Section 4.3.

For further intuition on ϕ, we can conduct a Taylor expansion of $U'(C_{T,0})$ around $C_{T,1}$ as $U'(C_{T,0}) = U'(C_{T,1}) + U''(C_{T,1})(C_{T,0} - C_{T,1}) + \mathcal{O}\big((C_{T,0} - C_{T,1})^2\big)$. Then rearranging and using (7), to first order,

$$\gamma(C_{T,1}) \equiv -\frac{C_{T,1}U''(C_{T,1})}{U'(C_{T,1})} = \frac{\phi - 1}{(C_{T,0} - C_{T,1})/C_{T,1}}. \tag{8}$$

Relative risk aversion γ depends on the ratio of marginal utilities across states ϕ relative to the percent consumption gap across states. Thus up to a scaling constant $(C_{T,0} - C_{T,1})/C_{T,1}$, the value ϕ can be thought of as an index of risk aversion.

Given that beliefs are not directly inferable from asset valuations without knowledge of ϕ, one cannot in general apply Proposition 1 in the current context. We instead aim to derive bounds on excess movement in *risk-neutral (RN)* beliefs, which are defined (following convention) as

$$\begin{aligned}\pi_t^*(H_t) &\equiv \frac{q_t(1|H_t)}{q_t(0|H_t) + q_t(1|H_t)} \\ &= \frac{U'(C_{T,1})}{\mathbb{E}_t[U'(C_T)]}\pi_t(H_t) = \frac{\phi\pi_t(H_t)}{1 + (\phi - 1)\pi_t(H_t)},\end{aligned} \tag{9}$$

with the second line following from (5) and (7). RN beliefs for $\theta = 0$ can be similarly defined as $\frac{q_t(0|H_t)}{q_t(0|H_t)+q_t(1|H_t)} = 1 - \pi_t^*$, so the two states' RN beliefs are positive and sum to 1 by construction. Risk-neutral beliefs are so named because they can be interpreted as the subjective beliefs for a fictitious risk-neutral agent, as they coincide with actual subjective beliefs in the case that the person is risk-neutral. In general, they represent a pseudo-belief distribution, reflecting a combination of beliefs and risk preferences. Given $\phi \geqslant 1$, the bad-state RN belief π_t^* in general exceeds the person's subjective belief π_t: the person is willing to pay more than the actuarially fair value for a bad-state consumption claim given her high marginal utility in that state, and vice versa.

Returning to Table 1, the sixth column provides RN beliefs for the three example DGPs under the calibration $\phi = 3$. The person has a prior of .25 in each DGP, so using (9), the RN prior is $\pi_0^* = .5$. Intuitively, the person perceives state 0 as three times as likely as state 1, but values a marginal dollar in state 1 three times as much as in state 0, so $q_t(0|H_t) = q_t(1|H_t)$ and $\pi_0^* = .5$.

One can also invert (9) to solve for π_t as a function of π_t^* and ϕ, the solution to which we denote by $\pi_t(\pi_t^*, \phi)$:

$$\pi_t(\pi_t^*, \phi) = \frac{\pi_t^*}{\phi + (1-\phi)\pi_t^*}. \tag{10}$$

While π_t is a time-varying unobservable, (10) clarifies that it can be represented as a function of the observable π_t^* and a single unknown parameter ϕ.

We define the RN belief stream $\boldsymbol{\pi}^*(H_t)$, RN belief movement $\mathrm{m}^*(\boldsymbol{\pi}^*)$, RN initial uncertainty $\mathrm{u}_0^*(\boldsymbol{\pi}^*)$, and RN excess movement $X^*(\boldsymbol{\pi}^*)$ as in Section 2.1, but with RN beliefs π_t^* in the place of physical beliefs π_t. It will also be useful to define the *RN measure* as

$$\mathbb{P}^*(H_T) \equiv \begin{cases} \mathbb{P}(H_T)\frac{\pi_0^*}{\pi_0} & \text{if } \pi_T(H_T) = 1 \\ \mathbb{P}(H_T)\frac{1-\pi_0^*}{1-\pi_0} & \text{if } \pi_T(H_T) = 0, \end{cases} \tag{11}$$

where $\mathbb{P}(H_T)$ is the probability of observing history H_T under *DGP*. We show in Section 3 that (11) follows from the usual definition of the RN measure in a general asset-pricing setting (see Lemma A.1 in Appendix A). For now, it suffices to think of $\mathbb{P}^*$ as representing the change of measure that adjusts the frequency of each path of signal realizations such that a person with RN beliefs has rational expectations: defining $\mathbb{E}^*[\cdot]$ to be the expectation under $\mathbb{P}^*$, we have from (11) that

$$\pi_t^*(H_t) = \mathbb{E}^*[\pi_{t+1}^*(H_{t+1})|H_t]. \tag{12}$$

RN beliefs are not well-calibrated under the physical measure; for instance, in the first example in Table 1, the person has a RN prior of .5 but ends up with a belief of 1 only 25% of the time. Instead, the RN measure adjusts the weights on signal realizations to .375, .125, .125, and .375, respectively, so that the person ends up with a belief of 1 with 50% RN probability and is well-calibrated.

Given (12), the expectation of movement of RN beliefs under the RN measure must equal RN

initial uncertainty and therefore expected excess RN movement must be zero:

$$\mathbb{E}^*[X^*] = 0. \tag{13}$$

We wish to make statements similar to Proposition 1, but applicable to the observable asset values. The fundamental challenge in doing so is twofold. First, asset values allow for the calculation of RN beliefs rather than physical beliefs; that is, we observe π_t^* rather than π_t, so we cannot construct $\mathbb{E}[X]$ in equation (4). But second, the frequency of the observed RN belief streams used to construct expected excess movement is determined by the physical measure rather than the RN measure; that is, we cannot directly observe an empirical counterpart to $\mathbb{E}^*[X^*]$ in equation (13).

Instead, we can observe a counterpart to $\mathbb{E}[X^*]$. But the distortion in RN relative to subjective beliefs can cause RN movement to differ from RN initial uncertainty on average, so $\mathbb{E}^*[X^*] \neq 0$. Thus even with rational physical beliefs, one can observe, for example, what appears to be excess movement in RN beliefs implied by valuations. (The third example in Table 1, discussed further below, is one such case.) So if we naïvely test for RE using Proposition 1 on observed RN (rather than actual) beliefs, we may spuriously conclude that beliefs are excessively volatile.

2.3 Risk-Neutral Beliefs: Results

We now proceed to derive theoretical bounds for $\mathbb{E}[X^*]$ given a particular value of the unobservable parameter ϕ. For any observed value for $\mathbb{E}[X^*]$, we can then place restrictions on the admissible values of ϕ under the null of RE by inverting the bounds. We sketch derivations for our main results in the text, and full formal proofs for all results are provided in Appendix A.

Main Results

To understand the object $\mathbb{E}[X^*]$, we start with a key observation using (11):

$$\mathbb{P}^*(H_T|\theta) = \mathbb{P}(H_T|\theta). \tag{14}$$

Compared to the physical measure, the RN measure places higher likelihood of all signal histories resolving in state 1, but does so proportionally, so that likelihoods of signal histories *conditional on state 1* do not change.[9] While this is implied directly by the definition (11), it again applies under the usual definition of the RN measure in a general setting (see Lemma A.2 in Appendix A).

Equation (14) implies that conditional expectations under the two respective measures are equal, $\mathbb{E}^*[X^*|\theta] = \mathbb{E}[X^*|\theta]$, which implies

$$\begin{aligned}\mathbb{E}^*[X^*] &= \pi_0^* \cdot \mathbb{E}^*[X^*|\theta = 1] + (1 - \pi_0^*) \cdot \mathbb{E}^*[X^*|\theta = 0] \\ &= \pi_0^* \cdot \mathbb{E}[X^*|\theta = 1] + (1 - \pi_0^*) \cdot \mathbb{E}[X^*|\theta = 0] = 0,\end{aligned} \tag{15}$$

[9] For example, in the first DGP of Table 1, the signal histories $[l, H]$ and $[h, H]$ that imply state 1 occur in the physical measure with probability .0625 and .1875 respectively, so that $\mathbb{P}([l, H]|\theta = 1) = .25$. In the RN measure, these are increased by a factor of $\pi_0^*/\pi_0 = 2$ to .125 and .375, respectively. Consequently, $\mathbb{P}^*([l, H]|\theta = 1)$ remains .25.

where the last equality applies (13). For $\mathbb{E}[X^*]$, it is useful to separate this expectation as in (15):

$$\mathbb{E}[X^*] = \pi_0 \cdot \mathbb{E}[X^*|\theta = 1] + (1 - \pi_0) \cdot \mathbb{E}[X^*|\theta = 0]. \tag{16}$$

Finally, it will be helpful to define the difference in conditional expected X^* as

$$\triangle \equiv \mathbb{E}[X^*|\theta = 0] - \mathbb{E}[X^*|\theta = 1]. \tag{17}$$

As movement must always be positive, using equation (15),[10] we have that

$$\triangle \leqslant \pi_0^*. \tag{18}$$

Combining (15) and (16), one can express $\mathbb{E}[X^*]$ as follows. This expression in turn generates a set of bounds relating objects from observable asset valuations to the unobservable structural parameter ϕ. We assume throughout that Assumptions 1–3 hold.

PROPOSITION 2. *For any DGP,*

$$\begin{aligned}\mathbb{E}[X^*] &= (\pi_0^* - \pi_0)(\triangle) \\ &= \left(\pi_0^* - \frac{\pi_0^*}{\phi + (1-\phi)\pi_0^*}\right)(\mathbb{E}[X^*|\theta = 0] - \mathbb{E}[X^*|\theta = 1]).\end{aligned}$$

PROPOSITION 3. *For any DGP and any value for $\triangle$,*

$$\mathbb{E}[X^*] \leqslant (\pi_0^* - \pi_0)\pi_0^*.$$

Equivalently,

$$\mathbb{E}[X^*] \leqslant \left(1 - \frac{1}{\phi + (1-\phi)\pi_0^*}\right)\pi_0^{*2}. \tag{19}$$

COROLLARY 1. *For any DGP and any values $\triangle$ and ϕ,*

$$\mathbb{E}[X^*] \leqslant \pi_0^{*2}.$$

Additionally, if one is willing to make an assumption on the sign of $\triangle$ (discussed shortly), the following stronger bound applies.

COROLLARY 2. *If $\mathbb{E}[X^*|\theta = 0] \leqslant \mathbb{E}[X^*|\theta = 1]$, for any DGP and any value for ϕ,*

$$\mathbb{E}[X^*] \leqslant 0.$$

Proposition 2 shows that the expectation of RN excess movement $\mathbb{E}[X^*]$ is equal to the product

[10] Equation (15) can be rewritten as $\pi_0^* \cdot \mathbb{E}[m^*|\theta = 1] + (1 - \pi_0^*) \cdot \mathbb{E}[m^*|\theta = 0] = \pi_0^*(1 - \pi_0^*)$. Thus, fixing π_0^*, $\mathbb{E}[m^*|\theta = 0]$ decreases monotonically in $\mathbb{E}[m^*|\theta = 1]$. Given that $\mathbb{E}[m^*|\theta = 1] \geqslant 0$, $\mathbb{E}[m^*|\theta = 0]$ must be bounded above by π_0^*. Therefore, $\triangle$, which can be rewritten as $\mathbb{E}[m^*|\theta = 0] - \mathbb{E}[m^*|\theta = 1]$, must be bounded above by π_0^*.

of $\pi_0^* - \pi_0$ and $\triangle \equiv \mathbb{E}[X^*|\theta = 0] - \mathbb{E}[X^*|\theta = 1]$. Note that if $\phi = 1$, then $\pi_0^* = \pi_0$ and therefore $\mathbb{E}[X^*] = 0$. As ϕ rises, π_0 drops further below π_0^* and $\mathbb{E}[X^*]$ deviates from 0: the greater is risk aversion, the more one can observe excess RN movement differ from zero on average under RE. The size of this deviation, and whether it is positive or negative, depends on $\triangle$. In other words, all DGPs with the same $\triangle$ will produce the same level of $\mathbb{E}[X^*]$ for a given ϕ.

Regardless of the signal distribution, $\triangle$ is bounded above as in (18). Plugging this into Proposition 2 yields Proposition 3. The version of the bound expressed in (19) is one of our main results. It gives a bound for $\mathbb{E}[X^*]$ as a function of π_0^* and ϕ, regardless of $\triangle$. Equivalently, it relates the unobserved structural parameter ϕ, which corresponds to the slope of the SDF across the two states as in (6), to a set of observable values. Under risk neutrality ($\phi = 1$), this upper bound becomes zero, following Proposition 1. But this bound is otherwise positive, and the *admissible excess movement* in RN beliefs given by the right side of the inequality increases monotonically in ϕ. Movement in RN beliefs must still correspond on average to the agent learning something about the true terminal state, but the bias in RN beliefs relative to subjective beliefs induced by risk aversion allows for positive excess movement in those observed beliefs under RE. This result thus formalizes a more general notion of the admissible amount of belief volatility under rationality, this time as an increasing function of risk aversion across the two states.

The bound in Proposition 3 is very conservative, as it implicitly holds under a "worst-case" DGP with an extreme value for $\triangle$. Higher RN priors yield more "room" for downward belief movement, so the worst case $\triangle = \pi_0^*$ rises monotonically. For $\phi > 1$, this generates asymmetry in the bound around $\pi_0^* = .5$, as lower values of π_0^* generally yield a lower upper bound for $\mathbb{E}[X^*]$.[11] But the bound in (19) is also non-monotonic in π_0^* for $1 < \phi < \infty$ (in particular for $\pi_0^* > 0.5$): for a given ϕ, the first term $\pi_0^* - \pi_0$ does not rise monotonically, exerting a countervailing force relative to the monotonic increase in the second term π_0^*.[12]

While the bound in Proposition 3 maps between observed values and the slope of the SDF required to rationalize those values, the fact that RN beliefs are bounded between 0 and 1 by construction implies that the bound is well-defined even for infinitely large risk aversion: there is only so far that RN beliefs can be distorted relative to subjective beliefs. Taking $\phi \to \infty$ in (19) yields Corollary 1, which provides our most conservative bound for $\mathbb{E}[X^*]$. In the limit as $\phi \to \infty$, $\pi_0 \to 0$ for any π_0^*, so $\pi_0^* - \pi_0 \to \pi_0^*$ and the bound in Proposition 3 approaches ${\pi_0^*}^2$. That is, $\mathbb{E}[X^*] > {\pi_0^*}^2$ simply cannot be rationalized under RE given constant ϕ. For low π_0^*, this bound again does not allow large positive $\mathbb{E}[X^*]$. For example, if $\pi_0 = .2$, RN excess movement can only rise to .04, regardless of ϕ or the DGP. That is, even arbitrarily high risk aversion does not allow for large amounts of excess movement in RN beliefs implied by asset valuations.

Taken together, Proposition 3 and Corollary 1 characterize the maximal admissible excess movement in RN beliefs as a function of ϕ for any RN prior. Figure 1 provides a graphical illustration of these bounds. Starting from the bottom of the chart, the thick purple line corresponds

[11] For example, with $\pi_0^* = .2$ and $\phi = 2$, (19) gives $\mathbb{E}[X^*] \leqslant .018$. This implies the maximal expected RN movement, $\max \mathbb{E}[\mathsf{m}^*] = 0.178$, is barely above $\mathbb{E}[\mathsf{m}] = 0.16$ for physical beliefs in the analogous case with $\pi_0 = .2$.

[12] For example, take $\phi = 3$. If $\pi_0^* = .5$, then $\pi_0 = .25$ and $\pi_0^* - \pi_0 = .25$. But when $\pi_0^* = .99$, $\pi_0^* - \pi_0 = .02$.

to the bound for $\phi = 1$: in this case, $\mathbb{E}[X^*] = \mathbb{E}[X] = 0$ regardless of the prior or DGP, from Proposition 1. The thin dashed gray lines correspond to arbitrarily selected DGPs in the case of $\phi = 3$. While there can be positive RN excess movement, this is not necessarily the case for all possible DGPs. Taking the envelope over all of these processes for $\phi = 3$ yields the bound from Proposition 3, which is shown in the thick blue line. It is asymmetric around 0.5 as well as non-monotonic, as described above. Finally, the thick red line shows the bound for the limiting case $\phi \to \infty$, which is equal to the squared RN prior from Corollary 1.

Finally, when $\triangle < 0$, $\mathbb{E}[X^*]$ is decreasing in ϕ and therefore $\mathbb{E}[X^*] < 0$ for any $\phi > 1$. Consequently, as formalized in Corollary 2, the highest excess movement is $\mathbb{E}[X^*] = 0$.

The above conclusions can also be seen in the examples in Table 1. For the first example, the RN prior is .5 and the signals are symmetric, so that $\triangle = 0$ and $\mathbb{E}[X^*] = 0$. For the second example, the DGP is asymmetric with large movements upwards, so that expected movement conditional on $\theta = 1$ is large (.405) compared to $\theta = 0$ (.095), so that $\triangle$ is negative (-.31) and therefore $\mathbb{E}[X^*]$ is negative. The converse occurs in the third example, which is asymmetric in the opposite way and therefore leads to $\triangle = -.31$. These examples point to the relationship of π_0^* and the DGP with the sign of $\triangle$. We now study this relationship in more detail.

What Determines $\triangle$?

Proposition 2 tells us that the deviation of $\mathbb{E}[X^*]$ from 0 depends on the product of $\pi_0^* - \pi_0$ and $\triangle \equiv \mathbb{E}[X^*|\theta = 0] - \mathbb{E}[X^*|\theta = 1]$. The difference $\pi_0^* - \pi_0$ is always positive (strictly positive, for $\phi > 1$) and increases in ϕ. But how is the sign and magnitude of $\triangle$ related to the RN prior and DGP? To answer this question, we provide two theoretical results and briefly summarize a set of numerical simulations discussed in detail in Appendix B.2.

First, given the arbitrary labeling of the two states, there is no reason to expect under RE that $\triangle$ should take a particular sign:

PROPOSITION 4. *Fixing ϕ, for every RN prior and DGP that leads to a given $\triangle$, there exists a different RN prior and DGP that leads to $-\triangle$.*

Intuitively, for any π_0^* and DGP with some $\triangle$, the RN prior $1 - \pi_0^*$ with the "reversed" DGP will necessarily lead to $-\triangle$. Consequently, there is no reason to assume that $\mathbb{E}[X^*]$ is more likely than not to be positive given $\phi > 1$ under RE.

Next, we summarize a set of numerical simulations for a large family of DGPs as described in Appendix B.2. The results there suggest some intuitive comparative statics. First, when the DGP changes such that downward movements become larger, expected RN movement conditional on state 0 rises, so that $\triangle$ rises and $\mathbb{E}[X^*]$ rises. Second, when upward movements become larger, the opposite occurs and $\mathbb{E}[X^*]$ falls. Third, as π_0^* rises, $\triangle$ rises. Intuitively, a higher π_0^* allows more movement potential toward state 0 (and less toward state 1), so that expected movement given state 0 rises (and movement given state 1 falls), leading to a higher $\triangle$. For example, when the DGP is symmetric, $\triangle < 0$ if $\pi_0^* < .5$, $\triangle = 0$ if $\pi_0^* = .5$, and $\triangle > 0$ if $\pi_0^* > .5$. Combining

these statements suggests that, for example, low π_0^* should lead to a negative $\triangle$ unless the DGP is extremely asymmetric.[13] And as we show later, the empirical DGPs in our setting appear largely symmetric, so that we estimate $\triangle < 0$ for $\pi_0^* < .5$.

These results also suggest that extreme values of $\triangle$ only occur in highly asymmetric DGPs where movements in one direction are large and movements in the other direction are tiny. In fact, our upper bound is attainable (in the limit) given the most asymmetric DGP possible:

PROPOSITION 5. *There exists a sequence of DGPs, indexed by T, for which $\mathbb{E}[X^*]$ approaches the bound in Proposition 3 as $T \to \infty$. For each DGP in this sequence, downward movements ($\pi_{t+1}^* < \pi_t^*$) are resolving ($\pi_{t+1}^* = 0$) and thus as large as possible, while upward movements are small ($\pi_{t+1}^* - \pi_t^* \to 0$ as $T \to \infty$). Meanwhile, the bound holds with strict inequality for any $T < \infty$ as long as $\phi > 1$ and $\pi_0^* \in (0,1)$.*

One implication of this result is that the bound in Proposition 3 is approximately tight, as one can construct a DGP for which $\mathbb{E}[X^*]$ is close to the bound for large T. Perhaps more important, though, is that it points to the bound's conservatism: it holds under a somewhat perverse DGP that can be thought of as a "rare bonanzas" process, where with small probability the person receives news that the bad state ($\theta = 1$) will not be realized (so $\pi_{t+1}^* = 0$), and otherwise there is mostly uninformative bad news that increases π_{t+1}^* slightly. More reasonable DGPs, or $T \ll \infty$, will give lower $\mathbb{E}[X^*]$. That said, the conservative bound has the advantage of being very simple and not requiring any estimation of $\triangle$. And as we show below, empirical excess movement is in fact so high that even these conservative bounds are often violated for reasonable values of ϕ.

What Generates Bound Violations?

We have assumed throughout that Assumption 1 holds, which in general requires both a correct subjective prior $\pi_0 = \mathbb{P}_0(\theta = 1)$ and rational updating using the true signal distribution $DGP(s_t \mid \theta, H_{t-1})$ as the likelihood. One natural question is whether an incorrect prior by itself can generate violations of the upper bound for excess RN movement, or whether excessive movement in general requires incorrect updating.[14] The following proposition, which is straightforward to show, makes clear that the latter is likely to be necessary for a violation of the bound.

PROPOSITION 6. *Assume that Assumptions 2–3 continue to hold. In place of Assumption 1, however, assume that the person has an incorrect prior, $\pi_0 \neq \mathbb{P}_0(\theta = 1)$, but updates correctly, in the sense that $\pi_t(H_t) \propto \pi_{t-1}(H_{t-1})DGP(s_t \mid \theta = 1, H_{t-1})$. Define $\check{\phi} \equiv \phi L$, where the person's ϕ is as in (6) and where $L \equiv \frac{\pi_0/(1-\pi_0)}{\mathbb{P}_0(\theta=1)/(1-\mathbb{P}_0(\theta=1))}$ indexes the prior belief distortion, with $0 < L < \infty$. Then:*

(i) For all H_t, the person's RN beliefs π_t^* are equivalent to the RN beliefs of a fictitious agent whose physical beliefs $\{\check{\pi}_t\}$ satisfy Assumption 1 but who has $\check{\phi}$ in place of ϕ.

[13] For example, in the large class of DGPs we simulate, only 2% of DGPs have a positive $\triangle$ when $\pi_0^* < .25$. For $\pi_0^* < .5$, it is 11%. When π_0^* is low, the only DGPs in which $\triangle > 0$ are asymmetric and extreme. For example, when $\pi_0^* = .25$, $\triangle > 0$ only occurs if the likelihood ratio of the downward signals is at least two times the likelihood ratio of the upward signals. Thus Corollary 2 (which requires $\triangle < 0$) appears likely to be applicable for low values of π_0^*.

[14] One might think that the distinction between the person's prior and likelihood is artificial both formally and empirically, but we think the result here nonetheless helps clarify what kinds of belief streams violate the bounds.

(ii) *If $\check{\phi} \geqslant 1$, then the restrictions on $\mathbb{E}[X^*]$ above (Propositions 2–3, Corollaries 1–2) continue to hold, with $\check{\phi}$ in place of ϕ and $\check{\pi}_0$ in place of π_0. In particular, one cannot in this case have $\mathbb{E}[X^*] > \pi_0^{*2}$.*

(iii) *If $\check{\phi} < 1$ so that $\pi_0^* < \mathbb{P}_0(\theta = 1) = \check{\pi}_0$, then Proposition 2 continues to hold, but the bound in Proposition 3 becomes $\mathbb{E}[X^*] \leqslant (\check{\pi}_0 - \pi_0^*)(1 - \pi_0^*)$, and Corollary 1 becomes $\mathbb{E}[X^*] \leqslant (1 - \pi_0^*)^2$. Thus regardless of $\check{\phi}$, it must be the case that $\mathbb{E}[X^*] \leqslant \max(\pi_0^{*2}, (1 - \pi_0^*)^2)$.*

Part (i) formalizes that risk aversion is isomorphic to an incorrect prior, in that both have the same effect on π_t^* relative to the true $\mathbb{P}_t(\theta = 1)$. Thus with a suitably altered value of ϕ, the bounds generally cover the case of an incorrect prior, as in part (ii). The only case in which this argument requires slight amendment is when the prior is so downwardly distorted that $\pi_0^* < \mathbb{P}_0(\theta = 1)$. Even in this case, though, a slightly altered version of Corollary 1 still applies, as in part (iii). An incorrect prior acts as a one-time belief distortion; while reverting to the correct belief in this case does require some excess movement, this is generally not sufficient for a full violation of the bound in Proposition 3. In general, then, incorrect updating behavior must be present in such a violation, and the restriction imposed by our bound implies that this incorrect updating behavior necessarily entails excessive volatility in beliefs relative to the degree of uncertainty resolution over time.

3. Generalized Theoretical Results for Equilibrium Asset Prices

While the setting considered in the previous section is useful for providing intuition, it is also artificial: one cannot obtain a single person's valuation of Arrow-Debreu claims in observational data alone; there are more than two possible states; and the realized state determines more than just consumption. We thus now consider a general many-state framework for equilibrium asset prices and show how our results extend to this case. We first set up the framework and define notation (Section 3.1), then discuss the main assumption used to derive our bounds (Section 3.2), and finally describe how our results apply in this case (Section 3.3). This generalized setting will also allow us to more thoroughly consider the robustness and interpretation of our results, as we do in Section 4.

3.1 Setup and Notation

Preliminaries: Probability Space, Prices, and Risk-Neutral Probabilities

Time is again indexed by $t \in \{0, 1, 2, \ldots\}$, and we consider a discrete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with filtration $\{H_t\}$.[15] A realization of the elementary state is denoted by $\omega \in \Omega$. To make our results empirically implementable, we will be concerned with the ex-dividend value of the market index, $V_t^m : \Omega \to \mathbb{R}_+$, on some option expiration date T. (We will later extend the notation to allow for multiple option expiration dates.) A European call option on the index with strike price K has date-T payoff $(V_T^m - K)^+ = \max(V_T^m - K, 0)$, and its time-$t$ price is $q_{t,K}^m$. Assume without loss of

[15] We could extend to continuous state spaces with additional technicalities, but do not do so given that empirical implementation requires discretization and that our results are better understood for probabilities than densities. We note as well that objects analogous to those in Section 2 are given the same denotation in this section.

generality that these option prices are observable for some set of strike prices $\mathcal{K} \subseteq \mathbb{R}_+$ beginning at date 0. Assuming the absence of arbitrage, there exists a strictly positive *stochastic discount factor (SDF)* $M_{t,T}$ such that option prices satisfy $q^m_{t,K} = \mathbb{E}_t[M_{t,T}(V^m_T - K)^+]$, where $\mathbb{E}_t[\cdot] \equiv \mathbb{E}^{\mathbb{P}}[\,\cdot \mid H_t]$.

Option prices will be of interest for inferring a distribution over the change in value of the market index from 0 to T (rather than consumption, for which options are not directly traded). We say that *return state* $\theta \in \Theta \subset \mathbb{R}_+$ is realized for the market as of date T if $R^m_T \equiv V^m_T / V^m_0 = \theta$. The measure $\mathbb{P}\colon \mathcal{F} \to [0,1]$ governs the *objective* or *physical* probabilities of these return states. In this general case, the *risk-neutral (RN) measure* is defined by the change of measure

$$\left.\frac{d\mathbb{P}^*}{d\mathbb{P}}\right|_{H_t} = \frac{M_{t,T}}{\mathbb{E}_t[M_{t,T}]}, \tag{20}$$

and expectations under $\mathbb{P}^*$ are denoted by $\mathbb{E}^*[\cdot]$.[16] Using this definition of $\mathbb{P}^*$, the RN probability of return state θ is

$$\mathbb{P}^*_t(R^m_T = \theta) = \frac{\mathbb{E}_t[M_{t,T} \mid R^m_T = \theta]}{\mathbb{E}_t[M_{t,T}]}\,\mathbb{P}_t(R^m_T = \theta). \tag{21}$$

The RN pricing equation $q^m_{t,K} = \mathbb{E}^*_t[(V^m_T - K)^+]/R^f_{t,T}$ (see footnote 16) can be used to show that the date-t option prices $\{q^m_{t,K}\}_{K\in\mathcal{K}}$ reveal the set of RN probabilities $\{\mathbb{P}^*_t(R^m_T = \theta)\}_{\theta\in\Theta}$. Assume that the set of return states Θ is ordered such that $\theta_1 < \theta_2 < \cdots < \theta_J$, and assume for notational simplicity that the set of traded option strikes $\mathcal{K}$ coincides with the set of possible date-T index values (i.e., $\mathcal{K} = \{K_j\}_{j=1,\ldots,J}$, with $K_j = V^m_0\theta_j$). We can then back out RN probabilities from option prices as follows:

$$\mathbb{P}^*_t(R^m_T = \theta_j) = R^f_{t,T}\left[\frac{q^m_{t,K_{j+1}} - q^m_{t,K_j}}{K_{j+1} - K_j} - \frac{q^m_{t,K_j} - q^m_{t,K_{j-1}}}{K_j - K_{j-1}}\right]. \tag{22}$$

See Appendix A.2 for a derivation of this result, which follows directly from a discrete-state application of the classic result of Breeden and Litzenberger (1978).[17]

Beliefs

Aside from assuming no arbitrage, we have not yet taken a stance on the market structure or subjective beliefs underlying prices and RN probabilities. We could in principle pursue a strict mapping from Section 2 to the current case, by assuming a setting in which all individual traders in the economy have common beliefs satisfying RE with respect to the return state. The assumptions required to generate such an equilibrium are well studied in the literature on information and asset

[16] Since $\text{Price}_t = \mathbb{E}_t[M_{t,T}\text{Payoff}_T]$ for any asset, $R^f_{t,T} = \mathbb{E}_t[M_{t,T}]^{-1}$, where $R^f_{t,T}$ is the gross risk-free rate corresponding to an asset with payoff 1 at T. One can thus rewrite $\text{Price}_t = \mathbb{E}^*_t[\text{Payoff}_T]/R^f_{t,T}$, so $\mathbb{P}^*$ incorporates the risk adjustment needed to discount T-payoffs at the risk-free rate from t to T. This definition of $\mathbb{P}^*$ is sometimes referred to as the T-forward measure (e.g., Jamshidian, 1989, Qin and Linetsky, 2017).

[17] As can be seen from the result, $\mathcal{K}$ need not coincide with $V^m_0 \times \Theta$ to back out $\mathbb{P}^*_t(R^m_T = \theta_j)$, as strikes at (near) $V^m_0\theta_{j-1}$, $V^m_0\theta_j$, and $V^m_0\theta_{j+1}$ are sufficient for exact (approximate) measurement.

prices following Radner (1979) and Milgrom and Stokey (1982).[18] But rather than maintaining this microeconomic focus on testing the rationality of individual beliefs, we prefer an interpretation in which the "agent" in question is the market as a whole (or alternatively, the marginal trader); such an interpretation requires no auxiliary assumptions, and our resultant tests are informative about the efficiency of market valuations. We thus cannot and do not seek to rule out explanations for our empirical results based on belief heterogeneity, where all individual traders begin with different priors (about which they agree to disagree) and update rationally, but where the identity of the marginal trader changes with the wealth distribution (as, e.g., in Martin and Papadimitriou, 2021).

We therefore assume prices correspond to valuations for an agent ("the market") who, at the beginning of each period, observes a signal vector $s_t \in \mathcal{S}$ drawn from the distribution $DGP(s_t \mid \theta, H_{t-1}) = \mathbb{P}_{t-1}(s_t|\theta)$, where θ is the return state realized at T and $H_t = \sigma(s_\tau, 0 \leqslant \tau \leqslant t)$ is the Borel σ-algebra representing the history of signal realizations. The agent's subjective belief distribution over return states is denoted by $\Pi_{t,T} = \{\pi_t(R_T^m = \theta)\}_{\theta\in\Theta}$, where $\pi_t(R_T^m = \theta) \geqslant 0 \ \forall \theta \in \Theta$ and $\sum_{\theta\in\Theta} \pi_t(R_T^m = \theta) = 1$. More generally, for any random variable $Y(\omega)$, the agent attaches subjective probability $\pi_t(Y = y)$ to the outcome $Y = y$. We generalize Assumption 1 as:

Assumption 1′ (*RE*)**.** For any random variable Y, beliefs satisfy $\pi_t(Y = y) = \mathbb{P}_t(Y = y)$ with probability 1 for all t.

An immediate implication of this assumption, from the law of iterated expectations, is that beliefs satisfy a martingale condition analogous to Assumption 1, $\pi_t(R_T^m = \theta) = \mathbb{E}_t[\pi_{t+1}(R_T^m = \theta)]$ for all $\theta \in \Theta$. As in Section 2, this martingale condition for beliefs over returns is all that is required for our main results to carry through. The full-RE generalization stated in Assumption 1′, though, will be useful for streamlining some of the remaining discussion, as it further implies that *all* conditional expectations — including over the SDF — are a martingale under the objective measure.

Given Assumption 1′, we can define the RN belief distribution without explicitly restricting the agent's utility or constraint set by applying the same change of measure as in (20), using the general SDF $M_{t,T}$. This yields the RN belief-distribution $\Pi^*_{t,T} = \{\pi^*_t(R_T^m = \theta)\}_{\theta\in\Theta}$ such that $\pi^*_t(R_T^m = \theta) = \frac{\mathbb{E}_t[M_{t,T} \mid R_T^m = \theta]}{\mathbb{E}_t[M_{t,T}]} \pi_t(R_T^m = \theta)$ as in (21), and thus (22) tells us that option prices reveal the agent's RN beliefs as given here.

