Excess Movement in Option-Implied Beliefs*

Ned Augenblick and Eben Lazarus

SEPTEMBER 2025

Abstract

We derive bounds on the rational variation in option-implied beliefs about market returns. While risk preferences distort option-implied (or *risk-neutral*) beliefs away from subjective beliefs, one can nonetheless bound risk-neutral belief movement under a general assumption on the stochastic discount factor. The resulting test requires no knowledge of the objective distribution and allows significantly more flexibility in discount rates than standard volatility tests. Implementing our test empirically using index options, we find that there is so much movement in risk-neutral beliefs that the bounds are routinely violated. Our results imply significant excess index-price volatility.

KEYWORDS: Option prices, volatility bounds, beliefs, rational expectations, risk aversion

JEL CODES: D84, G13, G14, G41

*Previous versions of this paper circulated under the titles "Restrictions on Asset-Price Movements Under Rational Expectations: Theory and Evidence" and "A New Test of Excess Movement in Asset Prices." We are grateful to Jonathan Lewellen, two anonymous referees, and an associate editor for very helpful feedback, as well as Francesca Bastianello, Laura Blattner, Jaroslav Borovička, John Campbell, Gabriel Chodorow-Reich, Emmanuel Farhi, Nicola Fusari, Xavier Gabaix, Niels Gormsen, Sam Hanson, Bryan Kelly, Leonid Kogan, David Laibson, Chen Lian, Ian Martin, Jonathan Parker, Matthew Rabin, Andrei Shleifer, Juan Sotes-Paladino, David Sraer, Jim Stock, Adrien Verdelhan, Jessica Wachter, and seminar participants at Harvard, MIT, Berkeley, Chicago Booth, Yale SOM, Stanford GSB, Northwestern Kellogg, Duke Fuqua, LSE, University of Sydney, Bocconi University, City University of Hong Kong, the AEA Annual Meeting, SITE, FIRN, the Chicago Fed Rookie Conference, the NBER Behavioral Finance Meeting, and the NBER Summer Institute Asset Pricing Meeting. Augenblick and Lazarus are at the Haas School of Business, University of California, Berkeley. Contact: ned@haas.berkeley.edu and lazarus@berkeley.edu.

1. Introduction

Option markets are a rich source of information about the perceived distribution of future stock returns. Like the overall market itself, option-implied beliefs about future returns are highly volatile over time. Testing whether they are too volatile relative to a rational benchmark, though, is a difficult task. First, subjective beliefs are not directly observable, as option prices also depend on unobservable risk and time preferences. Second, even if subjective beliefs were observable, we do not know the objectively correct conditional probabilities to use as a benchmark for comparison.

In this paper, we show that it is nonetheless possible to place bounds on the rational variation in option prices, even without knowledge of the correct beliefs and with significant flexibility in preferences and discount rates. Implementing our bounds empirically, we find that there is so much price variation that the bounds are routinely violated. This implies that there must either be excess volatility in beliefs relative to rationality, or very large, high-frequency variation in the price of risk over small changes in the market index. In additional analysis, we find little evidence that rational variation in the price of risk is a key driver of our results, indicating that excess volatility in beliefs likely plays a role. Our findings imply significant excess volatility in the underlying price.

We focus our analysis on index options. As these options are effectively bets on the future value of the index, their prices can be used to measure a market-implied belief distribution over future index returns. Specifically, we use standard methods to transform option prices into so-called *risk-neutral (RN) beliefs*, which are a function of (i) the relative likelihood of different return states and (ii) the relative value of a marginal dollar in each state (which depends on risk preferences). Importantly, as RN beliefs are constructed using the *relative* prices of options with different strike prices but the same maturity, they strip out variation arising from time discounting and common unobservable shocks (e.g., to marginal utility or the quantity of risk) that have the same effect on options with different strikes. We can accordingly make informative statements about rational price variation with significantly weaker assumptions than in past volatility tests for the price of the underlying index.[1]

To introduce the logic of our test, we begin in a stripped-down setting and build up to an empirically realistic environment piece by piece. In order to isolate the first component of RN beliefs, assume for now that we can directly observe a person's subjective (or *physical*) beliefs $\pi_t(\theta = 1)$ over some binary outcome $\theta \in \{0, 1\}$. Suppose we observe π_t repeatedly oscillating from 0.90 to 0.10 and back as t progresses. While it is of course possible to rationalize this movement ex post by constructing a particular set of signal realizations from some data-generating process (DGP), this amount of movement appears intuitively "rare" for someone with rational expectations (RE). Formalizing this idea, 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 *initial uncertainty* ($\pi_0(1 - \pi_0)$) under RE, *regardless* of the DGP.

[1] For example, Shiller (1981) famously documented excess volatility in stock prices relative to a measure of fundamental value, but his measure requires that discount rates over future cash flows remain constant over time. As discount rates appear to vary significantly, what to make of his results remains contested (Fama, 1991).

Equivalently, expected *excess movement* — belief movement minus initial uncertainty, which we denote by $\mathbb{E}[X]$ — must always be zero. Intuitively, belief movements imply learning, and rational learning implies a concomitant reduction of uncertainty (from the initial level $\pi_0(1-\pi_0)$ to 0) on average. Therefore, if one observes a set of empirical belief streams for which excess movement is far from zero on average, the null of rationality can be statistically rejected at some confidence level. In other words, one can make statements about excess volatility in belief streams even without knowledge of the correct beliefs period by period.

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 (π_t^*) in the general case in which only risk-neutral rather than physical beliefs are observable. Our task becomes considerably more difficult in this case because risk-preference distortions allow RN excess movement $\mathbb{E}[X^*]$ to be non-zero, and its sign and magnitude depend on the degree of risk aversion and the precise form of the DGP. Put differently, non-zero excess movement in risk-neutral beliefs may in fact be consistent with rational updating. But even while allowing meaningful degrees of freedom in risk preferences and the form of the DGP, we show that the admissible $\mathbb{E}[X^*]$ can be bounded as a simple function of risk aversion over the terminal states (or, more generally, the slope of the stochastic discount factor across states).

The broad intuition carries over from the previous case. Suppose, for illustration, that a person's valuation for a binary option that pays \$1 in state $\theta = 1$ at time T oscillates between \$0.90 and \$0.10, while her valuation for an option that pays \$1 in state $\theta = 0$ oscillates between \$0.10 and \$0.90. Risk-neutral beliefs π_t^* here are also moving back and forth between 0.90 and 0.10. But physical beliefs are not pinned down, as relative marginal utility (MU) in the two states is unknown. For example, the person might see the states as equally likely ex ante ($\pi_0 = 0.50$), but value the first option nine times higher because she values \$1 in state $\theta = 1$ nine times more. But if this marginal-utility ratio — which we call ϕ — is constant over time, then a movement in the RN belief from 0.90 to 0.10 means that physical beliefs are oscillating between 0.50 and 0.01, which is still rare for a Bayesian. The same logic applies for any given value of ϕ as long as this parameter is relatively stable within a belief stream, an assumption we make to begin with and return to below.[2] Intuitively, extreme fluctuations in RN beliefs imply physical belief movements that must be rare regardless of the exact degree of risk aversion.

Our bounds for risk-neutral excess movement $\mathbb{E}[X^*]$ formalize this idea. The first bound we derive is highly conservative: it holds under a particular "worst case" DGP that produces the highest possible movement for a given ϕ. This bound therefore requires no knowledge of the true statistical DGP for option prices, as it applies uniformly over the space of all possible DGPs. We then provide a set of tighter bounds that hold given additional information about the DGP. In both cases, the bounds are simple formulas that depend on the observable RN prior and are increasing in ϕ. Consequently, any observed average excess movement in the data is informative as to the

[2]Note that if ϕ can move arbitrarily, it becomes difficult to make statements about rational price variation. But as will be discussed further, arbitrary variation in ϕ is also difficult to make sense of under RE.

minimal risk-aversion value ϕ required in order for the bound to be satisfied.

How does this abstract setting apply to financial markets? In our empirical setting, the state θ depends on the value of the market index as of option expiration date T, or equivalently the return on the market over the life of the option. For example, suppose that the index value on date 0 is $S_0 = 100$. Define state $\theta = 1$ as occurring if S_T is equal to (or in a neighborhood close to) 100, or equivalently if the index gain R_T is close to 0%. Similarly, say $\theta = 0$ occurs if $S_T \simeq 105$. The RN belief π_t^* is then defined as the option-implied probability that $\theta = 1$ (i.e., $S_T \simeq 100$) conditional on $\theta \in \{0, 1\}$, which can be measured from the cross-section of options with different strike prices and expiration date T. This π_t^* plays the same role as in the abstract setting above, and its excess movement must satisfy our bounds. Importantly, the defined outcome θ stays fixed over the life of the option. This *fixed-event* setting is key for our results. It implies that the underlying subjective belief π_t must be a martingale under RE, and that learning must converge at date T: the index will either be close to 100 or 105 (i.e., $R_T = 0\%$ or 5%), or neither state will be realized.[3]

In this asset-pricing context, the object that maps physical to RN beliefs is now $\phi = \frac{\mathbb{E}[M_T \mid R_T = 0\%]}{\mathbb{E}[M_T \mid R_T = 5\%]}$, where M_T is the stochastic discount factor (SDF). This ratio of expected SDF realizations can be thought of as the market's local relative risk aversion, or price of risk, in a neighborhood of a 0% return realized from 0 to T. The assumption that ϕ is stable within a belief stream — which we maintain in deriving our main bounds — is therefore equivalent to assuming that the local price of risk in this neighborhood is constant. This is significantly weaker than the assumption of constant discount rates used, for example, by Shiller (1981): it allows for time-varying discount rates arising from multiple sources, including general changes in interest rates or the quantity of risk.[4] It is also met either exactly or approximately under many standard modeling frameworks.

That said, the assumption that the local price of risk ϕ is constant within a belief stream is a knife-edge restriction that is unlikely to hold exactly, and it is also violated in some leading option pricing models with priced stochastic volatility. We therefore also consider how our bounds change given variation in ϕ. We find, both theoretically and in a set of simulations, that rational variation in ϕ has a very limited impact on RN belief movement. This may seem counterintuitive: one might imagine that if ϕ oscillates between 3.1 and 2.9, for example, then RN belief movement can be unbounded even with no movement in physical beliefs. While this is true, it overlooks that these oscillations in ϕ are *also* inconsistent with rational updating: they imply predictable mean reversion in the expected marginal value of \$1 in a fixed terminal state. A bound violation thus suggests that either physical beliefs or risk prices are excessively volatile. We return to the question of their relative importance in our empirical analysis.

After providing and discussing our theoretical results, we then take our bounds to the data. We

[3] By contrast, in many common settings, the fundamental value is a moving target. For example, the time-t subjective distribution over the *fixed-horizon* outcome $S_{t+\tau}$ may have lower variance than the $t+1$ distribution over $S_{t+1+\tau}$. See Nordhaus (1987) for discussion of fixed-event vs. fixed-horizon tests, and see Kelly et al. (2024) for a recent example.

[4] For example, suppose that the second-period option prices in the example above were \$0.05 and \$0.45 (for states 1 and 0) rather than \$0.10 and \$0.90, due to a shift in the time-t value of money. This shift induces no change in risk-neutral beliefs relative to the example, because the relative prices are unchanged. Meanwhile, such a change would be problematic for any single-asset volatility test assuming constant discount rates.

obtain S&P 500 index option prices from OptionMetrics, and we use standard methods to infer the risk-neutral distribution over index returns for each option expiration date in the sample. 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 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 ranges for the S&P return, matching the example above.) We then implement our theoretical bounds, which allow us to infer the minimal local risk-aversion value ϕ (at each point in the return distribution) needed to rationalize the observed variation in RN beliefs over the index return.[5]

Our main empirical finding is that there is so much excess movement in RN beliefs that extremely high risk aversion is needed in order to rationalize the data under our maintained assumptions. 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. This suggests that many leading frameworks capable of explaining variation in underlying asset prices have difficulty rationalizing the large degree of observed variation in RN beliefs.

To better understand the implications of our results, we conduct three further analyses. First, we simulate two benchmark option pricing models to see whether they generate RN belief movement close to our empirical estimates. While they produce a moderate degree of excess movement, it is far below what we observe in the data. So despite the models' ability to explain many empirical regularities in option markets, they have difficulty generating the strong and predictable mean-reversion in relative option prices that we find. We are of course not the first to provide evidence that well-known models are inconsistent with various moments, but our results offer progress in understanding what specifically can be deduced from the observed volatility in asset prices.

Second, to begin to disentangle possible explanations for our results — predictable mean reversion in either (i) physical beliefs or (ii) the price of risk, or alternatively (iii) option-market frictions — we conduct a set of reduced-form regressions for observed excess movement. We find a strong relationship between excess movement and proxies for overreaction in expectations, and little explanatory power from proxies for risk-price variability or market frictions. This suggests that overreaction by the marginal investor is likely a necessary feature in explaining our results.

Finally, to conclude our analysis, we conduct a rough quantification of how our results on RN beliefs translate to excess volatility in the underlying index. We do so by constructing an alternative index price from the RN distribution in which we exclude tail return states, and we show how excess movement in this index helps discipline excess volatility at the index level. We estimate that a sizable share (roughly on the order of 1/3) of the observed volatility in the S&P 500 is excessive relative to a martingale-belief benchmark. Our results on RN beliefs therefore imply meaningful excess volatility in the index itself.

[5]Given that we conduct our estimation using variation in index option prices, we must also account for the effect of non-fundamental or microstructure noise. To do so, we estimate the variance of the noise component of observed RN beliefs using a set of intraday option prices and the microstructure noise variance estimator proposed by Li and Linton (2022). We then construct an empirical noise correction, removing the effect of noise from X^* before we conduct our estimation. Our tests are therefore constructed to be robust to idiosyncrasies specific to the option market.

Relation to previous literature. In addition to Shiller (1981), we follow, among others, LeRoy and Porter (1981), West (1988), and Stein (1989) in testing for excess volatility in asset prices relative to measures of fundamental value. 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). By contrast, our test allows for variation in discount rates and does not require any measure of fundamental value.

While our framework requires significantly weaker theoretical assumptions than classic volatility tests, one cost of this generality is that we rely on option prices rather than measuring the behavior of the underlying index directly. In doing so, we follow a long line of work using options for information about more general expectations or quantities of interest.[6] Our work particularly complements Giglio and Kelly (2018), who document excess volatility in long-maturity claims on volatility, inflation, commodities, and interest rates. They achieve identification by parameterizing the DGP for cash flows on the underlying, whereas we restrict the SDF. Their parameterization — an 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.[7]

Our results also complement evidence obtained from survey data, as, for example, in Greenwood and Shleifer (2014) and De la O and Myers (2021). Another set of related literature endeavors to measure physical probabilities indirectly from asset prices. For example, Aït-Sahalia, Wang, and Yared (2001) test whether option-implied return distributions are well-calibrated, and Polkovnichenko and Zhao (2013) consider the probability weighting functions consistent with option-price data given an assumption on the form of the weighting function. Our approach differs from this and related work in that we need not measure physical beliefs at all or know the true DGP for returns to conduct our tests. This semi-parametric approach loosely ties our theoretical contribution to a line of work including, among many others, Hansen and Jagannathan (1991) and Alvarez and Jermann (2005), which consider the identification of structural parameters from different moments of the observable data than the ones we consider here.

Organization. Section 2 introduces our theory in a two-state setting, which allows for clear derivations and intuition for our main results. Section 3 extends these results to a general asset-pricing setting. Section 4 discusses our identifying assumptions and provides robustness results. We implement our bounds empirically in Section 5, and in Section 6 we discuss the interpretation of our results. Section 7 concludes. Proofs of our main results are provided in Appendix A, and an Internet Appendix contains the remaining proofs and additional technical material.

[6] A recent sample includes Backus, Chernov, and Martin (2011) regarding disaster risk, Martin (2017) and Beason and Schreindorfer (2022) on the equity premium, and Haddad, Moreira, and Muir (2024) on the impact of policy promises.

[7] See also Gandhi, Gormsen, and Lazarus (2025) for recent evidence on the term structure of the equity premium.