Localization: Conditional Beliefs, SDF Ratio, and Excess Movement

To align with the analysis in Section 2, we consider the behavior of *conditional* RN beliefs over adjacent pairs of return states. That is, rather than directly considering the full distribution $\Pi^*_{t,T}$, we instead consider restrictions on the behavior of the individual entries in $\{\widetilde{\pi}^*_{t,j}\}_{j=1,\ldots,J-1}$, defined by

$$\widetilde{\pi}^*_{t,j} \equiv \pi^*_t(R_T^m = \theta_j \mid R_T^m \in \{\theta_j, \theta_{j+1}\}) = \frac{\pi^*_t(R_T^m = \theta_j)}{\pi^*_t(R_T^m = \theta_j) + \pi^*_t(R_T^m = \theta_{j+1})}, \tag{23}$$

[18] Complete markets and a common-prior assumption, for example, are sufficient: prices in general reveal information (including private signals) in a rational expectations equilibrium, giving common posteriors. Results under alternative conditions have also been studied extensively (to take one example, see Blume, Coury, and Easley, 2006).

for $\pi_t^*(R_T^m = \theta_j) + \pi_t^*(R_T^m = \theta_{j+1}) > 0$. In words, $\widetilde{\pi}_{t,j}^*$ is the RN belief that return state θ_j will be realized, conditional on either θ_j or θ_{j+1}. This binary localization will be useful for two reasons: (i) it will allow us to apply results from the two-state setting of Section 2, and (ii) the main identifying assumption used to derive our tests is less restrictive than it would be without such a transformation (as discussed in Section 3.2 below). Conditional physical beliefs $\widetilde{\pi}_{t,j}$ are defined analogously, and the expectation under the conditional physical measure is $\widetilde{\mathbb{E}}_t[\,\cdot\,] \equiv \mathbb{E}_t[\,\cdot \mid R_T^m \in \{s_j, s_{j+1}\}]$.

In this context, the analogue to ϕ_t as defined in (6) is

$$\phi_{t,j} \equiv \frac{\mathbb{E}_t[M_{t,T} \mid R_T^m = \theta_j]}{\mathbb{E}_t[M_{t,T} \mid R_T^m = \theta_{j+1}]}, \tag{24}$$

which encodes the slope of the SDF across the adjacent return states θ_j, θ_{j+1}. In a representative-agent economy with consumption C_t, time-separable utility, and time discount factor β, the SDF is $M_{t,T} = \beta^{T-t} U'(C_T)/U'(C_t)$, and thus $\phi_{t,j} = \mathbb{E}_t[U'(C_T) \mid R_T^m = \theta_j]/\mathbb{E}_t[U'(C_T) \mid R_T^m = \theta_{j+1}]$, similar to the marginal rate of substitution in (7). But the representation in (24) is general and does not require a representative-agent structure (though we make periodic reference to such an economy for interpretation). Return state θ_j corresponds to state 1 (vs. state 0 for θ_{j+1}) in Section 2, which motivates our use of the index j to refer to variables conditioned on the pair (θ_j, θ_{j+1}). We maintain the convention of labeling θ_j as the "bad" (i.e., low) return state, so that Assumption 2 becomes:

Assumption 2′ (*Labeling of Return States*)**.** $\phi_{t,j} \geqslant 1$ with probability 1 for all t, j, where the set of return states Θ is ordered such that $\theta_1 < \theta_2 < \cdots < \theta_J$.

Using (23)–(24), we can write the RN odds ratio as $\frac{\widetilde{\pi}_{t,j}^*}{1-\widetilde{\pi}_{t,j}^*} = \phi_{t,j} \frac{\widetilde{\pi}_{t,j}}{1-\widetilde{\pi}_{t,j}}$, from which it is clear that (9) and (10) hold in this context for $\widetilde{\pi}_{t,j}^*$, $\widetilde{\pi}_{t,j}$, and $\phi_{t,j}$. Given the resolving RN belief stream $\boldsymbol{\pi}_j^* = [\widetilde{\pi}_{0,j}^*, \ldots, \widetilde{\pi}_{T,j}^*]$, RN belief movement m_j^*, RN initial uncertainty $\mathsf{u}_{0,j}^*$, and RN excess movement X_j^* are as defined in (1)–(3), with $\widetilde{\pi}_{t,j}^*$ in place of π_t. We commonly suppress the dependence on j (writing, e.g., X^*) when considering an arbitrary state pair.

3.2 Restriction on the SDF

We now confront the joint hypothesis problem, restating the main assumption (Assumption 3) we make alongside the maintained hypothesis of RE (Assumption 1). While the usefulness of this assumption was described briefly in Section 2, discussing it in this context will clarify why it is appealing economically: while it does impose a restriction on the SDF, the restriction is sufficiently general to allow for identification in a broad class of models. First, restating the assumption:

Assumption 3′ (*Constant SDF Ratio,* or *Conditional Transition Independence*)**.** We say the SDF satisfies *conditional transition independence (CTI)* for the return-state pair (θ_j, θ_{j+1}) if $\phi_{t,j}$ is constant with probability 1 for all t. We assume CTI is satisfied for all *interior* state pairs, $j = 2, 3, \ldots, J-2$.

Remarks:

1. As in Section 2, this assumption requires that changes in the RN odds ratio arise from changes

in the subjective belief $\widetilde{\pi}_{t,j}$ rather than changes in the expected relative severity of the adjacent return states θ_j and θ_{j+1}.

2. We exclude the extreme state pairs (θ_1, θ_2) and (θ_{J-1}, θ_J) from the constant-ϕ requirement. This will be useful theoretically: thinking of θ_1 and θ_J as tail return states, we are allowing for time-varying disaster (or positive-jump) severity.[19] This accordingly does *not* require that all changes in the underlying RN belief distribution arise from changes in subjective beliefs $\pi_t(R_T^m = \theta_j)$: there may be simultaneous time variation in the values in the numerator and denominator in (24), and the value $\mathbb{E}_t[M_{t,T}]$ need not be constant.
3. The assumption corresponds to a notion of transition or path independence because it implies that the realization of $M_{t,T}$ in return state θ_j, relative to θ_{j+1}, depend in expectation only on the return state pair and not on the path of any variables realized between t and T. This intuition is formalized in Lemma A.1, and it motivates the CTI label we use to reference it.[20]
4. If, analogous to the case in Section 2, a representative agent's utility depends only on the maturity value of the market index, then $\phi_{t,j}$ is constant.[21] But the assumption holds in more general cases as well. For example, permanent shocks to the SDF that raise $M_{t,T}$ in all states are admissible under the assumption. Given that permanent shocks to the SDF are empirically much more important for SDF variation than transitory shocks (Alvarez and Jermann, 2005), the assumption appears reasonable to a first order. (In addition, some transitory shocks are also admissible; see the first two examples below.)

We now turn to a set of examples to illustration the restriction concretely.

EXAMPLE 1. Assume an economy with a one-dimensional state variable $A_t \colon \Omega \to \mathbb{R}$ (e.g., productivity, consumption, volatility), with $dV_t^m / dA_t > 0$. This process may be non-stationary but is assumed to satisfy the Markov property. Assume further that there exists a representative agent with time-separable utility over the consumption process $\{C_t(A_t)\}$ and that the market index pays dividends $\{D_t(A_t)\}$, where these are arbitrary but yield a stationary price-dividend ratio. Then CTI holds for any two adjacent return states. If, in addition, consumption or consumption growth is i.i.d. over t, then CTI holds as well if the agent instead has Epstein–Zin recursive utility.

While the assumption of a scalar Markov forcing process in this example is restrictive, it nonetheless encompasses some simple leading cases that are useful for intuition. For example, with log consumption as the state variable ($A_t = c_t \equiv \log(C_t)$), we could have $c_t = g + \rho c_{t-1} + h(c_{t-1})\varepsilon_t$ where ε_t is i.i.d. with arbitrary distribution. This example allows for time variation in discount rates and risk premia (see Appendix B.1), and it also illustrates that both temporary and permanent

[19] Empirically, return states will be defined as narrow intervals for the index return with the exception of θ_1 and θ_J, which contain the left and right tails of the distribution, respectively; it is likely that, e.g., $\mathbb{E}[M_{t,T} \mid R_T^m \leqslant -20\%]$ changes over time as the distribution of returns conditional on $R_T^m \leqslant -20\%$ changes.

[20] "Conditional" is used to underscore that it requires only constancy of the ratio of conditional expectations of the SDF, rather than requiring deterministic state-contingent SDF realizations (as in the transition independence assumptions used by, among others, Ross, 2015, which are thus more restrictive). Borovička, Hansen, and Scheinkman (2016) show that Ross's assumptions rule out permanent shocks to the SDF, while ours does not.

[21] This holds more generally as long as there is some agent whose indirect utility can be written as a function only of the terminal index value (e.g., an investor retiring at date T with savings fully invested in the market).

shocks to the SDF are admissible in principle.[22] The same is true of the following example, which considers a well-known (and more fully specified) model featuring time-varying risk premia and permanent SDF shocks (see Bakshi and Chabi-Yo, 2012), and which has been advanced as a rationalization of the excess-volatility puzzle.

EXAMPLE 2. Consider the variable rare-disasters model of Gabaix (2012), described fully in Appendix B.3. Given any market-index option horizon T and any (small) positive value δ, there exists a return state $\underline{\theta}$ such that for all $\theta_j \geqslant \underline{\theta}$, the conditional probability of having realized at least one disaster over the life of the option is negligible: $\mathbb{P}_0\big(\sum_{t=1}^{T} \mathbb{1}\{\text{disaster}_t\} > 0 \,\big|\, R_T^m \geqslant \underline{\theta}\big) < \delta$. For all $\theta_j \geqslant \underline{\theta}$, CTI holds for the state pair (θ_j, θ_{j+1}) up to a negligible error, as $\phi_{t,j} = \phi_j + \eta_t$ for ϕ_j constant and $\eta_t = o_p(1)$ as $\delta \to 0$.

This result implies that for an economy described by this model, we need only focus attention on RN beliefs across return states (θ_j, θ_{j+1}) for which there is negligible probability of having realized a disaster conditional on reaching θ_j. This helps illustrate the use of the localization provided by considering conditional beliefs, and this insight guides our empirical analysis in Section 5.

Finally, our last example helps illustrate the types of settings in which CTI does *not* hold.

EXAMPLE 3. Consider the external habit-formation model of Campbell and Cochrane (1999), described fully in Appendix B.4. Under the assumptions stated there, CTI fails to hold for all j.

With habit formation, the path of consumption always matters in a manner not fully accounted for by conditioning on the return state. We note, though, that nothing about our theoretical framework requires considering beliefs over *return* states: all our results would apply for beliefs over elementary states in Ω, or, in the Campbell–Cochrane case, over the joint realization of the terminal consumption and *surplus* consumption values. Empirical implementation in this case, though, is infeasible. We consider return states specifically in our theory because this allows us to map directly between the theory and data. We later consider the effects of the violation of CTI implied by the Campbell–Cochrane model in a calibrated simulation study.

3.3 Main Results in the General Setting

Having completed the formal setup and discussion of our main assumption, we turn now to our main results in this more general asset-pricing setting. The bulk of the work in this case is, it turns out, in the setup and notation, as all our main results apply with appropriate relabeling:

PROPOSITION 7. *Under no arbitrage and Assumptions 1′–3′, for $j = 2, 3, \ldots, J-2$, Propositions 1–6 and Corollaries 1–2 continue to hold, with $\widetilde{\pi}^*_{t,j}$ replacing π^*_t, $\widetilde{\pi}_{t,j}$ replacing π_t, X^*_j replacing X^*, ϕ_j replacing ϕ, $\widetilde{\mathbb{E}}_0[\cdot]$ replacing $\mathbb{E}[\cdot]$, and with $\triangle_j \equiv \widetilde{\mathbb{E}}_0[X^*_j \mid R^m_T = \theta_{j+1}] - \widetilde{\mathbb{E}}_0[X^*_j \mid R^m_T = \theta_j]$ replacing $\triangle$.*

The main theoretical complication in applying the results in Section 2 to this setting is in proving that the RN measure, defined in a standard manner in (20), satisfies the property that pathwise

[22] For example, with CRRA utility and i.i.d. consumption-growth shocks, there are only permanent shocks to marginal utility; meanwhile, an economy with i.i.d. consumption has purely transitory shocks and a fixed $\mathbb{E}_t[U'_T \mid R^m_T = \theta_j]$ $\forall j$.

probabilities share the same value $\widetilde{\mathbb{P}}^*(H_T)/\widetilde{\mathbb{P}}(H_T)$ conditional on R_T^m, as in (11). This result — along with its main implication, that $\widetilde{\mathbb{E}}^*[X_j^* \mid R_T^m = \theta_j] = \widetilde{\mathbb{E}}[X_j^* \mid R_T^m = \theta_j]$, as after (14) — is proven to hold in Appendix A.2 for a measure that is observationally equivalent to the RN measure, which is sufficient for our purposes.

The economic intuition for these results is largely identical to the intuition discussed in Section 2. Two additional points, though, are worth discussing in this setting. First, in the case of a bound violation $\widetilde{\mathbb{E}}[X_j^*] > \widetilde{\pi}_{0,j}^{*\,2}$, the conclusion would not necessarily be that there are violations of the no-arbitrage condition; instead, there would be no rational-expectations SDF process satisfying CTI capable of generating the observed excess movement in RN beliefs. The *actual* SDF process translating between objective probabilities and RN beliefs would in this case include a market-level belief distortion that induces excess volatility.[23]

Second, while the results above are convenient to express in terms of the SDF slope ϕ_j given that this allows for closed-form solutions that can be applied across a wide range of structural models regardless of the origin of the SDF, the results also admit an interpretation in terms of the approximate required risk-aversion value for a fictitious representative agent with utility over the terminal value of the market index. Analogous to equation (8) in Section 2, we have the following result. To avoid repetition, for the remaining results we will continue to assume no arbitrage and that Assumptions 1′–3′ hold for $j = 2, 3, \ldots, J-2$, unless stated otherwise.

PROPOSITION 8. *Assume additionally that there is a representative agent with (indirect) utility over time-T wealth, with wealth equal to the market index value, and denote $V_j^m \equiv V_0^m \theta_j$. Then local relative risk aversion $\gamma_j \equiv -V_j^m U''(V_j^m)/U'(V_j^m)$ is given to a first order around return state θ_j by*

$$\gamma_j = \frac{\phi_j - 1}{(V_{j+1}^m - V_j^m)/V_j^m}. \tag{25}$$

As in Section 2, γ_j is proportional to $\phi_j - 1$, as this gives the percent decrease in marginal utility in moving from low-return state θ_j to high-return state θ_{j+1}. To calculate relative risk aversion, this change in marginal utility must be normalized by the percent wealth increase $(V_{j+1}^m - V_j^m)/V_j^m$ in moving from θ_j to θ_{j+1}, which is also equal to the percent return deviation $(\theta_{j+1} - \theta_j)/\theta_j$ between the two states. If, for example, $\theta_j = 1$, $\theta_{j+1} = 1.05$, then a value $\phi_j = 1.5$ implies $\gamma_j = 10$.

4. Theoretical Results on Robustness and Empirical Implementation

Building on Section 3, we now turn to a set of additional robustness results. First, to obtain a direct mapping from theory to data, we consider the observability of the statistics in our bounds (Section 4.1) and how to account for possible mismeasurement or market microstructure noise (Section 4.2). We then extend our analysis to consider cases in which CTI is violated (Section 4.3).

[23] Such a finding would be close in spirit to a violation of the "good-deal bounds" of Cochrane and Saá-Requejo (2000): even if the no-arbitrage condition holds, there would be an investment strategy with a large Sharpe ratio under the objective measure — in particular, betting on mean reversion in RN beliefs — that is nonetheless not traded away.

4.1 Aggregating Over Belief Streams

The discussion to this point has taken as given that all the values in our bounds for expected excess RN movement are observable aside from ϕ_j. But this has elided one issue: the bounds are stated with respect to date-0 expectations conditional on the RN prior, but we observe only one draw X_j^* per expiration date rather than the ex-ante expectation of this statistic for a given $\widetilde{\pi}_{0,j}^*$. We now show how one can aggregate over multiple RN belief streams — i.e., over multiple option expiration dates — so as to implement the bounds empirically. In doing so, we allow ϕ_j to change as well across belief streams, though we continue to assume that it is fixed within a given stream.[24]

To do so, we must further generalize the environment slightly. We now assume that we can observe index option prices for some set of N option expiration dates $\mathcal{T} \equiv \{T_i\}_{i=1,\ldots,N}$, so that belief streams (and their DGPs) are indexed by i. We accordingly add the subscript i to refer to objects corresponding to expiration date T_i: $\phi_{i,j}$ is the SDF ratio for date T_i and state pair (θ_j, θ_{j+1}), RN beliefs are $\widetilde{\pi}_{t,i,j}^*$, and RN excess movement is $X_{i,j}^*$. To simplify notation for now, in order to focus specifically on the issue of aggregation for an arbitrary state pair, we maintain the i subscripts but otherwise revert to the notation from Section 2 (e.g., using X_i^* in place of $X_{i,j}^*$).

Due to Jensen's inequality, we cannot simply insert $\mathbb{E}[\phi_i]$ in place of ϕ_i or $\mathbb{E}[\pi_{0,i}^*]$ in place of $\pi_{0,i}^*$ (where these expectations are implicitly taken over DGPs i) when taking the expectation of both sides of the results in Propositions 2 and 3 (or their counterparts in Proposition 7). However, we can show that the following modified generalizations do hold:

PROPOSITION 9. *Define* $\overline{\phi} \equiv \max_{\pi_{0,i}^*} \mathbb{E}[\phi_i \mid \pi_{0,i}^*]$. *We have:*

(i) GENERALIZATION OF PROPOSITION 2: *If* $\mathrm{Cov}(\pi_{0,i}, \triangle_i) = 0$, *and* $\pi_{0,i}^*$ *is constant across* i *(i.e., fixing a given* $\pi_{0,i}^*$*), then over all streams,*

$$\mathbb{E}[X_i^*] \leqslant \max\left\{0, \left(\pi_{0,i}^* - \frac{\pi_{0,i}^*}{\mathbb{E}[\phi_i] + (1 - \mathbb{E}[\phi_i])\pi_{0,i}^*}\right)\mathbb{E}[\triangle_i]\right\}. \tag{26}$$

(ii) GENERALIZATION OF PROPOSITION 3: *Over all streams, we have without any additional assumptions that*

$$\mathbb{E}[X_i^*] \leqslant \mathbb{E}\left[\left(\pi_{0,i}^* - \frac{\pi_{0,i}^*}{\overline{\phi} + (1 - \overline{\phi})\pi_{0,i}^*}\right)\pi_{0,i}^*\right], \tag{27}$$

or, fixing a given $\pi_{0,i}^*$, $\mathbb{E}[X_i^*] \leqslant \left(\pi_{0,i}^* - \frac{\pi_{0,i}^*}{\mathbb{E}[\phi_i] + (1 - \mathbb{E}[\phi_i])\pi_{0,i}^*}\right)\pi_{0,i}^*$.

(iii) GENERALIZATION OF COROLLARY 1: *Over all streams, without any additional assumptions,*

$$\mathbb{E}[X_i^*] \leqslant \mathbb{E}\left[{\pi_{0,i}^*}^2\right].$$

[24] While it might be palatable to assume that ϕ_j is constant *within* a belief stream (over the course of months), it becomes more challenging to assume that ϕ_j is constant *across* belief streams (over the course of years).

(iv) GENERALIZATION OF COROLLARY 2: *If* $\triangle_i \leqslant 0$ *for all* i*, then over all streams,*

$$\mathbb{E}[X_i^*] \leqslant 0.$$

First consider part (ii), which generalizes Proposition 3. This result effectively applies Jensen's inequality for one of several variables, using the fact that the second partial derivative of the bound with respect to $\phi_{i,j}$ is negative. The bound is accordingly even more conservative than the original bound as stated for a single stream, as it effectively accounts for the worst-case possible relationship between RN priors and ϕ_i. We also state the bound fixing a given value of $\pi_{0,i}^*$, because in some of our empirical results we report $\mathbb{E}[X^*]$ across streams conditional on a particular value of this RN prior.[25] This bound is quite conservative as well; for example, if $\pi_{0,i}^* = 0.5$ for $i = 1, 2$, with $\phi_1 = 1$ and $\phi_2 = 99$, then under the maximizing DGP for $\mathbb{E}[X^*]$ (from Proposition 5), the bound would lead us to infer $\mathbb{E}[\phi_i] \geqslant 2.9$, when in fact $\mathbb{E}[\phi_i] = 50$.

Taking $\overline{\phi} \to \infty$ in (27) generates the result in part (iii), which indicates that Corollary 1 does not require any modification in taking unconditional expectations over streams. Similarly, Corollary 2 does not depend on ϕ and therefore is robust to ϕ changing across streams as long as $\triangle_i \leqslant 0$.

Finally, returning to part (i), the analogue to Proposition 2 requires an additional assumption. The original formula includes the product of $\pi_0^* - \pi_0$ and $\triangle$. Therefore, when averaging over DGPs, the covariance between $\pi_{0,i}$ and $\triangle_i$ across DGPs affects the bounds (which otherwise follow the logic for part (ii)). For simplicity, the proposition assumes that this covariance is zero (though one could adjust this bound under different assumed values for the covariance). To understand this assumption, recall from Section 2.3 that, fixing π_0^*, $\triangle$ is determined by the asymmetry of the likelihood ratios of the upward and downward signals in the DGP. Assuming zero covariance is thus equivalent to assuming no relationship between the asymmetry of the DGP and ϕ. Although we have no reason to presume a particular relationship between the DGP asymmetry and ϕ, our main empirical results apply the more conservative bound in part (ii), which has the benefit of requiring no further assumptions. Part (i) also holds only for a fixed $\pi_{0,i}^*$; this is sufficient for our purposes, as our empirical results for this less-conservative bound will generally be conditional on a given $\pi_{0,i}^*$, in particular in Figure 2.

The bounds stated in the proposition are now empirically implementable: in particular, we can measure a sample counterpart of the average excess-movement statistic on the left side and the RN priors on the right side, and the minimum $\overline{\phi}$ that solves the bound in (27) is then a conservative estimate of the maximal conditional-mean SDF slope for the return-state pair in question.[26] If no such $\overline{\phi}$ solves the bound, then the bound in part (iii) is violated. Even simpler, the bound in part (iv) only requires comparing $\mathbb{E}[X_i^*]$ to zero.

The results and discussion above simplified notation by dropping the state-pair subscript j, but the values $\overline{\phi}_j$ are likely to vary over j. The same argument used in Proposition 9 to take expectations

[25] These constant-$\pi_{0,i}^*$ bounds can equivalently be read as bounds for the conditional expectation $\mathbb{E}[X_i^* \mid \pi_{0,i}^*]$ given $\mathbb{E}[\phi_i \mid \pi_{0,i}^*]$. We write them in the simpler format presented above ("fixing a given $\pi_{0,i}^*$") only to limit notational complexity.

[26] Further, if $\mathbb{E}[\phi_i \mid \pi_{0,i}^*] = \mathbb{E}[\phi_i]$, as might be expected to hold approximately, then $\overline{\phi} = \mathbb{E}[\phi_i]$.

over i, though, can also be applied to take expectations over j, thereby obtaining a single estimate $\overline{\boldsymbol{\phi}}$ aggregated over both streams *and* return states (for all states meeting CTI) if desired.

4.2 Robustness to Measurement Error

The final issue we confront in making our bounds empirically implementable is the possibility of mismeasurement or microstructure noise in RN beliefs. Our bounds provide a minimum value of ϕ required to rationalize the observed variation in RN beliefs; if some of this variation is in fact arising due to noise, then we may overestimate this required ϕ. A simple correction can be applied to our bounds to account for this issue, as shown in the following result. Given that noise arises period-by-period, we first define one-period analogues for our statistics: denote RN movement between t and $t+1$ by $\mathsf{m}^*_{t,t+1} \equiv (\pi^*_t - \pi^*_{t+1})^2$, RN uncertainty at t by $\mathsf{u}^*_t \equiv (1-\pi^*_{t+1})\pi^*_t$, and RN excess movement between t and $t+1$ as $X^*_{t,t+1} \equiv \mathsf{m}^*_{t,t+1} - (\mathsf{u}^*_t - \mathsf{u}^*_{t+1})$ (where we now omit i and j subscripts for clarity). Similar to AR (Section II.E), we then have the following:

PROPOSITION 10. *Assume that the observed $\widehat{\pi}^*_t$ is measured with error with respect to the true π^*_t:*

$$\widehat{\pi}^*_t = \pi^*_t + \epsilon_t, \tag{28}$$

where $\mathbb{E}[\epsilon_t] = 0$, $\mathbb{E}[\epsilon_t\,\epsilon_{t+1}] = 0$, and $\mathbb{E}[\epsilon_{t+k}\,\pi^*_{t+k'}] = 0$ for $k, k' \in \{0,1\}$. Denoting the observed one-period RN excess movement statistic by $\widehat{X}^*_{t,t+1}$, its relation to the true value $X^*_{t,t+1}$ in expectation is

$$\mathbb{E}\left[\widehat{X}^*_{t,t+1} - X^*_{t,t+1}\right] = 2\mathrm{Var}(\epsilon_t).$$

We can thus subtract $2\mathrm{Var}(\epsilon_{t,j})$ from each period's observed excess-movement statistic to obtain an unbiased true excess movement value, which can then be used in our bounds after summing over the full stream. If measurement error is positively correlated over time rather than uncorrelated, this will reduce the upward bias in measured X^*.[27] We discuss estimation of $\mathrm{Var}(\epsilon_t)$ in Section 5.

4.3 Robustness to Violations of CTI

Proposition 10 allows for a correction with respect to empirical misspecification; we turn now to a result that speaks to the possibility of theoretical misspecification in the form of violations of CTI. While this assumption of a constant ϕ within belief streams is met in some commonly used theoretical frameworks, it is a knife-edge restriction that is unlikely to be met exactly. In this section, we show theoretically that within a large class of environments, our main results continue to hold even with a time-varying ϕ_t. We also consider a set of simulations of other environments and find that realistic violations of CTI appear unlikely to affect our results more than minimally.

The fact that time-varying ϕ_t does not produce significant excess movement might seem

[27] One might worry about negatively correlated measurement error in the case of bid-ask bounce, but as we will see in our empirical estimation, the autocorrelation values for estimated microstructure noise have long died out at a one-day lag, and we use end-of-day data to construct our statistics.

unintuitive: one might worry that even small oscillations in ϕ_t could cause unbounded movement in RN beliefs and dramatically invalidate our conclusions. For example, suppose that π_t is constant at .5 and ϕ_t changes back and forth from 1 to 1.5 repeatedly. Without any movement in physical beliefs, π_0^* will vary repeatedly between .5 and .6, leading to unbounded movement as $T \to \infty$. Crucially, however, this argument overlooks a core insight of the paper: given rational expectations, ϕ_t cannot undergo such repeated oscillation if its variance is bounded, because it is itself a function of martingale conditional expectations.

More formally, taking ϕ_t as defined in (24) (again suppressing i, j subscripts for clarity), write $\phi_t = \phi_t^{(1)}/\phi_t^{(0)}$, where $\phi_t^{(1)} \equiv \mathbb{E}_t[M_{t,T} \mid R_T^m = \theta_j]$ is the numerator and $\phi_t^{(0)} \equiv \mathbb{E}_t[M_{t,T} \mid R_T^m = \theta_{j+1}]$ is the denominator of (24). We now drop Assumption 3′ (CTI) and let ϕ_t vary, but Assumption 1′ implies that $\phi_t^{(\ell)} = \mathbb{E}_t[\phi_{t+1}^{(\ell)}]$ for $\ell = 0, 1$. In spite of this restriction, though, allowing ϕ_t to vary meaningfully complicates the situation, as RN beliefs become a non-linear function of time-varying subjective beliefs *and* conditional SDF expectations (see, e.g., (9)), all of which have unrestricted DGPs. Surprisingly, though, our bounds still hold as stated over a large class of such DGPs:

PROPOSITION 11. *Assume no arbitrage and Assumptions 1′–2′ continue to hold, but in place of Assumption 3′, assume only that $Var(\phi_t) < \infty$. If ϕ_t evolves as a martingale or supermartingale ($\mathbb{E}_t[\phi_{t+1}] \leqslant \phi_t$), then the bounds in Proposition 3 and Corollary 1 (and their counterparts in Propositions 7 and 9) continue to apply, with ϕ_0 replacing ϕ.*

The proof of Proposition 11 demonstrates that the variation in π_0^* arising when ϕ_t is a non-degenerate supermartingale always *strictly lowers* the expected excess belief movement period by period, rendering the main bound in (19) even more conservative.[28] Whether ϕ_t is a supermartingale depends on the sign of $\mathrm{Cov}_t(\phi_{t+1}, \phi_{t+1}^{(0)})$: from the definition of this covariance, write

$$\mathbb{E}_t[\phi_{t+1}] = \phi_t - \frac{\mathrm{Cov}_t(\phi_{t+1}, \phi_{t+1}^{(0)})}{\phi_{t+1}^{(0)}},$$

so $\mathbb{E}_t[\phi_{t+1}] \leqslant \phi_t$ if and only if $\mathrm{Cov}_t(\phi_{t+1}, \phi_{t+1}^{(0)}) = \mathrm{Cov}_t(\phi_{t+1}, \mathbb{E}_t[M_{t,T} \mid R_T^m = \theta_{j+1}]) \geqslant 0$. Thinking of $M_{t,T}$ as proportional to $U'(C_T)$ as in Section 2, this requires that expected terminal risk aversion (encoded in ϕ_{t+1}, as in Proposition 8) be positively correlated with expected marginal utility (MU) in the high-consumption state. This is an intuitively reasonable restriction, as it implies bad news (in the form of higher expected MU) generally arrives at the same time for both states, with the expected low-consumption MU increasing more than its high-consumption counterpart. The converse ($\mathrm{Cov}_t(\phi_{t+1}, \phi_{t+1}^{(0)}) < 0$), by contrast, requires risk aversion to increase in general in response to *good* news about MU in the good state.[29]

If one wishes to consider such a setting in which ϕ_t does not evolve as a supermartingale, though, it is difficult to make analytical statements given the complexity of such a setting. Instead,

[28] Further, given the continuity of our bounds, if ϕ_t is a submartingale but close to a martingale, then our results also hold up to a small additive error.

[29] For further discussion of such a restriction in a different context, see Lazarus (2019).

we numerically simulate a set of binary DGPs featuring strict submartingale ϕ_t processes alongside learning about the return state. Results for this exercise are reported in Appendix B.5. Even in a parameterization with extremely high ϕ uncertainty — in which ϕ can vary between 1 and 9, corresponding to γ varying between 0 and 160 — $\mathbb{E}[X^*]$ rises by at most 0.015, and it breaches the bound for $\phi = 3$ (equal to ϕ_0) less than one percent of the time in such DGPs.

Finally, to consider how time-varying ϕ_t affects our results in a less abstract environment, we simulate the Campbell and Cochrane (1999) habit-formation model, which violates CTI, using the baseline calibration proposed by Campbell and Cochrane. We find in this case that ϕ_t is closely approximated by a martingale and therefore does not generate additional excess movement. Our bounds thus continue to hold and continue to be conservative. See Appendix B.6 for details.

Given our theoretical and numerical results, we conclude that allowing ϕ_t to change within a stream is unlikely to impact our results and bounds meaningfully. That said, we can imagine that there exists a set of perverse distributions with negatively correlated values of $\phi_{t+1}^{(1)}$ and $\phi_{t+1}^{(0)}$ which, combined with near-worst-case DGPs for information about the return state, might further increase excess movement. Our strong suspicion based on the above results, though, is that any realistic DGP will have a small impact.

5. Empirical Estimation and Results

Our theory leads to bounds on the variation in risk-neutral beliefs over the value of the market index, which we proceed now to measure in the data. We begin by describing how we map from theory to data (Section 5.1) and how we estimate microstructure noise (Section 5.2), and then we summarize the data (Section 5.3) before turning to our empirical results (Section 5.4). We conclude with a brief discussion of the reduced-form correlates of RN excess movement (Section 5.5).

5.1 Data and Risk-Neutral Distribution

Data. Our main source for S&P 500 index options data is the OptionMetrics database, which provides end-of-day prices for European call and put options for all strike prices and option expiration dates traded on the Chicago Board Options Exchange (CBOE). The sample runs from January 1996 through December 2018. We augment this data with intraday price quotes obtained directly from the CBOE for a subset of trading days in our sample, in order to account for market microstructure noise; this additional data is described further in Section 5.2.

We apply standard filters to remove outliers and options with poor trading liquidity from the OptionMetrics data, with details provided in Appendix B.7.[30] Two aspects of this data cleaning bear mention here. First, while our bounds apply for belief streams of arbitrary length, we follow past literature (e.g., Christoffersen, Heston, and Jacobs, 2013; Martin, 2017) and consider options with

[30]The raw dataset contains 12.4 million option prices; the filtered data, which are then used to measure the RN distribution, contain 4.3 million option prices corresponding to 5,537 trading dates and 991 option expiration dates. The majority of the difference is attributable to our use of only out-of-the-money call and put strikes.

maturity of at most one year. Second, after transforming the filtered prices to RN beliefs (described below), we keep only conditional RN belief observations $\widetilde{\pi}^*_{t,i,j}$ for which the non-conditional beliefs satisfy $\pi^*_t(R^m_{T_i} = \theta_j) + \pi^*_t(R^m_{T_i} = \theta_{j+1}) \geqslant 5\%$, as conditional beliefs $\widetilde{\pi}^*_{t,i,j}$ are likely to be particularly susceptible to mismeasurement when the underlying beliefs are close to zero.

Empirical return space. For our baseline estimation, we define the return state space Θ in terms of log excess return intervals:

$$\Theta = R^f_{0_i,T_i} \exp\big\{(-\infty, -0.2], (-0.2, -0.15], (-0.15, -0.1], \ldots, (0.1, 0.15], (0.15, 0.2], (0.2, \infty)\big\},$$

where $R^f_{0_i,T_i}$ is the gross risk-free rate from 0_i to T_i.[31] In words, return state 1 is realized if the log excess S&P return from 0_i to T_i is less than -0.2 $\approx$ -20%; state 2 is realized if the excess return is in the five-percentage-point bin between -0.2 and -0.15; and so on. Abusing notation slightly, we typically refer to states by the right end of their associated log excess-return bin: $\theta_1 = -0.2$, $\theta_2 = -0.15, \ldots, \theta_9 = 0.2, \theta_{10} = \infty$. Following the labeling convention in Assumption 2′, we define the binary conditional beliefs to be used in our tests as $\widetilde{\pi}^*_{t,i,j} \equiv \pi^*_t(R^m_{T_i} = \theta_j \mid R^m_{T_i} \in \{\theta_j, \theta_{j+1}\})$, so $\widetilde{\pi}^*_{t,i,j}$ corresponds to the probability that the low state j (e.g., $\theta_2 = -15\%$) will be realized, conditional on j or $j+1$ being realized (in this case, conditional on an excess return between -20% and -10%). We again use of the right end of the return bin for j as shorthand in referencing the conditional belief and excess movement for state pair (θ_j, θ_{j+1}), corresponding to the midpoint of the two return bins.