2. Theoretical Framework: Introduction in a Simple Setting

To introduce our framework, we first examine risk-neutral (RN) belief movement in a simple setting with a single person and two terminal states. This setting helps clarify (i) the economics underlying restrictions on belief movement under RE when the individual is risk-neutral; (ii) how risk aversion complicates this analysis; and (iii) how we can nonetheless bound RN belief movement. This step-by-step discussion sets the stage for our generalized framework in Section 3.

2.1 Setup and Initial Results

Time is discrete and indexed by $t = 0, 1, 2, \ldots, T$. At the beginning of each period, a person observes a signal $s_t \in S$ regarding two mutually exclusive and exhaustive 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 is 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 H_t. We often suppress the dependence of these objects on H_t to simplify notation. Given our empirical setting, we focus 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.

Throughout the paper, we maintain the assumption that the person's beliefs over the terminal state satisfy rational expectations (RE).

ASSUMPTION 1 (*RE*)**.** Beliefs satisfy $\pi_t(H_t) = \mathbb{E}[\theta \mid H_t]$ for any H_t.

This states that the agent's beliefs coincide period by period with the true conditional probability of realizing state $\theta = 1$, and the assumption will be satisfied by a person with a correct prior who updates using Bayes' rule according to the true DGP. The assumption is stronger than necessary for our main results: we could instead assume the weaker martingale restriction that $\pi_t = \mathbb{E}[\pi_{t+1} \mid \pi_t]$, which is implied by Assumption 1. Assuming RE directly, though, helps streamline exposition.

Given our asset-pricing setting, we assume that the DGP and the person's physical beliefs cannot be observed directly. Instead, we assume that the econometrician can observe, period by period, the person's willingness to pay for an Arrow-Debreu security that pays \$1 (one unit of the numeraire consumption good) in period T if state θ is realized. 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 empirically observable using options data for suitably defined states, but we postpone this additional formalism to Section 3.

We start by considering the simple case in which the person values consumption in all periods and states equally (i.e., she is risk-neutral and does not discount future consumption). In this case, beliefs are in fact directly inferable from asset values: $q_t(1|H_t) = \pi(H_t)$ and $q_t(0|H_t) = 1 - \pi(H_t)$. Testable restrictions on prices are thus equivalent in this case to restrictions on physical beliefs.

We keep track of the following objects related to the physical belief stream, and we discuss

shortly how these objects are restricted under RE. First, total *belief movement* of $\boldsymbol{\pi}$ is defined as the sum of squared changes (or quadratic variation) in beliefs across all periods:

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

Second, *initial uncertainty* of $\boldsymbol{\pi}$ is defined as the variance of the Bernoulli random variable $\mathbb{1}\{\theta = 1\}$ as of $t = 0$:

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

Given that we focus on resolving belief streams for which $\pi_T \in \{0, 1\}$, final uncertainty is always zero ($\mathsf{u}_T = (1 - \pi_T)\pi_T = 0$). 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 will be the difference between movement and initial uncertainty, which — for reasons that will become clear — we call *excess movement*:

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

Belief movement and initial uncertainty are related under RE according to the following lemma. This lemma restates a main result used by Augenblick and Rabin (2021), which has appeared in other forms in past literature (e.g., Kyle, 1985, p. 1329; Barndorff-Nielsen and Shephard, 2001, p. 180; Collin-Dufresne and Fos, 2016, footnote 5).

Lemma 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:*

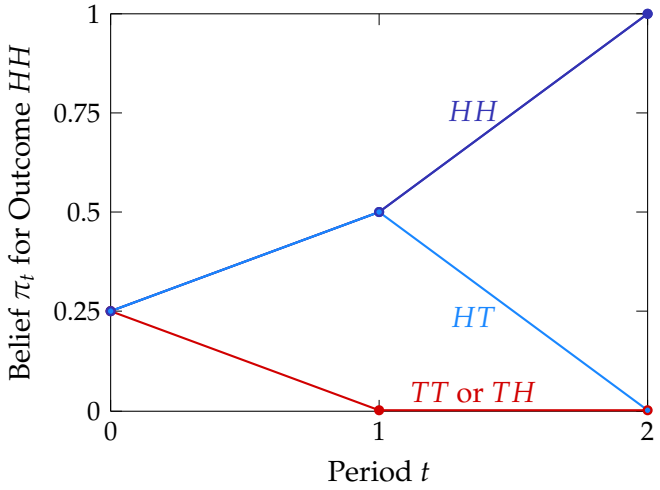
$$\mathbb{E}[X] = 0.$$

Lemma 1 is a straightforward implication of the assumption of martingale beliefs.[8] The result formalizes a notion of the "correct" amount of belief volatility under RE, and it motivates referring to X as *excess movement*. 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 the lemma would still reject the null of RE if, for instance, low values of π_t ($\pi_t < 0.5$) tend to be revised upward ($\pi_{t+1} - \pi_t > 0$), and high values tend to be revised downward.

To clarify our setting and the above lemma, Figure 1 plots belief streams under a two-period DGP. Two fair coins are flipped sequentially at $t = 1$ and 2, generating two possible signals (H or T)

[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.

Figure 1: Physical Beliefs and Excess Movement: Two-Period Example



STATISTICS

Signals (s_1, s_2)	Frequency $\mathbb{P}(s_1, s_2)$	Movement m
HH	0.25	0.3125
HT	0.25	0.3125
TT or TH	0.5	0.0625

$$\mathbb{E}[\mathsf{m}] = 0.1875 = \mathsf{u}_0$$
$$\Longrightarrow \mathbb{E}[X] = 0$$

Notes: Signals (s_1, s_2) are generated by sequential fair coin flips. State $\theta = 1$ is realized if two heads occur: $(s_1, s_2) = (H, H)$, labeled HH; otherwise $\theta = 0$. Each path in the figure corresponds to a possible belief stream over $\theta = 1$ for a person with RE. In the table, m, u_0, and X are as defined in (1)–(3), and $\mathbb{E}[\mathsf{m}] = 0.25 \times 0.3125 \times 2 + 0.5 \times 0.0625$.

at each of these dates. If two heads occur (HH), then state $\theta = 1$ is realized; otherwise, $\theta = 0$ is realized. As the figure shows, the person's prior π_0 is equal to 0.25 under RE (equal to the probability of two heads). If the first coin flip is tails (shown in the red path), then $\pi_1 = \pi_2 = 0$; if heads then tails are flipped, then $\pi_1 = 0.5$, $\pi_2 = 0$ (light blue); if two heads are flipped, then $\pi_1 = 0.5$, $\pi_2 = 1$ (dark blue). The accompanying table shows belief movement $\mathsf{m} = (\pi_1 - \pi_0)^2 + (\pi_2 - \pi_1)^2$ for each possible stream. Weighting the paths by their relative frequencies, expected belief movement is $\mathbb{E}[\mathsf{m}] = 0.1875$. This is exactly equal to initial uncertainty $\mathsf{u}_0 = (1 - 0.25) \times 0.25 = 0.1875$, illustrating the restriction that $\mathbb{E}[X] = 0$. This applies beyond this two-period example: the lemma implies that *any* DGP for which $\pi_0 = 0.25$ would generate expected movement of 0.1875 under RE.

2.2 Risk-Neutral Beliefs: Setup and Identification Challenge

Lemma 1 shows that one can make statements about excess belief movement under RE when beliefs are directly observable or inferable from asset prices. But identifying excess movement in asset valuations becomes significantly more complicated when the person is risk averse, as valuations no longer correspond directly to beliefs in this more general case. To see this, assume now that the agent has time-separable utility, with concave period utility function $U(C_t)$, and that she exponentially discounts future consumption with discount factor β. Assume that the state θ determines period-T consumption $C_{T,\theta}$. Valuations for the two Arrow-Debreu securities at time t are now

$$q_t(1|H_t) = \frac{\beta^{T-t} U'(C_{T,1})}{U'(C_t)} \pi_t, \qquad q_t(0|H_t) = \frac{\beta^{T-t} U'(C_{T,0})}{U'(C_t)} (1 - \pi_t). \tag{4}$$

Valuations thus no longer directly reveal beliefs. Instead, they are distorted by the relative value of a marginal dollar in state θ in period T versus one in period t. As $q_t(1|H_t)$ and $q_t(0|H_t)$ are similarly distorted by β^{T-t} and $U'(C_t)$, it is useful to focus on their relative valuations. This logic

leads to the consideration of *risk-neutral (RN)* beliefs,

$$\pi_t^*(H_t) \equiv \frac{q_t(1|H_t)}{q_t(0|H_t)+q_t(1|H_t)} = \frac{U'(C_{T,1})}{\mathbb{E}_t[U'(C_T)]}\pi_t(H_t) = \frac{\phi\pi_t(H_t)}{1+(\phi-1)\pi_t(H_t)}, \tag{5}$$

$$\text{where } \phi \equiv \frac{U'(C_{T,1})}{U'(C_{T,0})}. \tag{6}$$

The definition in (5) follows the usual convention for RN beliefs, and the remaining expressions follow from (4). Like π_t, the RN belief π_t^* corresponds to state $\theta = 1$. The RN belief 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. Define the RN belief stream $\boldsymbol{\pi}^*(H_t)$, RN belief movement $\mathsf{m}^*(\boldsymbol{\pi}^*)$, RN initial uncertainty $\mathsf{u}_0^*(\boldsymbol{\pi}^*)$, and RN excess movement $X^*(\boldsymbol{\pi}^*)$ as in Section 2.1, but with RN beliefs π_t^* in place of π_t.

Risk-neutral beliefs are so named because they can be interpreted as the subjective beliefs for a fictitious risk-neutral agent. In general, they represent a pseudo-belief distribution, reflecting a combination of the person's physical beliefs π_t and risk preferences as indexed by ϕ. This object ϕ will be particularly important in our environment. It represents the (constant) marginal rate of substitution across primitive states, as can be seen in (6). Up to a scaling constant, it also serves as an index of relative risk aversion: conducting a Taylor expansion of $U'(C_{T,0})$ around $C_{T,1}$ gives $U'(C_{T,0}) = U'(C_{T,1}) + U''(C_{T,1})(C_{T,0} - C_{T,1}) + \mathcal{O}\big((C_{T,0} - C_{T,1})^2\big)$, and thus, to first order,

$$\gamma(C_{T,1}) \equiv -\frac{C_{T,1}U''(C_{T,1})}{U'(C_{T,1})} = \frac{\phi - 1}{(C_{T,0} - C_{T,1})/C_{T,1}}. \tag{7}$$

Relative risk aversion γ thus depends on the ratio of marginal utilities across states ϕ relative to the percent consumption gap across states. Finally, in asset-pricing terms, ϕ is equivalent to the ratio of *stochastic discount factor (SDF)* realizations $M_{t,T}(\theta)$ across the two states:

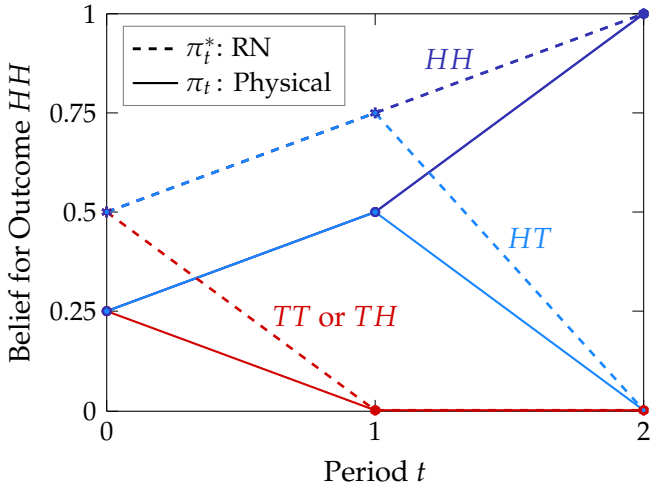
$$\phi = \frac{M_{t,T}(1)}{M_{t,T}(0)}, \text{ where } M_{t,T}(1) \equiv \frac{q_t(1|H_t)}{\pi_t(H_t)}, \; M_{t,T}(0) \equiv \frac{q_t(0|H_t)}{1-\pi_t(H_t)}.$$

This SDF-based representation of ϕ will be discussed in detail in the general setting in Section 3.

To complete the setup for the current consumption-based setting, we assume that consumption in state 1 is weakly less than in state 0, $C_{T,1} \leqslant C_{T,0}$. This is without loss of generality for now, as the states can be relabeled arbitrarily. With concave utility, this labeling implies that $U'(C_{T,1}) \geqslant U'(C_{T,0})$ and thus $\phi \geqslant 1$. Given $\phi \geqslant 1$, the RN belief π_t^* in general exceeds the subjective belief π_t: the person is willing to pay relatively more for a bad-state consumption claim given her high marginal utility in that state, upwardly biasing the bad-state RN belief relative to π_t.

We wish to make statements similar to Lemma 1, but for observable RN beliefs rather than physical beliefs. Under a risk-neutral expectation $\mathbb{E}^*[\cdot]$ defined such that $\pi_t^*(H_t) = \mathbb{E}^*[\pi_{t+1}(H_{t+1}) \,|\, H_t]$, one could in fact apply Lemma 1 directly, as $\mathbb{E}^*[X^*] = 0$. But the frequency of observed RN belief streams is determined by the physical measure rather than the RN measure; that is, we can only estimate $\mathbb{E}[X^*]$ rather than $\mathbb{E}^*[X^*]$. And even under the maintained assumption that physical

Figure 2: RN Beliefs and Excess Movement: Two-Period Example



STATISTICS

Signals (s_1, s_2)	Frequency $\mathbb{P}(s_1, s_2)$	RN Movement m^*
HH	0.25	0.125
HT	0.25	0.625
TT or TH	0.5	0.25

$$\mathbb{E}[\mathsf{m}^*] = 0.3125 > 0.25 = \mathsf{u}_0^*$$
$$\implies \mathbb{E}[X^*] = 0.0625 > 0$$

Notes: See Figure 1 for description of signals and physical beliefs. Observed RN beliefs are calculated using (5), with the assumption that $\phi = 3$. In the table, m^*, u_0^*, and X^* are calculated as in (1)–(3), with RN beliefs π_t^* in place of π_t.

beliefs π_t satisfy RE, the distortion in π_t^* relative to π_t can cause RN movement to differ from RN initial uncertainty on average, so $\mathbb{E}[X^*] \neq 0$. This is the fundamental identification challenge.

For an illustration of this issue, Figure 2 returns to the coin-flip example from Figure 1. In addition to physical beliefs, this figure now shows RN belief streams with $\phi = 3$.[9] From (5), the physical prior $\pi_0 = 0.25$ corresponds to an RN prior of $\pi_0^* = 0.5$. Intuitively, the person perceives state $\theta = 0$ as three times as likely as $\theta = 1$ (HH), but values a marginal dollar in $\theta = 1$ three times more than in $\theta = 0$, so $q_0(0) = q_0(1)$ and $\pi_0^* = 0.5$. The same calculations are then applied to obtain π_1^* and π_2^* for each stream shown in the figure. The accompanying table shows RN movement m^* for each stream. Comparing this to the table in Figure 1, it is apparent that $\mathsf{m}^* > \mathsf{m}$ for streams in which the RN belief has large downward revisions (HT, TT, and TH): π_t^* is biased upward relative to π_t, so these large downward revisions generate more RN movement than physical movement. These streams cause RN excess movement to be positive on average: $\mathbb{E}[X^*] = 0.0625 > 0$.

As this example illustrates, even with rational physical beliefs, one can observe what appears to be excess movement in RN beliefs implied by valuations. So if we naively test for RE using Lemma 1 on observed RN beliefs, we may spuriously conclude that beliefs are excessively volatile.

2.3 Risk-Neutral Beliefs: Results

The fact that there can be excess movement in RN beliefs even under RE would seem to pose an intractable challenge. But this turns out not to be the end of the story. RN beliefs are not arbitrarily distorted relative to physical beliefs: as in (5), they are linked through the single unobserved parameter ϕ. And regardless of the value of ϕ, RN beliefs must lie between 0 and 1 by definition. These insights will allow us to bound $\mathbb{E}[X^*]$ over all possible DGPs for a given value of ϕ, and over all possible values of ϕ. Because our main results are most straightforwardly stated in the current

[9] Under risk neutrality, $\phi = 1$. This was implicitly assumed to be the case for Figure 1.

section's two-state setting, we state them here before discussing how they generalize in Section 3.

Main Results

We first define two useful additional objects. First, we invert (5) to solve for π_t as a function of π_t^* and ϕ, the solution to which we denote by $\pi_t(\pi_t^*, \phi)$:

$$\pi_t(\pi_t^*, \phi) = \frac{\pi_t^*}{\phi + (1-\phi)\pi_t^*}. \tag{8}$$

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

$$\triangle \equiv \mathbb{E}[X^*|\theta = 0] - \mathbb{E}[X^*|\theta = 1]. \tag{9}$$

Given these definitions, we can now provide a number of expressions and bounds for $\mathbb{E}[X^*]$. We assume throughout that Assumption 1 holds.

PROPOSITION 1. *For any DGP,*

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

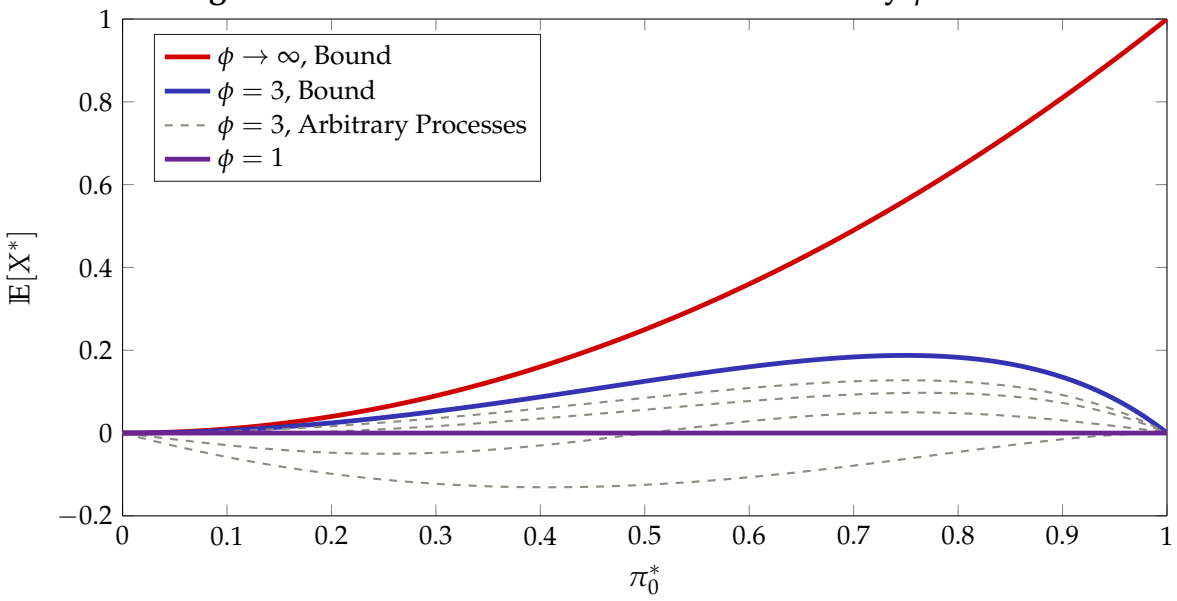
PROPOSITION 2. *For any DGP and any value for* $\triangle$*,*

$$\begin{aligned}\mathbb{E}[X^*] &\leqslant (\pi_0^* - \pi_0)\pi_0^* \\ &\leqslant \left(1 - \frac{1}{\phi + (1-\phi)\pi_0^*}\right)\pi_0^{*2}.\end{aligned} \tag{10}$$

Proofs for this section's main results are provided in Appendix A. Proposition 1 starts from the fact that $\mathbb{E}^*[X^*] = 0$. We then connect $\mathbb{E}^*[X^*]$ to $\mathbb{E}[X^*]$. The key step is to show that *conditional* expectations of X^* under both measures are equal, $\mathbb{E}^*[X^*|\theta] = \mathbb{E}[X^*|\theta]$ for $\theta = 0, 1$, which leads to the stated result. Note that if $\phi = 1$, then $\pi_0^* = \pi_0$ and therefore $\mathbb{E}[X^*] = 0$, as in Lemma 1. As ϕ rises, π_0 drops further below π_0^*, and $\mathbb{E}[X^*]$ departs from 0: the greater is risk aversion, the more one can observe excess RN movement differ from zero on average under RE. The sign and magnitude of this deviation depend on $\triangle$, as explored in more detail later in this section.

Across all DGPs, $\triangle$ is bounded above by π_0^*, from which Proposition 2 follows. The version of the bound in (10) is one of our main results. It gives a bound for $\mathbb{E}[X^*]$ as a function of π_0^* and ϕ, regardless of $\triangle$. Under risk neutrality ($\phi = 1$), this upper bound again becomes zero. But the bound is otherwise positive, and the *admissible excess movement* in RN beliefs given by the right side of the inequality increases monotonically in ϕ. 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.

Figure 3: RN Excess Belief Movement vs. Prior by ϕ Under RE



Note: Theoretical bounds are obtained from the formulas in Proposition 2 and Corollary 1.

Intuitively, movement in RN beliefs must still correspond on average to the agent learning something about the true terminal state, but the bias in RN relative to subjective beliefs induced by risk aversion allows for positive excess movement in those observed beliefs under RE. This is reflected in the first term in the bound, $(\pi_0^* - \pi_0)$, as this difference increases in ϕ. For the second term (π_0^*), higher RN priors yield more "room" for downward RN belief movement. This increases admissible excess movement along the lines of the example considered in Figure 2. The bound in Proposition 2 is conservative, as it holds under a "worst-case" DGP with an extreme value for $\triangle$.

Because of the monotonicity of the bound in ϕ, the inequality can also be read as providing a *lower* bound for the unobserved structural parameter ϕ as a function of observables. The observed excess movement in RN beliefs thus provides information on the minimal value of risk aversion necessary for the data to be consistent with RE. But more still can be said: taking $\phi \to \infty$ in (10) generates the following maximally conservative bound for $\mathbb{E}[X^*]$, which applies for *any* ϕ.

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

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

Corollary 1 exploits that RN beliefs are bounded between 0 and 1 by construction: there is only so far that RN beliefs can be distorted relative to subjective beliefs, so the bound is well-defined even for arbitrarily high risk aversion.[10] It states that price movements for which $\mathbb{E}[X^*] > \pi_0^{*2}$ simply cannot be rationalized under RE given constant ϕ. Despite its maximal conservatism, this bound still imposes a meaningful limit on admissible RN excess movement, especially for low π_0^*. For example, with $\pi_0^* = 0.2$, RN excess movement is at most 0.04, regardless of ϕ or the DGP.

[10] In the limit as $\phi \to \infty$, $\pi_0 \to 0$ for any π_0^*, so $\pi_0^* - \pi_0 \to \pi_0^*$.

Together, Proposition 2 and Corollary 1 characterize the maximal admissible excess movement in RN beliefs as a function of ϕ for any RN prior. Figure 3 graphically illustrates these bounds. Starting from the bottom of the chart, the thick purple line shows the bound for $\phi = 1$: in this case, $\mathbb{E}[X^*] = \mathbb{E}[X] = 0$ regardless of the prior or DGP, from Lemma 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 2, which is shown in the thick blue line. It is asymmetric around 0.5 as well as non-monotonic in π_0^*.[11] 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.

Additional Results

The Internet Appendix contains a number of additional theoretical results (Appendix C.1) and simulations (C.2), which we summarize briefly here. First, we characterize how the DGP determines the sign and magnitude of RN excess movement $\mathbb{E}[X^*]$. Given that we observe positive excess movement in the data, most of our focus in the main text is on the positive upper bounds for $\mathbb{E}[X^*]$. But as shown in Propositions C.1–C.2, asset-pricing settings with risk aversion do not necessarily entail positive excess movement in observed beliefs. Further, given the arbitrary labeling of the two states, there is no a priori reason to expect under RE that $\mathbb{E}[X^*]$ should take a particular sign.

Second, we conduct numerical simulations to further explore how RN excess movement depends on the DGP. When the DGP for beliefs is symmetric (with equal movements up or down, as is likely to roughly apply for options), we find that $\mathbb{E}[X^*]$ takes on the same sign as $\pi_0^* - .5$. This suggests, for example, that for priors close to .5, excess movement should be small unless the DGP is asymmetric. Indeed, we also find that extreme values of $\mathbb{E}[X^*]$ occur only in highly asymmetric DGPs where movements in one direction are large and movements in the other direction are tiny.

Finally, in Proposition C.3, we show that the bound in Proposition 2 is approximately tight: there exists a DGP for which $\mathbb{E}[X^*]$ is close to the bound for large T. The proof is constructive, as we characterize the form of the maximizing DGP. In line with the simulations, this DGP takes a somewhat perverse form, indicating further that the bound is quite conservative.

3. Generalized Theoretical Results for Equilibrium Asset Prices

While the binary-state setting considered in the previous section is useful for exposition, 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 empirically relevant case.

[11] The second term in the bound (π_0^*) of course increases monotonically, generating asymmetry in the bound around $\pi_0^* = 0.5$ for $\phi > 1$. But the first term in the bound $(\pi_0^* - \pi_0)$ does not increase monotonically for $1 < \phi < \infty$, exerting a countervailing force that generates the observed non-monotonicity.

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 $\{H_t\}$.[12] A realization of the elementary state is denoted by $\omega \in \Omega$. To make our results empirically implementable, we will be concerned with the spot price of the market index, $S_t \colon \Omega \to \mathbb{R}_+$, 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 $(S_T - K)^+ = \max(S_T - K, 0)$, and its time-$t$ price is $q_{t,K}$. 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 $q_{t,K} = \mathbb{E}_t[M_{t,T}(S_T - K)^+]$, where $\mathbb{E}_t[\cdot] \equiv \mathbb{E}^{\mathbb{P}}[\cdot \mid H_t]$. The SDF can equivalently be written as $M_{t,T} = M_T / M_t$ for strictly positive $\{M_t\}$, with $M_0 = 1$.

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 $R_T \equiv S_T / S_0 = \theta$. The measure $\mathbb{P} \colon \mathcal{F} \to [0, 1]$ governs the *objective* or *physical* probabilities of these return states.[13] In this general case, the *risk-neutral (RN) measure* is defined by the change of measure

$$\left.\frac{d\mathbb{P}^*}{d\mathbb{P}}\right|_{H_t} = \frac{M_{t,T}}{\mathbb{E}_t[M_{t,T}]} = \frac{M_T}{\mathbb{E}_t[M_T]}, \tag{11}$$

and expectations under $\mathbb{P}^*$ are denoted by $\mathbb{E}^*[\cdot]$.[14] Using this definition of $\mathbb{P}^*$, the RN probability of return state θ is

$$\mathbb{P}_t^*(R_T = \theta) = \frac{\mathbb{E}_t[M_T \mid R_T = \theta]}{\mathbb{E}_t[M_T]} \mathbb{P}_t(R_T = \theta). \tag{12}$$

The RN pricing equation $q_{t,K} = \mathbb{E}_t^*[(S_T - K)^+] / R_{t,T}^f$ can be used to show that the date-t option prices $\{q_{t,K}\}_{K \in \mathcal{K}}$ reveal the set of RN probabilities $\{\mathbb{P}_t^*(R_T = \theta)\}_{\theta \in \Theta}$. 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., $\mathcal{K} = \{K_j\}_{j=1}^J$, with $K_j = S_0 \theta_j$). We can then back out RN probabilities from option prices as follows:

$$\mathbb{P}_t^*(R_T = \theta_j) = R_{t,T}^f \left[\frac{q_{t,K_{j+1}} - q_{t,K_j}}{K_{j+1} - K_j} - \frac{q_{t,K_j} - q_{t,K_{j-1}}}{K_j - K_{j-1}} \right]. \tag{13}$$

See Internet Appendix B.1 for a derivation of this result, which follows from a discrete-state

[12] We could consider continuous states with additional technicalities.

[13] It is without loss to define this measure over arbitrary fixed return states: we are simply discretizing the set of possible index values as of expiration date T and keeping track of the distribution over these fixed outcomes.

[14] As $\text{Price}_t = \mathbb{E}_t[M_{t,T}\text{Payoff}_T]$ for any asset, we have $R_{t,T}^f = \mathbb{E}_t[M_{t,T}]^{-1}$, where $R_{t,T}^f$ is the gross risk-free rate from t to T, and $\text{Price}_t = \mathbb{E}_t^*[\text{Payoff}_T] / R_{t,T}^f$. This $\mathbb{P}^*$ is sometimes referred to as the T-forward measure.

application of the classic result of Breeden and Litzenberger (1978). As in Section 2.2, the above RN probabilities are a more convenient object of analysis than raw option prices.

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 have common beliefs satisfying RE. But our framework accommodates less strict requirements as well. As long as there is an unconstrained agent with rational beliefs satisfying our three main assumptions below, then the tests we derive should not be rejected in the data. One can therefore think of our framework as testing for the possible existence of such an agent.[15] One might also, more loosely, prefer an interpretation in which the "agent" in question is the market as a whole.

Regardless of the interpretation of the agent in question, assume prices correspond to valuations for an agent who, at the beginning of each period, observes a signal vector $s_t \in \mathcal{S}$ drawn from the distribution $DGP(s_t \,|\, \theta, H_{t-1}) = \mathbb{P}_{t-1}(s_t|\theta)$, 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 $\Pi_{t,T} = \{\pi_t(R_T = \theta)\}_{\theta \in \Theta}$, where $\pi_t(R_T = \theta) \geqslant 0 \ \forall \theta \in \Theta$ and $\sum_{\theta \in \Theta} \pi_t(R_T = \theta) = 1$. 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 follows.

Assumption 2 (*RE*)**.** For any random variable Y, beliefs satisfy $\pi_t(Y = y) = \mathbb{P}_t(Y = y)$ with probability 1 for all t.

This assumption again implies that beliefs satisfy $\pi_t(R_T = \theta) = \mathbb{E}[\pi_{t+1}(R_T = \theta) \,|\, \pi_t(R_T = \theta)]$ 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 2 is useful for streamlining some of the remaining discussion, as it further implies that *all* conditional expectations — including over the SDF — are martingales with respect to H_t.

Given Assumption 2, 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 (11). This yields the RN belief-distribution $\Pi^*_{t,T} = \{\pi^*_t(R_T = \theta)\}_{\theta \in \Theta}$ such that $\pi^*_t(R_T = \theta) = \frac{\mathbb{E}_t[M_T \,|\, R_T = \theta]}{\mathbb{E}_t[M_T]} \pi_t(R_T = \theta)$ as in (12), and thus (13) 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 $\Pi^*_{t,T}$,

[15] We thank an anonymous referee for suggesting this point.

we instead consider restrictions on the behavior of the binarized beliefs $\{\widetilde{\pi}^*_{t,j}\}_{j=1}^{J-1}$, defined by

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

for $\pi^*_t(R_T = \theta_j) + \pi^*_t(R_T = \theta_{j+1}) > 0$. 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 4.1 below). Conditional physical beliefs $\widetilde{\pi}_{t,j}$ are defined analogously, and the expectation under the conditional physical measure $\widetilde{\mathbb{P}}_t$ is $\widetilde{\mathbb{E}}_t[\,\cdot\,] \equiv \mathbb{E}_t[\,\cdot \mid R_T \in \{\theta_j, \theta_{j+1}\}]$.

In this context, the analogue to ϕ as defined in Section 2 is

$$\phi_{t,j} \equiv \frac{\mathbb{E}_t[M_T \mid R_T = \theta_j]}{\mathbb{E}_t[M_T \mid R_T = \theta_{j+1}]}, \tag{14}$$

which encodes the slope of the SDF across the adjacent return states θ_j, θ_{j+1}. In a representative-agent economy with SDF $M_{t,T} = \beta^{T-t}\frac{U'(C_T)}{U'(C_t)}$, the above definition implies $\phi_{t,j} = \frac{\mathbb{E}_t[U'(C_T) \mid R_T = \theta_j]}{\mathbb{E}_t[U'(C_T) \mid R_T = \theta_{j+1}]}$, akin to (6). But (14) is general and does not require a representative-agent structure (though we make periodic reference to such an economy for interpretation). This SDF slope can also be thought of loosely as the local price of risk around index return state θ_j.