Our use of five-percentage-point ranges for return states reflects a desire to balance (i) measurement accuracy for the RN beliefs and (ii) plausibility of our assumption of constant ϕ_j (CTI). Wider bins lead to greater measurement accuracy, but make it less likely that there are no changes in the expected SDF realization conditional on a given return state θ_j relative to θ_{j+1}. We report empirical estimates below for all adjacent state pairs for completeness, but as discussed in Remark 2 after Assumption 3′, it is unlikely that CTI is met for the extreme state pairs (θ_1 relative to θ_2, and $\theta_9 = \theta_{J-1}$ vs. $\theta_{10} = \theta_J$). Our focus is thus on the *interior* state pairs with low-return states $\theta_2, \ldots, \theta_8$; in particular, when we aggregate our state-by-state estimates of ϕ_j required to rationalize the data into a single average value $\overline{\phi}$ across states, we use only these interior states.[32]

Risk-neutral beliefs. To extract a risk-neutral distribution over the return states in Θ from the observed option cross-sections, we use standard tools from the option-pricing literature. Our starting point is equation (22), which tells us how to map from option prices to RN beliefs. We use this to construct a smooth RN distribution for returns, largely following the technique proposed by Malz (2014); Appendix B.7 provides a detailed description. With the RN beliefs $\pi^*_{t,i,j}$ in hand, we can then calculate conditional beliefs straightforwardly as $\widetilde{\pi}^*_{t,i,j} = \pi^*_{t,i,j}/(\pi^*_{t,i,j} + \pi^*_{t,i,j+1})$. We then use the resulting conditional RN belief streams to calculate the excess movement statistics $X^*_{i,j}$

[31] We use excess returns for convenience of interpretation. Following van Binsbergen, Diamond, and Grotteria (2022), we measure $R^f_{0_i,T_i}$ directly from the options prices by applying the put-call parity relationship; again see Appendix B.7.

[32] This yields an additional de facto data filter, as we are effectively considering only option strikes with moneyness between 0.8 and 1.2 (following, e.g., Constantinides, Jackwerth, and Savov, 2013).

needed to implement our bounds. Our general results in Section 3 restrict the expectation of $X^*_{i,j}$ conditional on state θ_j or θ_{j+1} being realized — in particular, Propositions 7 and 9 give bounds for $\widetilde{\mathbb{E}}[X^*_{i,j}]$ — and we accordingly keep only observations for which $\widetilde{\pi}^*_{T_i,i,j} = 0$ or 1 ex post; for example, if the total excess return on the market over the life of option contract i is -14%, then we keep only $X^*_{i,2}$ (θ_2 ranges from -20% to -15% return, so $\widetilde{\pi}^*_{T_i,i,2} = 0$) and $X^*_{i,3}$ ($\widetilde{\pi}^*_{T_i,i,3} = 1$).

Simplifying notation. To this point, we have taken care in this section to use detailed notation to clarify that our conditional statistics ($\widetilde{\pi}^*_{t,i,j}$, $X^*_{i,j}$) depend on the contract i and state pair j. Having clarified this, though, in what follows we generally suppress the cumbersome use of i, j, and $\widetilde{\cdot}$, and simply write π^*_t in place of $\widetilde{\pi}^*_{t,i,j}$, X^* for $X^*_{i,j}$, and so on (as, e.g., in Section 4.2). Similarly, we often drop the "conditional" qualifier when referring to conditional RN belief statistics.

5.2 Noise Estimation

As in Proposition 10, we also wish to account for measurement error stemming from possible microstructure noise in our estimation.[33] That result shows that unlike in the classical errors-in-variables regression model (which leads to attenuation), measurement error in our case can increase the observed variation X^* and thereby lead to an upward bias in the estimated SDF slope needed to rationalize the data. With noise described by $\widehat{\pi}^*_t = \pi^*_t + \epsilon_t$ as in (28), Proposition 10 tells us that we must estimate $\mathrm{Var}(\epsilon_t)$ in order to eliminate this bias. We turn to a sample of high-frequency option prices to estimate this noise variance in our RN beliefs data.

Specifically, we obtain minute-by-minute price quotes on S&P index options for a subset of trading days directly from the CBOE. For each available option expiration date on each such trading day, we recalculate the RN belief distribution at the end of each minute using exactly the same procedure as described in Section 5.1. As this requires calculating 390 sets of RN beliefs for each trading day (9:30 AM–4:00 PM), this procedure would be computationally infeasible if applied to our entire sample of 5,537 trading days (each of which has an average of 11 available option expiration dates, generating 60,543 (t, T) combinations). We accordingly select 30 trading days at random from within our available sample period, and use the minute-by-minute quotes to calculate intraday RN distributions for these days.[34]

We then use tools from the literature on microstructure noise to estimate $\mathrm{Var}(\epsilon_t)$ using these intraday data. The basic intuition for this estimation strategy — as described, for example, by Zhang, Mykland, and Aït-Sahalia (2005) — is best understood by assuming temporarily that the noise ϵ_t in (28) is i.i.d., while the true π^*_t changes smoothly over time (i.e., $\eta_{t+h} \equiv \pi^*_{t+h} - \pi^*_t \to 0$ and $\mathrm{Var}(\eta_{t+h}) \to 0$ as $h \to 0$, as would be the case for an Itô process). Imagine calculating movement using the observed beliefs, $(\widehat{\pi}^*_{t+h} - \widehat{\pi}^*_t)^2$, with less and less time h between consecutive observations. As one decreases h to 0, the noise swamps the true variation: since $(\pi^*_{t+h} - \pi^*_t)^2 \to 0$,

[33] For example, transient demand pressure in the option market (Bollen and Whaley, 2004) may cause variation in observed RN beliefs unrelated to the underlying index dynamics.

[34] This yields an intraday data set roughly twice as large as the original one, as $30 \times 390 \times 11 \approx 130{,}000$.

we have $\mathbb{E}[(\widehat{\pi}^*_{t+h} - \widehat{\pi}^*_t)^2] \to 2\text{Var}(\epsilon_t)$. Thus in this simple example, $\text{Var}(\epsilon_t)$ can be estimated by calculating the quadratic variation in RN beliefs sampled at a high frequency.

In practice, one would expect the data to contain both non-i.i.d. noise ϵ_t and jumps in the true process π^*_t, and it is desirable to use a noise estimation method that is robust to these features. One such estimator for $\text{Var}(\epsilon_t)$ is the ReMeDI ("Realized moMents of Disjoint Increments") estimator proposed by Li and Linton (2021). This estimator effectively takes the average product of *disjoint increments* of the observed process, $(\widehat{\pi}^*_t - \widehat{\pi}^*_{t-h})(\widehat{\pi}^*_t - \widehat{\pi}^*_{t+h})$.[35] The idea is that even if the true process features jumps so that $\mathbb{E}[(\pi^*_{t+h} - \pi^*_t)^2] > 0$, its increments over non-overlapping windows are still approximately uncorrelated. Li and Linton (2021) show that this estimator is consistent for $\text{Var}(\epsilon_t)$ under general semimartingale processes for π^*_t (i.e., π^*_t can feature jumps but must be of bounded variation), and for quite general dependent noise processes as long as the autocovariances for ϵ_t decay to 0 sufficiently quickly relative to the increment width (see their Theorem 4.1 for precise conditions).[36] It also performs well in simulations and empirical applications.

Using this ReMeDI estimator on our minute-by-minute data, we estimate $\text{Var}(\epsilon_t) = \text{Var}(\epsilon_{t,i,j})$ separately for each combination of trading day t, expiration date T_i, and return state pair j in our intraday sample.[37] We then match the noise estimates (which are obtained for a subsample of days) to the observed excess movement observations in our original data; see Appendix B.8 for details on this procedure. Finally, we subtract $2\widehat{\text{Var}}(\epsilon_{t,i,j})$ from $\widehat{X}^*_{t,t+1,i,j}$ to obtain a *noise-adjusted* estimate of one-day excess movement following Proposition 10, and we sum these noise-adjusted one-day values over the full stream to obtain noise-adjusted estimates of $X^*_{i,j}$.

We discuss the magnitude of the noise estimates in the next subsection alongside descriptive statistics for the excess movement values. The ReMeDI procedure also allows for estimation of the intraday autocovariances of the noise ϵ_t. These autocovariances are estimated to be positive for small lag values, but they die out quickly and are precisely estimated near zero for noise observations more than an hour apart. This justifies the assumption in Proposition 10 that end-of-day noise observations are uncorrelated, $\mathbb{E}[\epsilon_t \epsilon_{t+1}] = 0$, as ultimately we care about noise only to the extent that it affects our excess movement statistics at a daily frequency.

Our main results in Section 5.4 use the noise-adjusted excess movement data. All standard errors and confidence intervals are based on a bootstrap procedure (detailed in Section 5.4) that accounts for the sampling uncertainty in the above noise estimation and averaging procedure.

[35] More formally and specifically, our ReMeDI estimator (following the replication code provided kindly by Li and Linton) is $\widehat{\text{Var}}(\epsilon_t) = \frac{1}{N_{\epsilon,n}} \sum_{i=2k_n}^{N_{\epsilon,n}-k_n} (\widehat{\pi}^*_{t_i} - \widehat{\pi}^*_{t_i-2k_n})(\widehat{\pi}^*_{t_i} - \widehat{\pi}^*_{t_i+k_n})$, where $N_{\epsilon,n}$ is the number of observations over a fixed time span (in our case, one trading day) and k_n is a tuning parameter, with $k_n, N_{\epsilon,n} \to \infty$ and $k_n / N_{\epsilon,n} \to 0$ as $n \to \infty$. We select k_n for each return state using the algorithm in Section F.1 of the Online Appendix of Li and Linton (2021).

[36] While jumps induce upward bias in a naïve estimator based on $(\widehat{\pi}^*_{t+h} - \widehat{\pi}^*_t)^2$, positively autocorrelated noise goes in the other direction. To address this, the ReMeDI estimator uses a window width (k_n in footnote 35) that is sufficiently wide even as one decreases the time h between observations. The second (downward) bias in the naïve estimator seems predominant in our data: the ReMeDI noise estimates are about 50% higher than the naïve estimates.

[37] For this exercise, to increase our available observations, we do not condition on the ex post state being θ_j or θ_{j+1}.

5.3 Excess Movement: Descriptive Statistics and Figures

We first provide summary statistics and plots describing the RN beliefs and excess movement values observed in our data. Table 2 summarizes the average excess movement values $\overline{X}^*$ overall (across all interior state pairs and expiration dates) and by subsample. Excess movement is difficult to interpret without some normalization; we should, for example, expect movement to increase alongside initial uncertainty under Bayesian updating. The first two columns thus divide $\overline{X}^*$ by average initial uncertainty $\overline{u}_0$. As in Augenblick and Rabin (2021, Section II.D), the resulting normalized statistic can be interpreted as the percent by which average movement exceeds initial uncertainty and thus uncertainty resolution. (For example, a value of 1 corresponds to 100% more movement than uncertainty resolution.) These values are quite high in our data: for the noise-adjusted statistics, there is on average 123% more movement than initial uncertainty. The splits by bin show that these values decrease for return states in the middle of the distribution. The beginning of the sample features high but noisy X^* statistics, but these averages remain high until the most recent subsample. Finally, higher priors π_0^* correspond with greater excess movement, which accords with our theoretical bounds.

The next two columns of Table 2 instead normalize $\overline{X}^*$ by the average contract length $\overline{T}$, so the resulting statistics can be interpreted as excess movement per day. For a rough understanding of the actual variation in RN beliefs corresponding to these values, consider a pair $(t, t+1)$ for which $\pi_{t+1}^* = 1 - \pi_t^*$, so that there is no uncertainty resolution from t to $t+1$. One-day excess movement and movement therefore coincide, so that excess movement is equal to the squared change in beliefs; for example, the noise-adjusted average of 0.0038 corresponds to a raw change of $\sqrt{0.0038} \approx 0.06$ (continuing the example above, $\pi_t^* = 0.47$, and then $\pi_{t+1}^* = 0.53$). Under this normalization, there is now no clear pattern for average excess movement across bins, as the more-extreme return states also tend to have longer contract lengths.[38] This pattern is in fact quite consistent in our data: longer contract lengths tend to coincide with more excess movement, as RN beliefs bounce up and down over the length of a contract. This general pattern is inconsistent with the null of Bayesian updating, which tells us that excess movement in subjective beliefs should not depend at all on the horizon at which uncertainty is resolved. Meanwhile, for the splits by date and by prior, the basic patterns discussed above are still present under this normalization by $\overline{T}$.

Comparing the raw and noise-adjusted values makes clear that despite the substantial excess movement in the noise-adjusted statistics, noise does represent a meaningful portion (about 1/3) of the raw X^* data. The raw and noise-adjusted mean for $\overline{X}^*/\overline{T}$ differ by $0.0059 - 0.0038 \approx 0.002$, so $\widehat{\text{Var}}(\epsilon_t) \approx 0.002/2 = 0.001$. The standard deviation of ϵ_t is thus roughly 0.03 per day. For the splits by state, noise variances tend to be lowest for return states near the center of the distribution, as is intuitive.

Next, Figure 2 provides a visual summary of the X^* statistics relative to the bounds. The blue curves describe the raw and noise-adjusted local-average X^* statistics as one varies the RN prior π_0^*,

[38] Recall that an observation for a given state pair is only included conditional on the realized return being in one of the two states. Longer contracts are likelier to generate greater absolute returns, explaining this positive covariance.

calculated using all interior state pairs and streams; these curves are the same in both panels. As one would expect, there is very little excess movement for RN priors near 0 or 1, but excess movement is positive for intermediate π_0^* values for which there is greater initial uncertainty. We compare these values to the theoretical bounds under different levels of ϕ (or more precisely, $\mathbb{E}[\phi_i]$) for each value of π_0^*, which are shown in gray. Panel (a) uses the tighter bound from Proposition 9(i), which we implement by estimating $\triangle$ using local averages for the conditional expectations $\mathbb{E}[X^* \mid \theta, \pi_0^*]$ in (17) over π_0^*.[39] Panel (b) uses the conservative bound from the second inequality in Proposition 9(ii), so the gray lines in this panel align with the solid bound lines in Figure 1.

Across the two panels, it is evident that the values of X^* observed in the data exceed both sets of bounds, other than for high RN priors π_0^* and for large values of ϕ. For panel (a), we estimate $\widehat{\triangle} = 0$ for $\pi_0^* \approx 0.5$, $\widehat{\triangle} < 0$ for π_0^* below this cutoff, and $\widehat{\triangle} > 0$ above it, as is evident from the bounds crossing zero at $\pi_0^* \approx 0.5$. As discussed in Section 2.3 (see pages 15–16), this indicates that the DGP is close to symmetric, with equally informative signals for the two states (θ_j, θ_{j+1}) on average (and thus roughly equal-sized upward and downward movements of π_t^*). As the bounds in Proposition 9 apply for each possible π_0^*, the positive point estimates for $\mathbb{E}[X^* \mid \pi_0^*]$ clearly violate the bounds for $\pi_0^* < 0.5$, which are (at most) 0 for all ϕ. And while the empirical curves are closer to the more-conservative bounds in panel (b), the noise-adjusted estimates still exceed π_0^{*2} (the bound for $\phi = \infty$) for π_0^* less than about $2/3$.

These figures do not, however, integrate over π_0^* to provide a single statistic summarizing excess movement relative to the bounds over all streams. More importantly, they do not include any measures of statistical uncertainty necessary to make inferential statements or conduct hypothesis tests. To address these two issues, we move on to our main estimation and results.

5.4 Main Results

We turn now to the empirical implementation of our theoretical bounds. Given our sample of noise-adjusted excess movement statistics and corresponding RN priors, each possible value of $\overline{\phi}$ leads to a *residual* excess movement value $e_i(\overline{\phi}) = X_i^* - \text{bound}(\pi_{0,i}^*, \overline{\phi})$ for contract i (where we continue to suppress dependence on state pair j). We calculate two versions of this residual corresponding to the bound in part (i) and the unconditional bound in part (ii) of Proposition 9 (equations (26) and (27), respectively):

$$\begin{aligned} e_i^{\triangle}(\overline{\phi}) &= X_i^* - \max\left\{0, \left(\pi_{0,i}^* - \frac{\pi_{0,i}^*}{\overline{\phi} + (1-\overline{\phi})\pi_{0,i}^*}\right)\widehat{\triangle}_i\right\}, \\ e_i^{\text{main}}(\overline{\phi}) &= X_i^* - \left(\pi_{0,i}^* - \frac{\pi_{0,i}^*}{\overline{\phi} + (1-\overline{\phi})\pi_{0,i}^*}\right)\pi_{0,i}^*. \end{aligned} \tag{29}$$

[39] We slightly modify the formula in Proposition 9(i) for the figure: we use $\mathbb{E}[X_i^*] \leqslant (\pi_{0,i}^* - \frac{\pi_{0,i}^*}{\mathbb{E}[\phi_i] + (1-\mathbb{E}[\phi_i])\pi_{0,i}^*})\mathbb{E}[\triangle_i]$, rather than cutting the bounds off at 0, to clarify that $\widehat{\triangle} < 0$ for $\pi_0^* < 0.5$.

The first version corresponds to the tighter bound from Proposition 2, which requires a smoothed estimate of $\triangle_i$ as calculated for Figure 2. This version accounts for the estimated DGP (through $\widehat{\triangle}_i$) and thus conveys some useful preliminary information, but strictly speaking, part (i) of the proposition applies only conditional on a given $\pi^*_{0,i}$ and only under the unverifiable condition that $\text{Cov}(\pi_{0,i}, \triangle_i) = 0$.[40] Thus the second version, which implements the more conservative bound from Proposition 3 and which applies unconditionally, is the basis for our main set of results. Both residuals can be directly calculated using our noise-adjusted data for each possible value of $\overline{\phi}$.

We first present sample averages of these residual statistics for a range of values of $\overline{\phi}$, both for each individual return state pair j and aggregated across all interior state pairs. As these values have no natural scaling, we present them in the form of t-statistics, $t_{\bar{e}} = \overline{e_i(\phi)} / \widehat{SE}_{\bar{e}}$, where $\overline{e_i(\phi)}$ is the sample average of residual excess movement $e_i(\overline{\phi})$ and $\widehat{SE}_{\bar{e}}$ is its standard error. We calculate these standard errors using a block bootstrap with a block length of one month, with each block containing (i) raw excess movement statistics and priors for all contracts expiring in a given month, and (ii) noise variance estimates for any trading days in our intraday sample that fall in the same month. For each resampled data set, we use the set of $(X^*_i, \widetilde{\pi}^*_{0,i}, \{\widehat{\text{Var}}(\epsilon_{t,i})\})$ values to recalculate noise-adjusted excess movement and residual values $e_i(\overline{\phi})$.[41] The bootstrap accordingly accounts for sampling uncertainty in all statistics used to calculate noise-adjusted excess movement and $e_i(\overline{\phi})$. We conduct 10,000 such draws, from which we calculate standard errors as the standard deviation of $\overline{e_i(\phi)}$ across bootstrap draws.

These residual t-statistics are presented in Table 3 for the two versions of the residual in (29), both overall (for all interior state pairs) and by individual return state pair. Since $\mathbb{E}[e_i(\overline{\phi})] = 0$ under RE given a correctly specified value $\overline{\phi}$, these t-statistics tell us how far the residuals are from being consistent with any hypothesized null value for the SDF slope. Positive numbers correspond to the data exhibiting too much excess movement to be consistent with a given value of $\overline{\phi}$, and vice versa for negative numbers. In panel (a), the only negative t-statistics for the tighter bound are for RN beliefs over $(\theta_j, \theta_{j+1}) = ([0\%, 5\%], [5\%, 10\%])$ (i.e., $j = 6$) for $\overline{\phi} > 10$; all other t-statistics, including the overall values, are positive (and generally large in magnitude), indicating no value of $\overline{\phi}$ is consistent with the degree of observed excess RN movement. In panel (b), the average conservative bound residuals are smaller in magnitude but also generally positive, other than for large $\overline{\phi}$ (with magnitudes discussed below) and for return-state pairs in the middle of the distribution.

As admissible excess movement $\mathbb{E}[X^*_i]$ is monotonically increasing in the unobserved parameter $\overline{\phi}$, our main empirical exercise is to estimate the lower bound for this SDF slope such that the bound for $\mathbb{E}[X^*_i]$ is satisfied. This lower bound for $\overline{\phi}$ is estimated as the minimal value for which the average residual value $\overline{e_i(\phi)}$ is zero, so that we are effectively finding the root of the function traced out in Table 3. Given that the tighter bound is generally not satisfied even for $\overline{\phi} = \infty$ (and since it holds only under restrictive assumptions), we now confine attention to the conservative

[40] It is also stated for $\mathbb{E}[\phi_i]$ rather than $\overline{\phi}$, but the bound still applies using the more conservative value $\overline{\phi}$ as well.

[41] For the residual $e^{\triangle}_i(\overline{\phi})$ corresponding to the tighter bound, we also re-estimate $\widehat{\triangle}_i$ in each bootstrap draw by calculating the same local average (with respect to $\pi^*_{0,i}$) as in Figure 2(a), and then evaluating it at the observed $\pi^*_{0,i}$.

bound from Proposition 9(ii). The estimated $\overline{\phi}$ can be interpreted as the minimal SDF slope for which the amount of observed excess movement in RN beliefs can be rationalized; it is accordingly an index of the restrictiveness of the joint assumptions of RE and CTI (or, given Proposition 11, RE and supermartingale ϕ_t). As above, we estimate this SDF slope both for each individual state pair ($\overline{\phi} = \overline{\phi}_j$) and over all interior state pairs ($\overline{\phi} = \overline{\boldsymbol{\phi}}$). To make the estimates for $\overline{\phi}$ more readily interpretable, we also use Proposition 8 to translate them into local relative risk aversion values $\overline{\gamma}$ for a fictitious representative agent who consumes the market.

Table 4 presents our main results. Whenever there is *no* value of $\overline{\phi}$ for which the bound for $\mathbb{E}[X_i^*]$ is satisfied — i.e., from Proposition 9(iii), when we estimate $\mathbb{E}[X_i^*] > \mathbb{E}[{\pi_{0,i}^*}^2]$ — we write $\overline{\phi} = \infty$. In brackets below each point estimate, we provide the lower bound of a one-sided 95% confidence interval (CI) for the parameter in question.[42] Starting from the same bootstrap resampling procedure as described just above, we obtain these CIs by inverting a one-sided test: the CI lower bound $\widehat{\phi}_{LB}$ is the minimal $\overline{\phi}$ such that $\overline{e_i^{\text{main}}(\phi)} = 0$ is not rejected at the 5% level using the bootstrapped data (see Appendix B.9 for further details). The first column of the table provides estimates for the minimal SDF slopes $\overline{\phi}$. The second column translates these figures into relative risk aversion values using $\overline{\gamma} = \frac{(\overline{\phi}-1)}{0.05} = (\overline{\phi} - 1) \times 20$ from (25), since adjacent return states have midpoints 5 percent (more precisely, 5 log points) apart by construction.

Starting with the overall estimates in the first row: the point estimate for the conservative lower bound for $\overline{\boldsymbol{\phi}}$ is slightly greater than 50, in line with the t-statistic in Table 3(b) for $\overline{\boldsymbol{\phi}} = 50$ (which is slightly positive but close to zero). This translates to an extraordinarily high estimated lower bound for $\overline{\gamma}$ of 1,075. Values of $\overline{\boldsymbol{\phi}}$ below 9.8 (for $\overline{\gamma}$, below 175) are rejected at the 5% level. We conclude that extremely high risk aversion is needed to rationalize the large degree of excess movement in RN beliefs observed in the data.

Moving to the local estimates for the individual return-state pairs, all but two of the point estimates of the lower bounds for $\overline{\phi}_j$ and $\overline{\gamma}_j$ are infinite, indicating that no amount of utility curvature (or SDF slope) can rationalize the observed excess movement such that the bounds are satisfied. For many of the states ($j = 2, 3, 4$, and 8), their associated confidence intervals also have lower bounds of ∞ (or more precisely, are empty), indicating outright model rejection. Only for RN beliefs over the state pairs in the middle of the distribution — in particular, $j = 5$ and 6, corresponding to excess-return midpoints of 0% and 5%, respectively — have finite point estimates and confidence intervals that contain reasonable risk-aversion values of about 20. RN beliefs over these intermediate return states are thus comparatively well-behaved; for all other states, there is so much mean-reverting variation in RN beliefs that the bounds are only met for implausibly large values of $\overline{\phi}_j$, if at all.

In light of Proposition 6, we conclude that belief revisions are excessively volatile in all cases for which the data cannot be rationalized with finite risk aversion, as these findings cannot in general be produced solely by miscalibrated priors. Further, the large local risk-aversion bounds at every point of the return distribution and the extremely large overall estimates (which use only beliefs

[42] The set identification implied by our theoretical bound motivates our use of these one-sided intervals.

over the excess-return states between -20% and +20%) imply that no feature of the true underlying data-generating process (e.g., volatility in the left tail of the return distribution) can by itself be responsible for these findings of excess belief volatility.

5.5 Suggestive Evidence on the Correlates of RN Excess Movement

We conclude our empirical analysis by briefly considering reduced-form evidence on the macroeconomic and financial correlates of RN excess movement. Online Appendix Table A.1 shows the results of a set of time-series regressions to this end. The dependent variable is the quarterly average of noise-adjusted RN excess movement $X^*_{t,t+1,i,j}$. We consider as regressors: (1) proxies for option liquidity and trading activity; (2) measures of volatility and macroeconomic uncertainty; and (3) statistics related to index returns and valuations. Across all specifications, the liquidity- and noise-related variables (bid-ask spreads and volume) have coefficients that are both economically and statistically small, which gives further evidence that factors specific to the option market (or mismeasurement of RN beliefs) are not driving our results. By contrast, excess movement has a significant positive relationship with the VIX, as is intuitive (and further evidence that our measure is in fact reflective of excess volatility). Lagged S&P 500 returns and valuation ratios are also positively related to excess movement. The R^2 value for the regression with all right-hand-side variables included is 0.61, indicating that these statistics are capable of jointly accounting for much of the quarterly variation in excess movement. That said, this regression evidence is only suggestive and reduced-form.

6. Conclusion

We consider a general theoretical framework in which we show that the assumption of rational expectations imposes testable restrictions on the time variation in risk-neutral beliefs as expressed in asset values. Unlike in much of the previous literature, these results do not require any restrictive assumptions on the data-generating process, and they allow for time variation in discount rates. Further, by using asset values, we do not require direct measures of physical beliefs over future outcomes, and our bounds exploit intertemporal consistency requirements of rational beliefs without the need for the econometrician to know what agents' beliefs "should" be under RE.

When taken to the data using observed asset prices, our bounds provide evidence on the rationality of the market as a whole. Using risk-neutral beliefs over the future return on the S&P 500 index measured from options data, we find that under our assumption of conditional transition independence, extremely high risk aversion is needed to rationalize the variation in these beliefs. We conclude that the RE assumption appears to be quite restrictive. In many cases, no amount of risk aversion (or SDF slope) is capable of rationalizing the behavior of RN beliefs.

While our bounds are informative in ruling out classes of models (or modeling assumptions), this paper only provides basic suggestive evidence as to the potential causes of our bound violations. We believe, though, that there are numerous feasible ways to make progress on this

question. Conducting additional tests on the empirical correlates of excess movement, as well as generalizing our analysis to alternative asset classes, may provide useful additional information. Further, detailed data on changes in individual portfolios could allow for tests on the rationality of individual beliefs, which would help distinguish between micro and macro explanations for the observed excess movement in RN beliefs. We leave these possibilities to future work.

References

ALVAREZ, F. AND U. J. JERMANN (2005): "Using Asset Prices to Measure the Persistence of the Marginal Utility of Wealth," *Econometrica*, 73, 1977–2016.

AUGENBLICK, N. AND M. RABIN (2021): "Belief Movement, Uncertainty Reduction, and Rational Updating," *Quarterly Journal of Economics*, 136, 933–985.

BAKSHI, G. AND F. CHABI-YO (2012): "Variance Bounds on the Permanent and Transitory Components of Stochastic Discount Factors," *Journal of Financial Economics*, 105, 191–208.

BARNDORFF-NIELSEN, O. E. AND N. SHEPHARD (2001): "Non-Gaussian Ornstein-Uhlenbeck-Based Models and Some of Their Uses in Financial Economics," *Journal of the Royal Statistical Society, Series B*, 63, 167–241.

BARRO, R. J. (2006): "Rare Disasters and Asset Markets in the Twentieth Century," *Quarterly Journal of Economics*, 121, 823–866.

VAN BINSBERGEN, J. H., M. BRANDT, AND R. KOIJEN (2012): "On the Timing and Pricing of Dividends," *American Economic Review*, 102, 1596–1618.

VAN BINSBERGEN, J. H., W. F. DIAMOND, AND M. GROTTERIA (2022): "Risk-Free Interest Rates," *Journal of Financial Economics*, 143, 1–29.

BLUME, L., T. COURY, AND D. EASLEY (2006): "Information, Trade and Incomplete Markets," *Economic Theory*, 29, 379–394.

BOLLEN, N. P. AND R. E. WHALEY (2004): "Does Net Buying Pressure Affect the Shape of Implied Volatility Functions?" *Journal of Finance*, 59, 711–753.

BORDALO, P., N. GENNAIOLI, Y. MA, AND A. SHLEIFER (2020): "Overreaction in Macroeconomic Expectations," *American Economic Review*, 110, 2748–2782.

BOROVIČKA, J., L. P. HANSEN, AND J. A. SCHEINKMAN (2016): "Misspecified Recovery," *Journal of Finance*, 71, 2493–2544.

BREEDEN, D. T. AND R. H. LITZENBERGER (1978): "Prices of State-Contingent Claims Implicit in Option Prices," *Journal of Business*, 51, 621–651.

CAMPBELL, J. Y. AND J. H. COCHRANE (1999): "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107, 205–251.

CAMPBELL, J. Y. AND R. J. SHILLER (1987): "Cointegration and Tests of Present Value Models," *Journal of Political Economy*, 95, 1062–1088.

CHRISTOFFERSEN, P., S. HESTON, AND K. JACOBS (2013): "Capturing Option Anomalies with a Variance-Dependent Pricing Kernel," *Review of Financial Studies*, 26, 1962–2006.

COCHRANE, J. H. (2011): "Presidential Address: Discount Rates," *Journal of Finance*, 66, 1047–1108.

COCHRANE, J. H. AND J. SAÁ-REQUEJO (2000): "Beyond Arbitrage: Good-Deal Asset Price Bounds in Incomplete Markets," *Journal of Political Economy*, 108, 79–119.

CONSTANTINIDES, G. M., J. C. JACKWERTH, AND A. SAVOV (2013): "The Puzzle of Index Option Returns," *Review of Asset Pricing Studies*, 3, 229–257.

D'ARIENZO, D. (2020): "Maturity Increasing Overreaction and Bond Market Puzzles," *Working Paper*.

DE BONDT, W. F. M. AND R. THALER (1985): "Does the Stock Market Overreact?" *Journal of Finance*, 40, 793–805.

DE LA O, R. AND S. MYERS (2021): "Subjective Cash Flow and Discount Rate Expectations," *Journal of Finance*, 76, 1339–1387.

FAMA, E. F. (1991): "Efficient Capital Markets: II," *Journal of Finance*, 46, 1575–1617.

GABAIX, X. (2012): "Variable Rare Disasters: An Exactly Solved Framework for Ten Puzzles in Macro-Finance," *Quarterly Journal of Economics*, 127, 645–700.

GARCIA, R., E. GHYSELS, AND É. RENAULT (2010): "The Econometrics of Option Pricing," in *Handbook of Financial Econometrics*, ed. by Y. Aït-Sahalia and L. P. Hansen, Amsterdam: Elsevier, vol. 1, chap. 9, 479–552.

GIGLIO, S. AND B. KELLY (2018): "Excess Volatility: Beyond Discount Rates," *Quarterly Journal of Economics*, 133, 71–127.

GREENWOOD, R. AND A. SHLEIFER (2014): "Expectations of Returns and Expected Returns," *Review of Financial Studies*, 27, 714–746.

HANSEN, L. P. AND R. JAGANNATHAN (1991): "Implications of Security Market Data for Models of Dynamic Economies," *Journal of Political Economy*, 99, 225–262.

JAMSHIDIAN, F. (1989): "An Exact Bond Option Formula," *Journal of Finance*, 44, 205–209.

KLEIDON, A. W. (1986): "Variance Bounds Tests and Stock Price Valuation Models," *Journal of Political Economy*, 94, 953–1001.

LAZARUS, E. (2019): "Horizon-Dependent Risk Pricing: Evidence from Short-Dated Options," *Working Paper*.

LEROY, S. F. AND R. D. PORTER (1981): "The Present-Value Relation: Tests Based on Implied Variance Bounds," *Econometrica*, 49, 555–574.

LI, Z. M. AND O. B. LINTON (2021): "A ReMeDI for Microstructure Noise," *Forthcoming, Econometrica*.

MALZ, A. M. (2014): "A Simple and Reliable Way to Compute Option-Based Risk-Neutral Distributions," *Federal Reserve Bank of New York Staff Report No. 677*.

MARSH, T. A. AND R. C. MERTON (1986): "Dividend Variability and Variance Bounds Tests for the Rationality of Stock Market Prices," *American Economic Review*, 76, 483–498.

MARTIN, I. (2017): "What Is the Expected Return on the Market?" *Quarterly Journal of Economics*, 132, 367–433.

MARTIN, I. AND D. PAPADIMITRIOU (2021): "Sentiment and Speculation in a Market with Heterogeneous Beliefs," *Working Paper*.

MEHRA, R. AND E. C. PRESCOTT (1985): "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, 15, 145–161.

MILGROM, P. AND N. STOKEY (1982): "Information, Trade, and Common Knowledge," *Journal of Economic Theory*, 26, 17–27.