Using this definition of $\phi_{t,j}$, RN beliefs satisfy $\frac{\widetilde{\pi}^*_{t,j}}{1-\widetilde{\pi}^*_{t,j}} = \phi_{t,j}\frac{\widetilde{\pi}_{t,j}}{1-\widetilde{\pi}_{t,j}}$ and thus $\widetilde{\pi}^*_{t,j} = \frac{\phi_{t,j}\widetilde{\pi}_{t,j}}{1+(\phi_{t,j}-1)\widetilde{\pi}_{t,j}}$, as in (5). Given a 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 often suppress the dependence on j (writing, e.g., X^*) when considering an arbitrary state pair.

3.2 Identifying Assumptions on ϕ

The analysis in Section 2 considered Arrow-Debreu claims on primitive states (in that case, consumption states). This section's analysis instead considers claims on index-return states, with an eye toward empirical implementation. We must thus confront the joint hypothesis problem. The additional identifying assumptions to be tested jointly with Assumption 2 take the form of two restrictions on $\phi_{t,j}$. We introduce these assumptions here, and we discuss them in much greater detail in the following sections after presenting our main results below.

We first impose an ordering assumption on the states. Return state θ_j here corresponds to state 1 in Section 2 (vs. state 0 for θ_{j+1}). We thus maintain the convention of labeling θ_j as the "bad" state, so that $\phi_{t,j} \geqslant 1$. While this is an innocuous labeling convention in theory, empirical implementation requires taking a stand on how to distinguish θ_j from θ_{j+1}. We make the intuitive assumption that the bad state θ_j corresponds to the lower return.

Assumption 3 (*Positive Risk Aversion in Index Return*). $\phi_{t,j} \geqslant 1$ with probability 1 for all t, j, where the set of return states Θ is ordered such that $\theta_1 < \theta_2 < \cdots < \theta_J$.

We discuss the empirical plausibility of this assumption, and how our results can be modified if it is violated (as might be a concern given the so-called pricing kernel puzzle), in Section 4.4.

Second, and more substantively, we must impose an assumption on the evolution of the unobserved parameter $\phi_{t,j}$. In Section 2's setting, ϕ is naturally constant over t: the terminal states index consumption and thus marginal utility, so the marginal rate of substitution across these primitive states is fixed. For our main analysis, we impose an analogous assumption on $\phi_{t,j}$.

ASSUMPTION 4 (*Constant SDF Ratio*, or *Conditional Transition Independence*)**.** We say the SDF satisfies *conditional transition independence (CTI)* for 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, $j = 2, 3, \ldots, J-2$, and we write $\phi_{t,j} = \phi_j$ for these states.

This assumption imposes that the *relative* expected "severity" of the adjacent return states is constant over a contract, so that the expected SDF (or marginal utility) realization in the low return state θ_j is a constant multiple of that of θ_{j+1}. This is akin to assuming a constant local price of risk, as discussed further in Section 4. It corresponds to a notion of transition independence because it implies that the expected relative SDF realizations depend only on the return state pair and not on the path of variables realized between t and T (see Lemma A.3 in Appendix A.2). Note that we exclude the extreme state pairs (θ_1, θ_2) and (θ_{J-1}, θ_J) from the constant-ϕ requirement: thinking of θ_1 and θ_J as tail returns, we are allowing for time-varying disaster (or positive jump) risk.

As we discuss in Section 4, this assumption is weaker than the joint assumptions imposed in past volatility tests, and we view it as a reasonable starting point for our analysis. But it is unlikely to hold perfectly in reality, so that section also provides results characterizing RN excess movement when the assumption is violated.

3.3 Main Results in the General Setting

Having completed the formal setup and description of our main assumptions, 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 3. *Under no arbitrage and Assumptions 2–4, for $j = 2, 3, \ldots, J-2$, Lemma 1, Propositions 1–2, and Corollary 1 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 $\triangle_j \equiv \widetilde{\mathbb{E}}_0[X^*_j \mid R_T = \theta_{j+1}] - \widetilde{\mathbb{E}}_0[X^*_j \mid R_T = \theta_j]$ replacing $\triangle$.*

The main theoretical complication in applying the results in Section 2 to this setting is in proving that $\widetilde{\mathbb{E}}^*[X^*_j \mid R_T = \theta_j] = \widetilde{\mathbb{E}}[X^*_j \mid R_T = \theta_j]$. The economic intuition for these results is largely identical to the intuition discussed in Section 2. And 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 regardless of the origin of the SDF, the results also admit an interpretation in terms of the approximate required risk aversion for a fictitious representative agent who consumes the market index (analogous to (7)). To avoid repetition, for the remaining results we will continue to assume no arbitrage and that Assumptions 2–4 hold for $j = 2, 3, \ldots, J-2$, unless stated otherwise.

PROPOSITION 4. *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 $S_j \equiv S_0\theta_j$. Then local relative risk aversion $\gamma_j \equiv -S_j U''(S_j)/U'(S_j)$ is given to a first order around return state θ_j by*

$$\gamma_j = \frac{\phi_j - 1}{(S_{j+1} - S_j)/S_j}.$$

This result formalizes the sense in which ϕ_j reflects the local price of risk, as it corresponds to the market's effective relative risk aversion in a neighborhood around return θ_j. 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 $(S_{j+1} - S_j)/S_j$ 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. Interpreting and Relaxing the Test's Assumptions

Our analysis has proceeded under the joint null implied by Assumptions 2–4, which we now discuss in more detail. What do these assumptions entail specifically? How informative is a rejection of the joint null, and how should it be interpreted? We begin by considering CTI, which we consider the most important assumption imposed alongside RE. We first discuss settings in which it does (and doesn't) hold; provide a theoretical result relaxing the assumption; and conduct simulations of models in which it does not hold. We then briefly discuss the other two assumptions and provide a result accounting for possible mismeasurement or microstructure noise.

4.1 CTI Generates an Informative Test

We note first that the constant-ϕ assumption is significantly weaker than the assumption of constant discount rates. Our framework allows for any variation in the physical distribution of R_T, which can be thought of as changes in the *quantity* of risk. Further, by considering RN beliefs implied by options, we allow arbitrary variation in risk-free discount rates (since risk-free discounting affects all option prices proportionally). Both sources of variation are ruled out in tests with constant discount rates. We formalize these statements and discuss the relationship between RN beliefs and discount rates further in Internet Appendix C.3.

That said, the assumption of constant ϕ is substantively restrictive in one way: the price of risk must be locally constant over the life of a given option contract. We are assuming, for example, that the market's effective risk aversion over whether the return will be 0% or 5% is fixed over a span of weeks or months. Thus a violation of our bounds can be interpreted as reflecting either excess volatility in the marginal investor's beliefs, or variation in this local price of risk. A constant local price of risk, though, does not require that the *aggregate* price of risk is constant. For example, if risk aversion is decreasing with R_T, then bad news about the return distribution can increase aggregate

risk aversion without affecting *local* risk aversion in a neighborhood of a 0% return.

More generally, CTI is satisfied in a range of theoretical settings, with a few notable exceptions. We make this point not to assert that CTI must hold in reality, but instead to demonstrate that a violation of our bounds is informative in ruling out a range of well-studied frameworks:[16]

1. Any shocks that affect the expectation of M_T proportionally in all return states (e.g., multiplying $\mathbb{E}_t[M_T \mid R_T = \theta_j]$ by 1.2 for all j) are trivially ruled in by the definition of CTI. More generally, if $M_T = \delta_T f(R_T)$ for some function $f(\cdot)$, and shocks to δ_T are orthogonal to $f(R_T)$, then CTI holds. This rules in at least a subset of permanent SDF shocks as classified by Alvarez and Jermann (2005), though we note that it does not cover all possible permanent shocks.[17]

2. If a representative agent's utility depends only on the terminal index value, then CTI holds. Similarly, if there is any agent whose indirect utility can be written as a function only of the time-T index value — for example, an unconstrained investor with horizon T (i.e., Utility $= f(\text{Wealth}_T)$) who is fully invested in the market — then CTI holds. This includes settings in which an unconstrained log investor holds the market, which is the leading case studied by Martin (2017) in measuring the equity premium (see also Gandhi, Gormsen, and Lazarus, 2025).

3. In the variable rare disasters model of Gabaix (2012), CTI holds for all market return-state pairs (θ_j, θ_{j+1}) for which there is negligible probability of having realized a disaster conditional on reaching θ_j.[18] This illustrates the usefulness of the localization provided by considering conditional beliefs: if we are concerned that time-varying disaster risk may affect $\phi_{t,j}$ for states in the tail of the return distribution, we can ignore these states and confine attention to the center of the return distribution, where $\phi_{t,j}$ can be expected to be approximately constant.

4. If a representative agent has Epstein–Zin (1989) utility and holds the market, then CTI holds if either: (i) relative risk aversion is $\gamma = 1$, and the elasticity of intertemporal substitution (EIS) ψ and DGP are arbitrary; (ii) γ is arbitrary, the EIS is $\psi = 1$, and log consumption growth follows an AR(1); or (iii) γ and ψ are arbitrary, and consumption growth is i.i.d. (a less interesting case). As Martin (2017, p. 378) notes, case (i) is "considered (and not rejected) by Epstein and Zin (1991) and Hansen and Jagannathan (1991)." Case (ii) approximates Case I of the Bansal and Yaron (2004) long-run risks model (without stochastic volatility), with $\psi = 1$ and common shocks to current and expected future consumption. Finally, case (iii) is considered by Martin (2013).

These examples illustrate further that the assumption of CTI is meaningfully weaker than the assumption of constant discount rates. This should again not be taken as evidence *in favor* of CTI in fact holding; instead, it allows for an informative joint test whose null includes a range of models that have been advanced as rationalizations of the excess-volatility puzzle.

That said, there are also well-known models under which CTI does *not* hold. As suggested by

[16] Derivations for statements 3–4 and the subsequent examples are in Internet Appendix B.2, and 1–2 are immediate.

[17] Meanwhile, Borovička, Hansen, and Scheinkman (2016) show that permanent SDF shocks are ruled out in the framework proposed by Ross (2015) for recovering physical probabilities from state prices.

[18] Formally, for any δ>0, $\exists\, \underline{\theta}$ s.t. $\forall\, \theta_j \geqslant \underline{\theta}$, $\mathbb{P}_0\big(\sum_{t=1}^{T} \mathbb{1}\{\text{disaster}_t\} > 0 \,\big|\, R_T \geqslant \underline{\theta}\big) < \delta$, so the conditional probability of having realized a disaster is negligible. For all $\theta_j \geqslant \underline{\theta}$, CTI holds for the state pair (θ_j, θ_{j+1}) up to a negligible error.

statement 4, models featuring priced stochastic volatility generally pose an issue for CTI. Since priced volatility risk is a central feature of most leading option pricing models, we conduct a detailed simulation study of how much RN belief variation such models generate in Section 6.1.

Other models in which CTI is violated further illustrate its content. CTI fails in the habit model of Campbell and Cochrane (1999): the path of consumption matters in a manner not fully accounted for by conditioning on the return state. In addition, CTI can fail to hold in heterogeneous-belief models if there is no rational investor who is unconstrained with probability one. For example, in the Basak (2000) model with heterogeneous beliefs and extraneous (non-fundamental) risk, CTI fails to hold if extraneous risk is perceived to matter by all agents. We view the fact that CTI rules out stochastic volatility and habit models to be a downside of the framework, which we address with robustness results below. The fact that we rule out some (though not all) models of belief volatility induced by heterogeneity, though, is a feature: time variation in ϕ in these cases is equivalent to the as-if representative agent having excessively volatile beliefs, which represents a meaningful alternative to our null.

4.2 Relaxing CTI: Theory and Simulations

While CTI allows for an informative null in our main analysis, it is also a knife-edge restriction that is unlikely to hold exactly. We thus now investigate how such a violation affects RN excess movement. One might worry that even small fluctuations in ϕ could generate dramatic violations of our bounds. For example, suppose that π_t is constant at 0.5 and ϕ_t changes back and forth from 1 to 1.5 repeatedly. Without any movement in physical beliefs, π_t^* will vary repeatedly between 0.5 and 0.6, leading to unbounded movement as $T \to \infty$. But this argument overlooks a core insight: this repeated oscillation is *also* inconsistent with RE as long as the variance of ϕ_t is bounded, because it implies predictable mean reversion in the expected marginal value of \$1. That is, the fact that ϕ_t is itself a function of martingale conditional expectations restricts its evolution under RE. We pursue this logic more formally here, both theoretically and in a set of simulations.

First, in the theory, dropping the constant-ϕ_t assumption comes at the cost of much of the parsimony in our analysis: SDF expectations non-linearly map to ϕ_t, which non-linearly combines with π_t to determine π_t^*, which is non-linearly mapped into X^*. Perhaps surprisingly, though, we can still make theoretical statements for a meaningful subset of DGPs. Consider ϕ_t as defined in (14) (continuing to suppress j subscripts). The agent now also receives signals $s_{\phi_t} \in \mathcal{S}_{\phi_t}$ to learn about ϕ_t over time. For simplicity, assume that uncertainty over ϕ_t evolves on a binomial tree ($|\mathcal{S}_{\phi_t}| = 2$), though this is not necessary for the following statements. We also assume that π_t and ϕ_t do not change (relative to their $t-1$ values) together in the same period.[19] Then the following holds.

PROPOSITION 5. *If ϕ_t evolves as a martingale or supermartingale ($\mathbb{E}_t[\phi_{t+1}] \leqslant \phi_t$) and $Var(\phi_t) < \infty$, then the bounds in Proposition 2, Corollary 1, and Proposition 3 continue to apply, with ϕ_0 replacing ϕ.*

[19] In a previous version of this manuscript (Augenblick and Lazarus, 2022), we prove the result for arbitrary $|\mathcal{S}_{\phi_t}|$. For the second assumption, one can split each period into two sub-periods, with ϕ_t changing in the first half and π_t in the second. The proof becomes intractable without this assumption, but we provide simulation evidence relaxing it below.

The proof of Proposition 5 demonstrates that the variation in π_0^* arising when ϕ_t is a non-degenerate (super)martingale always strictly lowers excess RN belief movement, rendering the main bound in (10) even more conservative. Changes in ϕ_t do cause π_t^* to change, which adds movement. But when ϕ_t is a supermartingale, this movement works against the upward bias in RN beliefs themselves and decreases $\pi_t^* - \pi_t$ in expectation. This reduces potential future excess movement more than enough to offset the increase in period t.

As for the interpretation of the supermartingale restriction, from the definition of covariance, $\phi_{t,j}$ is a supermartingale if and only if $\text{Cov}_t(\phi_{t+1,j}, \mathbb{E}_t[M_T \mid R_T = \theta_{j+1}]) \geqslant 0$. Interpreting the SDF M_T as proportional to $U'(C_T)$, this requires that risk aversion (encoded in $\phi_{t+1,j}$) be positively correlated with expected marginal utility (MU) in the high-C_T state. This is an intuitively reasonable restriction, as it implies bad news (higher expected MU) generally arrives at the same time for both states; see Lazarus (2022) for further discussion. That said, the requirement that π_t and ϕ_t do not vary contemporaneously may be more restrictive. And given the complexity of the setting, it is difficult to make analytical statements when ϕ_t does not satisfy the assumptions of Proposition 5.

We therefore next turn to a set of simulations in which ϕ_t can vary with more freedom. For now, we consider a set of abstract DGPs. We also conduct simulations of specific option pricing models with varying ϕ_t; we compare these quantitatively with our empirical results, and accordingly postpone discussion of these simulations to the next section. In our abstract simulations, the agent learns over time about the terminal state and the SDF realization in both the j and $j+1$ states. Each of these three objects takes on one of two values at T, with different combinations of ex ante uncertainty and signal strengths for all three objects. We describe the setting further and present full results in Internet Appendix C.4, and we outline the results of these simulations here.

The simulations all start with $\pi_0^* = 0.5$ and $\phi_0 = 3$ but differ in the variation in ϕ_t, with the ex ante standard deviation of ϕ_T ranging from $\text{SD}_0(\phi_T) = 0$ to 3. Using Proposition 4, if states θ_j and θ_{j+1} differ by 5% (as in our data), this corresponds to relative risk aversion of $\gamma_0 = 40$ and a standard deviation for γ_T between 0 and 60. Figure C.2 plots distributions for the estimated $\mathbb{E}[\mathsf{m}^*]$ across simulations. It shows that uncertainty in ϕ_T has limited impact on estimated movement in these DGPs. Even in the case with the highest uncertainty in ϕ_T, movement $\mathbb{E}[\mathsf{m}^*]$ only increases by 0.015, or by 6% relative to the constant-ϕ value of $\mathbb{E}[\mathsf{m}^*] = \mathsf{u}_0^* = 0.25$. Further, $\mathbb{E}[X^*]$ rarely exceeds the bound for $\phi = 3$ (in just a fraction of 1% of the high-uncertainty simulations). And in all cases, the bound in Corollary 1 for $\phi \to \infty$ holds for 100% of the simulations.

We conclude, somewhat surprisingly, that even significant uncertainty in ϕ_T does not meaningfully affect our bounds, at least for these DGPs. When the agent updates her SDF expectations, these updates must still respect RE. Thus even large values for σ_ϕ do not allow for arbitrary oscillations of ϕ_t, as information about ϕ_T is revealed gradually over time. That said, these binary DGPs are somewhat difficult to map to more fully specified option pricing models, which may produce richer time variation in ϕ_t than considered in these preliminary simulations. We therefore return to simulations of these option pricing models after presenting our empirical results.

4.3 Aggregating Over Belief Streams

The previous subsection considers the effect of time variation in ϕ_j *within* a given RN belief stream (i.e., within a single option contract, lasting weeks or months). But even if the effects of such higher-frequency variation are small, it may be less palatable to assume that ϕ_j is constant *across* belief streams: there may be lower-frequency variation over the span of years. We now ask how such variation affects our bounds when considering multiple belief streams.

Answering this question is also important for empirical implementation. The bounds in Proposition 3 are stated as date-0 expectations conditional on the RN prior, but we observe only one draw X_j^* per expiration date rather than the expectation of this statistic for a given $\widetilde{\pi}_{0,j}^*$. Estimation thus requires aggregating over multiple streams with different $\widetilde{\pi}_{0,j}^*$ and potentially different ϕ_j, especially since our estimation groups different streams together using the same fixed return states.

Thus, generalizing slightly, assume now that we can observe index options for N expiration dates $\mathcal{T} \equiv \{T_i\}_{i=1}^N$, so i indexes belief streams (and their DGPs). We use $\phi_{i,j}$ for the SDF ratio for expiration date T_i and state pair (θ_j, θ_{j+1}); RN beliefs are $\widetilde{\pi}_{t,i,j}^*$; and RN excess movement is $X_{i,j}^*$. We again often suppress j for an arbitrary state pair. Due to Jensen's inequality, we cannot simply use $\mathbb{E}[\phi_i]$ in place of ϕ_i or $\mathbb{E}[\pi_{0,i}^*]$ in place of $\pi_{0,i}^*$ when taking the expectation of both sides of the results in Propositions 1–3 over all i. However, the following generalizations do hold:

PROPOSITION 6. *Define* $\overline{\phi} \equiv \max_{\pi_{0,i}^*} \mathbb{E}[\phi_i \mid \pi_{0,i}^*]$. *We have:*

(i) GENERALIZATION OF PROPOSITION 1: *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,*

$$\mathbb{E}[X_i^*] \leqslant \max\left\{0, \left(\pi_{0,i}^* - \frac{\pi_{0,i}^*}{\mathbb{E}[\phi_i] + (1 - \mathbb{E}[\phi_i])\pi_{0,i}^*}\right)\mathbb{E}[\triangle_i]\right\}. \tag{15}$$

(ii) GENERALIZATION OF PROPOSITION 2: *Over all streams, without any additional assumptions,*

$$\mathbb{E}[X_i^*] \leqslant \mathbb{E}\left[\left(\pi_{0,i}^* - \frac{\pi_{0,i}^*}{\overline{\phi} + (1 - \overline{\phi})\pi_{0,i}^*}\right)\pi_{0,i}^*\right], \tag{16}$$

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,*

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

Only the analogue to Proposition 1 requires an additional assumption. The original bound is $(\pi_0^* - \pi_0)\triangle$, so the covariance between $\pi_{0,i}$ and $\triangle_i$ over DGPs affects the generalized bound. For simplicity, we set this covariance to zero.[20] Part (ii), meanwhile, generalizes (10); as can be seen in

[20] This is equivalent to assuming no relationship between the asymmetry of the DGP and ϕ. This part also holds fixing $\pi_{0,i}^*$; this is sufficient for us, as our empirical results for this bound will generally be conditional on a given $\pi_{0,i}^*$.

the proof of the result, it is more conservative than the original bound for a single stream. Finally, the bound that does not depend on $\phi_{i,j}$ (part (iii)) applies as previously stated.

These bounds are now empirically implementable, and the minimum $\overline{\phi}$ that solves (16) is a conservative estimate of the maximal conditional-mean SDF slope for the return-state pair in question. Finally, reintroducing dependence on the state pair j, it is likely that the values $\overline{\phi}_j$ vary over j. But the same steps used for Proposition 6 to take expectations over i 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) when desired.

4.4 Robustness to Additional Assumptions

Having considered Assumption 4 in detail, we turn to the other two assumptions to continue our discussion of how to interpret a rejection of the joint null. To conserve space, we relegate details to Internet Appendix C.5 and provide a brief summary here.

First, considering Assumption 2, this assumption requires both a correct subjective prior and rational updating using the true signal distribution as the likelihood. A natural question is whether an incorrect prior by itself can generate violations of the upper bound for X^*, or whether excessive movement requires incorrect updating. We show in Proposition C.4 that because an incorrect prior acts as a one-time belief distortion, it is generally not sufficient for a full violation of the bound in Proposition 2. In general, then, incorrect updating behavior must be present in such a violation.[21]

We now turn to Assumption 3, which imposes that $\phi_j \geqslant 1$. This requires that the expected SDF realization in the low-return state be greater than in the high-return state. A line of work beginning with Jackwerth (2000), however, argues that the SDF does not decrease monotonically with the index return in options data. This finding, often referred to as the "pricing kernel puzzle," would imply a violation of Assumption 3 for at least some subset of the return space. Given such a violation, it turns out that there is still a well-defined theoretical upper bound for RN excess movement: we show in Proposition C.5 that our main bounds apply with minor modification (effectively flipping the role of the two states when $\phi < 1$), and for *any* value of ϕ, $\mathbb{E}[X^*] \leqslant \max({\pi_0^*}^2, (1-\pi_0^*)^2)$. So excess movement is bounded regardless of ϕ. While this is useful theoretically, the bound in the $\phi < 1$ case would have to be estimated separately from our main bounds derived for the $\phi \geqslant 1$ case. So for the sake of simplicity, we empirically estimate only the main bounds. We do so in part because we estimate the bounds separately for multiple points in the return space. Thus even if one is concerned about the pricing kernel puzzle affecting our results, one can confine empirical attention to our estimates for return ranges for which the puzzle does not emerge.[22]

[21] Note that even if ϕ is constant under the agent's subjective belief, incorrect updating will induce a probability distortion such that the *actual* SDF (assessed under $\mathbb{P}$) will feature a time-varying ϕ_t.

[22] For example, a robust finding in this literature is that the estimated SDF declines monotonically for the range of negative index returns; see, for example, Driessen, Koëter, and Wilms (2024, Fig. 4) or Schreindorfer and Sichert (2023, p. 4). And as will be seen in Section 5, our empirical estimates of ϕ_j for these negative return states are somewhat higher than for positive return states, indicating that the results are unlikely to be driven by violations of Assumption 3.

4.5 Robustness to Measurement Error

The robustness results above speak to the possibility of theoretical misspecification. As the final step in making our bounds implementable, we now consider how to account for possible *empirical* misspecification arising from mismeasurement or microstructure noise in RN beliefs. The 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 $X^*_{t,t+1} \equiv \mathsf{m}^*_{t,t+1} - (\mathsf{u}^*_t - \mathsf{u}^*_{t+1})$. Similar to Augenblick and Rabin (2021, Section II.E), we then have the following:

PROPOSITION 7. *Assume that the observed $\widehat{\pi}^*_t$ is measured with error with respect to the true π^*_t:*

$$\widehat{\pi}^*_t = \pi^*_t + \epsilon_t, \tag{17}$$

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. If measurement error is positively correlated over time rather than uncorrelated, this will reduce the upward bias in measured X^*. We discuss estimation of $\mathrm{Var}(\epsilon_t)$, and autocovariance statistics for ϵ_t, in Section 5.2.

5. Empirical Estimation and Results

Our theory leads to bounds on the variation in RN beliefs over the index return, which we proceed now to measure in the data. We begin by describing how we map from theory to data and how we estimate microstructure noise; we then summarize the data and present our main results, along with a set of robustness checks.

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.[23] We augment this data with intraday price quotes obtained

[23]We use the same sample as in an earlier version of the paper (Augenblick and Lazarus, 2022), as it aligns with our intraday data for noise estimation. The 2018 cutoff precludes the volatility observed in 2020 from affecting our estimates.

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; see Internet Appendix C.6.[24] Two aspects of this 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 maturity of at most one year, and streams with at least $T = 5$ observations. Second, after transforming the filtered prices to RN beliefs, we keep only conditional RN belief observations $\widetilde{\pi}^*_{t,i,j}$ for which the non-conditional beliefs satisfy $\pi^*_t(R_{T_i} = \theta_j) + \pi^*_t(R_{T_i} = \theta_{j+1}) \geqslant 5\%$, as conditional beliefs 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:

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

where $R^f_{0_i,T_i}$ is the gross risk-free rate from 0_i to T_i. 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 if the excess return is in the five percentage point (pp) bin between -0.2 and -0.15; and so on.

More concretely, if the index value on the start date 0_i is $S_{0_i} = 100$ and the net risk-free rate is 0, then the return states will be defined so that state 1 is realized if S_{T_i} is less than (roughly) 80, state 2 if S_{T_i} is between 80 and 85, and so on. These bins are then fixed for all t until expiration, so that beliefs are over fixed events. We do not annualize returns when defining return states so that the states have the same meaning in terms of percent wealth differences regardless of the length of a given belief stream. Of course, the meaning of, e.g., a -15% excess return may be different in one year than it is the next. But Proposition 6 tells us that when grouping them together in estimation, our bound will still be conservative for the maximal value of ϕ_j across all panels.

Our 5 pp bin choice 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 less measurement noise but make it less likely that there are no changes in the expected SDF realization conditional on a given state θ_j relative to θ_{j+1}. We discuss alternative bin choices after presenting our main results.

For notation, we often refer to states by the right end of their return bin: $\theta_1 = -0.2$, $\theta_2 = -0.15$, and so on. The binary beliefs to be used in our tests are $\widetilde{\pi}^*_{t,i,j} \equiv \pi^*_t(R_{T_i} = \theta_j \,|\, R_{T_i} \in \{\theta_j, \theta_{j+1}\})$, 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 $j+1$ (here, conditional on an excess return between -20% and -10%). We again use the right end of the return bin for j in discussing statistics for pair (θ_j, θ_{j+1}), corresponding to the midpoint of the two bins. We report empirical estimates for all adjacent state pairs for completeness, but as discussed after Assumption 4, it is unlikely that CTI is met for the extreme state pairs (θ_1 relative

[24] There are 12.4 million option prices in the raw data, and 4.3 million (for 5,537 trading dates and 991 expiration dates) after filtering. The bulk of the difference is attributable to our use of only out-of-the-money call and put strikes.

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 $\theta_2, \ldots, \theta_8$. And when we aggregate our state-by-state estimates of ϕ_j required to rationalize the data into a single average value $\overline{\boldsymbol{\phi}}$ across states, we use only these interior states.

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. We follow Malz (2014) and fit an interpolating spline to the observed prices (in implied volatility, or IV, space). Then using equation (13), we map fitted prices to RN beliefs by calculating numerical derivatives over very small finite differences. See Appendix C.6 for details. With the RN beliefs $\pi^*_{t,i,j}$ in hand, we then calculate conditional beliefs as $\widetilde{\pi}^*_{t,i,j} = \pi^*_{t,i,j}/(\pi^*_{t,i,j} + \pi^*_{t,i,j+1})$. We use the resulting conditional RN belief streams to calculate the excess movement statistics $X^*_{i,j}$ 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 3 and 6 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. Having clarified how the relevant empirical objects ($\widetilde{\pi}^*_{t,i,j}$, $X^*_{i,j}$) depend on the contract i and state pair j, in what follows we generally drop the cumbersome use of i, j, and $\widetilde{\cdot}$, and again write π^*_t for $\widetilde{\pi}^*_{t,i,j}$, X^* for $X^*_{i,j}$, and so on. Similarly, we often drop the "conditional" qualifier when referring to conditional RN belief statistics.

5.2 Noise Estimation

As in Proposition 7, we also wish to account for measurement error stemming from possible non-fundamental or microstructure noise. With $\widehat{\pi}^*_t = \pi^*_t + \epsilon_t$ as in (17), Proposition 7 tells us that we must estimate $\mathrm{Var}(\epsilon_t)$ in order to eliminate the bias in X^* arising from any such noise. 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 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 would be computationally infeasible if applied to our entire sample of 5,537 trading days (each of which has 11 available expiration dates on average). We thus select 30 trading days at random from our sample period and calculate intraday RN distributions for these days, yielding an intraday data set roughly twice as large as the original one.

We then use tools from the literature on microstructure noise to estimate $\mathrm{Var}(\epsilon_t)$ using these data. The intuition for this strategy is best understood by assuming temporarily that the noise ϵ_t in (17) is i.i.d., while the true π^*_t changes smoothly over time (e.g., Zhang, Mykland, and Aït-Sahalia, 2005). Given high-frequency option data, imagine calculating movement using the observed beliefs, $(\widehat{\pi}^*_{t+h} - \widehat{\pi}^*_t)^2$. As one decreases the time h between consecutive observations, the noise swamps the

true variation: since $(\pi^*_{t+h} - \pi^*_t)^2 \to 0$, we have $\mathbb{E}[(\widehat{\pi}^*_{t+h} - \widehat{\pi}^*_t)^2] \to 2\text{Var}(\epsilon_t)$. Thus in this example, $\text{Var}(\epsilon_t)$ can be estimated using the quadratic variation in RN beliefs sampled at a high frequency.

In practice, one would expect both non-i.i.d. noise ϵ_t and jumps in the true process π^*_t. One estimator for $\text{Var}(\epsilon_t)$ that is robust to these features is the ReMeDI ("Realized moMents of Disjoint Increments") estimator proposed by Li and Linton (2022). It takes the average product of *disjoint increments* of the observed process, $(\widehat{\pi}^*_t - \widehat{\pi}^*_{t-h})(\widehat{\pi}^*_t - \widehat{\pi}^*_{t+h})$.[25] Intuitively, even if the true process features jumps so that $\mathbb{E}[(\pi^*_{t+h} - \pi^*_t)^2] > 0$, its non-overlapping increments are still approximately uncorrelated. Li and Linton show that this estimator is consistent for $\text{Var}(\epsilon_t)$ for quite general noise and underlying processes, and it 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 state pair j in our intraday sample. We then match the noise estimates (which are obtained for a subsample of days) to the X^* observations in our original data.[26] 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 7, and we sum these noise-adjusted one-day values over the full stream to obtain noise-adjusted $X^*_{i,j}$.

We discuss the magnitude of the noise estimates in the next subsection. 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 7 that end-of-day noise observations are uncorrelated, $\mathbb{E}[\epsilon_t \epsilon_{t+1}] = 0$.

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.

5.3 Excess Movement: Descriptive Statistics and Figures

Table 1 summarizes the average RN excess movement $\overline{X}^*$ overall (across all interior state pairs and expiration dates) and by subsample. Excess movement is difficult to interpret without some normalization. The first two columns thus divide $\overline{X}^*$ by average initial uncertainty $\overline{u}^*_0$. As in Augenblick and Rabin (2021), this normalized statistic can be interpreted as the percent by which movement exceeds initial uncertainty and thus 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. These values decrease for return states in the middle of the distribution. The early sample has high but noisy X^* statistics,[27] but these averages remain high until the most recent subsample. And higher priors π^*_0 correspond with greater X^*, in line with our bounds given $\phi \geqslant 1$.