QIN, L. AND V. LINETSKY (2017): "Long-Term Risk: A Martingale Approach," *Econometrica*, 85, 299–312.

RADNER, R. (1979): "Rational Expectations Equilibrium: Generic Existence and the Information Revealed by Prices," *Econometrica*, 47, 655.

ROSS, S. (2015): "The Recovery Theorem," *Journal of Finance*, 70, 615–648.

SHILLER, R. J. (1981): "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?" *American Economic Review*, 71, 421–436.

STEIN, J. (1989): "Overreactions in the Options Market," *Journal of Finance*, 44, 1011–1023.

ZHANG, L., P. A. MYKLAND, AND Y. AÏT-SAHALIA (2005): "A Tale of Two Time Scales," *Journal of the American Statistical Association*, 100, 1394–1411.

Tables and Figures

Table 1: Example DGPs, Beliefs, Movement and Uncertainty Resolution Statistics

Signals	Stream Likelihood	Physical Beliefs	Movement & Init. Uncertainty		RN Beliefs ($\phi = 3$)	RN Movement & Init. Uncertainty	
$[s_1, s_2, \ldots]$	$\mathbb{P}([s_1, \ldots])$	$[\pi_0, \pi_1, \ldots]$	m	u_0	$[\pi_0^*, \pi_1^*, \ldots]$	m^*	u_0^*
(i) Symmetric signals with $L_l \equiv \frac{\mathbb{P}[s_t=l\|\theta=0]}{\mathbb{P}[s_t=l\|\theta=1]} = 3$ and $L_h \equiv \frac{\mathbb{P}[s_t=h\|\theta=1]}{\mathbb{P}[s_t=h\|\theta=0]} = 3$, and then resolution							
$[l, L]$	.5625	$[.25, .1, 0]$	.0325	.1875	$[.5, .25, 0]$	.125	.25
$[l, H]$	.0625	$[.25, .1, 1]$	.8325	.1875	$[.5, .25, 1]$	.625	.25
$[h, L]$	.1875	$[.25, .5, 0]$	.3125	.1875	$[.5, .75, 0]$	.625	.25
$[h, H]$	.1875	$[.25, .5, 1]$	.3125	.1875	$[.5, .75, 1]$	.125	.25
		$\mathbb{E}[\mathsf{m}] = .1875 = \mathsf{u}_0$ $\implies \mathbb{E}[X] = 0$			$\mathbb{E}[\mathsf{m}^*] = .25 = \mathsf{u}_0^*$ $\implies \mathbb{E}[X^*] = 0$, $\triangle = 0$		
(ii) Asymmetric signals with $L_l = 3$ and $L_h \to \infty$, and then resolution							
$[l, l, L]$	.75	$[.25, .1, .04, 0]$	.0279	.1875	$[.5, .25, .1, 0]$	.095	.25
$[H]$	.1667	$[.25, 1, 1, 1]$	.5625	.1875	$[.5, 1, 1, 1]$	.250	.25
$[l, H]$	.0567	$[.25, .1, 1, 1]$	.8325	.1875	$[.5, .25, 1, 1]$	.625	.25
$[l, l, H]$	.0278	$[.25, .1, .04, 1]$	.9565	.1875	$[.5, .25, .1, 1]$	.895	.25
		$\mathbb{E}[\mathsf{m}] = .1875 = \mathsf{u}_0$ $\implies \mathbb{E}[X] = 0$			$\mathbb{E}[\mathsf{m}^*] = .1725$, $\mathsf{u}_0^* = .25$ $\implies \mathbb{E}[X^*] =$ -.0775, $\triangle =$ -.31		
(iii) Asymmetric signals with $L_l \to \infty$ and $L_h = 3$, and then resolution							
$[h, h, H]$	.25	$[.25, .5, .75, 1]$	.1875	.1875	$[.5, .75, .9, 1]$	.095	.25
$[L]$	.5	$[.25, 0, 0, 0]$	.0625	.1875	$[.5, .0, 0, 0]$	.250	.25
$[h, L]$	.1667	$[.25, .5, 0, 0]$	.3125	.1875	$[.5, .75, 0, 0]$	.625	.25
$[h, h, L]$	.0833	$[.25, .5, .75, 1]$	.6875	.1875	$[.5, .75, .9, 0]$	.895	.25
		$\mathbb{E}[\mathsf{m}] = .1875 = \mathsf{u}_0$ $\implies \mathbb{E}[X] = 0$			$\mathbb{E}[\mathsf{m}^*] = .3275$, $\mathsf{u}_0^* = .25$ $\implies \mathbb{E}[X^*] = .0775$, $\triangle = .31$		

Note: This table provides theoretical examples for physical and risk-neutral belief statistics under rational expectations for three possible DGPs, as described in Section 2.1.

Table 2: Descriptive Statistics for Excess Movement

	$\overline{X}^*/\overline{u}_0$		$\overline{X}^*/\overline{T}$				
	Raw	Noise-Adj.	Raw	Noise-Adj.	$\overline{u}_0$	$\overline{T}$	N (Obs.)
Overall mean:	1.89	1.23	0.0059	0.0038	0.18	56	1,809
[Bootstrapped SE]	[0.25]	[0.22]	[0.0015]	[0.0013]	[0.00]	[2]	
By return state:							
1 (-20%)	5.83	4.83	0.0049	0.0041	0.17	200	26
	[1.18]	[1.05]	[0.0027]	[0.0024]	[0.01]	[20]	
2 (-15%)	11.61	5.70	0.0180	0.0088	0.22	141	19
	[3.32]	[3.06]	[0.0096]	[0.0083]	[0.01]	[25]	
3 (-10%)	5.76	2.37	0.0151	0.0062	0.21	81	49
	[0.99]	[1.07]	[0.0059]	[0.0051]	[0.01]	[12]	
4 (-5%)	2.67	1.39	0.0088	0.0046	0.14	42	272
	[0.59]	[0.50]	[0.0038]	[0.0029]	[0.01]	[5]	
5 (0%)	0.70	0.47	0.0045	0.0030	0.23	37	700
	[0.16]	[0.14]	[0.0019]	[0.0017]	[0.00]	[2]	
6 (+5%)	1.71	1.14	0.0039	0.0026	0.11	49	567
	[0.35]	[0.34]	[0.0015]	[0.0014]	[0.01]	[3]	
7 (+10%)	3.87	2.92	0.0053	0.0040	0.18	129	144
	[1.00]	[1.03]	[0.0023]	[0.0023]	[0.01]	[9]	
8 (+15%)	5.65	5.26	0.0060	0.0056	0.21	200	58
	[1.48]	[1.48]	[0.0027]	[0.0027]	[0.01]	[11]	
9 (+20%)	3.44	2.09	0.0032	0.0020	0.22	232	36
	[0.89]	[1.27]	[0.0015]	[0.0020]	[0.01]	[9]	
By date:							
1996–2000	10.89	9.67	0.0211	0.0187	0.21	107	109
	[2.24]	[2.17]	[0.0074]	[0.0072]	[0.01]	[11]	
2001–2005	1.75	0.55	0.0042	0.0013	0.22	90	112
	[0.51]	[0.40]	[0.0021]	[0.0015]	[0.01]	[11]	
2006–2010	1.25	0.68	0.0065	0.0035	0.17	32	502
	[0.22]	[0.19]	[0.0026]	[0.0021]	[0.00]	[4]	
2011–2015	1.75	1.09	0.0050	0.0031	0.19	67	530
	[0.36]	[0.28]	[0.0024]	[0.0019]	[0.00]	[5]	
2016–2018	0.36	-0.11	0.0011	-0.0003	0.16	50	556
	[0.21]	[0.14]	[0.0017]	[0.0009]	[0.00]	[3]	
By π_0^**:*							
0–0.25	1.01	0.30	0.0055	0.0017	0.09	16	185
	[0.60]	[0.55]	[0.0048]	[0.0047]	[0.01]	[2]	
0.25–0.5	1.58	0.91	0.0067	0.0039	0.23	55	883
	[0.21]	[0.17]	[0.0017]	[0.0014]	[0.00]	[3]	
0.5–0.75	2.84	2.19	0.0053	0.0041	0.23	123	284
	[0.61]	[0.58]	[0.0020]	[0.0019]	[0.00]	[7]	
0.75–1	2.54	1.88	0.0048	0.0036	0.06	31	457
	[0.95]	[0.90]	[0.0031]	[0.0029]	[0.00]	[2]	

Notes: Empirical conditional means of risk-neutral excess movement $\overline{X}^* \equiv \widehat{\mathbb{E}}[X^*_{i,j}]$ are calculated over all interior state pairs $j = 2, \ldots, 8$, aside from averages by bin, which are calculated for each state pair separately. Standard errors are estimated using a block bootstrap for the normalized statistic $\overline{X}^*/\overline{u}_0$ or $\overline{X}^*/\overline{T}$, with a block size of one month (where contracts are classified by the month in which they expire) and 10,000 draws.

Table 3: Residual Excess Movement t-Statistics for Different $\overline{\phi}$

(a) Tighter Bound from Proposition 9*(i)*

	$\overline{\phi} = 1$	2	5	10	50	∞
Overall t-stat.:	5.19	5.15	4.98	4.64	3.03	0.78
By return state:						
1 (-20%)	4.06	3.69	3.04	2.57	1.94	1.72
2 (-15%)	2.13	2.13	2.13	2.13	2.14	2.14
3 (-10%)	2.23	2.23	2.23	2.23	2.23	2.23
4 (-5%)	2.78	2.78	2.77	2.77	2.76	2.76
5 (0%)	3.09	3.09	3.08	3.08	3.06	3.06
6 (+5%)	3.26	2.66	1.27	0.04	-1.71	-2.42
7 (+10%)	2.82	2.73	2.51	2.29	1.70	1.23
8 (+15%)	3.53	3.49	3.42	3.37	3.29	3.25
9 (+20%)	1.63	1.58	1.52	1.48	1.44	1.43

(b) Conservative Bound from Proposition 9*(ii)*

	$\overline{\phi} = 1$	2	5	10	50	∞
Overall t-stat.:	5.19	4.08	2.68	1.75	0.07	-1.48
By return state:						
1 (-20%)	4.06	3.58	2.79	2.22	1.47	1.21
2 (-15%)	2.13	2.03	1.94	1.90	1.87	1.86
3 (-10%)	2.23	2.03	1.86	1.79	1.72	1.70
4 (-5%)	2.78	2.51	2.28	2.18	2.09	2.06
5 (0%)	3.09	1.86	0.74	0.26	-0.19	-0.31
6 (+5%)	3.26	1.96	-0.12	-1.77	-5.22	-8.39
7 (+10%)	2.82	2.38	1.73	1.28	0.50	0.06
8 (+15%)	3.53	3.20	2.77	2.51	2.21	2.10
9 (+20%)	1.63	1.36	1.05	0.89	0.72	0.68

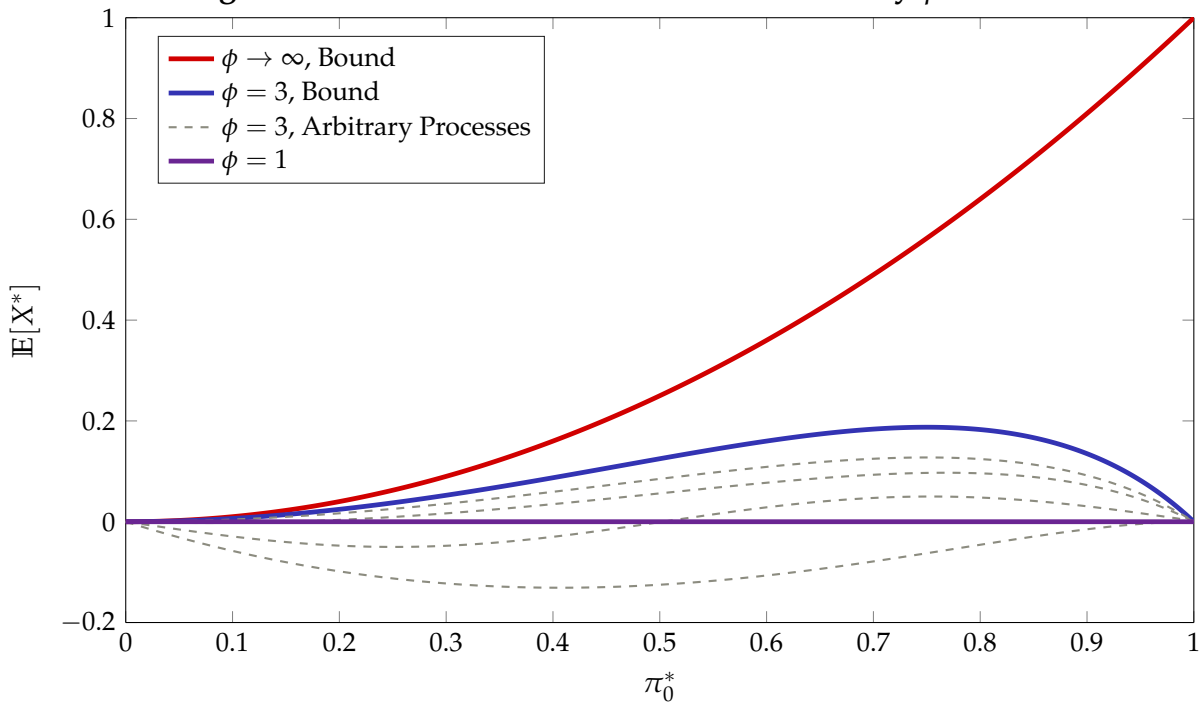
Notes: Both panels report t-statistics for the average residuals $\overline{e_i(\phi)}$ in (29) for different values of $\overline{\phi}$. Panel (a) uses $e_i^{\triangle}(\overline{\phi})$, with $\triangle_i$ estimated as in panel (a) of Figure 2. Panel (b) uses $e_i^{\text{main}}(\overline{\phi})$. Standard errors are estimated using a block bootstrap with block size of one month and 10,000 draws. All statistics are calculated using conditional means of noise-adjusted excess movement.

Table 4: Main Estimation Results

	Conservative Lower Bound for:	
	SDF Slope $\overline{\phi}$	RRA $\overline{\gamma}$
Overall bound:	54.7	1,075
[95% CI Lower Bound]	[9.8]	[175]
By return state:		
1 (-20%)	∞ [24.2]	∞ [464]
2 (-15%)	∞ [∞]	∞ [∞]
3 (-10%)	∞ [∞]	∞ [∞]
4 (-5%)	∞ [∞]	∞ [∞]
5 (0%)	19.4 [2.1]	368 [22]
6 (+5%)	4.8 [2.2]	75 [24]
7 (+10%)	∞ [4.6]	∞ [73]
8 (+15%)	∞ [∞]	∞ [∞]
9 (+20%)	∞ [1.0]	∞ [1]

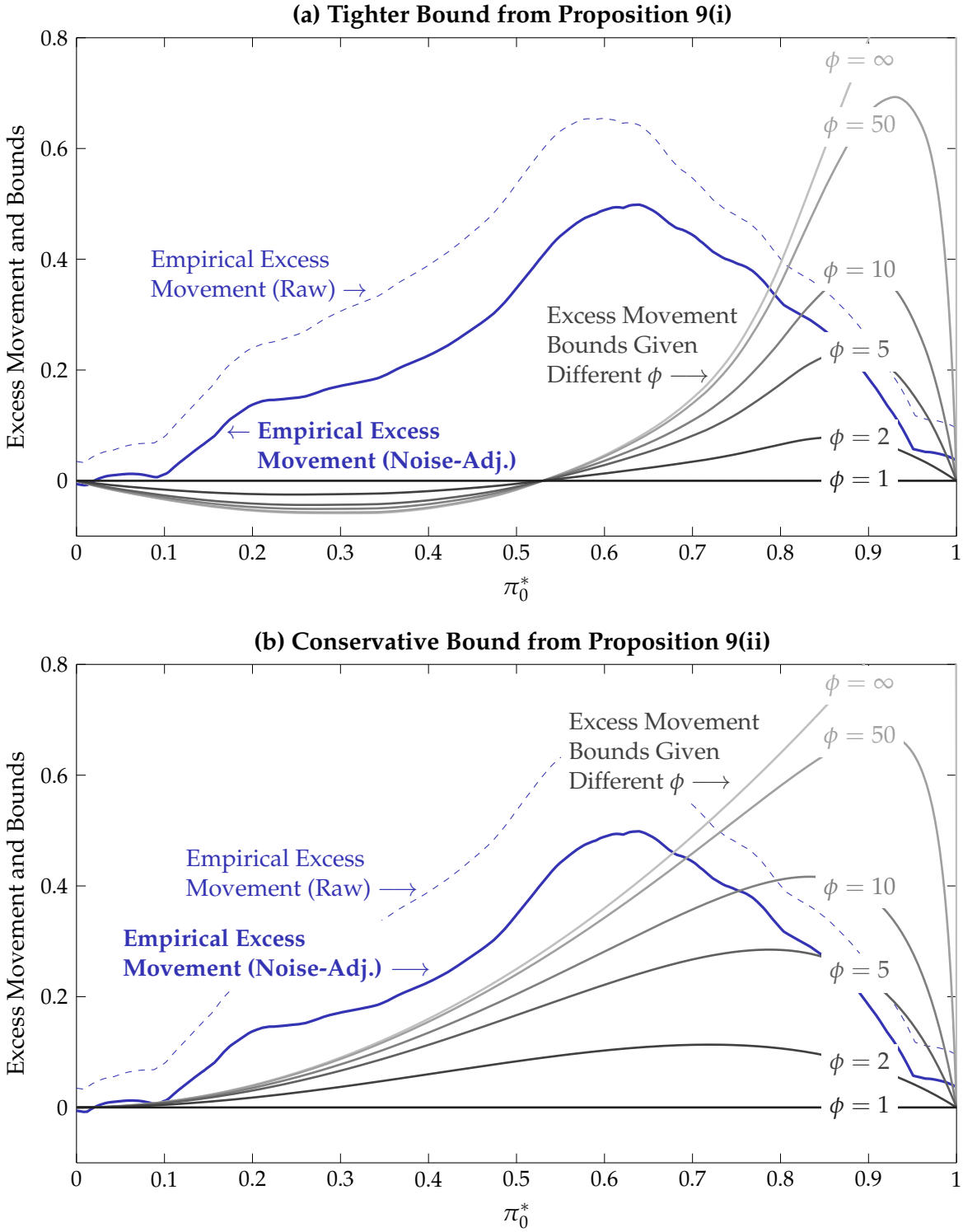
Notes: The first column reports estimates for the minimal value of $\overline{\phi}$ satisfying the conservative bound for excess movement in Proposition 9(ii). These estimates are translated to relative risk aversion values $\overline{\gamma}$ using Proposition 8, as shown in the second column. Point estimates are obtained by finding the value $\overline{\phi}$ such that $\overline{e_i^{\text{main}}(\phi)} = 0$ in (29). Confidence interval lower bounds are obtained by inverting a test for $\overline{\phi}$ using bootstrapped data; see Appendix B.9 for details. All estimates use conditional means of noise-adjusted excess movement.

Figure 1: RN Excess Belief Movement vs. Prior by ϕ Under RE



Note: Theoretical bounds are obtained from the formulas in Proposition 3 and Corollary 1.

Figure 2: Excess Movement vs. RN Prior: Data and Theoretical Bounds



Notes: Empirical excess movement curves are kernel-weighted local averages (Epanechnikov kernel, bandwidth for π_0^* of 0.07) over all interior state pairs $j = 2, \ldots, 8$. All statistics are estimates of conditional means $\widetilde{\mathbb{E}}[\cdot]$ for RN beliefs $\widetilde{\pi}^*_{t,i,j}$, and theoretical curves correspond to $\phi \equiv \widetilde{\mathbb{E}}[\phi_{i,j}]$, with notation simplified for clarity. Bounds for (a) obtain $\widehat{\triangle}$ using a kernel-weighted local average over π_0^* for each of the two terms in (17), with $\widehat{\triangle}$ then plugged into the inequality in Proposition 9(i). Bounds for (b) use only the second inequality in Proposition 9(ii).

FOR ONLINE PUBLICATION

Online Appendix for:

Restrictions on Asset-Price Movements Under Rational Expectations: Theory and Evidence*

Ned Augenblick
UC Berkeley Haas

Eben Lazarus
MIT Sloan

JANUARY 2022

Contents

*Contact: ned@haas.berkeley.edu and elazarus@mit.edu.

A. Proofs of Theoretical Results

A.1 Proofs for Section 2

Section 2.1

Proof of Proposition 1. Following Augenblick and Rabin (2021), it is useful to define period-by-period movement, uncertainty reduction, and excess movement, respectively, as

$$\begin{aligned}
\mathsf{m}_{t,t+1}(\boldsymbol{\pi}) &\equiv (\pi_{t+1} - \pi_t)^2, \\
\mathsf{r}_{t,t+1}(\boldsymbol{\pi}) &\equiv \pi_t(1-\pi_t) - \pi_{t+1}(1-\pi_{t+1}), \\
X_{t,t+1}(\boldsymbol{\pi}) &\equiv \mathsf{m}_{t,t+1}(\boldsymbol{\pi}) - \mathsf{r}_{t,t+1}(\boldsymbol{\pi}).
\end{aligned}$$

Given the definitions of movement, initial uncertainty, and excess movement in the text, note that

$$\mathsf{m}(\boldsymbol{\pi}) = \textstyle\sum_{t=0}^{T-1} \mathsf{m}_{t,t+1}(\boldsymbol{\pi}), \quad \mathsf{u}_0(\boldsymbol{\pi}) = \sum_{t=0}^{T-1} \mathsf{r}_{t,t+1}(\boldsymbol{\pi}), \quad X(\boldsymbol{\pi}) = \sum_{t=0}^{T-1} x_{t,t+1}(\boldsymbol{\pi}),$$

where the second equality relies on the fact that $\pi_T \in \{0,1\}$ and therefore $\pi_T(1-\pi_T) = 0$ for any belief stream $\boldsymbol{\pi}$. We have that

$$\begin{aligned}
\mathbb{E}[X_{t,t+1}|H_t] &= \mathbb{E}[\mathsf{m}_{t,t+1} - \mathsf{r}_{t,t+1}|H_t] \\
&= \mathbb{E}[(\pi_{t+1} - \pi_t)^2 - ((\pi_t(1-\pi_t) - (\pi_{t+1}(1-\pi_{t+1}))|H_t] \\
&= \mathbb{E}[(2\pi_t - 1)(\pi_t - \pi_{t+1})|H_t] \\
&= (2\pi_t(H_t) - 1)(\pi_t(H_t) - \mathbb{E}[\pi_{t+1}|H_t]) \\
&= (2\pi_t(H_t) - 1) \cdot 0 \\
&= 0,
\end{aligned}$$

where the fifth line uses the martingale beliefs assumption (Assumption 1). Summing and applying the law of iterated expectations (LIE),

$$\mathbb{E}[X] = \sum_{t=0}^{T-1} \mathbb{E}[X_{t,t+1}] = \sum_{t=0}^{T-1} \mathbb{E}[\mathbb{E}[X_{t,t+1}|H_t]] = 0. \qquad \square$$

Sections 2.2–2.3

Proof of Equations (12)–(14). For the physical measure,

$$\begin{aligned}
\mathbb{P}(H_T) &= \mathbb{P}(\theta = 1) \cdot \mathbb{P}(H_T|\theta = 1) + \mathbb{P}(\theta = 0) \cdot \mathbb{P}(H_T|\theta = 0) \\
&= \pi_0 \cdot \mathbb{P}(H_T|\theta = 1) + (1 - \pi_0) \cdot \mathbb{P}(H_T|\theta = 0), \qquad \text{(A.1)}
\end{aligned}$$

where the second line uses that $\pi_0 = \mathbb{E}[\pi_T] = \mathbb{E}[\theta] = \mathbb{P}(\theta = 1)$ (as shown after Assumption 1). Meanwhile, for the RN measure, we have from (11) and (A.1) that

$$\begin{aligned}\mathbb{P}^*(H_T) &= \frac{\pi_0^*}{\pi_0} \cdot \pi_0 \cdot \mathbb{P}(H_T|\theta = 1) + \frac{1-\pi_0^*}{1-\pi_0} \cdot (1-\pi_0) \cdot \mathbb{P}(H_T|\theta = 0) \\ &= \pi_0^* \cdot \mathbb{P}(H_T|\theta = 1) + (1-\pi_0^*) \cdot \mathbb{P}(H_T|\theta = 0). \end{aligned} \tag{A.2}$$

For any H_T such that $\pi_T = 1$, we have as well from (11) that $\mathbb{P}^*(H_T) = \frac{\pi_0^*}{\pi_0}\mathbb{P}(H_T)$, which implies $\mathbb{P}^*(\theta = 1) = \frac{\pi_0^*}{\pi_0}\mathbb{P}(\theta = 1)$. Thus by the definition of conditional probability, $\mathbb{P}^*(H_T|\theta = 1) = \mathbb{P}(H_T|\theta = 1)$. A similar argument gives $\mathbb{P}^*(H_T|\theta = 0) = \mathbb{P}(H_T|\theta = 0)$, and thus (14) holds. Then (A.2) becomes

$$\mathbb{P}^*(H_T) = \pi_0^* \cdot \mathbb{P}^*(H_T|\theta = 1) + (1-\pi_0^*) \cdot \mathbb{P}^*(H_T|\theta = 0).$$

Summing over all possible H_T for which $\theta = 1$ gives $\pi_0^* = \mathbb{P}^*(\theta = 1)$, so that $\mathbb{P}^*$ is a valid probability distribution for which LIE holds. Then noting $\mathbb{P}^*(\theta = 1) = \mathbb{E}^*[\theta] = \mathbb{E}^*[\pi_T] = \mathbb{E}^*[\pi_T^*]$, equation (12) follows.[1] Equation (13) then follows from Proposition 1. □

Proof of Equation (18). Footnote 10 provides a brief derivation. For a full derivation, first write

$$\begin{aligned}\triangle &\equiv \mathbb{E}^*[X^*|\theta = 0] - \mathbb{E}^*[X^*|\theta = 1] \\ &= \mathbb{E}^*[\mathsf{m}^*|\theta = 0] - \mathsf{u}_0^* - (\mathbb{E}^*[\mathsf{m}^*|\theta = 0] - \mathsf{u}_0^*) \\ &= \mathbb{E}^*[\mathsf{m}^*|\theta = 0] - \mathbb{E}^*[\mathsf{m}^*|\theta = 1]. \end{aligned} \tag{A.3}$$

Further, using equation (15),

$$\begin{aligned}0 &= \pi_0^* \cdot \mathbb{E}[X^*|\theta = 1] + (1-\pi_0^*) \cdot \mathbb{E}[X^*|\theta = 0] \\ &= \pi_0^* \cdot (\mathbb{E}[\mathsf{m}^*|\theta = 1] - \mathsf{u}_0^*) + (1-\pi_0^*) \cdot (\mathbb{E}[\mathsf{m}^*|\theta = 0] - \mathsf{u}_0^*),\end{aligned}$$

so using the definition of u_0^*,

$$\pi_0^* \cdot \mathbb{E}[\mathsf{m}^*|\theta = 1] + (1-\pi_0^*) \cdot \mathbb{E}[\mathsf{m}^*|\theta = 0] = \pi_0^*(1-\pi_0^*). \tag{A.4}$$

Solving for $\mathbb{E}[\mathsf{m}^*|\theta = 0]$ gives

$$\mathbb{E}[\mathsf{m}^*|\theta = 0] = \pi_0^* - \frac{\pi_0^*}{1-\pi_0^*} \cdot \mathbb{E}[\mathsf{m}^*|\theta = 1].$$

[1] Lemma A.1 provides a more detailed algebraic derivation of this fact.

Using this in (A.3),

$$\begin{aligned}\triangle &= \pi_0^* - \frac{\pi_0^*}{1-\pi_0^*} \cdot \mathbb{E}[\mathsf{m}^*|\theta=1] - \mathbb{E}^*[\mathsf{m}^*|\theta=1] \\ &= \pi_0^* - \frac{1}{1-\pi_0^*} \cdot \mathbb{E}[\mathsf{m}^*|\theta=1] \end{aligned} \tag{A.5}$$

Given that $\frac{1}{1-\pi_0^*} \geqslant 0$ and $\mathbb{E}[\mathsf{m}^*|\theta=1] \geqslant 0$, $\triangle$ is bounded above by π_0^*. □

Proof of Proposition 2. Start from equation (16) and apply equation (15):

$$\begin{aligned}\mathbb{E}[X^*] &= \pi_0 \cdot \mathbb{E}[X^*|\theta=1] + (1-\pi_0) \cdot \mathbb{E}[X^*|\theta=0] - 0 \\ &= \pi_0 \cdot \mathbb{E}[X^*|\theta=1] + (1-\pi_0) \cdot \mathbb{E}[X^*|\theta=0] - (\pi_0^* \cdot \mathbb{E}[X^*|\theta=1] + (1-\pi_0^*) \cdot \mathbb{E}[X^*|\theta=0]) \\ &= (\pi_0^* - \pi_0)(\mathbb{E}[X^*|\theta=0] - \mathbb{E}[X^*|\theta=1]) \\ &= (\pi_0^* - \pi_0)(\triangle), \end{aligned} \tag{A.6}$$

as stated. Then the second equality holds using equation (10) and the definition of $\triangle$. □

Proof of Proposition 3. From the proof of equation (18) above, we have $\triangle \leqslant \pi_0^*$. Further, equation (10) implies

$$\begin{aligned}\pi_0^* - \pi_0 &= \pi_0^* - \frac{\pi_0^*}{\pi_0^* + \phi(1-\pi_0^*)} \\ &= \pi_0^* \left(1 - \frac{1}{\pi_0^* + \phi(1-\pi_0^*)}\right) \geqslant 0, \end{aligned} \tag{A.7}$$

where the last inequality uses $\pi_0^* + \phi(1-\pi_0^*) \geqslant 0$ since $\phi \geqslant 1$. Using these two inequalities in the expression for $\mathbb{E}[X^*]$ in (A.6),

$$\mathbb{E}[X^*] = (\pi_0^* - \pi_0)(\triangle) \leqslant (\pi_0^* - \pi_0)\pi_0^*. \tag{A.8}$$

Plugging in the expression for $\pi_0^* - \pi_0$ in (A.7) then gives equation (19). □

Proof of Corollary 1. This is an immediate implication of (A.8) and $\pi_0 \geqslant 0$. □

Proof of Corollary 2. As in (A.7), we have $\pi_0^* - \pi_0 \geqslant 0$. Using this in the equality in (A.8) alongside the assumption that $\triangle = \mathbb{E}^*[\mathsf{m}^*|\theta=0] - \mathbb{E}^*[\mathsf{m}^*|\theta=1] \leqslant 0$ gives $\mathbb{E}[X^*] \leqslant 0$. □

Proof of Proposition 4. Consider a given ϕ, RN prior π_0^*, and signal DGPs $DGP(s_t|\theta=0,H_{t-1})$ and $DGP(s_t|\theta=1,H_{t-1})$ that lead to some $\mathbb{E}[X^*|\theta=0]$, $\mathbb{E}[X^*|\theta=1]$, and $\triangle$. Now consider the "reversed" DGP $\widehat{DGP}$ in which we modify the DGP by relabeling state 1 as state 0 and state 0 as state 1. That is, $\widehat{DGP}(s_t|\theta=0,H_{t-1}) \equiv DGP(s_t|\theta=1,H_{t-1})$ and $\widehat{DGP}(s_t|\theta=1,H_{t-1}) \equiv DGP(s_t|\theta=0,H_{t-1})$. Similarly, we consider the "reversed" RN prior $\widehat{\pi}_0^* = 1-\pi_0^*$ implied by the physical prior $\widehat{\pi}_0 = \frac{1-\pi_t^*}{\phi+(1-\phi)(1-\pi_t^*)}$.

As a result of this relabeling, if the RN belief in the original DGP following history H_t is $\pi_t^*(H_t)$, then the RN belief in the reversed $\widehat{DGP}$ with RN prior $1-\pi_0^*$ must be $\widehat{\pi}_t^*(H_t) = 1-\pi_t^*(H_t)$. Thus $\mathbb{E}^*[\widehat{X}^*|\theta=0] = \mathbb{E}^*[X^*|\theta=1]$ and $\mathbb{E}^*[\widehat{X}^*|\theta=1] = \mathbb{E}^*[X^*|\theta=0]$. Using that $\mathbb{E}^*[X^*|\theta] = \mathbb{E}[X^*|\theta]$ by equation (14) as proven above, this gives $\mathbb{E}[\widehat{X}^*|\theta=0] = \mathbb{E}[X^*|\theta=1]$ and $\mathbb{E}[\widehat{X}^*|\theta=1] = \mathbb{E}[X^*|\theta=0]$. We conclude that for $\widehat{DGP}$, $\widehat{\triangle} \equiv \mathbb{E}[\widehat{X}^*|\theta=0] - \mathbb{E}[\widehat{X}^*|\theta=1] = -\triangle$. □

Proof of Proposition 5. Consider a sequence of binary resolving DGPs indexed by T. There are two possible signals in each period, l and h, and assume that for any history,

$$DGP(s_t = h|\theta=1) = 1, \tag{A.9}$$

$$DGP(s_t = h|\theta=0) = \frac{\pi_{t-1}^*(1-\pi_{t-1}^*-\epsilon)}{(1-\pi_{t-1}^*)(\pi_{t-1}^*+\epsilon)}, \quad \text{with } \epsilon \equiv \frac{1-\pi_0^*}{T}. \tag{A.10}$$

Since $DGP(s_t = l|\theta=1) = 0$ from (A.9), beliefs (both physical and RN) update to 0 given any l signal. Meanwhile, after seeing h (and assuming no l signals through $t-1$), Bayes' rule gives that physical beliefs update to

$$\pi_t(\{s_1 = h, \ldots, s_t = h\}) = \frac{\pi_{t-1}}{\pi_{t-1} + (1-\pi_{t-1})DGP(s_t = h|\theta=0)}.$$

Applying the transformation (10) to the π_{t-1} values on the right side of this equation, we have after some algebra that

$$\pi_t(\{s_1 = h, \ldots, s_t = h\}) = \frac{\pi_{t-1}^*}{\pi_{t-1}^* + (1-\pi_{t-1}^*)\phi DGP(s_t = h|\theta=0)}.$$

Now applying the transformation (9), we obtain that π_t^* given an only-h signal history (suppressing the dependence on this history for simplicity) is, after additional tedious but straightforward algebra,

$$\pi_t^* = \frac{\pi_{t-1}^*}{\pi_{t-1}^* + (1-\pi_{t-1}^*)DGP(s_t = h|\theta=0)}.$$

Now using (A.10), we obtain after further algebra that

$$\pi_t^* - \pi_{t-1}^* = \epsilon.$$

Note given the definition of ϵ, then, that this DGP is resolving for any T: given any l signal at any t,

beliefs resolve to 0, while given only h signals, beliefs increase slowly ($\pi_t^* = \pi_0^* + t\epsilon$) and resolve to 1 at period T. We thus have

$$\mathbb{E}[\mathsf{m}^*|\theta = 1] = T\epsilon^2 = T\left(\frac{1-\pi_0^*}{T}\right)^2 = \frac{(1-\pi_0^*)^2}{T} \xrightarrow{T\to\infty} 0.$$

Thus for such a sequence, using equation (A.5),

$$\triangle = \pi_0^* - \frac{1}{1-\pi_0^*} \cdot \mathbb{E}[\mathsf{m}^*|\theta = 1] \xrightarrow{T\to\infty} \pi_0^*.$$

Using this in equation (A.6) gives $\mathbb{E}[X^*] \to (\pi_0^* - \pi_0)\pi_0^*$ as $T \to \infty$, as stated. And as further stated, the sequence of DGPs is constructed such that any downward movement is resolving and any upward movement is small ($\pi_t^* - \pi_{t-1}^* = \epsilon \to 0$). We have thus proven the first two statements in the proposition.

For the final statement, given $\phi > 1$ and $0 < \pi_0^* < 1$, the inequality in (A.7) is strict, so that $\pi_0^* - \pi_0 > 0$. Further, the only way to obtain $m^* = 0$ for finite T is if $\pi_0^* = \pi_1^* = \ldots = \pi_T^*$, which is ruled out by $0 < \pi_0^* < 1$ since $\pi_T^* = 0$ or 1 with probability 1, and therefore $\mathbb{E}[\mathsf{m}^*|\theta = 1] > 0$. Thus in (A.5), we have the strict inequality $\triangle < \pi_0^*$ for fixed $T < \infty$. Combining these in (A.6) gives $\mathbb{E}[X^*] < (\pi_0^* - \pi_0)\pi_0^*$ for fixed T, as stated. □

Proof of Proposition 6. For part (i), first define the likelihood of a prior π_0 as

$$\mathcal{L}(\pi_0) \equiv \frac{\pi_0}{1-\pi_0}, \tag{A.11}$$

and the likelihood of a signal s_t as

$$\mathcal{L}(s_t) \equiv \frac{DGP(s_t|\theta = 1)}{DGP(s_t|\theta = 0)},$$

where the dependence of the latter on the history H_{t-1} is left implicit for simplicity. The likelihood for any belief π_t is defined as well following (A.11). The above likelihoods are well-defined for interior priors (as we assume given finite L in the proposition) and for $DGP(s_t|\theta = 0, H_{t-1}) > 0$ (we return to the situation in which $DGP(s_t|\theta = 0, H_{t-1}) = 0$ shortly). Bayes' rule gives that beliefs satisfy

$$\mathcal{L}(\pi_t) = \mathcal{L}(\pi_0) \cdot \mathcal{L}(s_1) \cdot \mathcal{L}(s_2) \cdots \mathcal{L}(s_t).$$

Now note from (9) that

$$\mathcal{L}(\pi_0^*) \equiv \frac{\pi_0^*}{1-\pi_0^*} = \phi\frac{\pi_0}{1-\pi_0},$$

from which it follows that under Bayesian updating,

$$\begin{aligned}\mathcal{L}(\pi_t^*) &= \mathcal{L}(\pi_0^*) \cdot \mathcal{L}(s_1) \cdot \mathcal{L}(s_2) \cdots \mathcal{L}(s_t) \\ &= \phi\mathcal{L}(\pi_0) \cdot \mathcal{L}(s_1) \cdot \mathcal{L}(s_2) \cdots \mathcal{L}(s_t).\end{aligned}$$

For a fictitious agent with a rational prior, one could replace $\mathcal{L}(\pi_0)$ with $\mathcal{L}(\mathbb{P}_0(\theta = 1))$. In our case, given the incorrect prior (but correct Bayesian updating), we have

$$\frac{\pi_t^*}{1-\pi_t^*} = \check{\phi}\frac{\mathbb{P}_0(\theta = 1)}{1-\mathbb{P}_0(\theta = 1)},$$

where $\check{\phi} \equiv \phi L$, with L defined as in the proposition. We can therefore write

$$\mathcal{L}(\pi_t^*) = \check{\phi}\mathcal{L}(\mathbb{P}_0(\theta = 1)) \cdot \mathcal{L}(s_1) \cdot \mathcal{L}(s_2) \cdots \mathcal{L}(s_t).$$

As the likelihood ratio of the RN beliefs in this case are equal to those of a fictitious agent with a correct prior $\check{\pi}_0 = \mathbb{P}_0(\theta = 1)$ and $\check{\phi}$ in place of ϕ, we conclude that the RN beliefs are as well. Finally, for the case in which $DGP(s_t|\theta = 0, H_{t-1}) = 0$ and this signal s_t is observed, the person will update to $\pi_t = 1$, matching the belief of a rational agent again. We have thus shown part (i).

We can thus treat the agent with the incorrect prior as if she were fully rational (satisfying Assumption 1) but with $\check{\phi}$ in place of ϕ. We know as well that $\check{\phi}$ satisfies Assumption 3, since L is constant and ϕ is constant by that assumption as well. For part (ii) of the proposition, if $\check{\phi} \geqslant 1$, then Assumption 2 holds as well, so all three assumptions are satisfied, and all the stated results carry through.

For part (iii), assuming $0 < \check{\phi} < 1$ (so Assumption 2 no longer holds for the fictitious rational agent), note first that the proof of Proposition 2 never employs Assumption 2 and therefore still holds straightforwardly, as we can write $\mathbb{E}[X^*] = (\pi_0^* - \check{\pi}_0)\triangle$ without use of this assumption. For Proposition 3, the result as stated for a rational agent requires that $\pi_0^* > \check{\pi}_0$, which is not true for $\check{\phi} < 1$. An alternative bound, though, can be shown for this case, by obtaining a lower bound for $\triangle$ similar to the upper bound in equation (18). Starting from (A.4) but solving now for $\mathbb{E}[\mathsf{m}^*|\theta = 1]$, we have

$$\mathbb{E}[\mathsf{m}^*|\theta = 1] = (1 - \pi_0^*) - \frac{1-\pi_0^*}{\pi_0^*} \cdot \mathbb{E}[\mathsf{m}^*|\theta = 0].$$

Using this in (A.3),

$$\begin{aligned}\triangle &= \mathbb{E}[\mathsf{m}^*|\theta = 0] - \left((1 - \pi_0^*) - \frac{1-\pi_0^*}{\pi_0^*} \cdot \mathbb{E}[\mathsf{m}^*|\theta = 0]\right) \\ &= \frac{1}{\pi_0^*} \cdot \mathbb{E}[\mathsf{m}^*|\theta = 0] - (1 - \pi_0^*).\end{aligned}$$

Then, given that $\frac{1}{\pi_0^*} \geqslant 0$ and $\mathbb{E}[\mathsf{m}^*|\theta = 0] \geqslant 0$, $\triangle$ must be bounded below by $-(1 - \pi_0^*)$. Returning

to the formula from Proposition 3, if $\check{\phi} < 1$, then $\pi_0^* - \check{\pi}_0 \leqslant 0$, which gives

$$\begin{aligned}\mathbb{E}[X^*] &= (\pi_0^* - \check{\pi}_0)(\triangle) \\ &\leqslant (\check{\pi}_0 - \pi_0^*)(1 - \pi_0^*).\end{aligned}$$

Further, as $\check{\pi}_0 \leqslant 1$,

$$\begin{aligned}\mathbb{E}[X^*] &\leqslant (\check{\pi}_0 - \pi_0^*)(1 - \pi_0^*) \\ &\leqslant (1 - \pi_0^*)(1 - \pi_0^*) = (1 - \pi_0^*)^2.\end{aligned}$$

Thus taking (ii) and (iii) together, we have that $\mathbb{E}[X^*] \leqslant \max(\pi_0^{*2}, (1 - \pi_0^*)^2)$ given an incorrect prior. □

A.2 Proofs for Section 3

Section 3.1

Proof of Equation (22). This follows from a discrete-state application of Breeden and Litzenberger (1978), or see Brown and Ross (1991) for a general version. To review why the stated equation holds, the risk-neutral pricing equation for options can be written

$$q_{t,K}^m = \frac{1}{R_{t,T}^f}\mathbb{E}_t^*[\max\{V_T^m - K, 0\}] = \frac{1}{R_{t,T}^f}\left[\sum_{j\,:\,K_j \geqslant K}(K_j - K)\underbrace{\mathbb{P}_t^*(V_T^m = K_j)}_{\mathbb{P}_t^*(R_T^m = \theta_j)}\right].$$

This implies that for two adjacent return states θ_{j-1} and θ_j,

$$\begin{aligned}q_{t,K_j}^m - q_{t,K_{j-1}}^m &= \frac{1}{R_{t,T}^f}\left[\sum_{j' \geqslant j}(K_{j'} - K_j)\,\mathbb{P}_t^*(V_T^m = K_{j'}) - \sum_{j' \geqslant j-1}(K_{j'} - K_{j-1})\,\mathbb{P}_t^*(V_T^m = K_{j'})\right] \\ &= \frac{1}{R_{t,T}^f}\left[\sum_{j' \geqslant j}(K_{j-1} - K_j)\,\mathbb{P}_t^*(V_T^m = K_{j'})\right] = \frac{1}{R_{t,T}^f}(K_{j-1} - K_j)\left[1 - \mathbb{P}_t^*(V_T^m < K_j)\right].\end{aligned}$$

Rearranging,

$$R_{t,T}^f \frac{q_{t,K_j}^m - q_{t,K_{j-1}}^m}{K_j - K_{j-1}} = \mathbb{P}_t^*(V_T^m < K_j) - 1.$$

Repeating this analysis for the pair θ_j and θ_{j+1}, we obtain $R_{t,T}^f \frac{q_{t,K_{j+1}}^m - q_{t,K_j}^m}{K_{j+1} - K_j} = \mathbb{P}_t^*(V_T^m < K_{j+1}) - 1$. Subtracting the preceding equation from this equation and using $\mathbb{P}_t^*(R_T^m = \theta_j) = \mathbb{P}_t^*(V_T^m = K_j)$ yields equation (22). □

Section 3.2

Proof of Example 1. We prove the statement separately for the two assumptions on the form of the utility function:

(i) Time-separable utility: Denote $V_j^m \equiv V_0^m \theta_j$ and $V_{j+1}^m \equiv V_0^m \theta_{j+1}$, so the event $R_T^m = \theta_j$ is equivalent to $V_T^m = V_j^m$, and similarly for θ_{j+1} and V_{j+1}^m. Since $dV_T^m / dA_T > 0$ (and with $\mathbb{P}(V_T^m = V_j^m) > 0$, $\mathbb{P}(V_T^m = V_{j+1}^m) > 0$), there exist unique values a_j and a_{j+1} such that $V_T^m = V_j^m$ if and only if $A_T = a_j$, and $V_T^m = V_{j+1}^m$ if and only if $A_T = a_{j+1}$. Then with $M_{t,T} = \beta^{T-t} U'(C_T)/U'(C_t)$ given the assumptions for this example, we have

$$\begin{aligned}\phi_{t,j} \equiv \frac{\mathbb{E}_t[M_{t,T} \mid R_T^m = \theta_j]}{\mathbb{E}_t[M_{t,T} \mid R_T^m = \theta_{j+1}]} &= \frac{\mathbb{E}_t[M_{t,T} \mid A_T = a_j]}{\mathbb{E}_t[M_{t,T} \mid A_T = a_{j+1}]} \\ &= \frac{U'(C_T(a_j))}{U'(C_T(a_{j+1}))},\end{aligned}$$

which is almost surely constant, as required for CTI to hold.

(ii) Epstein–Zin (1989) utility: The Epstein–Zin (1989) preference recursion is

$$U_t = \left\{ (1-\beta) C_t^{1-\frac{1}{\psi}} + \beta \left(\mathbb{E}_t \left[U_{t+1}^{1-\gamma} \right] \right)^{\frac{1-\frac{1}{\psi}}{1-\gamma}} \right\}^{\frac{1}{1-\frac{1}{\psi}}}. \tag{A.12}$$

It can be shown (e.g., Campbell, 2018, p. 178) that given such preferences the SDF evolves according to

$$M_{t,t+1} = \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\frac{1}{\psi}} \left(\frac{U_{t+1}}{\mathbb{E}_t \left[U_{t+1}^{1-\gamma} \right]^{\frac{1}{1-\gamma}}} \right)^{-\left(\gamma - \frac{1}{\psi}\right)},$$

which gives that

$$\begin{aligned} M_{t,T} &= M_{t,t+1} M_{t+1,t+2} \cdots M_{T-1,T} \\ &= \beta^{T-t} \left(\frac{C_T}{C_t} \right)^{-\frac{1}{\psi}} \prod_{\tau=t}^{T-1} \left(\frac{U_{\tau+1}}{\mathbb{E}_\tau \left[U_{\tau+1}^{1-\gamma} \right]^{\frac{1}{1-\gamma}}} \right)^{-\left(\gamma - \frac{1}{\psi}\right)} \qquad \text{(A.13)} \\ &= \beta^{T-t} \left(\frac{C_T}{C_t} \right)^{-\gamma} \prod_{\tau=t}^{T-1} \left(\frac{U_{\tau+1}}{C_{\tau+1}} \right)^{-\left(\gamma - \frac{1}{\psi}\right)} \mathbb{E}_\tau \left[\left(\frac{C_{\tau+1}}{C_\tau} \right)^{1-\gamma} \left(\frac{U_{\tau+1}}{C_{\tau+1}} \right)^{1-\gamma} \right]^{\frac{\gamma - \frac{1}{\psi}}{1-\gamma}}. \qquad \text{(A.14)} \end{aligned}$$

Denote a_j and a_{j+1} as in part (i). From the first representation of $M_{t,T}$, equation (A.13),

it follows immediately that with i.i.d. consumption (or i.i.d. innovations to an otherwise predetermined consumption path),

$$\begin{aligned}\phi_{t,j} &= \frac{\mathbb{E}_t[M_{t,T} \mid A_T = a_j]}{\mathbb{E}_t[M_{t,T} \mid A_T = a_{j+1}]} \\ &= \left(\frac{C_T(a_j)}{C_T(a_{j+1})}\right)^{-\frac{1}{\psi}} \left(\frac{U_T(a_j)}{U_T(a_{j+1})}\right)^{-\left(\gamma-\frac{1}{\psi}\right)},\end{aligned}$$

which is almost surely constant given the definition (A.12) and that $E_\tau[U_{T+1}^{1-\gamma}]$ is constant given the i.i.d. assumption. When consumption growth C_t/C_{t-1} is i.i.d., note that the scale independence of Epstein–Zin (1989) utility in (A.12) allows us to guess and verify that U_t/C_t is constant almost surely. Then from the second representation of $M_{t,T}$, equation (A.14), we have in this case that

$$\phi_{t,j} = \left(\frac{C_T(a_j)}{C_T(a_{j+1})}\right)^{-\gamma}. \qquad \square$$

Proof of Example 2. Gabaix (2012, Theorem 1) shows that

$$V_t^m = \frac{D_t}{1-e^{-\beta_m}}\left(1 + \frac{e^{-\beta_m - h_*}\widehat{H}_t}{1-e^{-\beta_m-\phi_H}}\right),$$

where $h_* \equiv \log(1+H_*)$ and $\beta_m \equiv \beta - g_d - h_*$ (where β is the agent's time discount factor). Thus for any value θ and given H_0, there exists some value d_θ and function $f(d_\theta, \widehat{H}_T)$, which is strictly increasing in the first argument and strictly decreasing in the second argument, such that, by Bayes' rule,

$$\begin{aligned}&\mathbb{P}_0\left(\sum_{t=1}^T \mathbb{1}\{\text{disaster}_t\} > 0 \,\middle|\, R_T^m \geqslant \theta\right) \\ &= \frac{\mathbb{P}_0\left(R_T^m \geqslant \theta \mid \sum_{t=1}^T \mathbb{1}\{\text{disaster}_t\} > 0\right)\mathbb{P}_0\left(\sum_{t=1}^T \mathbb{1}\{\text{disaster}_t\} > 0\right)}{\mathbb{P}_0(R_T^m \geqslant \theta)} \\ &= \frac{\mathbb{P}_0\left(D_T \geqslant f(d_\theta, \widehat{H}_T) \,\middle|\, \sum_{t=1}^T \mathbb{1}\{\text{disaster}_t\} > 0\right)\mathbb{P}_0\left(\sum_{t=1}^T \mathbb{1}\{\text{disaster}_t\} > 0\right)}{\mathbb{P}_0\left(D_T \geqslant f(d_\theta, \widehat{H}_T)\right)}.\end{aligned}$$

Note now that (i) the innovation to $\widehat{H}_{t+1}$ is independent of the disaster realization; (ii) F_{t+1} (the exponential of the disaster shock to D_t) has support $[0,1]$; and (iii) $\mathbb{P}_t(\varepsilon_{t+1}^d \geqslant \epsilon) = o(e^{-\epsilon^2})$ as $\epsilon \to \infty$.[2] Thus $\mathbb{P}_0(D_T \geqslant f(d_\theta, \widehat{H}_T) \mid \sum_{t=1}^T \mathbb{1}\{\text{disaster}_t\} > 0) = o(\mathbb{P}_0(D_T \geqslant f(d_\theta, \widehat{H}_T)))$ as $d_\theta \to \infty$,

[2]To see why point (iii) holds, denote $\sigma_d \equiv \text{Var}(\varepsilon_t^d)$ and then note that $\int_\epsilon^\infty \exp(-x^2/(2\sigma_d^2))/\sqrt{2\pi\sigma_d^2}\,dx <$

from which the first statement given in the example follows. Denote the value δ in that statement by $\delta = \delta_0$. Then it follows immediately that for any $t > 0$ (with $t < T$), for any $\delta_t > 0$, there exists an $\underline{\theta}$ such that $\mathbb{P}_t(\sum_{\tau=1}^{T} \mathbb{1}\{\text{disaster}_t\} > 0 \mid R_T^m \geqslant \underline{\theta}) < \delta_t$ asymptotically $\mathbb{P}_0$-a.s. as $\delta_0 \to 0$.

Thus moving to the second statement, given a value $\delta_t > 0$, consider θ_j, θ_{j+1} large enough that $\mathbb{P}_t(\sum_{\tau=1}^{T} \mathbb{1}\{\text{disaster}_t\} > 0 \mid R_T^m \in \{\theta_j, \theta_{j+1}\}) < \delta_t$. We then have from (24) that

$$\begin{aligned}
\phi_{t,j} &= \frac{\mathbb{E}_t[M_{t,T} \mid R_T^m = \theta_j]}{\mathbb{E}_t[M_{t,T} \mid R_T^m = \theta_{j+1}]} \\
&= \frac{\begin{array}{c}\mathbb{E}_t[M_{t,T} \mid R_T^m = \theta_j, \sum_{\tau=1}^{T} \mathbb{1}\{\text{disaster}_\tau\} = 0]\, \mathbb{P}_t(\sum_{\tau=1}^{T} \mathbb{1}\{\text{disaster}_\tau\} = 0 \mid R_T^m = \theta_j) \\ + \mathbb{E}_t[M_{t,T} \mid R_T^m = \theta_j, \sum_{\tau=1}^{T} \mathbb{1}\{\text{disaster}_\tau\} > 0]\, \mathbb{P}_t(\sum_{\tau=1}^{T} \mathbb{1}\{\text{disaster}_\tau\} > 0 \mid R_T^m = \theta_j)\end{array}}{\begin{array}{c}\mathbb{E}_t[M_{t,T} \mid R_T^m = \theta_{j+1}, \sum_{\tau=1}^{T} \mathbb{1}\{\text{disaster}_\tau\} = 0]\, \mathbb{P}_t(\sum_{\tau=1}^{T} \mathbb{1}\{\text{disaster}_\tau\} = 0 \mid R_T^m = \theta_{j+1}) \\ + \mathbb{E}_t[M_{t,T} \mid R_T^m = \theta_{j+1}, \sum_{\tau=1}^{T} \mathbb{1}\{\text{disaster}_\tau\} > 0]\, \mathbb{P}_t(\sum_{\tau=1}^{T} \mathbb{1}\{\text{disaster}_\tau\} > 0 \mid R_T^m = \theta_{j+1})\end{array}} \\
&= \frac{\mathbb{E}_t[M_{t,T} \mid R_T^m = \theta_j, \sum_{\tau=1}^{T} \mathbb{1}\{\text{disaster}_\tau\} = 0](1 - \mathcal{O}(\delta_t)) + \mathcal{O}(\delta_t)}{\mathbb{E}_t[M_{t,T} \mid R_T^m = \theta_{j+1}, \sum_{\tau=1}^{T} \mathbb{1}\{\text{disaster}_\tau\} = 0](1 - \mathcal{O}(\delta_t)) + \mathcal{O}(\delta_t)} \\
&= \frac{\mathbb{E}_t[M_{t,T} \mid R_T^m = \theta_j, \sum_{\tau=1}^{T} \mathbb{1}\{\text{disaster}_\tau\} = 0]}{\mathbb{E}_t[M_{t,T} \mid R_T^m = \theta_{j+1}, \sum_{\tau=1}^{T} \mathbb{1}\{\text{disaster}_\tau\} = 0]} + \mathcal{O}(\delta_t).
\end{aligned}$$

Note that the fraction in the last expression is constant almost surely given that conditional on $\sum_{t=1}^{T} \mathbb{1}\{\text{disaster}_t\} = 0$, the conditions from Example 1 hold. Thus denoting

$$\phi_j \equiv \frac{\mathbb{E}_0[M_{t,T} \mid R_T^m = \theta_j, \sum_{t=1}^{T} \mathbb{1}\{\text{disaster}_\tau\} = 0]}{\mathbb{E}_0[M_{t,T} \mid R_T^m = \theta_{j+1}, \sum_{t=1}^{T} \mathbb{1}\{\text{disaster}_\tau\} = 0]},$$

we have $\phi_{t,j} = \phi_j + \mathcal{O}(\delta_t)$. Since we can take $\delta_t \to 0$ asymptotically $\mathbb{P}_0$-a.s. as $\delta_0 \to 0$, we have $\phi_{t,j} = \phi_j + o_p(1)$ for any sequence of values $\delta = \delta_0 \to 0$. □

Proof of Example 3. As in Campbell and Cochrane (1999), the SDF evolves according to

$$M_{t,t+1} = \beta \left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} \left(\frac{S_{t+1}^c}{S_t^c}\right)^{-\gamma},$$

with terms defined as in Appendix B.4, and thus

$$\frac{\mathbb{E}_t[M_{t,T} \mid R_T^m = \theta_j]}{\mathbb{E}_t[M_{t,T} \mid R_T^m = \theta_{j+1}]} = \frac{\mathbb{E}_t\left[\exp\left(\sum_{\tau=0}^{T-t-1} -\gamma\left(1 + \lambda(s_{t+\tau}^c)\right)\varepsilon_{t+\tau+1}\right) \,\middle|\, R_T^m = \theta_j\right]}{\mathbb{E}_t\left[\exp\left(\sum_{\tau=0}^{T-t-1} -\gamma\left(1 + \lambda(s_{t+\tau}^c)\right)\varepsilon_{t+\tau+1}\right) \,\middle|\, R_T^m = \theta_{j+1}\right]}.$$

$\int_\epsilon^\infty (x/\epsilon)\exp(-x^2/(2\sigma_d^2))/\sqrt{2\pi\sigma_d^2}\,dx = \sigma_d \exp(-\epsilon^2/(2\sigma_d^2))/(\sqrt{2\pi}\epsilon)$. A similar calculation can be used to derive a lower bound for the upper tail of the normal CDF. Then applying the previous upper-bound calculation to $\mathbb{P}_0(D_T \geqslant f(d_\theta, \widehat{H}_T) \mid \sum_{t=1}^{T} \mathbb{1}\{\text{disaster}_t\} > 0)$ and the lower-bound calculation to $\mathbb{P}_0(D_T \geqslant f(d_\theta, \widehat{H}_T))$, it follows that $\mathbb{P}_0(D_T \geqslant f(d_\theta, \widehat{H}_T) \mid \sum_{t=1}^{T} \mathbb{1}\{\text{disaster}_t\} > 0)/\mathbb{P}_0(D_T \geqslant f(d_\theta, \widehat{H}_T)) = o(1)$, as stated, since the distribution of the value in the denominator is shifted to the right relative to the distribution of the value in the numerator given (i)–(ii).

For a counterexample to the constant-ϕ_t restriction, set $T = 2$ and $c_t = d_t$ (i.e., for simplicity, consumption and dividends are identical, as in the simplest case considered by Campbell and Cochrane, 1999, so the market is a consumption claim). Note that a sufficient condition for time variation in ϕ_t is

$$\mathrm{Cov}_0(\phi_1, \mathbb{E}_1[M_{1,2} \mid R_2^m = \theta_{j+1}]) \neq 0, \tag{A.15}$$

as this implies $\mathbb{E}_0[\phi_1] \neq \phi_0$ (see after Proposition 11). As of $t = 0$, both ε_1 and ε_2 are relevant for R_2^m and $M_{0,2}$, as ε_1 determines s_1^c and thus the conditional volatility $\lambda(s_1^c)$ of surplus consumption s_2^c. Meanwhile, as of $t = 1$ (i.e., conditional on ε_1), the only source of uncertainty for both R_2^m and $M_{1,2}$ is ε_2: s_2^c and d_2 together determine R_2^m, and conditional on time-1 variables, these both depend only on ε_2. Thus write ε_j^1 for the realization of ε_2 needed to generate $R_2^m = \theta_j$ conditional on ε_1 — i.e., $\varepsilon_j^1 \equiv \{\varepsilon_2 : R_2^m = \theta_j \mid \varepsilon_1\}$ — and similarly write ε_{j+1}^1 for the realization of ε_2 needed for $R_2^m = \theta_{j+1}$ conditional on ε_1. We then have

$$\mathbb{E}_1[M_{1,2} \mid R_2^m = \theta_{j'}] = \exp\left(-\gamma\left(1 + \lambda(s_1^c)\right)\varepsilon_{j'}^1\right)$$

for $j' = j, j+1$, and thus

$$\phi_1 = \exp\left(-\gamma\left(1 + \lambda(s_1^c)\right)\left(\varepsilon_j^1 - \varepsilon_{j+1}^1\right)\right).$$

The covariance in (A.15) is therefore

$$\mathrm{Cov}_0\left(\exp\left(-\gamma\left(1 + \lambda(s_1^c)\right)\left(\varepsilon_j^1 - \varepsilon_{j+1}^1\right)\right), \exp\left(-\gamma\left(1 + \lambda(s_1^c)\right)\varepsilon_{j+1}^1\right)\right).$$

Given Gaussian ε_1, this value is generically non-zero. □

Additional Lemmas Used in Proofs for Section 3.3

Before proceeding to the proof of our main results, we provide two additional lemmas that are useful in proving those results. As usual, assume throughout that Assumptions 1′–3′ hold.

Lemma A.1. *For some return-state pair (θ_j, θ_{j+1}), with $\widetilde{\mathbb{P}} \equiv \mathbb{P}(\cdot \mid R_T^m \in \{\theta_j, \theta_{j+1}\})$ as per (23), define a new pseudo-risk-neutral measure $\widetilde{\mathbb{P}}^\diamond$ by*

$$\left.\frac{d\widetilde{\mathbb{P}}^\diamond}{d\widetilde{\mathbb{P}}}\right|_{H_t} = \frac{\widetilde{\pi}_{t,j}^*}{\widetilde{\pi}_{t,j}}\mathbb{1}\{R_T^m = \theta_j\} + \frac{1 - \widetilde{\pi}_{t,j}^*}{1 - \widetilde{\pi}_{t,j}}\mathbb{1}\{R_T^m = \theta_{j+1}\}. \tag{A.16}$$

Denote the conditional expectation under $\widetilde{\mathbb{P}}^\diamond$ by $\widetilde{\mathbb{E}}_t^\diamond[\,\cdot\,]$. If conditional transition independence holds for the return-state pair (θ_j, θ_{j+1}), and $\mathbb{P}_t(R_T^m \in \{\theta_j, \theta_{j+1}\}) > 0$, we have that $\widetilde{\mathbb{P}}^\diamond$ serves as a martingale measure for the risk-neutral belief in the sense that

$$\widetilde{\pi}_{t,j}^* = \widetilde{\mathbb{E}}_t^\diamond[\widetilde{\pi}_{t+1,j}^*]. \tag{A.17}$$

We conclude from Proposition 1 that

$$\widetilde{\mathbb{E}}_0^{\diamond}[X_j^*] = 0. \tag{A.18}$$

Proof of Lemma A.1. From (23)–(24), we have after some algebra that

$$\frac{\widetilde{\pi}_{t,j}^*}{\widetilde{\pi}_{t,j}} = \frac{\phi_j}{1 + \widetilde{\pi}_{t,j}(\phi_j - 1)}, \tag{A.19}$$

$$\frac{1 - \widetilde{\pi}_{t,j}^*}{1 - \widetilde{\pi}_{t,j}} = \frac{1}{1 + \widetilde{\pi}_{t,j}(\phi_j - 1)}. \tag{A.20}$$

Note therefore that $\widetilde{\mathbb{P}}^{\diamond}$ is absolutely continuous with respect to $\widetilde{\mathbb{P}}$.

Recall that $H_t = \sigma(s_\tau, 0 \leqslant \tau \leqslant t)$, where $\sigma(s_\tau, 0 \leqslant \tau \leqslant t)$ is the σ-algebra generated by the stochastic process $\{s_t\}$ and $s_t \in \mathcal{S}$ is the date-t signal vector. Denote $N_{\mathcal{S}} \equiv |\mathcal{S}|$, so that $s_t \in \{s_1, s_2, \ldots, s_{N_{\mathcal{S}}}\}$, and further denote

$$\begin{aligned}
\mathfrak{p}_{t,k} &\equiv \widetilde{\mathbb{P}}_t(s_{t+1} = \theta_k), \\
\varrho_{t,k} &\equiv \widetilde{\mathbb{P}}_t(R_T^m = \theta_j \,|\, s_{t+1} = s_k), \\
\varrho_{t,k}^* &\equiv \mathbb{P}_t^*(R_T^m = \theta_j \,|\, s_{t+1} = s_k, R_T^m \in \{\theta_j, \theta_{j+1}\}),
\end{aligned}$$

so that $\widetilde{\pi}_{t+1,j} = \varrho_{t,k}$ if $s_{t+1} = s_k$, and similarly $\widetilde{\pi}_{t+1,j}^* = \varrho_{t,k}^*$ if $s_{t+1} = s_k$.

Combining (A.16), (A.19), (A.20), and these definitions, we have

$$\begin{aligned}
\widetilde{\mathbb{E}}_t^{\diamond}[\widetilde{\pi}_{t+1,j}^*] &= \frac{\widetilde{\pi}_{t,j}^*}{\widetilde{\pi}_{t,j}} \sum_{k=1}^{N_{\mathcal{S}}} \mathfrak{p}_{t,k}\, \varrho_{t,k}^*\, \widetilde{\mathbb{E}}_t\big[\mathbb{1}\{R_T^m = \theta_j\} \,|\, s_{t+1} = s_k\big] \\
&\quad + \frac{1 - \widetilde{\pi}_{t,j}^*}{1 - \widetilde{\pi}_{t,j}} \sum_{k=1}^{N_{\mathcal{S}}} \mathfrak{p}_{t,k}\, \varrho_{t,k}^*\, \widetilde{\mathbb{E}}_t\big[\mathbb{1}\{R_T^m = \theta_{j+1}\} \,|\, s_{t+1} = s_k\big] \\
&= \frac{\phi_j}{1 + \widetilde{\pi}_{t,j}(\phi_j - 1)} \sum_{k=1}^{N_{\mathcal{S}}} \mathfrak{p}_{t,k}\, \frac{\phi_j \varrho_{t,k}}{1 + \varrho_{t,k}(\phi_j - 1)}\, \varrho_{t,k} \\
&\quad + \frac{1}{1 + \widetilde{\pi}_{t,j}(\phi_j - 1)} \sum_{k=1}^{N_{\mathcal{S}}} \mathfrak{p}_{t,k}\, \frac{\phi_j \varrho_{t,k}}{1 + \varrho_{t,k}(\phi_j - 1)}\, (1 - \varrho_{t,k}) \\
&= \frac{\phi_j}{1 + \widetilde{\pi}_{t,j}(\phi_j - 1)} \sum_{k=1}^{N_{\mathcal{S}}} \mathfrak{p}_{t,k}\, \frac{\varrho_{t,k}\big(1 + \varrho_{t,k}(\phi_j - 1)\big)}{1 + \varrho_{t,k}(\phi_j - 1)} \\
&= \frac{\phi_j}{1 + \widetilde{\pi}_{t,j}(\phi_j - 1)} \sum_{k=1}^{N_{\mathcal{S}}} \mathfrak{p}_{t,k}\, \varrho_{t,k} \\
&= \frac{\phi_j \widetilde{\pi}_{t,j}}{1 + \widetilde{\pi}_{t,j}(\phi_j - 1)}
\end{aligned}$$

$$= \widetilde{\pi}^*_{t,j},$$

where the second-to-last equality uses that $\widetilde{\pi}_{t,j} = \widetilde{\mathbb{E}}_t[\widetilde{\pi}_{t+1,j}]$, as can be seen from the law of iterated expectations given that $\widetilde{\pi}_{t,j} = \mathbb{E}_t[\mathbb{1}\{R^m_T = \theta_j\} \mid R^m_T \in \{\theta_j, \theta_{j+1}\}] = \widetilde{\mathbb{E}}_t[\mathbb{1}\{R^m_T = \theta_j\}] = \widetilde{\mathbb{E}}_t[\widetilde{\mathbb{E}}_{t+1}[\mathbb{1}\{R^m_T = \theta_j\}]] = \widetilde{\mathbb{E}}_t[\widetilde{\pi}_{t+1,j}]$, and the last equality above again uses (A.19). The fact that $\widetilde{\mathbb{E}}^\diamond_0[X^*_j] = 0$ then follows immediately from the proof of Proposition 1. □

LEMMA A.2. *For any return-state pair (θ_j, θ_{j+1}) meeting CTI, risk-neutral belief movement must satisfy the following for $j' = j, j+1$:*

$$\widetilde{\mathbb{E}}^\diamond_0[\mathsf{m}^*_j \mid R^m_T = \theta_{j'}] = \widetilde{\mathbb{E}}_0[\mathsf{m}^*_j \mid R^m_T = \theta_{j'}]. \tag{A.21}$$

Proof of Lemma A.2. The stream of risk-neutral beliefs is $\boldsymbol{\pi}^*_j \equiv (\widetilde{\pi}^*_{0,j}, \widetilde{\pi}^*_{1,j}, \ldots, \widetilde{\pi}^*_{T,j})$, and denote some arbitrary realization for that path by $\mathfrak{b}_j$. The realization of m^*_j depends on the path of risk-neutral beliefs, so denote $\mathsf{m}^*_j = \mathsf{m}^*_j(\boldsymbol{\pi}^*_j) = \sum_{t=1}^T (\widetilde{\pi}^*_{t,j} - \widetilde{\pi}^*_{t-1,j})^2$.