The next two columns of Table 1 instead normalize $\overline{X}^*$ by the average contract length $\overline{T}$, so the

[25] More formally, the estimator 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 span (in our case, one trading day) and k_n is a tuning parameter.

[26] We match these observations using the two best predictors of $\widehat{\text{Var}}(\epsilon_t)$ in our data: the state pair j and the sum $\pi^*_t(R_T = \theta_j) + \pi^*_t(R_T = \theta_{j+1})$. Further explanation and technical details can be found in Internet Appendix C.7.

[27] Excess movement peaks in this subsample during the 1998 Russian debt crisis.

resulting statistics can be interpreted as excess movement per day.[28] Under this normalization, there is now no clear pattern for average excess movement across bins: longer average contract lengths tend to coincide with more excess movement, as RN beliefs bounce up and down over the length of a contract. This is inconsistent with RE, under which excess movement in subjective beliefs should not depend at all on the horizon at which uncertainty is resolved. For the splits by date and by prior, the basic patterns discussed above are still present here.

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 about 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 return-state splits, noise tends to be lowest for returns near the center of the distribution, as is intuitive.

Next, Figure 4 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^*; 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 for each value of π_0^*, as shown in gray. Panel (a) uses the tighter bound from Proposition 6(i), with $\triangle$ estimated using local averages for the conditional expectations $\mathbb{E}[X^* \mid \theta, \pi_0^*]$ in (9) over π_0^*.[29] Panel (b) uses the conservative bound from the second inequality in Proposition 6(ii), so the gray lines in this panel align with the solid bound lines in Figure 3.

Across the two panels, the X^* values observed in the data exceed both sets of bounds, other than for high RN priors π_0^* and for high ϕ. As the bounds 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 ${\pi_0^*}^2$ (the bound for $\phi = \infty$) for π_0^* less than about 2/3.

These figures do not, however, integrate over π_0^*, nor do they include any measures of statistical uncertainty necessary to make inferential statements or conduct hypothesis tests. To address these 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. We calculate two versions of this residual corresponding to the bound in part (i) and the unconditional bound in

[28] For a rough idea of the actual variation in RN beliefs given these values, consider a pair $(t, t+1)$ for which $\pi_{t+1}^* = 1 - \pi_t^*$. One-day X^* and m^* then coincide (there is no uncertainty resolution), so, for example, the noise-adjusted mean of 0.0038 corresponds to a raw change of $\sqrt{0.0038} \approx 0.06$ (or $\pi_t^* = 0.47$, $\pi_{t+1}^* = 0.53$).

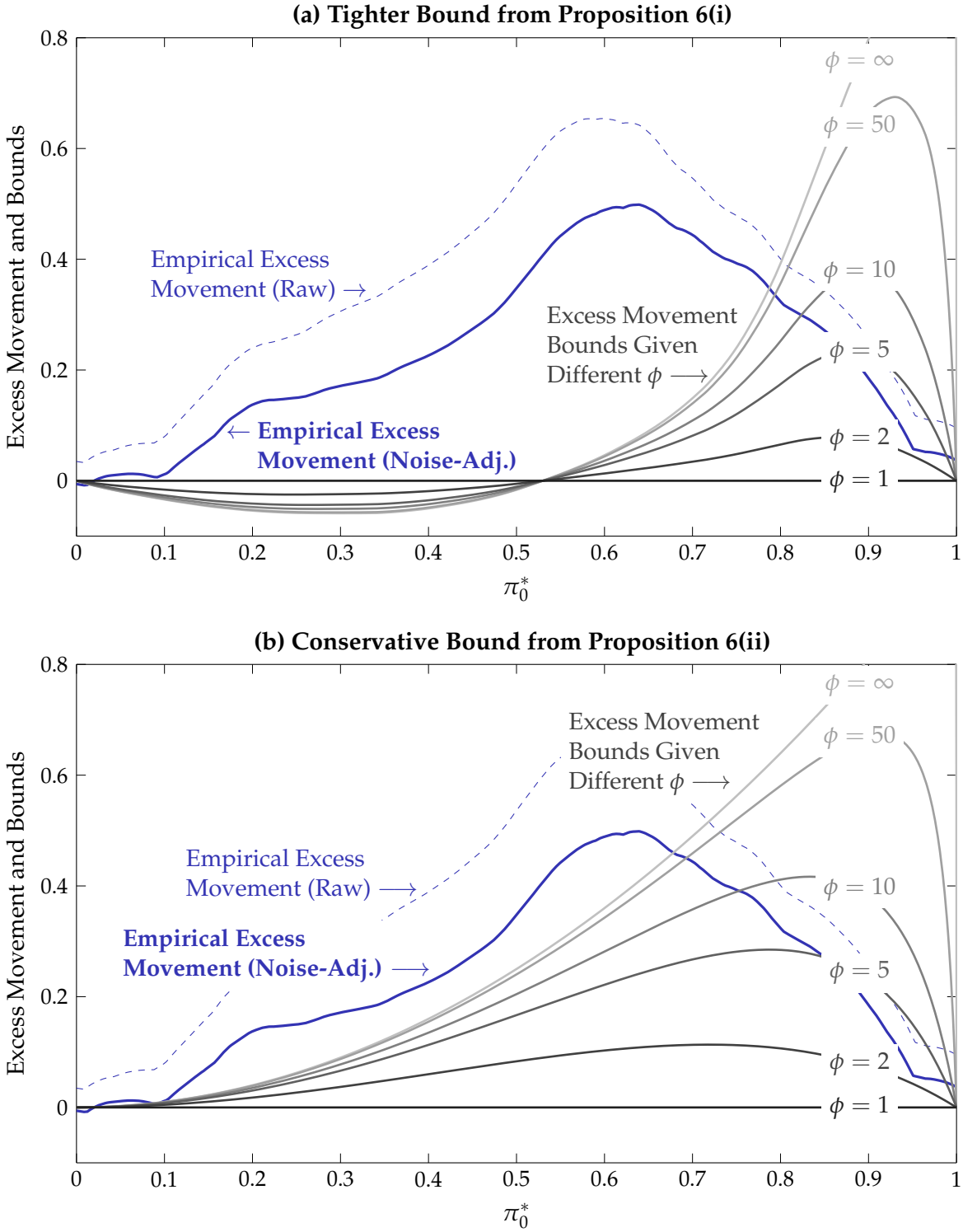
[29] To emphasize that $\widehat{\triangle} < 0$ for $\pi_0^* < 0.5$, the figure does not cut the bounds off at 0.

Table 1: 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.

Figure 4: 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 0.08) 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 (9), with $\widehat{\triangle}$ then plugged into the inequality in Proposition 6(i). Bounds for (b) use only the second inequality in Proposition 6(ii).

part (ii) of Proposition 6 (equations (15) and (16), respectively):

$$
\begin{aligned}
e_i^{\triangle}(\overline{\phi}) &= X_i^* - \max\left\{0, \left(\pi_{0,i}^* - \frac{\pi_{0,i}^*}{\overline{\phi} + (1-\overline{\phi})\pi_{0,i}^*}\right)\widehat{\triangle}_i\right\}, \\
e_i^{\text{main}}(\overline{\phi}) &= X_i^* - \left(\pi_{0,i}^* - \frac{\pi_{0,i}^*}{\overline{\phi} + (1-\overline{\phi})\pi_{0,i}^*}\right)\pi_{0,i}^*.
\end{aligned}
\tag{18}
$$

The first version corresponds to the tighter bound from Proposition 1, which requires a smoothed estimate of $\triangle_i$ as calculated for Figure 4. 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 $\text{Cov}(\pi_{0,i}, \triangle_i) = 0$. Thus the second version, which implements the more conservative bound from Proposition 2 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 state pair j and aggregated over all interior states. As these values have no natural scaling, we present them as t-statistics, $t_{\overline{e}} = \overline{e_i(\phi)} / \widehat{SE}_{\overline{e}}$, where $\overline{e_i(\phi)}$ is the sample average of $e_i(\overline{\phi})$ and $\widehat{SE}_{\overline{e}}$ is its standard error. Given the overlapping nature of option panels, we calculate standard errors with a block bootstrap. We use blocks 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 $e_i(\overline{\phi})$.[30] The bootstrap thus accounts for sampling uncertainty in all statistics used to calculate noise-adjusted X_i^* 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 draws.

These residual t-statistics are presented in Table 2 for the two versions of the residual in (18), both overall and by state pair. Since $\mathbb{E}[e_i(\overline{\phi})] = 0$ under RE given a correctly specified $\overline{\phi}$, these t-statistics tell us how far the residuals are from being consistent with any hypothesized null for the SDF slope. Positive numbers correspond to the data exhibiting too much excess movement to be consistent with a given value of $\overline{\phi}$. In the first six columns, 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 mostly large in magnitude), indicating no value of $\overline{\phi}$ is consistent with the degree of observed excess RN movement. In the remaining columns, the conservative bound t-statistics are somewhat smaller in magnitude but also generally positive, other than for large $\overline{\phi}$ and for returns 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

[30] For the residual $e_i^{\triangle}(\overline{\phi})$ 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 4(a), and then evaluating it at the observed $\pi_{0,i}^*$.

Table 2: Residual Excess Movement *t*-Statistics for Different $\overline{\phi}$

	(a) *Tighter Lower Bound from Proposition 6(i)*						(b) *Conservative Lower Bound from Proposition 6(ii)*					
$\overline{\phi}$:	1	2	5	10	50	∞	1	2	5	10	50	∞
Overall *t*-stat.:	5.19	5.15	4.98	4.64	3.03	0.78	5.19	4.08	2.68	1.75	0.07	-1.48
By return state:												
1 (-20%)	4.06	3.69	3.04	2.57	1.94	1.72	4.06	3.58	2.79	2.22	1.47	1.21
2 (-15%)	2.13	2.13	2.13	2.13	2.14	2.14	2.13	2.03	1.94	1.90	1.87	1.86
3 (-10%)	2.23	2.23	2.23	2.23	2.23	2.23	2.23	2.03	1.86	1.79	1.72	1.70
4 (-5%)	2.78	2.78	2.77	2.77	2.76	2.76	2.78	2.51	2.28	2.18	2.09	2.06
5 (0%)	3.09	3.09	3.08	3.08	3.06	3.06	3.09	1.86	0.74	0.26	-0.19	-0.31
6 (+5%)	3.26	2.66	1.27	0.04	-1.71	-2.42	3.26	1.96	-0.12	-1.77	-5.22	-8.39
7 (+10%)	2.82	2.73	2.51	2.29	1.70	1.23	2.82	2.38	1.73	1.28	0.50	0.06
8 (+15%)	3.53	3.49	3.42	3.37	3.29	3.25	3.53	3.20	2.77	2.51	2.21	2.10
9 (+20%)	1.63	1.58	1.52	1.48	1.44	1.43	1.63	1.36	1.05	0.89	0.72	0.68

Notes: This table shows *t*-statistics for the average residuals $\overline{e_i(\phi)}$ in (18) for different values of $\overline{\phi}$. The columns for (a) use $e_i^{\triangle}(\overline{\phi})$, with $\triangle_i$ estimated as in panel (a) of Figure 4. The columns for (b) use $e_i^{\text{main}}(\overline{\phi})$. 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 X^*.

the average residual $\overline{e_i(\phi)}$ is zero, so that we are effectively finding the root of the function traced out in Table 2. 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 bound from Proposition 6(ii). The estimated $\overline{\phi}$ is the minimal SDF slope for which the amount of observed excess movement in RN beliefs can be rationalized; it is thus an index of the restrictiveness of Assumptions 2–4 (or, given Proposition 5, Assumptions 2–3 and supermartingale ϕ_t). As above, we estimate this SDF slope both for each individual state pair ($\overline{\phi} = \overline{\phi}_j$) and overall ($\overline{\phi} = \overline{\boldsymbol{\phi}}$). To make the estimates for $\overline{\phi}$ more interpretable, we also use Proposition 4 to translate them into local relative risk aversion values $\overline{\gamma} = \frac{(\overline{\phi}-1)}{0.05} = 20(\overline{\phi} - 1)$ for a representative agent who consumes the market.

Table 3 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 6(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. To obtain these, we start from the same bootstrap resampling procedure as described for Table 2. We then search for and report the minimal value $\overline{\phi}$ such that we cannot reject that excess volatility ($\overline{e_i^{\text{main}}(\phi)}$) is zero at the 5% level in the bootstrapped data. (See Internet Appendix C.8 for further details.) We use one-sided intervals because our estimate for $\overline{\phi}$ is a lower bound for the SDF slope consistent with the data.

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 of 0.07 in Table 2(b) for $\overline{\boldsymbol{\phi}} = 50$. This translates to an extraordinarily high estimated lower bound for $\overline{\gamma}$ of 1,075. Values of

Table 3: 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 6(ii). These estimates are translated to relative risk aversion $\overline{\gamma}$ using Proposition 4, 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 (18). Confidence interval lower bounds are obtained by inverting a test for $\overline{\phi}$ using bootstrapped data; see Online Appendix C.8 for details. All estimates use conditional means of noise-adjusted excess movement.

$\overline{\boldsymbol{\phi}}$ below 9.8 (for $\overline{\gamma}$, below 175) are rejected at the 5% level. This is our main empirical result: under our maintained assumptions, extremely high risk aversion is needed to rationalize the large degree of excess movement in RN beliefs observed in the data.

For the individual return-state pairs, all but two of the point estimates 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 ($j = 2, 3, 4, 8$), their confidence intervals also have lower bounds of ∞ (or more precisely, are empty), indicating outright model rejection. Only for RN beliefs over state pairs in the middle of the distribution — i.e., for $j = 5$ and 6, with return midpoints of 0% and 5%, respectively — are there 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.

As discussed in Section 4.4, the pricing kernel puzzle tends to emerge only in the range of positive returns. One can thus focus attention on the interior states with negative returns, $j = 2$–5,

for estimates that are unlikely to be affected by the possibility that $\phi < 1$. For all four of these states, the null is fully rejected. It is thus unlikely that Assumption 3 is driving our results. Similarly, by Proposition C.4, our findings cannot in general be produced solely by miscalibrated priors. This indicates that our rejection must be arising either from excessively volatile physical belief revisions, or a strong violation of CTI arising from changes in the local price of risk.

5.5 Robustness Checks

To ensure our results are not overly sensitive to specific measurement choices, we re-estimate RN excess movement and our bounds with a range of different specifications, with full results in Internet Appendix C.9. First, we consider a finer partition when defining index return states: rather than our baseline of 5 pp, we use 2 pp bins. Excess movement is greater than our baseline, and the estimated bound gives $\overline{\phi} = \infty$ (CI lower bound: ∞). Second, we define states in terms of call-option delta $\frac{\partial q_{0,K}}{\partial S_0}$ as of the beginning of the option panel: state 1 corresponds to strikes with Black–Scholes delta between 1 and 0.9, state 2 between 0.9 and 0.8, and so on.[31] This similarly produces greater $\overline{X}^*$ and estimated bounds of ∞. Third, we consider alternative smoothing methods to construct RN beliefs from observed option prices. Our baseline method follows Malz (2014) and Neuberger (2012) in fitting a cubic spline interpolant through the observed prices (more specifically, their IVs). Other methods impose some degree of smoothness upfront by not requiring the fitted spline to pass through observed prices. Here, we use the approach in Bliss and Panigirtzoglou (2004), fitting a smoothing spline with roughly equal weights on goodness of fit and smoothness. We obtain slightly less excess movement (noise-adjusted $\overline{X}^*/\overline{u}_0^*$ of 117% rather than 123%) and estimated $\overline{\phi} = 26.1$ (CI bound: 7.7), corresponding to an RRA of $\overline{\gamma} = 501$ (CI: 134). Finally, we extrapolate beyond the observed strikes by pasting generalized extreme value tails onto the Bliss and Panigirtzoglou (2004) RN distribution, following Figlewski (2010); all previous tests use lognormal tails. This alternative extrapolation leads to very slightly lower X^*, with $\overline{\phi}$ estimated to be 25.4 (CI: 7.9). Overall, none of our main results are affected by any of these choices.