For any $\mathfrak{b}_j$ such that $\widetilde{\pi}^*_{T,j} = 1$ (i.e., $R^m_T = \theta_j$), the definition of $\widetilde{\mathbb{P}}^\diamond$ in (A.16) gives that

$$\widetilde{\mathbb{P}}^\diamond_0(\boldsymbol{\pi}^*_j = \mathfrak{b}_j) = \frac{\widetilde{\pi}^*_{0,j}}{\widetilde{\pi}_{0,j}} \widetilde{\mathbb{P}}(\boldsymbol{\pi}^*_j = \mathfrak{b}_j), \tag{A.22}$$

and further $\widetilde{\mathbb{P}}^\diamond_0(R^m_T = \theta_j) = (\widetilde{\pi}^*_{0,j}/\widetilde{\pi}_{0,j})\, \widetilde{\mathbb{P}}_0(R^m_T = \theta_j)$ trivially. Combining these two equations yields $\widetilde{\mathbb{P}}^\diamond_0(\boldsymbol{\pi}^*_j = \mathfrak{b}_j \mid R^m_T = \theta_j) = \widetilde{\mathbb{P}}_0(\boldsymbol{\pi}^*_j = \mathfrak{b}_j \mid R^m_T = \theta_j)$. (Intuitively, all paths ending in $\widetilde{\pi}^*_{T,j} = 1$ receive the same change of measure under $\widetilde{\mathbb{P}}^\diamond$ relative to $\widetilde{\mathbb{P}}$, so probabilities conditional on $R^m_T = \theta_j$ are preserved, and similarly for $R^m_T = \theta_{j+1}$, as was the case for the simpler version of the RN measure in Section 2.) Thus

$$\begin{aligned}
\widetilde{\mathbb{E}}^\diamond_0[\mathsf{m}^*_j \mid R^m_T = \theta_j] &= \sum_{\mathfrak{b}_j\,:\,\widetilde{\pi}^*_{T,j}=1} \mathsf{m}^*_j(\mathfrak{b}_j)\, \widetilde{\mathbb{P}}^\diamond_0\Big(\boldsymbol{\pi}^*_j = \mathfrak{b}_j \,\Big|\, R^m_T = \theta_j\Big) \\
&= \sum_{\mathfrak{b}_j\,:\,\widetilde{\pi}^*_{T,j}=1} \mathsf{m}^*_j(\mathfrak{b}_j)\, \widetilde{\mathbb{P}}_0\Big(\boldsymbol{\pi}^*_j = \mathfrak{b}_j \,\Big|\, R^m_T = \theta_j\Big) \\
&= \widetilde{\mathbb{E}}_0[\mathsf{m}^*_j \mid R^m_T = \theta_j].
\end{aligned}$$

The same steps apply for $R^m_T = \theta_{j+1}$: for any $\mathfrak{b}_j$ such that $\widetilde{\pi}^*_{T,j} = 0$, (A.22) now becomes $\widetilde{\mathbb{P}}^\diamond_0(\boldsymbol{\pi}^*_j = \mathfrak{b}_j) = (1 - \widetilde{\pi}^*_{0,j})/(1 - \widetilde{\pi}_{0,j})\, \widetilde{\mathbb{P}}(\boldsymbol{\pi}^*_j = \mathfrak{b}_j)$. We also have in this case that $\widetilde{\mathbb{P}}^\diamond_0(R^m_T = \theta_{j+1}) = (1 - \widetilde{\pi}^*_{0,j})/(1 - \widetilde{\pi}_{0,j})\, \widetilde{\mathbb{P}}_0(R^m_T = \theta_{j+1})$, so again $\widetilde{\mathbb{P}}^\diamond_0(\boldsymbol{\pi}^*_j = \mathfrak{b}_j \mid R^m_T = \theta_{j+1}) = \widetilde{\mathbb{P}}_0(\boldsymbol{\pi}^*_j = \mathfrak{b}_j \mid R^m_T = \theta_{j+1})$, and thus $\widetilde{\mathbb{E}}^\diamond_0[\mathsf{m}^*_j \mid R^m_T = \theta_{j+1}] = \widetilde{\mathbb{E}}_0[\mathsf{m}^*_j \mid R^m_T = \theta_{j+1}]$. □

Note that the definition in (A.16) aligns with the definition of the RN measure in equation (11), so the two lemmas above prove the statements in the text connecting the RN measure in the simple case in Section 2 to the general case in Section 3 (see after equation (11) and equation (14), as well

as the footnote in the proof of equation (13) above). Indeed, equation (A.17) is the precise analogue to equation (14) in the text; equation (A.18) is the analogue to equation (13); and equation (A.21) implies immediately that $\widetilde{\mathbb{E}}_0^{\diamond}[X_j^* \mid R_T^m] = \widetilde{\mathbb{E}}_0[X_j^* \mid R_T^m]$, which was the main implication of equation (14) used in deriving the results in Section 2. We will thus be able to directly apply those results in this case using the above two lemmas, by virtue of these three results, as follows.

Section 3.3

Proof of Proposition 7. No arbitrage gives the existence of a positive SDF for which equation (24) and Assumption 2′ are valid. We have

$$\begin{aligned}
\widetilde{\pi}_{t,j} &= \mathbb{E}_t[\widetilde{\pi}_{t+1,j}], \\
\widetilde{\pi}_{t,j}^* &= \widetilde{\mathbb{E}}_t^{\diamond}[\widetilde{\pi}_{t+1,j}^*], \\
\widetilde{\mathbb{E}}_0^{\diamond}[X_j^*] &= 0, \\
\widetilde{\mathbb{E}}_0^{\diamond}[X_j^* \mid R_T^m] &= \widetilde{\mathbb{E}}_0[X_j^* \mid R_T^m],
\end{aligned}$$

where the first equality uses LIE and the remainder use Lemmas A.1–A.2 as above. The last equation in addition implies, using the same argument as applied for equation (18), that $\triangle_j \leqslant \widetilde{\pi}_{0,j}^*$. Further, Equations (9)–(10) hold immediately for $\widetilde{\pi}_{t,j}, \widetilde{\pi}_{t,j}^*, \phi_j$. We have thus obtained all the conditions used to prove Propositions 1–6 and Corollaries 1–2 given Assumptions 1–3, and thus under Assumptions 1′–3′ (for $j = 2, 3, \ldots, J-2$), those results continue to hold, with $\widetilde{\pi}_{t,j}^*$ replacing π_t^*, $\widetilde{\pi}_{t,j}$ replacing π_t, X_j^* replacing X^*, ϕ_j replacing ϕ, $\widetilde{\mathbb{E}}_0[\cdot]$ replacing $\mathbb{E}[\cdot]$, and with $\triangle_j \equiv \widetilde{\mathbb{E}}_0[X_j^* \mid R_T^m = \theta_{j+1}] - \widetilde{\mathbb{E}}_0[X_j^* \mid R_T^m = \theta_j]$ replacing $\triangle$, as stated. □

Proof of Proposition 8. The result follows immediately from equation (8), with V_j^m and V_{j+1}^m replacing $C_{T,1}$ and $C_{T,0}$, respectively. □

A.3 Proofs for Section 4

Section 4.1

Proof of Proposition 9. In what follows, we will often use $\mathbb{E}_i[\cdot]$ to make explicit that we are taking expectations over DGPs indexed by i, and for now we will use the notational simplifications used in the statement of the proposition. First, for (i), start with the case fixing $\pi_{0,i}^* = \pi_0^*$ across i. Applying Proposition 2,

$$\begin{aligned}
\mathbb{E}_i[\mathbb{E}[X_i^*]] &= \mathbb{E}_i[(\pi_0^* - \pi_{0,i}) \cdot \triangle_i] \\
&= \mathbb{E}_i[\pi_0^* \cdot \triangle_i] - \mathbb{E}_i[\pi_{0,i} \cdot \triangle_i] \\
&= \pi_0^* \cdot \mathbb{E}_i[\triangle_i] - \mathbb{E}_i[\pi_{0,i}] \cdot \mathbb{E}_i[\triangle_i] \\
&= (\pi_0^* - \mathbb{E}_i[\pi_{0,i}]) \cdot \mathbb{E}_i[\triangle_i] \\
&= \mathbb{E}_i[\pi_0^* - \pi_{0,i}] \cdot \mathbb{E}_i[\triangle_i]
\end{aligned}$$

$$= \mathbb{E}_i \left[\pi_0^* - \frac{\pi_0^*}{\phi_i + (1-\phi_i)\pi_0^*} \right] \cdot \mathbb{E}_i[\triangle_i]$$

where the main step in line three from $\mathbb{E}_i[\pi_{0,i} \cdot \triangle_i]$ to $\mathbb{E}_i[\pi_{0,i}] \cdot \mathbb{E}[\triangle_i]$ follows from the assumption that $\text{Cov}(\pi_{0,i}, \triangle_i) = 0$.

Now consider $\zeta_1(\phi_i, \pi_0^*) \equiv \pi_0^* - \frac{\pi_0^*}{\phi_i + (1-\phi_i)\pi_0^*}$. This is a concave function in ϕ_i given that $\pi_0^* \in [0,1]$ and $\phi_i \geqslant 1$; the second derivative of this function is

$$\frac{\partial^2 \zeta_1}{\partial \phi_i^2} = \frac{-2\pi_0^*(1-\pi_0^*)^2}{(\pi_0^* + \phi(1-\pi_0^*))^3},$$

which is weakly negative if $\pi_0^* \in [0,1]$ and $\phi \geqslant 1$. Thus using Jensen's inequality, the expectation of the function over ϕ_i must be less than the function evaluated at the expectation of ϕ_i:

$$\mathbb{E}_i \left[\pi_0^* - \frac{\pi_0^*}{\phi_i + (1-\phi_i)\pi_0^*} \right] \leqslant \pi_0^* - \frac{\pi_0^*}{\underline{\phi} + (1-\underline{\phi})\pi_0^*},$$

where $\underline{\phi} \equiv \mathbb{E}_i[\phi_i]$. Now, returning to the equation above, suppose that $\mathbb{E}_i[\triangle_i] > 0$. In this case,

$$\begin{aligned} \mathbb{E}_i[\mathbb{E}[X_i^*]] &= \mathbb{E}_i[\pi_0^* - \frac{\pi_0^*}{\phi_i + (1-\phi_i)\pi_0^*}] \cdot \mathbb{E}_i[\triangle_i] \\ &\leqslant (\pi_0^* - \frac{\pi_0^*}{\underline{\phi} + (1-\underline{\phi})\pi_0^*}) \cdot \mathbb{E}_i[\triangle_i] \end{aligned}$$

Now assume that $\mathbb{E}_i[\triangle_i] \leqslant 0$. Then, as $\pi_0^* - \frac{\pi_0^*}{\phi_i + (1-\phi_i)\pi_0^*} = \pi_0^* - \pi_0 \geqslant 0$ under our maintained assumption that $\phi_i \geqslant 1$:

$$\begin{aligned} \mathbb{E}_i[\mathbb{E}[X_i^*]] &= \mathbb{E}_i[\pi_0^* - \frac{\pi_0^*}{\phi_i + (1-\phi_i)\pi_0^*}] \cdot \mathbb{E}_i[\triangle_i]. \\ &\leqslant 0 \end{aligned}$$

Taken together,

$$\mathbb{E}_i[\mathbb{E}[X_i^*]] \leqslant \max\{0, (\pi_0^* - \frac{\pi_0^*}{\underline{\phi} + (1-\underline{\phi})\pi_0^*}) \cdot \mathbb{E}_i[\triangle_i]\}.$$

For part (ii), first consider the situation in which $\pi_{0,i}^*$ is constant and equal to π_0^*. As above,

$$\begin{aligned} \mathbb{E}_i[\mathbb{E}[X_i^*]] &\leqslant \mathbb{E}_i[(\pi_0^* - \pi_{0,i}) \cdot \pi_0^*] \\ &\leqslant \mathbb{E}_i[\pi_0^* - \pi_{0,i}] \cdot \pi_0^* \\ &\leqslant \mathbb{E}_i \left[\pi_0^* - \frac{\pi_0^*}{\phi_i + (1-\phi_i)\pi_0^*} \right] \cdot \pi_0^* \end{aligned}$$

Following the same logic as above, given the concavity of $\zeta_2 \equiv \pi_0^* - \frac{\pi_0^*}{\phi_i + (1-\phi_i)\pi_0^*}$ with respect to ϕ_i

and the fact that $\pi_0^* \geqslant 0$,

$$\begin{aligned}\mathbb{E}_i[\mathbb{E}[X_i^*]] &\leqslant \mathbb{E}_i\left[\pi_0^* - \frac{\pi_0^*}{\phi_i + (1-\phi_i)\pi_0^*}\right] \cdot \pi_0^* \\ &\leqslant \left(\pi_0^* - \frac{\pi_0^*}{\underline{\phi} + (1-\underline{\phi})\pi_0^*}\right) \cdot \pi_0^*,\end{aligned}$$

as stated in the second inequality. Now allowing $\pi_{0,i}^*$ to vary, write the upper bound for $\mathbb{E}[X^*]$ in Proposition 3 as $\zeta_{2'}(\phi_i, \pi_{0,i}^*) \equiv \left(\pi_0^* - \frac{\pi_0^*}{\phi_i + (1-\phi_i)\pi_0^*}\right)\pi_{0,i}^*$. Again since $\partial^2\zeta_{2'}/\partial\phi_i^2 \leqslant 0$, for any arbitrary realization of $\pi_{0,i}^* = \varrho$, we have from the application of Jensen's inequality above (now dropping the dependence of $\mathbb{E}$ on i) that

$$\mathbb{E}[\zeta_{2'}(\phi_i, \pi_{0,i}^*) \mid \pi_{0,i}^*] \leqslant \zeta_{2'}\left(\mathbb{E}[\phi_i \mid \pi_{0,i}^* = \varrho], \varrho\right).$$

Now, using Proposition 3 and applying LIE to the above inequality,

$$\begin{aligned}\mathbb{E}[X_i^*] \leqslant \mathbb{E}[\zeta_{2'}(\phi_i, \pi_{0,i}^*)] &\leqslant \mathbb{E}\left[\zeta_{2'}\left(\mathbb{E}[\phi_i \mid \pi_{0,i}^*], \pi_{0,i}^*\right)\right] \\ &\leqslant \mathbb{E}\left[\zeta_{2'}(\overline{\phi}, \pi_{0,i}^*)\right], \qquad \text{(A.23)}\end{aligned}$$

where $\overline{\phi}$ is as in the proposition statement and where the second line uses $\partial\zeta_{2'}/\partial\phi_i \geqslant 0$. Substituting the definition of $\zeta_{2'}$ into this inequality yields equation (27).

For part (iii), as $(\pi_{0,i}^* - \frac{\pi_{0,i}^*}{\overline{\phi} + (1-\overline{\phi})\pi_{0,i}^*}) \leqslant \pi_{0,i}^*$ for any $\overline{\phi} \geqslant 1$,

$$\mathbb{E}[X_i^*] \leqslant \mathbb{E}[(\pi_{0,i}^* - 0)\pi_{0,i}^*] = \mathbb{E}[(\pi_{0,i}^*)^2],$$

as stated. (Equivalently, one can use (A.23) and note again that $\partial\zeta_{2'}/\partial\overline{\phi} \geqslant 0$, so that the bound is most slack as $\overline{\phi} \to \infty$, giving the same bound.)

Finally, for part (iv), Corollary 2 notes that if $\mathbb{E}[X^*|\theta = 0] \leqslant \mathbb{E}[X^*|\theta = 1]$, then $\mathbb{E}[X^*] \leqslant 0$. Therefore, if $\mathbb{E}[X_i^*|\theta = 0] \leqslant \mathbb{E}[X_i^*|\theta = 1]$ for all i, then $\mathbb{E}[X_i^*] \leqslant 0$ for all i and thus over all streams, completing the proof. □

Section 4.2

Proof of Proposition 10. Starting with measured belief movement, under the stated assumptions for ϵ_t,

$$\begin{aligned}\mathbb{E}[\widehat{\mathsf{m}}_{t,t+1}^*] &= \mathbb{E}[(\widehat{\pi}_{t+1}^* - \widehat{\pi}_t^*)^2] \\ &= \mathbb{E}\left[\left((\pi_{t+1}^* - \pi_t^*)^2 + (\epsilon_{t+1} - \epsilon_t)\right)^2\right] \\ &= \mathbb{E}[\mathsf{m}_{t,t+1}^*] + 2\mathbb{E}[\pi_{t+1}^*\epsilon_{t+1} - \pi_t^*\epsilon_{t+1} - \pi_{t+1}^*\epsilon_t + \pi_t^*\epsilon_t] + \mathbb{E}[(\epsilon_{t+1} - \epsilon_t)^2] \\ &= \mathbb{E}[\mathsf{m}_{t,t+1}^*] + \mathbb{E}[\epsilon_t^2 + \epsilon_{t+1}^2].\end{aligned}$$

For the measured counterpart of uncertainty resolution $\mathsf{r}^*_{t,t+1} \equiv (\mathsf{u}^*_t - \mathsf{u}^*_{t+1})$,

$$\begin{aligned}\mathbb{E}[\widehat{\mathsf{r}}^*_{t,t+1}] &= \mathbb{E}[(\pi^*_t + \epsilon_t)(1 - \pi^*_t - \epsilon_t) - (\pi^*_{t+1} + \epsilon_{t+1})(1 - \pi^*_{t+1} - \epsilon_{t+1})] \\ &= \mathbb{E}[\mathsf{r}^*_{t,t+1}] + \mathbb{E}[\epsilon^2_{t+1} - \epsilon^2_t].\end{aligned}$$

Combining these two, with $\text{Var}(\epsilon_t) \equiv \mathbb{E}[(\epsilon_t - \mathbb{E}[\epsilon_t])^2] = \mathbb{E}[\epsilon^2_t]$ and $X^*_{t,t+1} \equiv \mathsf{m}^*_{t,t+1} - \mathsf{r}^*_{t,t+1}$,

$$\mathbb{E}[\widehat{X}^*_{t,t+1}] = \mathbb{E}[X^*_{t,t+1}] + 2\text{Var}(\epsilon_t). \qquad \square$$

Section 4.3

Proof of Proposition 11. We work in the context of the example in Section 2 for simplicity of notation (and without loss of generality), so write ϕ_t as in equation (7) as $\mathbb{E}_t[U'(C_{T,1})]/\mathbb{E}_t[U'(C_{T,0})]$, where the two states are again $\theta = 0, 1$, but where ϕ_t can now vary. Regardless of the DGPs for θ, $U'(C_{T,1})$ and $U'(C_{T,1})$, we will show that if ϕ_t evolves as a martingale or supermartingale ($\mathbb{E}[\phi_{t+1}] \leqslant [\phi_t]$), then the bounds in Proposition 3 and Corollary 1 apply with ϕ_0 replacing ϕ.

Intuition: Given the length of the proof, it is useful to briefly outline the steps and intuition. (1) First, we focus on a particular situation in which a) ϕ_t can only take a high, medium, and low value, and b) ϕ_t evolves as a martingale. Then, we assume — for the sake of contradiction — that there exists a DGP in which ϕ moves in a way that "beats our bounds" (produces expected RN movement that is higher than that in our bounds). We then focus on the highest-movement DGP where ϕ changes and focus on the last meaningful movement of ϕ in this DGP. We show that expected RN movement is strictly increased if ϕ instead remains constant, leading to a contradiction. We conclude that there is not a DGP with expected RN movement that beats our bounds in which a) ϕ_t evolves as a martingale and b) ϕ_t only takes three values. Then, we expand this result to the general case. (2) First, we consider the situation in which ϕ_t can instead take an arbitrary number of values. We show that if there exists a general DGP that beats our bound in which ϕ_t is a martingale, there must exist a DGP that beats our bounds in which ϕ_t is a martingale and evolves into three values. But, given that we proved that this is not possible, we conclude that there is not a DGP that beats our bounds in which ϕ_t evolves as a martingale. (3) Finally, we expand this result to supermartingales. We show that if there exists a DGP in which ϕ_t evolves as a supermartingale and beats our bounds, there must be a DGP in which ϕ_t evolves as a martingale and beats our bounds. But, given that we proved that this is not possible, we conclude that there is not a DGP that beats our bounds.

Setup: Given that ϕ_t can change, we explicitly allow it to depend on the signal history. Therefore, RN beliefs are now denoted:

$$\pi^*_t(\mathbf{H}_t) = \frac{\phi_t(\mathbf{H}_t) \cdot \pi_t(\mathbf{H}_t)}{(\phi_t(\mathbf{H}_t) - 1)\pi_t(\mathbf{H}_t) + 1}$$

We still assume that the uncertainty about θ is resolved by period T. We allow more periods to

allow resolution of uncertainty about ϕ, although we now show this is inconsequential. Specifically, movement in ϕ_t is only consequential for RN movement when there is still uncertainty about θ. Specifically, as $\pi_T \in \{0,1\}$, then π_t must be constant for any $t \geqslant T$. Also, if $\pi_t \in \{0,1\}$, then $\pi_t^* = \pi_t$, so RN beliefs must also be constant for $t \geqslant T$. Therefore, there is no RN movement for $t \geqslant T$, regardless of whether ϕ_t for is changing over these periods. Given that ϕ_t then has no impact on RN movement for $t \geqslant T$, we can restrict our attention to DGPs in which ϕ_t is constant for $t \geqslant T$.

Proof: Now, assume for the sake of contradiction that there exists some DGP in which ϕ_t changes and expected RN movement is higher than the bounds in Proposition 3 for some T. Consider a DGP of this set with the highest expected RN movement. We now consider the last meaningful movement of ϕ in this DGP. Specifically, given that ϕ_t is assumed to change at some point, but ϕ_t is constant when $t \geqslant T$, there must exist some history H_t in which $\pi_t \in (0,1)$, ϕ_t can change between t and $t+1$ (i.e., there exists a signal s_{t+1} for which $\phi_{t+1}(\mathbf{H}_t \cup s_{t+1}) \neq \phi_t(\mathbf{H}_t)$) but for which ϕ_t is constant after $t+1$. We will now show that, in fact, expected RN movement is higher if ϕ_t is constant following H_t, contradicting the assumption that the ϕ-changing DGP has the highest expected RN movement.

(1) We start by focusing on a particular situation, drawing conclusions for this situation, and then showing how the results from this situation extend to the general case. We start by considering the case in which, following any H_t,

(a) $\phi_{t+1}(\mathbf{H}_t \cup s_{t+1})$ can only take three values:

- $\phi_{t+1}^H > \phi_t$ following signal s_{t+1}^H with probability $q^H > 0$
- $\phi_{t+1}^M = \phi_t$ following signal s_{t+1}^M with probability $q^M \geqslant 0$
- $\phi_{t+1}^L < \phi_t$ following signal s_{t+1}^L with probability $q^L > 0$

(b) ϕ_t evolves as a martingale: $\sum_{i\in\{L,M,H\}} q^i \cdot \phi_{t+1}^i = \phi_t$.

Given these assumptions and the maintained assumption that π_t does not evolve in the same period as ϕ_t and therefore is constant immediately following history H_t, $\pi_t^*(\mathbf{H}_t \cup s_{t+1})$ can take at most three values: $\pi_{t+1}^{*i} = \frac{\phi_{t+1}^i \cdot \pi_t}{(\phi_{t+1}^i - 1)\pi_t + 1}$ for $i \in \{L, M, H\}$.

Now, we will consider expected RN movement following H_t. From period t to $t+1$, given signal s_{t+1}^i, RN beliefs move from π_t^* to π_{t+1}^{*i}, leading to per-period RN movement

$$\begin{aligned}
\mathbb{E}[\mathsf{m}_{t,t+1}^* | H_t \cup s_{t+1}^i] &= (\pi_t^* - \pi_{t+1}^{*i})^2 \\
&= \left(\frac{\phi_t \cdot \pi_t}{(\phi_t - 1)\pi_t + 1} - \frac{\phi_{t+1}^i \cdot \pi_{t+1}}{(\phi_{t+1}^i - 1)\pi_{t+1} + 1}\right)^2 \\
&= \left(\frac{\phi_t \cdot \pi_t}{(\phi_t - 1)\pi_t + 1} - \frac{\phi_{t+1}^i \cdot \pi_t}{(\phi_{t+1}^i - 1)\pi_t + 1}\right)^2.
\end{aligned}$$

The second line is expressed in terms of π_t and ϕ_t rather than π_t^* as this will turn out to be easier later given that, unlike in the rest of the paper, ϕ_t is not constant. The third line uses our assumption that $\pi_t = \pi_{t+1}$.

After period $t+1$, ϕ_t is assumed to be constant, so our main bounds hold with π_0^* replaced with π_{t+1}^{*i} and ϕ replaced with ϕ_{t+1}^i. For example, expected RN movement (excess RN movement plus initial RN uncertainty) given signal s_{t+1}^i from period $t+1$ onward is then bounded above by:

$$\begin{aligned}
\mathbb{E}[\mathsf{m}_{t+1,T}^*|H_t \cup s_{t+1}^i] &= \mathbb{E}[X_{t+1,T}^*|H_t \cup s_{t+1}^i] + \mathbb{E}[\mathsf{r}_{t+1,T}^*|H_t \cup s_{t+1}^i] \\
&\leqslant (\pi_{t+1}^{i*} - \pi_{t+1}) \cdot \pi_{t+1}^{i*} + (1 - \pi_{t+1}^{i*}) \cdot \pi_{t+1}^{i*} \\
&= (1 - \pi_{t+1}) \cdot \pi_{t+1}^{i*} \\
&= (1 - \pi_{t+1}) \cdot \frac{\phi_{t+1}^i \cdot \pi_{t+1}}{(\phi_{t+1}^i - 1)\pi_{t+1} + 1} \\
&= (1 - \pi_t) \cdot \frac{\phi_{t+1}^i \cdot \pi_t}{(\phi_{t+1}^i - 1)\pi_t + 1},
\end{aligned}$$

where the first line is the definition of RN movement, the second line plugs in our bound for excess RN movement and uncertainty reduction given that uncertainty is zero at period T, the third line simplifies, the fourth line casks everything in terms of ϕ_t and π_t, and the final line uses our assumption that $\pi_t = \pi_{t+1}$.

Therefore, expected RN movement from period t onward following history H_t is then bounded above by:

$$\begin{aligned}
\mathbb{E}[\mathsf{m}_{t,T}^*|H_t] &= \mathbb{E}[\mathsf{m}_{t,t+1}^*|H_t] + \mathbb{E}[\mathsf{m}_{t+1,T}^*|H_t] \\
&\leqslant \sum_{i \in \{L,M,H\}} q^i \cdot \Big((\frac{\phi_t \cdot \pi_t}{(\phi_t - 1)\pi_t + 1} - \frac{\phi_{t+1}^i \cdot \pi_t}{(\phi_{t+1}^i - 1)\pi_t + 1})^2 + (1 - \pi_t) \cdot \frac{\phi_{t+1}^i \cdot \pi_t}{(\phi_{t+1}^i - 1)\pi_t + 1} \Big)
\end{aligned}$$

We now show that — given that ϕ_t evolves as a martingale — this DGP will have higher RN movement if ϕ_t is constant from H_t onward. To see this, consider the "worst-case" DGP noted in Proposition 5 in which ϕ remains constant at ϕ_t. In this case, RN movement is (arbitrarily close to):

$$\mathbb{E}_{maxDGP}[\mathsf{m}_{t,T}^*|H_t] = (1 - \pi_t) \cdot \frac{\phi_t \cdot \pi_t}{(\phi_t - 1)\pi_t + 1}$$

We now subtract the expected RN movement given ϕ changes ($\mathbb{E}[\mathsf{m}_{t,T}^*|H_t]$) from the worst-case RN movement ($\mathbb{E}_{maxDGP}[\mathsf{m}_{t,T}^*|H_t]$) and show it is positive given the assumption that ϕ_t evolves as a martingale. To start, note that as $\phi_{t+1}^M = \phi_t$, the martingale assumption can be rewritten:

$$\begin{aligned}
\sum_{i \in \{L,M,H\}} q^i \cdot \phi_{t+1}^i &= \phi_t \\
\sum_{i \in \{L,M,H\}} q^i \cdot \phi_{t+1}^i &= \sum_{i \in \{L,M,H\}} q^i \cdot \phi_t \\
\sum_{i \in \{L,H\}} q^i \cdot \phi_{t+1}^i &= \sum_{i \in \{L,H\}} q^i \cdot \phi_t
\end{aligned}$$

$$\sum_{i\in\{L,H\}} \frac{q^i}{q^H+q^L}\cdot\phi^i_{t+1} = \sum_{i\in\{L,H\}} \frac{q^i}{q^H+q^L}\cdot\phi_t$$
$$\sum_{i\in\{L,H\}} p^i\cdot\phi^i_{t+1} = \sum_{i\in\{L,H\}} p^i\cdot\phi_t$$
$$\sum_{i\in\{L,H\}} p^i\cdot\phi^i_{t+1} = \phi_t,$$

where $p^i \equiv \frac{q^i}{q^H+q^L}$. Similarly, the difference $\mathbb{E}_{maxDGP}[\mathsf{m}^*_{t,T}|H_t] - \mathbb{E}[\mathsf{m}^*_{t,T}|H_t]$ is positive if and only if:

$$(1-\pi_t)\cdot\frac{\phi_t\cdot\pi_t}{(\phi_t-1)\pi_t+1} - \sum_{i\in\{L,H\}} p^i\cdot\Big(\big(\frac{\phi_t\cdot\pi_t}{(\phi_t-1)\pi_t+1} - \frac{\phi^i_{t+1}\cdot\pi_t}{(\phi^i_{t+1}-1)\pi_t+1}\big)^2 + (1-\pi_t)\cdot\frac{\phi^i_{t+1}\cdot\pi_t}{(\phi^i_{t+1}-1)\pi_t+1}\Big) > 0.$$

Substituting this martingale equation into the modified difference equation above and simplifying yields $\mathbb{E}_{maxDGP}[\mathsf{m}^*_{t,T}|H_t] - \mathbb{E}[\mathsf{m}^*_{t,T}|H_t] > 0$ if and only if

$$\frac{\pi_t^3(1-\pi_t)^2(\phi^H_{t+1}-\phi_t)(\phi_t-\phi^L_{t+1})\big((\phi^H_{t+1}-\phi_t)+(\phi^L_{t+1}-1)+(\pi_t)(2+\pi_t(\phi_t-1))(\phi^H_{t+1}-1)(\phi^L_{t+1}-1)\big)}{(1+\pi_t(\phi_t-1))^2(1+\pi_t(\phi^H_{t+1}-1))^2(1+\pi_t(\phi^L_{t+1}-1))^2} > 0.$$

While the expression on the left side of the inequality is rather long, it is in fact straightforward to see that it is positive: every parentheses contains a positive value as $\phi^H_{t+1} > \phi_t > \phi^L_{t+1} \geqslant 1$ and $\pi_t \in (0,1)$. Therefore, we conclude that expected RN movement can be increased if ϕ_t remains constant following H_t rather than changing. But this gives us a contradiction, as it violates the assumption that the DGP with ϕ_t moving following H_t has the highest possible movement. Therefore, we conclude that there does not exist a DGP satisfying our assumptions (a) and (b) that produces more expected RN movement than the bound in Proposition 3.

(2) We now extend this observation to DGPs in which assumption (a) is relaxed. In particular, we now consider a DGP_{arb} in which ϕ_{t+1} following H_t can now take an arbitrary number of values (indexed by i). Following the proof above, our goal is to show that for every DGP_{arb}, $\mathbb{E}_{maxDGP}[\mathsf{m}^*_{t,T}|H_t] \geqslant \mathbb{E}_{DGP_{arb}}[\mathsf{m}^*_{t,T}|H_t]$. To do this, we will show that if there exists a DGP_{arb} such that $\mathbb{E}_{maxDGP}[\mathsf{m}^*_{t,T}|H_t] < \mathbb{E}_{DGP_{arb}}[\mathsf{m}^*_{t,T}|H_t]$ is true, there is a contradiction. Specifically, we show that this inequality would imply that there must exist a DGP satisfying assumption (a) in which trinary movements lead to $\mathbb{E}_{maxDGP}[\mathsf{m}^*_{t,T}|H_t] < \mathbb{E}[\mathsf{m}^*_{t,T}|H_t]$. But, given that we just showed this is not possible, we have a contradiction and it must be that the DGP_{arb} does not exist.

Intuition: To do this, we will show that the expected RN movement of DGP_{arb} at history H_t can be replicated with a DGP that only using trinary movements. Although the implementation is annoying clunky, the intuition is simple: rather that doing all of the arbitrary martingale movements of ϕ in period $t+1$, we "separate out" the movements into trinary martingale movements in sequential periods. For example, suppose that $\phi_t = 3$ and $\phi_t + 1$ takes values 1, 2, 4, and 5 with probability .25 and $\pi_t = \pi_{t+1} = 1/2$ That is, we have a martingale process on ϕ with a constant

π that leads $\pi_t^* = 3/4$ and π_{t+1}^* to take values $1/2$, $2/3$, $4/5$ and $5/6$ with equal probability, such that RN movement then takes values $(3/4 - 1/2)^2$, $(3/4 - 2/3)^2$, $(3/4 - 4/5)^2$, and $(3/4 - 5/6)^2$ with equal probability. Instead, consider a sequential DGP in which $\phi_t = 3$ but ϕ_{t+1} equals 1 with probability .25 and 5 with probability .25, but stays constant at 3 with probability .5. Then, following ϕ_{t+1} staying constant at 3, ϕ_{t+2} can take values 2 or 4 with equal probability (and, when ϕ_{t+1} equals 1, ϕ_{t+2} remains constant at 1 and when ϕ_{t+1} equals 5, ϕ_{t+2} remains constant at 5). Note that this new sequential DGP only used trinary martingale movements following every history. And, note that the likelihood of each outcome is the same: the sequential DGP has a equal probability of ending with ϕ equal to 1, 2, 4, and 5. Finally, the total RN movement of the sequential DGP between t and $t+2$ matches the RN movement of the original DGP between t and $t+1$. For example, there is a .25 probability that RN beliefs shift from $3/4$ to $1/2$ and then stay constant at $1/2$, for a total RN movement of $(3/4 - 1/2)^2 + (1/2 - 1/2)^2 = (1/2 - 3/4)^2$. Similarly, there is a .25 probability that RN beliefs stay constant at $3/4$ and then shifts from $3/4$ to $2/3$, for a total RN movement of $(3/4 - 3/4)^2 + (3/4 - 2/3)^2 = (1/2 - 2/3)^2$. Overall, then, RN movement still takes values $(3/4 - 1/2)^2$, $(3/4 - 2/3)^2$, $(3/4 - 4/5)^2$, and $(3/4 - 5/6)^2$ with equal probability. This simple idea is slightly clunky to implement because one needs an algorithm to separate the arbitrary number of movements into individual movements that satisfy the martingale property.

To understand the algorithm, consider a *full* list of the probabilities of each ϕ_{t+1} and the differences $\phi_{t+1} - \phi_t$, i.e. $\{(q^1, \phi_{t+1}^1 - \phi_t), (q^2, \phi_{t+1}^i - \phi_t), \ldots\}$. In the example DGP mentioned above, this difference list would be $\{(.25, -2), (.25, -1), (.25, 1), (.25, 2)\}$. For each two-part component in a list, we define the *product* of the component as $(q^i \cdot (\phi_{t+1}^i - \phi_t)$. As we assume that ϕ_t is a martingale, it must be that the sum of the product of the components of the full list is zero:

$$\sum_i q^i \cdot \phi_{t+1}^i = \phi_t$$
$$\sum_i q^i \cdot (\phi_{t+1}^i - \phi_t) = 0$$

Therefore, we call the full list *balanced*. Our first step is to remove from the list any component with a difference of 0 (that is, remove components where ϕ_{t+1}^i is equal to ϕ_t). Given this change, note that the list is still balanced as we removed a difference of zero. Also, note that the list must still contain some elements as we assumed that there existed some signal at history H_t for which $\phi_{t+1}^i \neq \phi_t$. Now, we consider an algorithm on this list to create j binary balanced lists, which each have two members. To do this, we start by separating the full list into two sub-lists depending on whether the difference is positive or negative. Note that as the main list was balanced (the sum of the products in the list was zero), the sum of the products in the positive and negative lists must be equal. For example, in the above example, the negative list would be $\{(.25, -2), (.25, -1)\}$ and the positive list would be $\{(.25, 2), (.25, 1)\}$. The algorithm then proceeds as such:

1. Enter with a positive and negative difference list in which the sum of the products is equal. Consider the first element of the current positive list $(q_{pos}^1, \phi_{t+1,pos}^1 - \phi_t)$ and negative list

$(q^1_{neg}, \phi^1_{t+1,neg} - \phi_t)$.

- If $q^1_{pos} \cdot (\phi^1_{t+1,pos} - \phi_t) - q^1_{neg} \cdot (\phi^1_{t+1,neg} - \phi_t) \geqslant 0$:
  - Let $q^* \leqslant q^1_{pos}$ solve $q^* \cdot (\phi^1_{t+1,pos} - \phi_t) - q^1_{neg} \cdot (\phi^1_{t+1,neg} - \phi_t) = 0$.
  - Add the balanced list $\{(q^*, \phi^1_{t+1,pos} - \phi_t), (q^1_{neg}, \phi^1_{t+1,neg} - \phi_t)\}$ to set of the binary difference lists.
  - Modify the current positive and negative difference lists: remove $(q^1_{neg}, \phi^1_{t+1,neg} - \phi_t)$ from the current negative list and replace $(q^1_{pos}, \phi^1_{t+1,pos} - \phi_t)$ in the current positive list with $(q^1_{pos} - q^*, \phi^1_{t+1,pos} - \phi_t)$. Note that as these subtractions are equal, the sum of the current negative and positive lists remains equal.
- If $q^1_{pos} \cdot (\phi^1_{t+1,pos} - \phi_t) - q^1_{neg} \cdot (\phi^1_{t+1,neg} - \phi_t) < 0$:
  - Let $q^* < q^1_{neg}$ solve $q^i_{pos} \cdot (\phi^1_{t+1,pos} - \phi_t) - q^* \cdot (\phi^1_{t+1,neg} - \phi_t) = 0$.
  - Add the balanced list $\{(q^1_{pos}, \phi^1_{t+1,pos} - \phi_t), (q^*, \phi^1_{t+1,neg} - \phi_t)\}$ to the binary lists.
  - Modify the current positive and negative lists: remove $(q^1_{pos}, \phi^1_{t+1,pos} - \phi_t)$ from the current positive list and replace $(q^1_{neg}, \phi^1_{t+1,neg} - \phi_t)$ in the current negative list with $(q^1_{neg} - q^*, \phi^1_{t+1,neg} - \phi_t)$. Note that as these subtractions are equal, the sum of the current negative and positive lists remains equal.

2. If there are no elements left in the current positive and negative list, end the algorithm. Otherwise, repeat.

We are left with a set of j balanced binary lists, each with two members. In the above example, there would be two binary lists: $\{(.25, -2), (.25, 2)\}$ and $\{(.25, -1), (.25, 1)\}$. Intuitively, we have taken the original balanced difference list $\{(.25, -2), (.25, -1), (.25, 1), (.25, 2)\}$ and broken it into a set of binary balanced lists. We now take our set of j binary balanced lists and use an algorithm to create a new sequential DGP starting at history H_t and lasting to period $t + j$.

1. In the initial period period $t + 1$:
   - With probability q^1_{pos}, let ϕ_{t+1} equal $\phi^1_{t+1,pos}$.
   - With probability q^1_{neg}, let ϕ_{t+1} equal $\phi^1_{t+1,neg}$.
   - With probability $1 - q^1_{pos} - q^1_{neg}$, let ϕ_{t+1} equal ϕ_t.

2. In period $t + k$, we enter the period with (1) histories in which ϕ has remained constant between period t and $t + k - 1$, which has occurred with probability $1 - \sum_{j=1}^{k-1} q^j_{neg} - \sum_{j=1}^{k-1} q^j_{pos}$ and (2) histories in which ϕ has changed between period t and $t + k - 1$, which has occured with probability $\sum_{j=1}^{k-1} q^j_{neg} + \sum_{j=1}^{k-1} q^j_{pos}$.
   - For histories in which ϕ has changed, let $\phi_{t+k} = \phi_{t+k-1}$ with probability 1.
   - For histories in which ϕ has remained constant:

- With probability $\frac{q^k_{pos}}{1-\sum_{j=1}^{k-1} q^j_{neg}-\sum_{j=1}^{k-1} q^j_{pos}}$, let ϕ_{t+k} equal $\phi^k_{t+1,pos}$.
- With probability $\frac{q^k_{neg}}{1-\sum_{j=1}^{k-1} q^j_{neg}-\sum_{j=1}^{k-1} q^j_{pos}}$, let ϕ_{t+k} equal $\phi^k_{t+1,neg}$.
- With probability $\frac{1-q^k_{pos}-q^k_{neg}}{1-\sum_{j=1}^{k-1} q^j_{neg}-\sum_{j=1}^{k-1} q^j_{pos}}$, let ϕ_{t+k} equal ϕ_t.

Therefore, we enter period $t+k+1$ with (1) histories in which ϕ has remained constant between period t and $t+k-1$, which has occurred with probability $1-\sum_{j=1}^{k} q^j_{neg}-\sum_{j=1}^{k} q^j_{pos}$ and (2) histories in which ϕ has changed between period t and $t+k$, which has occured with probability $\sum_{j=1}^{k} q^j_{neg}+\sum_{j=1}^{k} q^j_{pos}$.

This algorithm outputs a newly-constructed DGP that has a set of important characteristics between period t and $t+j$:

1. It is composed entirely of trinary movements from period t to period $t+j$.

2. At period $t+j$, the newly-constructed trinary DGP has the same probability distribution of ϕ_t+j as at period $t+j$ as the arbitrary DGP has for ϕ_{t+1} at period $t+1$ following history H_t. To see this, first focus on the values of ϕ_{t+1} that are positive. All of these values and probabilities are collected in the positive binary lists above. For the newly-constructed trinary DGP, in period $t+1$, there is a q^1_{pos} probability that ϕ_{t+1} equals $\phi^1_{t+1,pos}$. Then, following period $t+1$, there will be no movement in ϕ from period $t+1$ to period $t+j$ following the movement to $\phi^1_{t+1,pos}$. Therefore, in period $t+j$, there will be histories (which occur with probability q^1_{pos}) in which ϕ_{t+j} equals $\phi^1_{t+1,pos}$. Next, in period $t+2$, there will be a set of histories (which occur with probability $1-q^1_{pos}-q^1_{neg}$) in which there was no movement in ϕ from period t to $t+1$. Conditional on reaching this history, there is a $\frac{q^2}{1-q^1_{pos}-q^1_{neg}}$ probability that ϕ_{t+2} equals $\phi^2_{t+1,pos}$. Therefore, the unconditional probability is $(1-q^1_{pos}-q^1_{neg})\cdot\frac{q^2}{1-q^1_{pos}-q^1_{neg}}=q^2$. Then, following period $t+2$, there will be no movement in ϕ from period $t+2$ to period $t+j$ following the movement to $\phi^2_{t+1,pos}$. Therefore, in period $t+j$, there will be histories (which occur with probability q^2_{pos}) in which ϕ_{t+j} equals $\phi^2_{t+1,pos}$. The same argument extended to later periods $t+k$ shows that in period $t+j$, there will be histories (which occur with probability q^k_{pos}) in which ϕ_{t+j} equals $\phi^k_{t+1,pos}$. Therefore, we have replicated in the newly-constructed DGP the exact positive values and probabilities from the positive binary lists above created using the arbitrary DGP. The same argument holds for the negative values. Therefore, at period $t+j$, the newly-constructed DGP has the same probability distribution of ϕ_t+j as at period $t+j$ as the arbitrary DGP has for ϕ_{t+1} at period $t+1$.

3. The total expected movement between period t and $t+k$ must be the same in the newly constructed DGP as in the arbitrary DGP between periods t and $t+1$. To see this, note that in the arbitrary DGP, for every possible value of ϕ_{t+1}, the associated movement from period t to $t+1$ is $\frac{\phi_t\cdot\pi_t}{(\phi_t-1)\pi_t+1}-\frac{\phi^i_{t+1}\cdot\pi_t}{(\phi^i_{t+1}-1)\pi_t+1})^2$. In the newly-constructed DGP, there are j periods of

possible movement. However, movement in ϕ only occurs after histories in which there was no previous movement and is only followed by periods in which there is no future movement. Therefore, given a history in which we reach a given ϕ_{t+j}, the only associated movement from period t to period $t+j$ is $\frac{\phi_t \cdot \pi_t}{(\phi_t - 1)\pi_t + 1} - \frac{\phi^j_{t+j} \cdot \pi_t}{(\phi^j_{t+j} - 1)\pi_t + 1})^2$. But, as we just showed that the newly-constructed DGP has the same probability distribution of $\phi_t + j$ as at period $t+j$ as the arbitrary DGP has for ϕ_{t+1} at period $t+1$, the movements for each must be the same.

Following period $t+k$, the newly-constructed DGP can be designed given ϕ_{t+k} to match the arbitrary DGP after period $t+1$ with the same value of ϕ_{t+1}. That is, in the example above, if in the arbitrary DGP there is complete resolution following a realization of $\phi_{t+1} = 2$, then the newly-constructed DGP is designed to have complete resolution following $\phi_{t+k} = 2$. Therefore, following history H_t, the newly-constructed DGP has the same total expected movement as the arbitrary DGP. But recall that we were considering the existence of a DGP_{arb} such that $\mathbb{E}_{maxDGP}[\mathsf{m}^*_{t,T}|H_t] < \mathbb{E}_{DGP_{arb}}[\mathsf{m}^*_{t,T}|H_t]$. Given this existence, consider the DGP with the highest possible movement. Then (following the proof above), consider the history from this arbitrary DGP in which there is a meaningful final movement following history H_t. We can use the algorithm above to replicate the all of this movement for a trinary DGP following history H_t. But, from the previous proof, we know that there exists a DGP in which there is no movement in ϕ following H_t that produces more expected movement than any non-degenerate trinary DGP following H_t. But, then, as the constructed trinary DGP has the same movement following H_t as the arbitrary DGP, this non-movement DGP must also produce more expected movement than the arbitrary DGP. But, then we have a contradiction as we supposed that this arbitrary DGP had the highest possible movement.

(3) Finally, we now extend this observation to DGPs in which movement in ϕ is a supermartingale rather than a martingale. The logic here is relatively simple: if there exists a DGP where ϕ evolves as supermartingale and leads to expected movement that is higher than our bound, there there must exist a martingale that leads to higher expected movement. But, we just showed that this is not possible, and therefore we have a contradiction.

Following the same logic as the proofs above, we start by assuming that there exists a DGP_{super} in which ϕ evolves as a supermartingale such that the expected movement of this DGP is higher than our bound for a given T. Then, consider the supermartingale DGP with the maximum expected movement. We then focus on a history H_t with the last meaningful movement in which movement in ϕ is a strict supermartingale. If this period does not exist, the process is a martingale, and the previous results hold. Note that, following this movement, there cannot be further change in ϕ. If there were and the change in ϕ was a martingale, the previous proofs show that no change in ϕ would produce more expected movement, contradicting the assumption that this DGP produces the highest expected movement in the class. If instead there was movement and the change in ϕ was a strict supermartingale, it would contradict the assumption that the previous movement was the last meaningful movement of that type.

Now, we show that it is possible to adjust DGP_{super} following history H_t to increase the expected movement following H_t by making adjusting the change in ϕ from period t to period $t+1$ to be a

martingale rather than a supermartingale. Then, as we have previously shown that a martingale cannot have more movement than our bound, it must be that a supermartingale can also not have more movement.

To show this, we first show that any upward movement from ϕ_t to $\phi_{t+1} > \phi_t$ always leads to more total movement following history H_t than any downward movement from ϕ_t to $\phi_{t+1} < \phi_t$. To see this, consider the total expected movement from H_t onward given a change from ϕ_t to ϕ_{t+1}. Following above, this is:

$$\mathbb{E}[\mathsf{m}^*_{t,T}|H_t, \phi_t, \phi_{t+1}] = (\frac{\phi_t \cdot \pi_t}{(\phi_t - 1)\pi_t + 1} - \frac{\phi_{t+1} \cdot \pi_t}{(\phi_{t+1} - 1)\pi_t + 1})^2 + (1 - \pi_t) \cdot \frac{\phi_{t+1} \cdot \pi_t}{(\phi_{t+1} - 1)\pi_t + 1}.$$

Our claim is that this is higher if $\phi_{t+1} > \phi_t$ than if $\phi_{t+1} < \phi_t$. To see this, it is useful to compare the above with movement if $\phi_{t+1} = \phi_t$. In this case:

$$\mathbb{E}[\mathsf{m}^*_{t,T}|H_t, \phi_t = \phi_{t+1}] = (1 - \pi_t) \cdot \frac{\phi_t \cdot \pi_t}{(\phi_t - 1)\pi_t + 1}.$$

Subtracting the two and simplifying (and writing $\pi = \pi_t$ for simplicity) yields:

$$\begin{aligned}&\mathbb{E}[\mathsf{m}^*_{t,T}|H_t, \phi_t, \phi_{t+1}] - \mathbb{E}[\mathsf{m}^*_{t,T}|H_t, \phi_t = \phi_{t+1}] \\ &\quad = \frac{(\pi - 1)^2 \cdot \pi \cdot (1 + \pi \cdot (2 + \pi \cdot (\phi_t - 1)) \cdot (\phi_{t+1} - 1)) \cdot (\phi_t - \phi_{t+1})}{(1 + \pi(\phi - 1))^2 \cdot (1 + \pi(\phi_{t+1} - 1))^2}.\end{aligned}$$

As with the equation in part (a), while this is a complicated expression, it is easy to see that every component is weakly positive (as $0 < \pi < 1$ because the ϕ movement is meaningful and $\phi \geqslant 1$), except for $(\phi_t - \phi_{t+1})$. Therefore, this equation is positive if $\phi_{t+1} < \phi_t$ and negative if $\phi_{t+1} > \phi_t$. But then it must be that $\mathbb{E}[\mathsf{m}^*_{t,T}|H_t, \phi_t, \phi_{t+1}]$ is greater if $\phi_{t+1} > \phi_t$ than if $\phi_{t+1} < \phi_t$.

In this case, we can adjust the evolution of ϕ following history H_t — which was assumed to be a supermartingale — to be a martingale by taking a probability from downward change in ϕ and shifting it to an upward change in ϕ. Specifically, if ϕ_t is a strict supermartingale at H_t, there must be at least some probability on a realization of $\phi_{t+1} < \phi$. Consider the lowest possible realization of ϕ^L_{t+1} with associated probability q^L. There are two possibilities. First, there is some value $\phi^H_{t+1} > \phi$ such shifting the probability q^L from ϕ^L_{t+1} to ϕ^H_{t+1} makes ϕ a martingale. Second, there is some $q^H < q^L$ such that shifting q^H from ϕ^L_{t+1} to ϕ^H_{t+1} makes ϕ a martingale. In either case, we are shifting probability from $\phi^L_{t+1} < \phi_t$ to $\phi^H_{t+1} > \phi_t$. But, as just proven above, it must be that $\mathbb{E}[\mathsf{m}^*_{t,T}|H_t, \phi_t, \phi_{t+1}]$ is greater if $\phi_{t+1} > \phi_t$ than if $\phi_{t+1} < \phi_t$. But then the total movement of the change from ϕ at H_t must increase.

This implies that there exists a martingale process of ϕ at H_t that has higher expected movement than the proposed strict supermartingale process of ϕ at H_t. This contradicts the assumption that the strict supermartingale process has the highest movement in the class of supermartingale processes (which includes a martingale process), and we have a contradiction, completing the proof. □

B. Additional Material

B.1 Risk-Neutral Beliefs and Discount Rates

We again work in the context of the example in Section 2 for simplicity of exposition. The price of the terminal consumption claim is given in equilibrium in by $P_t(C_T) = \mathbb{E}_t\left[\beta_t^{T-t}\frac{U'(C_T)}{U'(C_t)}C_T\right]$, where β_t is now the agent's (possibly time-varying) time discount factor. Defining the gross return $R^C_{t,T} \equiv \frac{C_T}{P_t(C_T)}$, rearranging this equation for $P_t(C_T)$ yields

$$\begin{aligned}\mathbb{E}_t[R^C_{t,T}] &= \frac{1 - \text{Cov}_t\left(\beta_t^{T-t}\frac{U'(C_T)}{U'(C_T)}, C_T\right)}{\mathbb{E}_t\left[\beta_t^{T-t}\frac{U'(C_T)}{U'(C_t)}\right]} \\ &= \frac{\frac{U'(C_t)}{\beta_t^{T-t}} - \text{Cov}_t(U'(C_T), C_T)}{\mathbb{E}_t[U'(C_T)]},\end{aligned} \tag{B.1}$$

as usual. For full concreteness, we can write $\mathbb{E}_t[U'(C_T)] = \pi_t U'(C_{\text{low}}) + (1-\pi_t)U'(C_{\text{high}})$ in our two-state example, and $\text{Cov}_t(U'(C_T), C_T)$ can be similarly rewritten as a function of π_t, C_T, and $U'(C_T)$. This decomposition makes clear that intertemporal discount-rate variation can arise from four sources:

1. Changes in the time discount factor β_t.
2. Changes in contemporaneous marginal utility $U'(C_t)$.
3. Changes in the relative probability π_t.
4. Changes in state-contingent terminal consumption C_i and/or state-contingent marginal utility $U'(C_i)$.

Our framework thus allows for discount-rate variation arising from the first three sources, but not the last one. One might not consider this to be particularly restrictive in the context of this example; in theory, we can *define* the states such that the realization of the state fully determines consumption and marginal utility. But when taken to the data, we define states by the return on the market index, in which case this does become more restrictive. (We in fact slightly relax these assumptions and allow for independent consumption-growth or marginal-utility shocks for a given return state; Section 3.2 more fully discusses the models covered by our assumptions.)

Now consider a simple example in which a deterministic consumption stream for $t < T$ is given by $(C_0, C_1, C_2, C_3, \ldots, C_{T-1}) = (1, 1/2, 1, 1/2, \ldots)$ but π_t is constant at $\pi_t = \pi_0 = 0.5$ for $t < T$. Assume for simplicity that $\beta = 1$. Because the mapping between π_t and π_t^* is one-to-one for a given ϕ as in equation (10), measured risk-neutral beliefs would be constant for $t < T$ in this case: risk-neutral beliefs are invariant to changes in the risk-free rate arising from proportional changes to Arrow-Debreu state prices across the two states, as can be seen in equations (5) and (9), and all

discount-rate changes for the consumption claim are in fact driven by the risk-free rate in this case. The gross $(T-t)$-period risk-free rate with $\beta = 1$ is $R^f_{t,T} = \frac{U'(C_t)}{\mathbb{E}_t[U'(C_i)]}$ in equilibrium; we can thus rewrite (B.1) as

$$\mathbb{E}_t[R^C_{t,T}] = R^f_{t,T} - \frac{\text{Cov}_t(U'(C_T), C_T)}{\mathbb{E}_t[U'(C_T)]}, \tag{B.2}$$

and the second term is constant for $t < T$ under the current assumptions. But we need not restrict ourselves to settings in which all discount-rate variation arises due to changes in the risk-free rate. For example, with $\pi_0 = 0.3$, $C_t = \overline{C} = 1$ for $t < T$ and $\pi_1 = 0$ or 0.6 with equal probability, has no equity premium at $t = 1$ if $\pi_1 = 0$ since pricing is risk-neutral in this case (given that there is no risk); meanwhile, if $\pi_1 = 0.6$, then $\mathbb{E}_1[R^C_{1,T}] > R^f_{1,T}$ since the second term in (B.2) is positive. So the framework is capable of achieving identification in cases in which both the risk-free rate and risk premia are time-varying.

More generally, this example shows that the framework can handle cases in which an object that can be intuitively thought of as the quantity of aggregate risk is time-varying. As in Hansen and Jagannathan (1991), the conditional risk premium on any asset depends on the conditional volatility of the stochastic discount factor, which in this case is given for the horizon $T-t$ by $\text{Var}_t(\beta^{T-t}U'(C_T)/U'(C_t))$; we could rewrite (B.2) in terms of this value if desired. In the current example, this value is again equal to 0 at $t = 1$ if $\pi_1 = 0$, while it is positive if $\pi_1 = 0.6$. Further, while relative risk aversion (and thus the aggregate "price" of risk) is constant in the current example, nothing about the example restricts utility to take this form; we could, e.g., specify exponential utility and thus obtain time-varying relative risk aversion, and the analysis in Section 2.3 and here would nonetheless apply as well with slight modification.

Further, as discussed in Section 3.2, our framework in fact allows for much more general variation in discount rates; for example, permanent changes to the SDF are admissible, which (as discussed there) greatly broadens the scope of allowable variation relative to the constant-discount-rates framework.

B.2 Simulations for the Relationship of RN Prior and DGP with $\triangle$

As noted in Section 2.3, we run numerical simulations of a large number of DGPs and priors in order to understand the precise impact of the RN prior and DGP on $\triangle$ (and therefore $\mathbb{E}[X^*]$ for a given ϕ).

In particular, we consider the entire class of history-independent binary signal DGPs with a prior π^*_0 where $s_t \in \{l, h\}$ and $\mathbb{P}[s_t = h|\theta = 1]$ and (assumed lower) $\mathbb{P}[s_t = h|\theta = 0]$ are constant over t. These signal distributions imply likelihood ratios for the signals of $L_h \equiv \frac{\mathbb{P}[s_t=h|\theta=1]}{\mathbb{P}[s_t=h|\theta=0]} > 1$ and $L_l \equiv \frac{\mathbb{P}[s_t=l|\theta=0]}{\mathbb{P}[s_t=l|\theta=1]} > 1$. The simulations then allow the mapping of three variables L_h, L_l, and π^*_0 into a numerically estimated $\triangle$. We find:

CONCLUSIONS OF NUMERICAL SIMULATIONS:

1. When π^*_0 is low, $\triangle > 0$ is very unlikely: the percentage of DGPs with positive $\triangle$ given a

$\pi_0^* < .25$ is 2%. For $\pi_0^* < .5$, it is 11%.

2. When π_0^* is low, the only DGPs in which $\triangle > 0$ are very asymmetric and extreme. For example, when $\pi_0^* = .25$, $\triangle > 0$ only occurs if $\mathbb{P}[s_t = h|\theta = 1] > .95$ and $L_l > 2 \cdot L_h$.

3. The converse is true when π_0^* is high: $\triangle < 0$ is rare and only occurs given a very asymmetric and extreme DGP.

4. For symmetric DGPs ($L_h = L_l$), $\triangle \gtreqless 0$ when $\pi_0^* \gtreqless .5$.

5. Holding the DGP constant, $\triangle$ rises with π_0^*.

6. Holding all else constant, as L_h rises and the size of upward updates rises, $\triangle$ falls. As L_l rises and the size of upward-updates rises, $\triangle$ rises.

In order to present the results visually in a simple graph, we reduce the dimensionality of the mapping by focusing on the *likelihood ratio* $\frac{L_h}{L_l}$ rather than L_h and L_l individually (although this compression leads to a slightly messier graph). In particular, while the impact of both L_h and L_l on $\triangle$ appears monotonic, the impact of $\frac{L_h}{L_l}$ is only monotonic on average. For example, there are many combinations of L_h and L_l in which $\frac{L_h}{L_l} = 1$ but each combination leads to a different $\triangle$. Figure A.1 is a contour plot with the RN prior on the x-axis, with the y-axis stacking all of the DGP combinations in order of the likelihood ratio, and the contour colors showing the approximate value of $\triangle$ (darker colors corresponding to higher values) for each prior and DGP (with the dotted line highlighting the points at which $\triangle = 0$). For example, drawing a vertical line at a prior of $\pi_0^* = .25$ suggests that a large portion of DGPs produce a $\triangle < 0$ and the only DGPs that produce $\triangle > 0$ have extreme likelihood ratios.

We briefly note that these conclusions shows up in Table 1. For the first DGP, the RN prior is .5 and the signals are symmetric. Symmetry then requires that $\mathbb{E}[X^*|\theta = 0] = \mathbb{E}[X^*|\theta = 1]$ and $\triangle = 0$, with Proposition 2 then implying that $\mathbb{E}[X^*] = 0$. For the second DGP, the signals are such that updating sizes are asymmetric: updates upwards are large, whereas updates downward are small. Consequently, the expected movement given a state of 1 is large (.405) compared that given a state of 0 (.095), such that $\triangle$ is negative ($-.31$). Proposition 2 then implies that $\mathbb{E}[X^*]$ will also be negative. The opposite occurs in the third DGP, which is asymmetric in the opposite way and therefore leads to the opposite $\triangle = -.31$.

B.3 Description of Gabaix (2012) Rare Disasters Model for Example 2

Assume a representative agent with CRRA consumption utility, and assume that log consumption $c_t \equiv \log(C_t)$ and log dividends $d_t \equiv \log(D_t)$ evolve respectively according to

$$c_{t+1} = c_t + g_c + \varepsilon_{t+1}^c + \log(B_{t+1})\mathbb{1}\{\text{disaster}_{t+1}\},$$
$$d_{t+1} = d_t + g_d + \varepsilon_{t+1}^d + \log(F_{t+1})\mathbb{1}\{\text{disaster}_{t+1}\},$$

where $(\varepsilon^c_{t+1}, \varepsilon^d_{t+1})'$ is i.i.d. bivariate normal with mean zero and arbitrary covariance and is independent of all disaster-related variables,[3] and B_{t+1} and F_{t+1} are arbitrarily correlated random variables with support $[0,1]$ (or some discretization thereof) that affect consumption and dividends respectively in the case of a disaster in period $t+1$, which occurs with probability p_t. Define *resilience* H_t according to $H_t = p_t \mathbb{E}_t[B_{t+1}^{-\gamma} F_{t+1} - 1 \mid \mathbb{1}\{\text{disaster}_{t+1}\}]$, write $H_t = H_* + \widehat{H}_t$, and assume that the variable part follows

$$\widehat{H}_{t+1} = \frac{1+H_*}{1+H_t} e^{-\phi_H} \widehat{H}_t + \varepsilon^H_{t+1},$$

where $\mathbb{E}_t[\varepsilon^H_{t+1}] = 0$ and this shock is independent from all other shocks. Then the statements in Example 2 follow.

B.4 Description of Campbell–Cochrane (1999) Habit Formation Model for Example 3

Assume a representative agent with utility $\mathbb{E}_0\{\sum_{t=0}^{\infty} \beta^t [(C_t - H_t)^{1-\gamma} - 1]/(1-\gamma)\}$, where C_t is consumption and H_t is the level of habit, taken as exogenous by the agent. Defining the *surplus-consumption ratio* $S^c_t \equiv (C_t - H_t)/H_t$, assume that $s^c_t \equiv \log(S^c_t)$, $c_t \equiv \log(C_t)$, and log dividends $d_t \equiv \log(D_t)$ evolve respectively according to

$$\begin{aligned} s^c_{t+1} &= (1-\phi)\bar{s}^c + \phi s^c_t + \lambda(s^c_t)\varepsilon_{t+1}, \\ c_{t+1} &= g + c_t + \varepsilon_{t+1}, \\ d_{t+1} &= g + d_t + \eta_{t+1}, \end{aligned}$$

where $\varepsilon_{t+1} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2_\varepsilon)$ (see footnote 3), $\eta_{t+1} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2_\eta)$, $\text{Corr}(\varepsilon_{t+1}, \eta_{t+1}) = \rho$, and the *sensitivity function* $\lambda(s^c_t)$ is specified as

$$\lambda(s^c_t) = \left[\frac{1}{\overline{S}^c}\sqrt{1 - 2(s^c_t - \bar{s}^c)} - 1\right] \mathbb{1}\{s^c_t \leqslant s^c_{\max}\},$$

where $\overline{S}^c = \exp(\bar{s}^c) = \sigma_\varepsilon \sqrt{\gamma/(1-\phi)}$ is the assumed steady-state surplus-consumption ratio and $s^c_{\max} = \bar{s}^c + (1 - \overline{S}^c)^2/2$. Then the statement in Example 3 follows.

B.5 Simulations for Submartingale ϕ_t

For simplicity, we consider the framework in Section 2, and we simulate a situation in which the person learns about θ, $U'(C_{T,1})$ and $U'(C_{T,1})$ via constant binary DGPs with different combinations of signal strengths. Figure A.2 plots distributions for $\mathbb{E}[\mathsf{m}^*]$ (rather than $\mathbb{E}[X^*]$, since $\mathbb{E}[\mathsf{m}^*]$ in this case is what changes with the DGP) across these simulations.

First (using the very dark line), we consider the baseline situation in which $\pi^*_0 = .5$ and $\phi_t = 3$ for all t. This produces a symmetric distribution around 0. Minus smoothing and simulation errors,

[3] To be complete with respect to our discrete-state setting, we can assume $(\varepsilon^c_{t+1}, \varepsilon^d_{t+1})'$ is in fact an appropriately discretized normal distribution (e.g., a shifted binomial distribution).

$\mathbb{E}[X^*]$ never crosses .125 (the theoretical upper bound from Proposition 3 for $\mathbb{E}[X^*]$ given $\pi_0^* = .5$ and $\phi_t = 3$.)

Next, we allow additional uncertainty about $U'(C_{T,1})$ and $U'(C_{T,1})$ so that ϕ_t also evolves over time. To start, consider the slightly lighter-dark set of distributions with the label "low ϕ uncertainty." In this case, $U'(C_{T,1})$ is assumed to be 2.5 or 3.5 with equal probability and $U'(C_{T,0})$ is 0.833 or 1.167 with equal probability, so that $\phi_0 = 3$ but ϕ_T can vary from 2.14 to 4.2 (with a coefficient of variance of 12%). Each line then represents a different $\mathbb{E}[X]$ distribution given a different set of DGPs that reveal information about $U'(C_{T,1})$ and $U'(C_{T,0})$. In this case, changing ϕ has virtually no impact on the $\mathbb{E}[X^*]$ statistic regardless of how information is revealed about ϕ: the average $\mathbb{E}[X^*]$ rises by 0.0012, and the percentage of DGPs in which $\mathbb{E}[X^*]$ rises above the bound of .375 rises by just .00007. Similarly, in the gray set of distributions with the label "medium ϕ uncertainty," $U'(C_{T,1})$ is 2 or 4 and $U'(C_{T,0})$ is 0.667 or 1.333, so that $\phi_0 = 3$ but ϕ_T can vary from 1.5 to 6 (with a coefficient of variance of 54%). Even given this large uncertainty about ϕ_T, $\mathbb{E}[X^*]$ rises by 0.006 on average, and the percentage of DGPs above the bound rises by .0003. Finally, in the lightest-colored set of distributions with the label "high ϕ uncertainty," $U'(C_{T,1})$ is 1.5 or 4.5 and $U'(C_{T,0})$ is 0.5 or 1.5, so that $\phi_0 = 3$ but ϕ_T can vary from 1 to 9 (with a coefficient of variance of 100%). Given this extreme uncertainty about ϕ_T, average $\mathbb{E}[X^*]$ still only rises by 0.015, and the percentage of DGPs above the bound rises by .0012.