6. Interpreting the Results

Our results to this point appear quantitatively large: we find 123% more RN movement than uncertainty reduction, and extremely high risk aversion needed to rationalize this excess movement according to our bound. But because our statistics are based on non-linear transformations of option prices, it may be somewhat unclear how to interpret our estimates concretely. We take up this issue here using simulations, regressions, and a quantification of excess index volatility.

[31] Our estimation uses panels of varying lengths (from 5 to 252 trading days, with shorter streams becoming more frequent in recent years). Five pp return states are thus wide (relative to the shape of the distribution) for shorter-horizon panels. Defining states in terms of date-0 delta makes the states roughly equivalent in their share of the date-0 RN CDF.

6.1 Comparison to Simulated Option Pricing Models

A natural first question is whether leading option pricing models are capable of generating results close to the ones we find empirically. Many of these models feature pricing kernels that load on innovations to volatility, which can lead to violations of our CTI assumption as discussed in Section 4.1. While these models are typically not fully specified equilibrium models with explicit assumptions about agents' subjective beliefs, they can be interpreted as rational models with time-varying prices of risk.[32] We therefore conduct a set of simulations to test whether such existing models are capable of generating RN excess movement statistics similar to our estimates. We consider two leading models in these simulations.

First, we consider the GARCH model proposed by Christoffersen, Heston, and Jacobs (2013), or CHJ. This model features a variance-dependent pricing kernel, and it is able to match Stein's (1989) finding of apparent overreaction in long-maturity implied volatility. See Internet Appendix C.10 for a full description of the model. We conduct 10,000 simulations of this model using the baseline parameter values estimated by CHJ. Each simulation contains a burn-in period followed by 900 non-overlapping windows of three months (63 trading days) of simulated data. Each three-month window corresponds to a single option panel: at the beginning of a given window, there are $T = 63$ days to expiration (close to our sample mean $\overline{T}$), and we simulate spot prices every day to expiration ($t = T$), at which point the next panel starts.[33] Each day, we calculate index option prices using CHJ's solution for the spot-price characteristic function. We then construct RN beliefs over 5 pp excess return bins as in our empirical data, from which we calculate RN statistics. In each simulation, we estimate average RN excess movement and estimate $\overline{\phi}$ using our bound, and we then report the average estimates across simulations.

Second, we consider a stochastic volatility model in which volatility shocks are less directly tied to the index return than in CHJ. We use a model in the affine jump-diffusion class proposed by Duffie, Pan, and Singleton (2000), and in particular the version of their model with stochastic volatility with correlated jumps (SVCJ) in prices and volatility. We use estimates from Eraker, Johannes, and Polson (2003) for parameters under the physical distribution, and from Broadie, Chernov, and Johannes (2007) for parameters under the RN distribution; again see Internet Appendix C.10. The simulation setup and estimation otherwise parallel the description for the CHJ model above.[34]

We show results from the two sets of simulations in Table 4. The CHJ model produces 11% more movement than uncertainty reduction ($\overline{X}^*/\overline{u}_0^*$), far below the 123% noise-adjusted value reported in the first row of Table 1. The SVCJ model produces a significantly higher value of 27%, which is nonetheless still far below our empirical estimate.[35] The same is true for the daily excess

[32] In a previous version of this paper (Augenblick and Lazarus, 2023), we also reported simulation results for the Campbell and Cochrane (1999) habit model (see that version's Appendix C.4). We omit these here, as the simulated X^* statistics were low enough that the bound estimates were 1 (RRA $= 0$) for all interior states.

[33] We use 900 option panels per simulation because this is close to the number of expiration dates in our sample.

[34] We solve for prices here using code adapted from Nicola Fusari (Fusari, 2017), which we thank him for sharing.

[35] This implies that we find about 100% more excess variation in beliefs than is predicted by these models. While this may seem high, one possible point of comparison is provided by Giglio and Kelly (2018). They estimate variance ratio statistics of around 2 for many derivative term structures, relative to a set of affine models. A variance ratio of 2 implies

Table 4: Excess Movement Estimates for Simulated Option Pricing Models

			Bound Estimates		*True Values*	
	$\overline{X}^*/\overline{u}_0^*$	$\overline{X}^*/\overline{T}$	SDF Slope $\overline{\phi}$	RRA $\overline{\gamma}$	SDF Slope $\overline{\phi}_0$	RRA $\overline{\gamma}_0$
CHJ Model	0.11	0.0004	1.3	5.1	1.4	8.7
SVCJ Model	0.27	0.0010	1.7	13.7	1.6	12.4

Notes: All statistics are means across 10,000 simulations, with each simulation containing 900 option contracts of $T = \overline{T} = 63$ days with states defined as 5 pp excess return ranges. See the notes to Table 3 for details on the bound-based estimates of $\overline{\phi}$ and $\overline{\gamma}$. The true value $\overline{\phi}_0$ is estimated by calculating the ratio of RN to physical probabilities as of date 0 for each state pair satisfying $R_T \in \{\theta_j, \theta_{j+1}\}$, then averaging the value $\max_{\pi_{0,i}^*} \mathbb{E}[\phi_i \mid \pi_{0,i}^*]$ (from Proposition 6) across simulations. This is translated to $\overline{\gamma}_0$ using Proposition 4. See Internet Appendix C.10 for details on the models and simulations.

movement statistics $\overline{X}^*/\overline{T}$, and the SVCJ value of 0.0010 is again roughly 1/4 of the empirical value of 0.0038. For the bound estimates for the SDF slope and local relative risk aversion, both models produce enough RN excess movement to obtain non-negligible estimates for $\overline{\phi}$ and $\overline{\gamma}$.[36] Both models, however, produce estimates an order of magnitude lower than the empirical estimates in Table 3. Further, despite the violation of CTI in both models, the estimated bounds are still quite close to the models' true values of the maximal SDF slope (estimated from the ratio of RN to physical probabilities), as shown in the last two columns. The bound is slightly conservative in the CHJ model ($1.3 < 1.4$), and it only slightly overstates the true $\overline{\phi}_0$ in the SVCJ model ($1.7 > 1.6$).

Despite these models' ability to explain many empirical features of option prices and returns, they do not match the excess movement we document. Intuitively, the models are well designed to fit spot-price dynamics, features of option cross-sections (e.g., the IV smile), and average option returns. But our test exploits that option prices at different strikes are highly volatile (in the time series) relative to each other (in the cross-section of strikes); that is, there is predictable mean-reverting variation in the relative value of the RN density at different strikes, even in the interior of the distribution. It is this feature of prices that existing models have a difficult time matching.

6.2 Predictors of RN Excess Movement

The previous simulations indicate that it is difficult to rationalize our results with well-known models with time-varying prices of risk, but they do not help identify what ingredients a model would need in order to get closer to matching the data. So for our second analysis aimed at addressing how to interpret our results concretely, we conduct a set of regressions to determine whether there are known variables that comove strongly with observed RN excess movement. We provide details of the analysis in Internet Appendix C.11 and describe the results here.

Table C.4 shows results from time-series regressions in which the dependent variable is the monthly average of noise-adjusted RN excess movement $X^*_{t,t+1,i,j}$. We find that proxies for option illiquidity, trading activity, and intermediary constraints do comove positively with X^*, but only

about 100% more variation than under the null model, quite close to our estimates in a very different setting.

[36] Standard errors are very small; for example, for the SVCJ $\overline{\phi}$ estimate of 1.7, the simulation s.e. is 0.001.

very weakly so: the monthly average option bid-ask spread, option trading volume, and a measure of net put buying pressure are all insignificant as predictors of X^*, both statistically and economically. So while they help explain a small amount of excess movement (joint $R^2 = 0.08$), these results indicate that option-market frictions are unlikely to be the main drivers of our results.

We do, however, find evidence in favor of a few economically meaningful factors as predictors of X^*. First, the average length of RN belief streams (i.e., $\overline{T_i}$ for contracts traded in the given month) strongly positively predicts excess movement, consistent with the discussion in Section 5.3. There is thus a pattern of apparent overreaction to weak signals about outcomes at distant horizons, which lessens for stronger signals at shorter horizons.[37] Second, for volatility-related predictors, we find evidence that X^* comoves strongly with the quantity of market uncertainty (the VIX); slightly less strongly with the price of this uncertainty (the variance risk premium); and not at all with the volatility of risk aversion (using the risk-aversion proxy from Bekaert, Engstrom, and Xu, 2022), which can be thought of as a proxy for $\text{Var}(\phi_t)$. We thus find no evidence with this set of predictors for meaningful comovement between variation in the price of risk and RN excess movement. Third, we find that proxies for misvaluation and return reversals are significantly positively related to X^*. Together, the significant predictors explain 37% of the variation in X^*.

Taken together, these results suggest that RN excess movement is a meaningful and interpretable phenomenon: it is not significantly attributable to option-specific frictions, and it comoves strongly with variables that are intuitively related to aggregate equity valuations and excess volatility in expectations. While we find no evidence that variation in ϕ_t contributes substantially to X^*, it is of course possible that alternative models for time variation in ϕ_t are capable of generating significant excess movement, alongside alternative models of belief formation.

6.3 Quantifying the Effect on Overall Index Volatility

Finally, we conduct a back-of-the-envelope quantification to estimate how our results on RN beliefs translate to overall volatility in the underlying index itself. The challenge in doing so is that all our main results use binarized (i.e., conditional) beliefs over a pair of index outcomes, rather than the full distribution. The index, meanwhile, depends on the RN expectation over its future value, where the expectation is taken with respect to the full distribution.[38] The key advantage in using binarized beliefs is that we can restrict excess movement under the relatively weak assumption that the *local* price of risk over small index changes is constant. For the index itself, there may be a time-varying price or quantity of aggregate risk, including time-varying jump or variance risk.

To make some progress toward a quantification given these issues, we take two steps. First, we restrict attention to the center of the return distribution. For date t and option expiration T, we construct an alternative, *tail-free*, index value $S_{t,\text{alt}} = \mathbb{E}_t^*[S_T \,|\underline{S} \leqslant S_T \leqslant \overline{S}]$, where $\underline{S}$ is the lowest

[37] In a follow-up paper (Augenblick, Lazarus, and Thaler, 2025), we investigate this pattern in much greater detail, showing that it holds as well in sports betting markets and in controlled experiments. See Section 7 for further discussion.

[38] That is, $S_t = \mathbb{E}_t^*[S_T]$. Note that since we ignore the risk-free rate (other than for one definition below) and dividends, the "index" here is really the futures price for the dividend-stripped index. Given the horizon ($\overline{T} < 3$ months on average) and use of daily changes, dividend uncertainty and risk-free rate variation are unlikely to play a major role.

possible index value (corresponding to return $R_T = S_T/S_0 = \underline{\theta}$) and $\overline{S}$ is the highest ($\overline{\theta}$). This can be thought of as the price of a conditional forward contract with no tail exposure: the buyer gets net payoff $S_T - S_{t,\text{alt}}$ at T if $\underline{S} \leqslant S_T \leqslant \overline{S}$, and 0 otherwise. Extending the logic used for our main tests, excluding the tails makes us more comfortable assuming that the price of risk is constant and reasonably small, which we will use to get a rough estimate of excess index movement. We consider a range of different tail cutoffs (from $\pm 10\%$ down to $\pm 4\%$). For each cutoff, we use the option-implied histogram over 2 pp return bins to compute the no-arbitrage price $S_{t,\text{alt}}$ for each trading date t and expiration T in our sample.[39]

Next, under the assumption that $S_{t,\text{alt}}$ is a square-integrable martingale under RE for a sufficiently narrow range $[\underline{S}, \overline{S}]$, we can extend the standard martingale restriction in Lemma 1 to the expectation $S_{t,\text{alt}}$. Each option panel gives us an implied index stream from 0 to T, for which we define movement as $\mathsf{m}(S_{\text{alt}}) = \sum_{t=0}^{T-1}(S_{t+1,\text{alt}} - S_{t,\text{alt}})^2$. We also use the RN beliefs to calculate ex ante implied RN variance: $\mathsf{u}_0(S_{\text{alt}}) = \sum_{\underline{S} \leqslant s \leqslant \overline{S}} \mathbb{P}_0^*(S_T = s | \underline{S} \leqslant S_T \leqslant \overline{S}) \cdot (s - S_{0,\text{alt}})^2$. Excess movement is again $X(S_{\text{alt}}) = \mathsf{m}(S_{\text{alt}}) - \mathsf{u}_0(S_{\text{alt}})$, and the martingale restriction is that $\mathbb{E}[X(S_{\text{alt}})] = 0$. We measure the sample mean $\overline{X}(S_{\text{alt}})$ using our S&P options data. As in Table 1, we normalize this by the mean $\overline{\mathsf{u}}_0(S_{\text{alt}})$ to obtain a measure interpretable as the percent excess movement in the value of the index. To equalize the scale of $\mathsf{m}(S_{\text{alt}})$ and $\mathsf{u}_0(S_{\text{alt}})$ across option panels, we normalize each starting price to $S_{0,\text{alt}} = 1/R^f_{0,T}$ (so that all statistics are effectively in excess-return space, and $\underline{S} = \underline{\theta}$, $\overline{S} = \overline{\theta}$).

Note that $X(S_{\text{alt}})$ is equivalent to the payoff on an implementable variance swap contract: m is equal to realized variance, and u_0 is the swap strike (implied variance). We construct the contract and measure its payoff using the RN distribution, and the underlying has been constructed to remove tail exposure. These distinctions aside, our measurement otherwise parallels the approach used in past work for different types of variance swaps (e.g., Carr and Wu, 2009; Martin, 2017).

Moving now to our results, Table 5 shows the option-implied excess-movement statistics $\overline{X}(S_{\text{alt}})/\overline{\mathsf{u}}_0(S_{\text{alt}})$ for different choices of the central return range $[\underline{\theta}, \overline{\theta}]$ used to construct the tail-free index $S_{t,\text{alt}}$. The first entry shows that if we do not exclude the tails at all (setting $S_{t,\text{alt}} = S_t$), we estimate 14% too little index variance. This is unsurprising: it is a restatement of the fact that there is a positive variance risk premium (VRP), measured as average implied minus realized variance (i.e., the average return for the short side of a variance swap, $-\overline{X}$). It is reassuring that our measurement approach using the RN distribution finds evidence for the VRP consistent with past work.

In contrast, the remaining entries show that once the tails are excluded, we estimate significantly too much movement in the index, ranging from 15% to 45% as we narrow the part of the distribution used to construct $S_{t,\text{alt}}$. The tails are evidently the key contributor to the VRP; once they are removed, we observe realized variance in excess of initial uncertainty. So our rough estimate is that the (tail-hedged) S&P index exhibits on the order of 30% more variance than it would under RE. We cannot say for certain how much excess variation there is when including the tails: there, our estimate of $\overline{X}$ includes both excess variance *and* the VRP, and the latter dominates. But if we assume that the

[39] That is, $S_{t,\text{alt}} = \sum_{\underline{S} \leqslant s \leqslant \overline{S}} \mathbb{P}_t^*(S_T = s | \underline{S} \leqslant S_T \leqslant \overline{S}) \cdot s$. We set each s to the midpoint of the associated return bin; see Internet Appendix C.12 for details. We use 2 pp bins so that the midpoint approximates the expected value in that bin.