B.6 Solution Method and Simulations for Habit Formation Model

See Appendix B.4 for a description of the model, and the calibrated parameters are identical to those used by Campbell and Cochrane (1999, Table 1), converted to daily values, for the version of their model with imperfectly correlated consumption and dividends. We consider 90-day option-expiration horizons (i.e., $T_i - 0_i = 90$), and after solving the model for the price-dividend ratio, we then solve for the joint distribution for returns (from t to T_i) and the SDF at every point in a gridded state space as of $t = T_i - 1$, then $t = T_i - 2$, and so on, as below.

The initial market index value is normalized to $V_{0_i}^m = 1$, and the joint CDF for the SDF realization and the return as a function of the current surplus-consumption state is then solved by iterating backwards from T_i: after solving the model for the price-dividend ratio as a function of the surplus-consumption value, we then calculate the $T_i - 1$ CDF for any possible surplus-consumption value by integrating over the distributions of shocks to consumption (and thus surplus consumption) and dividends at T_i; we then project this CDF onto an interpolating cubic spline over the three dimensions $(S_{T_i-1}^c, M_{T_i}, \log(R_{T_i}^{m,e}))$; we then calculate the $T_i - 2$ CDF by integrating over the distribution of shocks at $T_i - 1$ *and* the projection solutions for the conditional distribution functions for $(T_i - 1) \to T_i$ obtained in the previous step; and so on. These CDFs are then used for the model simulations.

We conduct 25,000 simulations, where each simulation runs from 0_i to T_i, and for which the initial surplus-consumption state is drawn from its unconditional distribution. For each period in each simulation, we evaluate risk-neutral beliefs over return states at every point in the space

$\mathcal{S}_{\text{baseline}}$ as used above and use these to calculate the set of conditional risk-neutral beliefs $\{\widetilde{\pi}^*_{t,i,j}\}_j$. Further, we store the associated set of expected SDF slopes $\{\phi_{t,i,j}\}_j$. We can thus calculate the true average values of these objects of interest, $\overline{\phi}_{0_i,j} \equiv \widehat{\mathbb{E}}[\phi_{0_i,i,j}]$, where $\widehat{\mathbb{E}}[\cdot]$ denotes the expectation over all simulations i and we have fixed the state pair j. And using the risk-neutral beliefs series, we can naïvely apply our theoretical bound in Proposition 9 to obtain lower-bound estimates for those SDF slopes and compare those estimates to the true simulated values. Relative risk aversion for this model's representative agent does not match the definition used in Proposition 8, as this agent's utility does not depend only on terminal wealth (see Campbell and Cochrane, 1999, Section IV.B), so we accordingly present estimates for the SDF slope rather than for relative risk aversion.

Figure A.3 presents these simulation results. The blue circles show the true simulated average values of the SDF slopes $\overline{\phi}_{0_i,j}$, while the red triangles show the naïve lower-bound estimates of these values using our theoretical bound on the simulated risk-neutral beliefs data. Considering the first question posed at the outset of this subsection, it is clear in both cases that these SDF-slope values are far below those obtained from our empirical estimates above, so the model does not replicate the observed variation in risk-neutral beliefs even with the violation of CTI. We can understand the validity of the theoretical bound for the interior states by way of Proposition 11, which shows that the bounds hold approximately for violations of CTI for which the $\phi_{t,i,j}$ process is close to a martingale. In our simulations, the values $|\widehat{\mathbb{E}}[\phi_{t+1,i,j} - \phi_{t,i,j}]|$ for different state pairs j range from a minimum of 0.00002 to a maximum of 0.00011, which is not large enough to invalidate the theoretical bounds.

B.7 Data Cleaning and Measurement of Risk-Neutral Distribution

Before detailing measurement of the risk-neutral distribution, we note that we must collect additional data in order to follow the procedure below. In particular, in order to obtain the ex post return state for each option expiration date T_i (and thereby assign probability 1 to that state on date T_i, so that our streams are resolving), we need S&P 500 index prices used as option settlement values. Our first step in this exercise is therefore to obtain end-of-day index prices (which we take as well from OptionMetrics). But the settlement value for many S&P 500 options in fact reflects the opening (rather than closing) price on the expiration date; for example, the payoff for the traditional monthly S&P 500 option contract expiring on the third Friday of each month depends on the opening S&P index value on that third Friday morning, while the payoff for the more recently introduced end-of-month option contract depends on the closing S&P index value on the last business day of the month.[4] To obtain the ex-post return state for A.M.-settled options, we hand-collect the option settlement values for these expiration dates from the Chicago Board Options Exchange (CBOE) website, which posts these values.

In addition, in order to measure the risk-neutral distribution *and* to measure realized excess index returns, we need risk-free zero-coupon yields R^f_{t,T_i} for $t = 0_i, \ldots, T_i - 1$. To obtain these,

[4] See http://www.cboe.com/SPX for further detail. For our dataset, the majority (roughly 2/3) of option expiration dates correspond to A.M.-settled options.

we follow van Binsbergen, Diamond, and Grotteria (2022) and obtain the relevant yield directly from the cross-section of option prices by applying the put-call parity relationship. We apply their "Estimator 2," which obtains $R^f_{t,T_i} = \beta^{-1/T}$ from Theil–Sen (robust median) estimation of $q^{m,\text{put}}_{t,i,K} - q^{m,\text{call}}_{t,i,K} = \alpha + \beta K + \varepsilon_{t,i,K}$. This provides a very close fit to the option cross-sections (see van Binsbergen, Diamond, and Grotteria, 2022, for details) and thus produces a risk-free rate consistent with observed option prices, as is necessary to correctly back out the risk-neutral distribution.

Finally, for both the OptionMetrics end-of-day and CBOE intraday data, we apply standard filters (e.g., Christoffersen, Heston, and Jacobs, 2013; Constantinides, Jackwerth, and Savov, 2013; Martin, 2017) to the raw option-price data before estimating risk-neutral distributions. We drop any options with bid or ask price of zero (or less than zero), with uncomputable Black–Scholes implied volatility or with implied volatility of greater than 100 percent, with more than one year to maturity, or (for call options) with mid prices greater than the price of the underlying; we drop any option cross-section (i.e., the full set of prices for the pair (t, T_i)) with no trading volume on date t, with fewer than three listed prices across different strikes, or for which there are fewer than three strikes for which both call and put prices are available (as is necessary to calculate the forward price and risk-free rate); and after transforming the data to a risk-neutral distribution as below, we keep only conditional RN belief observations $\widetilde{\pi}^*_{t,i,j}$ for which the non-conditional beliefs satisfy $\pi^*_t(R^m_{T_i} = \theta_j) + \pi^*_t(R^m_{T_i} = \theta_{j+1}) \geqslant 5\%$. Our bounds can be calculated using data of arbitrary frequency, so we calculate $X^*_{i,j}$ using changes in RN beliefs over whatever set of trading days are left in the sample after this filtering procedure.

As introduced in Section 5.1, we measure the risk-neutral distribution for returns by applying the following steps to the remaining observed option prices (for which we use mid prices), following Malz (2014):

1. Transform the collections of call- and put-price cross-sections (for example, for call options on date t for expiration date T_i, this set is $\{q^m_{t,i,K}\}_{K\in\mathcal{K}}$) into Black–Scholes implied volatilities.

2. Discard the implied volatility values for in-the-money calls and puts, so that the remaining steps use data from only out-of-the-money put and call prices (as, e.g., in Martin, 2017). Moneyness is measured relative to the at-the-money-forward price, measured (again following Martin, 2017) as the strike K at which $q^{m,\text{put}}_{t,i,K} = q^{m,\text{call}}_{t,i,K}$.

3. Fit a cubic spline to interpolate a smooth function between the points in the resulting implied-volatility schedule for each trading date–expiration date pair. The spline is *clamped*: its boundary conditions are that the slope of the spline at the minimum and maximum values of the knot points $\mathcal{K}$ is equal to 0; further, to extrapolate outside of the range of observed knot points, set the implied volatilities for those unobserved strikes equal to the implied volatility for the closest observed strike (i.e., maintain a slope of 0 for the implied-volatility schedule outside the observed range).

4. Evaluate this spline at 1,901 strike prices, for S&P index values ranging from 200 to 4,000 (so that the evaluation strike prices are $K = 200, 202, \ldots, 4000$), to obtain a set of implied-volatility

values across this fine grid of possible strike prices for each (t, T_i) pair.[5]

5. Invert the resulting smoothed 1,901-point implied-volatility schedule for each (t, T_i) pair to transform these values back into call prices, and denote this fitted call-price schedule as $\{\hat{q}^m_{t,i,K}\}_{K\in\{200,202,\ldots,4000\}}$.

6. Calculate the risk-neutral CDF for the date-T_i index value at strike price K using $\mathbb{P}^*_t(V^m_{T_i} < K) = 1 + R^f_{t,T_i}(\hat{q}^m_{t,i,K} - \hat{q}^m_{t,i,K-2})/2$. (See the proof of equation (22) in Appendix A.2 for a derivation of this result; the index-value distance between the two adjacent strikes is equal to 2 given that we evaluate the spline at intervals of two index points.)

7. Defining $V^m_{i,j,\max}$ and $V^m_{i,j,\min}$ to be the date-T_i index values corresponding to the upper and lower bounds, respectively, of the bin defining return state θ_j,[6] we then calculate the risk-neutral probability that return state θ_j will be realized at date T_i, referred to with slight notational abuse as $\mathbb{P}^*_t(\theta_j)$, as

$$\mathbb{P}^*_t(\theta_j) = \mathbb{P}^*_t(V^m_{T_i} < V^m_{i,j,\max}) - \mathbb{P}^*_t(V^m_{T_i} < V^m_{i,j,\min}),$$

where the CDF values are taken from the previous step using linear interpolation between whichever two strike values $K \in \{200, 202, \ldots, 4000\}$ are nearest to $V^m_{i,j,\max}$ and $V^m_{i,j,\min}$, respectively.

Steps 1 and 2 represent the only point of distinction between our procedure and that of Malz, who assumes access to a single implied-volatility schedule without considering put or call prices directly; our procedure is accordingly essentially identical to his. Note that we transform the option prices into Black–Scholes implied volatilities simply for purposes of fitting the cubic spline and then transform these implied volatilities back into call prices before calculating risk-neutral beliefs, so this procedure does *not* require the Black–Scholes model to be correct.[7] The clamped cubic spline proposed by Malz (2014), and used in step 3 above, is chosen to ensure that the call-price schedule obtained in step 5 is decreasing and convex with respect to the strike price outside the range of observable strike prices, as required under the restriction of no arbitrage. Violations of these restrictions *inside* the range of observable strikes, as observed infrequently in the data, generate negative implied risk-neutral probabilities; in any case that this occurs, we set the associated risk-neutral probability to 0.

As noted in step 3, the clamped spline is an *interpolating* spline, as it is restricted to pass through all the observed data points so that the fitted-value set $\{\hat{q}^m_{t,i,K}\}$ contains the original values $\{q^m_{t,i,K}\}$. Some alternative methods for measuring risk-neutral beliefs use smoothing splines that are not constrained to exhibit such interpolating behavior. To check the robustness of our results to the

[5] This set of ∼1,900 strike prices is on average about 20 times larger than the set of strikes for which there are prices in the data, as there is a mean of roughly 90 observed values in a typical set $\{q^m_{t,i,K}\}_{K\in\mathcal{K}}$.

[6] That is, formally, $V^m_{i,j,\min} = R^f_{0_i,T_i} V^m_{T_0} \exp(\theta_j - 0.05)$ and $V^m_{i,j,\max} = R^f_{0_i,T_i} V^m_{0_i} \exp(\theta_j)$. For example, for excess return state θ_2, we have $V^m_{i,j,\min} = R^f_{0_i,T_i} V^m_{0_i} \exp(-0.2)$ and $V^m_{i,j,\max} = R^f_{0_i,T_i} V^m_{T_0} \exp(-0.15)$.

[7] We conduct this transformation following Malz (2014), as well as much of the related literature, which argues that these smoothing procedures tend to perform slightly better in implied-volatility space than in the option-price space given the convexity of option-price schedules; see Malz (1997) for a discussion.

choice of measurement technique, we have accordingly used one such alternative method proposed by Bliss and Panigirtzoglou (2004). Empirical results obtained using risk-neutral beliefs calculated in this alternative manner are unchanged as compared to the benchmark results in Section 5.4.

We have also conducted robustness tests with respect to the fineness of the grid on which we evaluate the spline in step 4 and calculate the risk-neutral CDF in step 6, with results from these exercises also indistinguishable from the benchmark results.

B.8 Matching Noise Variance Estimates to X^* Observations

After estimating $\text{Var}(\epsilon_t) = \text{Var}(\epsilon_{t,i,j})$ separately for each combination of trading day t, expiration date T_i, and return state pair j in our intraday sample following Section 5.2, we must then match the noise estimates (which are obtained only for a subsample of days) to the observed excess movement observations in our original daily data. To do so, we take advantage of the fact that the best predictors of $\widehat{\text{Var}}(\epsilon_{t,i,j})$ are (i) state pair j (we see more noise for tail states) and (ii) the observed RN belief of either θ_j or θ_{j+1} being realized, $\Sigma^*_{t,i,j} \equiv \pi^*_t(R^m_{T_i} = \theta_j) + \pi^*_t(R^m_{T_i} = \theta_{j+1})$ (conditional beliefs are noisier when the underlying sum $\Sigma^*_{t,i,j}$ is lower, as $\Sigma^*_{t,i,j}$ enters into the denominator of $\widetilde{\pi}^*_{t,i,j}$). We thus partition $\Sigma^*_{t,i,j}$ into 5-percentage-point bins ($[0, 0.05)$, $[0.05, 0.1]$, ...), and then calculate the average noise $\widehat{\sigma}_{\epsilon,j,\Sigma} \equiv \overline{\widehat{\text{Var}}(\epsilon_{t,i,j})}$ for each combination of state pair j and bin for $\Sigma^*_{t,i,j}$. We then match $\widehat{\sigma}_{\epsilon,j,\Sigma}$ to each observed one-day excess movement observation $\widehat{X}^*_{t,t+1,i,j}$ in our original end-of-day data, based on that observation's state j and total probability $\Sigma^*_{t,i,j}$.

B.9 Details of Bootstrap Confidence Intervals

Our block-bootstrap resampling procedure is described in Section 5.4, and we provide further details on how we construct our one-sided confidence intervals for Table 4 here. Fixing a given $\overline{\phi}$, denote the point estimate for $\overline{e_i^{\text{main}}(\phi)}$ by $\widehat{e}(\overline{\phi})$. The null that $\overline{e_i^{\text{main}}(\phi)} = 0$ is rejected at the 5% level if $2\widehat{e}(\overline{\phi}) - e^*_{(0.95)}(\overline{\phi}) > 0$, where $e^*_{(0.95)}(\overline{\phi})$ is the 95$^{\text{th}}$ percentile of the bootstrap distribution of $\overline{e_i^{\text{main}}(\phi)}$ statistics (i.e., it is rejected if it is outside of the one-sided 95% basic bootstrap CI for $\overline{e_i^{\text{main}}(\phi)}$). We conduct this procedure for all possible $\overline{\phi}$ values, and we obtain $\widehat{\phi}_{LB} = \min_{\overline{\phi}}$ s.t. $2\widehat{e}(\overline{\phi}) - e^*_{(0.95)}(\overline{\phi}) \leqslant 0$.

A more straightforward procedure for conducting inference on $\overline{\phi}$ would be to construct the basic bootstrap CI directly for $\overline{\phi}$ (i.e., $\widehat{\phi}_{LB} = 2\widehat{\phi} - \phi^*_{(0.95)}$). The challenge preventing us from doing so is that in nearly all cases, the 95$^{\text{th}}$ percentile of the bootstrap distribution for $\widehat{\phi}$ is ∞, given how large our point estimates are (and how much excess movement we observe in our data). This motivates our use of a test-inversion confidence interval using the residuals for different possible values of $\overline{\phi}$, which solves this problem. These CIs achieve asymptotic coverage of at least the nominal level under weak conditions (discussed further below), given the duality between testing and CI construction; see, e.g., Carpenter (1999). We find that our procedure performs quite well, with unbiased and symmetric bootstrap distributions around the full-sample point estimate.

We note that our bootstrap procedure fully preserves the groupings of return-state pairs (in-

dexed by $j = 1, \ldots, J-1$) for each set of observations indexed by i (corresponding to the option expiration date) within each block, as we split the observations into blocks only by time and not by return states. We do so in order to obtain valid inference for the aggregate value $\overline{\boldsymbol{\phi}}$, which uses observations for state pairs $(\theta_2, \theta_3), \ldots, (\theta_{J-2}, \theta_{J-1})$, in the face of arbitrary dependence for the observations across those state pairs and a fixed number of return states J (whereas we assume $N \to \infty$, and further the number of blocks $B \to \infty$ according to a sequence such that $(T_N + 1)/B \to \infty$). In this way our procedure is in fact a *panel* (or *cluster*) *block bootstrap*; see, for example, Palm, Smeekes, and Urbain (2011). Lahiri (2003, Theorem 3.2) provides a weak condition on the strong mixing coefficient of the relevant stochastic process — in our case, $\{(X^*_{i,j}, \widetilde{\pi}^*_{0,i,j}, \{\widehat{\mathrm{Var}}(\epsilon_{t,i,j})\})_{t,j}\}_i$ — under which the blocks are asymptotically independent and the bootstrap distribution estimator is consistent for the true distribution under the asymptotics above, so that our test-inversion confidence intervals have asymptotic coverage probability of at least 95% for the population parameters of interest in the presence of nearly arbitrary (stationary) autocorrelation and heteroskedasticity.[8] This coverage rate may in fact be greater than 95% given that we are estimating lower bounds for the parameters of interest rather than the parameters themselves, and this motivates our use of one-sided rather than two-sided confidence intervals, as in Section 5.4.

B.10 Regressions for RN Excess Movement

As discussed in Section 5.5, we consider reduced-form evidence on the macroeconomic and financial correlates of RN excess movement. Table A.1 shows the results of the regressions discussed in that section. The dependent variable is the quarterly average of noise-adjusted RN excess movement $X^*_{t,t+1,i,j}$, where the average is calculated across all available expiration dates and interior state pairs for all trading days in a quarter. We aggregate to the quarterly level given the frequency of data available for the regressors we consider, and we use quarterly averages as well for any dependent variables with data available at a higher frequency. Aside from the constant and time trend, all variables (both dependent and independent) are normalized to have unit standard deviation for purposes of interpretation, and we present heteroskedasticity- and autocorrelation-robust standard errors using the equal-weighted periodogram estimator of the long-run variance; see Lazarus, Lewis, and Stock (2021) for results on the optimality properties of this estimator.

Across all specifications — see columns (1), (4), and (5), in particular — the liquidity- and noise-related variables (bid-ask spreads and volume) have coefficients that are both economically and statistically small, which provides further evidence that factors specific to the option market (or mismeasurement of RN beliefs) are not driving our results. By contrast, excess movement has a significant positive relationship with the VIX in (2), as is intuitive. Lagged S&P 500 returns

[8] There are additional conditions required for the result of Lahiri (2003, Theorem 3.2) to hold, but they will hold trivially in our context under the RE null given the boundedness of the relevant belief statistics. Our block bootstrap is a non-overlapping block bootstrap (NBB); others (Künsch, 1989; Liu and Singh, 1992) have proposed a *moving* block bootstrap (MBB) using overlapping blocks, among other alternatives. While the MBB has efficiency gains relative to the NBB (Hall, Horowitz, and Jing, 1995), these are "likely to be very small in applications" (Horowitz, 2001, p. 3190), so we use the NBB for computational convenience.

and valuation ratios are also positively related to excess movement, depending on the particular specification. The R^2 value for the regression with all right-hand-side variables included is 0.61, indicating that these statistics are capable of jointly accounting for a significant portion of the quarterly variation in excess movement.

Additional References

AUGENBLICK, N. AND M. RABIN (2021): "Belief Movement, Uncertainty Reduction, and Rational Updating," *Quarterly Journal of Economics*, 136, 933–985.

VAN BINSBERGEN, J. H., W. F. DIAMOND, AND M. GROTTERIA (2022): "Risk-Free Interest Rates," *Journal of Financial Economics*, 143, 1–29.

BLACK, F. AND M. SCHOLES (1973): "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81, 637.

BLISS, R. R. AND N. PANIGIRTZOGLOU (2004): "Option-Implied Risk Aversion Estimates," *Journal of Finance*, 59, 407–446.

BREEDEN, D. T. AND R. H. LITZENBERGER (1978): "Prices of State-Contingent Claims Implicit in Option Prices," *Journal of Business*, 51, 621–651.

BROWN, D. J. AND S. A. ROSS (1991): "Spanning, Valuation and Options," *Economic Theory*, 1, 3–12.

CAMPBELL, J. Y. (2018): *Financial Decisions and Markets: A Course in Asset Pricing*, Princeton: Princeton University Press.

CAMPBELL, J. Y. AND J. H. COCHRANE (1999): "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107, 205–251.

CARPENTER, J. (1999): "Test Inversion Bootstrap Confidence Intervals," *Journal of the Royal Statistical Society, Series B*, 61, 159–172.

CHRISTOFFERSEN, P., S. HESTON, AND K. JACOBS (2013): "Capturing Option Anomalies with a Variance-Dependent Pricing Kernel," *Review of Financial Studies*, 26, 1962–2006.

CONSTANTINIDES, G. M., J. C. JACKWERTH, AND A. SAVOV (2013): "The Puzzle of Index Option Returns," *Review of Asset Pricing Studies*, 3, 229–257.

EPSTEIN, L. G. AND S. E. ZIN (1989): "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 57, 937–969.

GABAIX, X. (2012): "Variable Rare Disasters: An Exactly Solved Framework for Ten Puzzles in Macro-Finance," *Quarterly Journal of Economics*, 127, 645–700.

HALL, P., J. L. HOROWITZ, AND B.-Y. JING (1995): "On Blocking Rules for the Bootstrap with Dependent Data," *Biometrika*, 82, 561–574.

HANSEN, L. P. AND R. JAGANNATHAN (1991): "Implications of Security Market Data for Models of Dynamic Economies," *Journal of Political Economy*, 99, 225–262.

HOROWITZ, J. L. (2001): "The Bootstrap," in *Handbook of Econometrics*, ed. by J. J. Heckman and E. Leamer, Amsterdam: Elsevier, vol. 5, chap. 52, 3159–3228.

KÜNSCH, H. R. (1989): "The Jackknife and the Bootstrap for General Stationary Observations," *Annals of Statistics*, 17, 1217–1241.

LAHIRI, S. N. (2003): *Resampling Methods for Dependent Data*, New York: Springer.

LAZARUS, E., D. J. LEWIS, AND J. H. STOCK (2021): "The Size-Power Tradeoff in HAR Inference," *Econometrica*, 89, 2497–2516.

LAZARUS, E., D. J. LEWIS, J. H. STOCK, AND M. W. WATSON (2018): "HAR Inference: Recommendations for Practice," *Journal of Business & Economic Statistics*, 36, 541–559.

LIU, R. AND K. SINGH (1992): "Moving Blocks Jackknife and Bootstrap Capture Weak Dependence," in *Exploring the Limits of the Bootstrap*, ed. by R. LePage and L. Billard, New York: Wiley, chap. 11, 225–248.

MALZ, A. M. (1997): "Option-Implied Probability Distributions and Currency Excess Returns," *Federal Reserve Bank of New York Staff Report No. 32*.

——— (2014): "A Simple and Reliable Way to Compute Option-Based Risk-Neutral Distributions," *Federal Reserve Bank of New York Staff Report No. 677*.

MARTIN, I. (2017): "What Is the Expected Return on the Market?" *Quarterly Journal of Economics*, 132, 367–433.

PALM, F. C., S. SMEEKES, AND J.-P. URBAIN (2011): "Cross-Sectional Dependence Robust Block Bootstrap Panel Unit Root Tests," *Journal of Econometrics*, 163, 85–104.

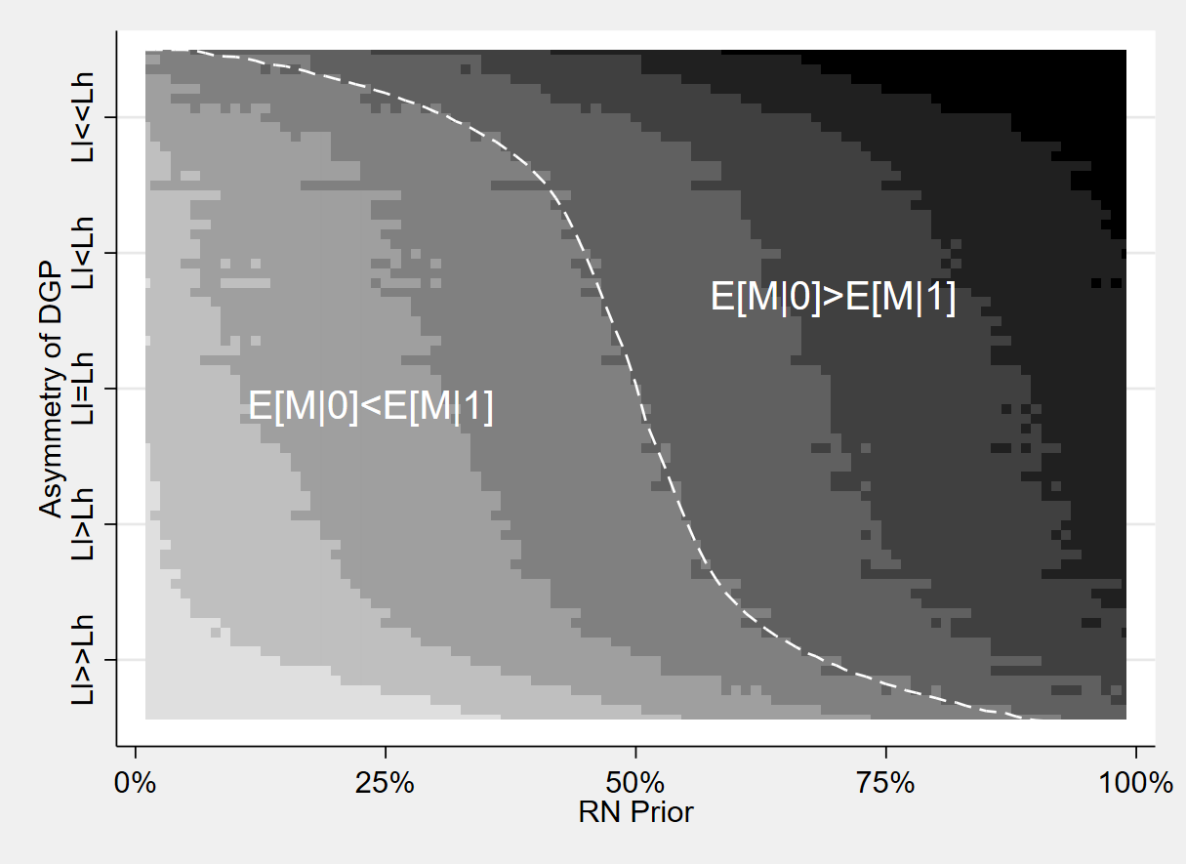
Additional Tables and Figures

Table A.1: Regressions for Quarterly Average of RN Excess Movement

	(1)	(2)	(3)	(4)	(5)
Bid-Ask Spread	0.22			-0.20	-0.26
	[0.12]			[0.13]	[0.15]
Option Volume	0.12			0.13	0.17
	[0.07]			[0.07]	[0.09]
VIX		0.55		0.84	0.91
		[0.16]		[0.17]	[0.18]
Baker–Bloom–Davis Uncertainty		-0.36		0.08	0.05
		[0.15]		[0.13]	[0.12]
12-Month S&P Return			0.03	0.50	0.50
			[0.09]	[0.12]	[0.12]
Price to 10-Year Earnings Ratio			0.40	0.34	0.38
			[0.11]	[0.11]	[0.12]
Time					0.00
					[0.00]
R^2	0.09	0.26	0.18	0.60	0.61
N	88	88	88	88	88

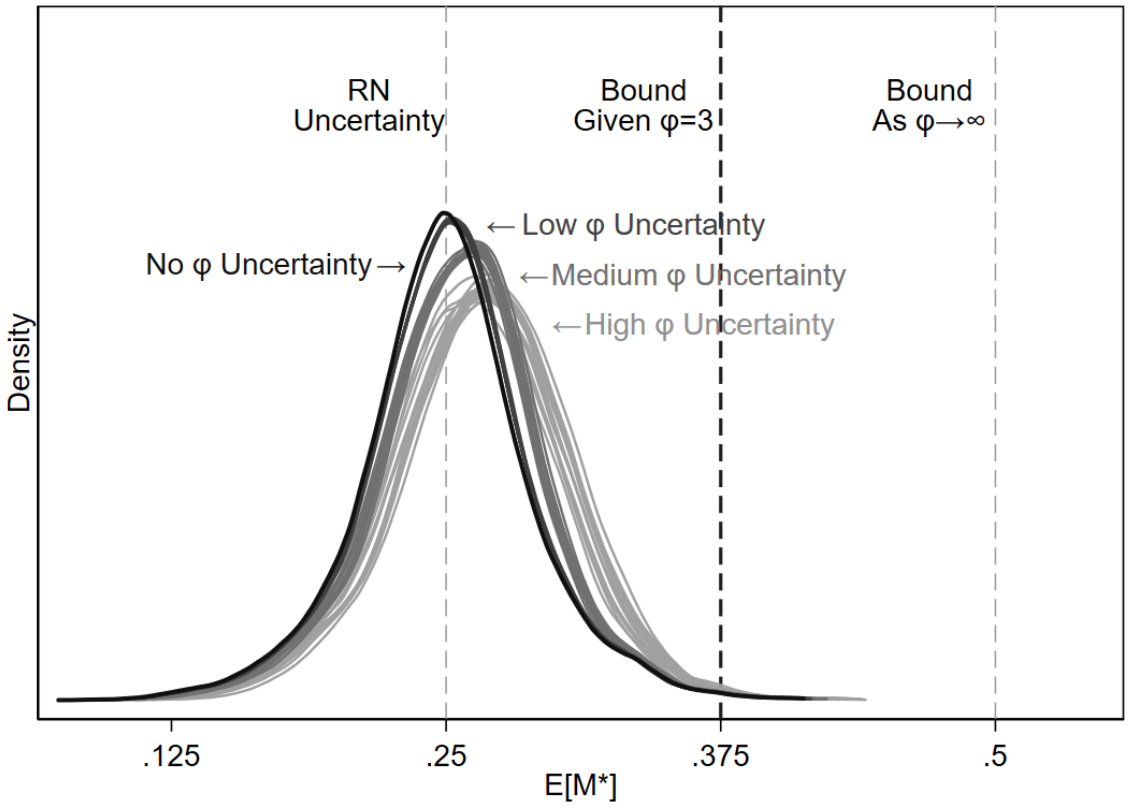
Notes: Dependent variable in all regressions is the empirical average $\widehat{\mathbb{E}}[X^*_{t,t+1,i,j}]$ calculated across all available expiration dates and interior state pairs, using all trading dates t within each given quarter. Regressors are correspondingly quarterly averages of each relevant series. All variables (dependent and independent, aside from time trend) are normalized to have unit standard deviation. Constant is included in each regression. Heteroskedasticity- and autocorrelation-robust standard errors are in parentheses, calculated using the equal-weighted periodogram orthonormal series estimator for the long-run variance with 8 degrees of freedom, following the formula recommended in Lazarus, Lewis, Stock, and Watson (2018).

Figure A.1: Contour Plot: Simulations for $\triangle$ by DGP and π_0^*



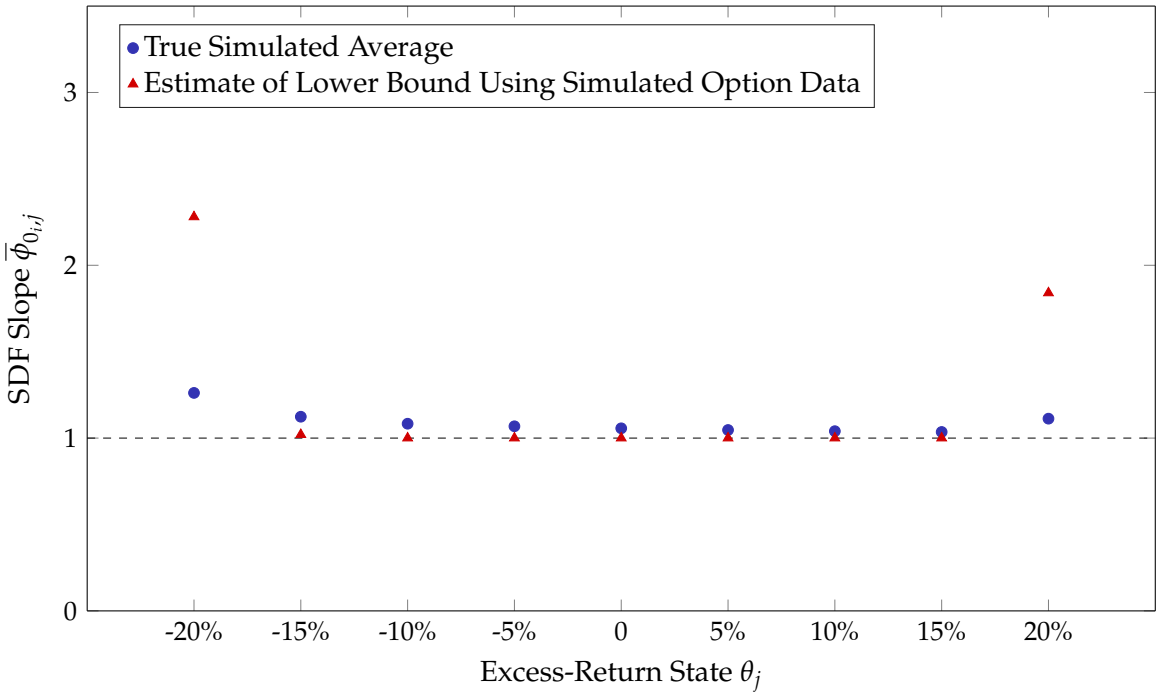
Note: See text in Appendix B.2 for description of simulations.

Figure A.2: RN Belief Movement Distributions with Time-Varying ϕ_t



Note: See text in Appendix B.5 for description of simulations.

Figure A.3: Estimates of SDF Slope in Habit Formation Model Simulations



Notes: See text in Appendix B.6 for description of simulations.