Table 5: Option-Implied Excess Index Variance

		Net Excess-Return Range $[\underline{\theta}, \overline{\theta}]$ *Used to Define the Tail-Free Index* $S_{t,alt}$			
	Full ($S_{t,\text{alt}} = S_t$)	$[-10\%, 10\%]$	$[-8\%, 8\%]$	$[-6\%, 6\%]$	$[-4\%, 4\%]$
$\overline{X}(S_{\text{alt}})/\overline{\mathsf{u}}_0(S_{\text{alt}})$	-0.14	0.15	0.19	0.33	0.45
[Bootstrapped SE]	[0.03]	[0.04]	[0.04]	[0.04]	[0.06]

Notes: Each entry shows average excess index movement normalized by initial uncertainty (implied variance). We construct the index $S_{t,\text{alt}}$ (and implied variance) for forward horizon T using the option-based RN distribution confined to the range $[\underline{\theta}, \overline{\theta}]$, and the first column uses the full distribution. For each T, movement is realized variance for $S_{t,\text{alt}}$ from 0 to T. Standard errors use a block bootstrap for the statistic $\overline{X}(S_{\text{alt}})/\overline{\mathsf{u}}_0(S_{\text{alt}})$, with a block size of one month and 10,000 draws. See Internet Appendix C.12 for details.

excess variance is the same in the tails as it is in the center of the distribution, this implies that the S&P overall has roughly 30% more variance than it would under a martingale-belief benchmark.[40]

The value $\overline{X}/\overline{\mathsf{u}}_0$ can be thought of as a variance ratio statistic minus one ($VR = \overline{\mathsf{m}}/\overline{\mathsf{u}}_0$). We can also normalize excess movement by observed movement, $\overline{X}/\overline{\mathsf{m}}$, to estimate what share of the observed index variance $\overline{\mathsf{m}}$ is excessive. This results in estimates ranging from $\overline{X}/\overline{\mathsf{m}} = 0.13$ (s.e. 0.03) for the $\pm 10\%$ window, to 0.31 (s.e. 0.03) for the $\pm 4\%$ window. Even the 13% estimate is sizable economically. It implies that the average annualized S&P return volatility of about $\sigma = 20\%$ would be reduced by $\sqrt{13\% \times (20\%)^2} = 7$ percentage points if our null were to hold. That is, close to 1/3 of the observed index volatility is excessive. This is of course a rough first-pass quantification under a set of assumptions. But it lends some support to the Shiller (1981) excess volatility hypothesis, using an approach that sidesteps the need to measure the index's fundamental value directly.

7. Conclusion

We derive new bounds on the admissible rational variation in risk-neutral beliefs as expressed in asset prices. Unlike in much of the previous literature, these results do not require any restrictions on the data-generating process, and they allow for meaningful time variation in discount rates. Further, by using asset prices, 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 risk-neutral beliefs over the return on the S&P 500 index, we find that these RN beliefs are so volatile that our bounds are routinely violated. This implies a violation of our joint assumptions. We provide evidence suggesting that RE violations by the marginal investor are likely to be at least partly responsible for this violation, though we remain open to alternative forms of variation in the local price of risk playing a part. But our framework allows

[40] If we instead assume that the tails exhibit *no* excess variance, this matters less than one might think: we estimate that 60–70% of overall index variance arises from the middle of the distribution (e.g. $\overline{\mathsf{m}} = 0.04$ for $\pm 10\%$, versus $\overline{\mathsf{m}} = 0.055$ for the full index). So if the tails have no excess variance, then index-level excess variance is roughly $(2/3) \times 30\% \approx 20\%$.

us to rule out other explanations for the excess volatility we document, providing a meaningful improvement in understanding the restrictions implied by the observed volatility in prices. Further, our results are not well-explained by leading option pricing models, and we show that they imply meaningful excess variation in the underlying index.

In follow-up work (Augenblick, Lazarus, and Thaler, 2025), we more directly consider what alternative models might better fit the data. We show that a model featuring overreaction to weak signals, which weakens and then reverses in response to sufficiently strong signals — as raised briefly in Section 6.2 — explains empirical patterns not just in option markets, but also in sports betting markets and in a range of controlled experiments. The fact that we find positive unconditional excess movement in the current paper suggests, according to this explanation, that day-to-day information is fairly uninformative about future index outcomes at the aggregate level. Applying such a model quantitatively to option prices, however, remains to be done.

We believe that there are numerous feasible ways to make additional progress in identifying the specific causes of our bound violations. Conducting additional tests on the empirical correlates of excess movement, for example, may provide useful information, as would generalizing our analysis to alternative asset classes. Finally, 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.

Appendix A. Proofs of Main Results

This appendix contains proofs for the main results in Sections 2–3. The remaining proofs can be found in the Internet Appendix, along with additional technical materials and discussions.

A.1. Proofs for Section 2

See Internet Appendix B.1 for a rederivation of Lemma 1. Some preliminaries are useful before proceeding to the main proofs. Start by defining the *RN measure* for this section as

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

where $\mathbb{P}(H_T)$ is the probability of observing history H_T under DGP. This $\mathbb{P}^*$ is a change of measure that adjusts the frequency of each signal history such that a person with RN beliefs has RE. (As shown in Lemma A.3, (A.1) is consistent with the usual definition in a general setting.)

Lemma A.1. *Define $\mathbb{E}^*[\cdot]$ to be the expectation under $\mathbb{P}^*$. Under Assumption 1 (RE):*

(i) For any H_T and θ, $\mathbb{P}^*(H_T|\theta) = \mathbb{P}(H_T|\theta)$.

(ii) For any H_t, $\pi_t^*(H_t) = \mathbb{E}^*[\pi_{t+1}^*(H_{t+1})|H_t]$.

(iii) For any DGP, $\mathbb{E}^*[X^*] = 0$.

Proof of Lemma A.1. For the physical measure,

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

where the second line uses that $\pi_0 = \mathbb{E}[\theta] = \mathbb{P}(\theta=1)$ by RE. For the RN measure, (A.1)–(A.2) give

$$\begin{aligned}\mathbb{P}^*(H_T) &= \frac{\pi_0^*}{\pi_0}\cdot\pi_0\cdot\mathbb{P}(H_T|\theta=1)+\frac{1-\pi_0^*}{1-\pi_0}\cdot(1-\pi_0)\cdot\mathbb{P}(H_T|\theta=0)\\ &= \pi_0^*\cdot\mathbb{P}(H_T|\theta=1)+(1-\pi_0^*)\cdot\mathbb{P}(H_T|\theta=0).\end{aligned}\tag{A.3}$$

For any H_T such that $\pi_T = 1$, (A.1) gives that $\mathbb{P}^*(H_T) = \frac{\pi_0^*}{\pi_0}\mathbb{P}(H_T)$, so $\mathbb{P}^*(\theta=1) = \frac{\pi_0^*}{\pi_0}\mathbb{P}(\theta=1)$. Thus from the definition of conditional probability, $\mathbb{P}^*(H_T|\theta=1) = \mathbb{P}(H_T|\theta=1)$. Similarly, $\mathbb{P}^*(H_T|\theta=0) = \mathbb{P}(H_T|\theta=0)$, proving part (i).

Given this, (A.3) 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 H_T for which $\theta=1$ gives $\pi_0^* = \mathbb{P}^*(\theta=1)$, so $\mathbb{P}^*$ is a valid distribution for which the law of iterated expectations holds. Part (ii) then follows from $\mathbb{P}^*(\theta=1) = \mathbb{E}^*[\theta] = \mathbb{E}^*[\pi_T] = \mathbb{E}^*[\pi_T^*]$.

Finally, given (ii), the same proof for Lemma 1 (see Appendix B.1) gives $\mathbb{E}^*[X^*]=0$, as expected RN movement under the RN measure must equal RN initial uncertainty. This proves part (iii). □

Part (i) says that 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. This implies that conditional expectations under the two respective measures are equal. Therefore, $\mathbb{E}^*[X^*|\theta] = \mathbb{E}[X^*|\theta]$ (see Lemma A.4 below for the analogue of this result in the general asset-pricing setting). This implies

$$\begin{aligned}\mathbb{E}^*[X^*] &= \pi_0^*\cdot\mathbb{E}^*[X^*|\theta=1]+(1-\pi_0^*)\cdot\mathbb{E}^*[X^*|\theta=0]\\ &= \pi_0^*\cdot\mathbb{E}[X^*|\theta=1]+(1-\pi_0^*)\cdot\mathbb{E}[X^*|\theta=0]=0,\end{aligned}\tag{A.4}$$

where the last equality applies part (iii) of the lemma. For $\mathbb{E}[X^*]$, it is useful to similarly write

$$\mathbb{E}[X^*] = \pi_0\cdot\mathbb{E}[X^*|\theta=1]+(1-\pi_0)\cdot\mathbb{E}[X^*|\theta=0].\tag{A.5}$$

Now, recall that $\triangle \equiv \mathbb{E}[X^*|\theta=0]-\mathbb{E}[X^*|\theta=1]$. It will be useful to bound $\triangle$:

LEMMA A.2. *For any DGP, $\triangle \leqslant \pi_0^*$.*

Proof of Lemma A.2. First, we can write $\triangle \equiv \mathbb{E}^*[X^*|\theta=0]-\mathbb{E}^*[X^*|\theta=1]$ from Lemma A.1, or

$$\triangle = \mathbb{E}^*[m^*|\theta=0]-u_0^*-(\mathbb{E}^*[m^*|\theta=0]-u_0^*) = \mathbb{E}^*[m^*|\theta=0]-\mathbb{E}^*[m^*|\theta=1].\tag{A.6}$$

Further, using (A.4),

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

so from the definition of u_0^*,

$$\pi_0^* \cdot \mathbb{E}[\mathsf{m}^*|\theta = 1] + (1 - \pi_0^*) \cdot \mathbb{E}[\mathsf{m}^*|\theta = 0] = \pi_0^*(1 - \pi_0^*). \tag{A.7}$$

Solving for $\mathbb{E}[\mathsf{m}^*|\theta = 0]$ gives $\mathbb{E}[\mathsf{m}^*|\theta = 0] = \pi_0^* - \frac{\pi_0^*}{1-\pi_0^*} \cdot \mathbb{E}[\mathsf{m}^*|\theta = 1]$. Using this in (A.6),

$$\triangle = \pi_0^* - \frac{\pi_0^*}{1 - \pi_0^*} \cdot \mathbb{E}[\mathsf{m}^*|\theta = 1] - \mathbb{E}^*[\mathsf{m}^*|\theta = 1] = \pi_0^* - \frac{1}{1 - \pi_0^*} \cdot \mathbb{E}[\mathsf{m}^*|\theta = 1]. \tag{A.8}$$

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 1. Start from equation (A.5) and apply equation (A.4):

$$\begin{aligned} \mathbb{E}[X^*] &= \pi_0 \cdot \mathbb{E}[X^*|\theta = 1] + (1 - \pi_0) \cdot \mathbb{E}[X^*|\theta = 0] - 0 \\ &= \pi_0 \cdot \mathbb{E}[X^*|\theta = 1] + (1 - \pi_0) \cdot \mathbb{E}[X^*|\theta = 0] - (\pi_0^* \cdot \mathbb{E}[X^*|\theta = 1] + (1 - \pi_0^*) \cdot \mathbb{E}[X^*|\theta = 0]) \\ &= (\pi_0^* - \pi_0)(\mathbb{E}[X^*|\theta = 0] - \mathbb{E}[X^*|\theta = 1]) = (\pi_0^* - \pi_0)\triangle, \end{aligned} \tag{A.9}$$

as stated. Then the second equality holds using equation (8) and the definition of $\triangle$. □

Proof of Proposition 2. From Lemma A.2 above, we have $\triangle \leqslant \pi_0^*$. Further, equation (8) implies

$$\pi_0^* - \pi_0 = \pi_0^* - \frac{\pi_0^*}{\pi_0^* + \phi(1 - \pi_0^*)} = \pi_0^* \left(1 - \frac{1}{\pi_0^* + \phi(1 - \pi_0^*)}\right) \geqslant 0, \tag{A.10}$$

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.9),

$$\mathbb{E}[X^*] = (\pi_0^* - \pi_0)\triangle \leqslant (\pi_0^* - \pi_0)\pi_0^*. \tag{A.11}$$

Plugging in the expression for $\pi_0^* - \pi_0$ in (A.10) then gives equation (10). □

Proof of Corollary 1. This is an immediate implication of (A.11) and $\pi_0 \geqslant 0$. □

A.2. Proofs for Section 3

Again see Internet Appendix B.1 for a derivation of equation (13). A pair of preliminary lemmas will also be useful in proving this section's main results. As usual, assume throughout that Assumptions 2–4 hold.

LEMMA A.3. *For some return-state pair (θ_j, θ_{j+1}), with $\widetilde{\mathbb{P}} \equiv \mathbb{P}(\cdot \mid R_T \in \{\theta_j, \theta_{j+1}\})$, define a new pseudo-risk-neutral measure $\widetilde{\mathbb{P}}^\diamond$ by*

$$\left.\frac{d\widetilde{\mathbb{P}}^\diamond}{d\widetilde{\mathbb{P}}}\right|_{H_t} = \frac{\widetilde{\pi}^*_{t,j}}{\widetilde{\pi}_{t,j}} \mathbb{1}\{R_T = \theta_j\} + \frac{1-\widetilde{\pi}^*_{t,j}}{1-\widetilde{\pi}_{t,j}} \mathbb{1}\{R_T = \theta_{j+1}\}. \tag{A.12}$$

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

$$\widetilde{\pi}^*_{t,j} = \widetilde{\mathbb{E}}^\diamond_t[\widetilde{\pi}^*_{t+1,j}]. \tag{A.13}$$

We conclude from Lemma 1 that

$$\widetilde{\mathbb{E}}^\diamond_0[X^*_j] = 0. \tag{A.14}$$

Proof of Lemma A.3. Following the discussion after equation (14), we have that

$$\frac{\widetilde{\pi}^*_{t,j}}{\widetilde{\pi}_{t,j}} = \frac{\phi_j}{1+\widetilde{\pi}_{t,j}(\phi_j - 1)}, \qquad \frac{1-\widetilde{\pi}^*_{t,j}}{1-\widetilde{\pi}_{t,j}} = \frac{1}{1+\widetilde{\pi}_{t,j}(\phi_j - 1)}. \tag{A.15}$$

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 $\{s_t\}$, with signals $s_t \in \mathcal{S}$. Denote $N_\mathcal{S} \equiv |\mathcal{S}|$, so $s_t \in \{s_1, s_2, \ldots, s_{N_\mathcal{S}}\}$, and further denote $\mathfrak{p}_{t,k} \equiv \widetilde{\mathbb{P}}_t(s_{t+1} = \theta_k)$, $\varrho_{t,k} \equiv \widetilde{\mathbb{P}}_t(R_T = \theta_j \mid s_{t+1} = s_k)$, and $\varrho^*_{t,k} \equiv \mathbb{P}^*_t(R_T = \theta_j \mid s_{t+1} = s_k, R_T \in \{\theta_j, \theta_{j+1}\})$, so that $\widetilde{\pi}_{t+1,j} = \varrho_{t,k}$ and $\widetilde{\pi}^*_{t+1,j} = \varrho^*_{t,k}$ if $s_{t+1} = s_k$. Combining (A.12), (A.15), and these definitions:

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

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 (LIE) given that $\widetilde{\pi}_{t,j} = \mathbb{E}_t[\mathbb{1}\{R_T = \theta_j\} \mid R_T \in \{\theta_j, \theta_{j+1}\}] = \widetilde{\mathbb{E}}_t[\mathbb{1}\{R_T = \theta_j\}] = \widetilde{\mathbb{E}}_t[\widetilde{\mathbb{E}}_{t+1}[\mathbb{1}\{R_T = \theta_j\}]] = \widetilde{\mathbb{E}}_t[\widetilde{\pi}_{t+1,j}]$, and the last equality above again uses (A.15). Then $\widetilde{\mathbb{E}}_0^\diamond[X_j^*] = 0$ follows immediately from the proof of Lemma 1. □

Lemma A.4. *For any return-state pair (θ_j, θ_{j+1}) meeting CTI, for $j' = j, j+1$, RN movement must satisfy*

$$\widetilde{\mathbb{E}}_0^\diamond[\mathsf{m}_j^* \mid R_T = \theta_{j'}] = \widetilde{\mathbb{E}}_0[\mathsf{m}_j^* \mid R_T = \theta_{j'}]. \tag{A.16}$$

Proof of Lemma A.4. The stream of RN beliefs is $\boldsymbol{\pi}_j^*$, and denote some arbitrary realization for that path by $\mathfrak{b}_j$. For any $\mathfrak{b}_j$ such that $\widetilde{\pi}_{T,j}^* = 1$ (i.e., $R_T = \theta_j$), the definition of $\widetilde{\mathbb{P}}^\diamond$ in (A.12) gives that

$$\widetilde{\mathbb{P}}_0^\diamond(\boldsymbol{\pi}_j^* = \mathfrak{b}_j) = \frac{\widetilde{\pi}_{0,j}^*}{\widetilde{\pi}_{0,j}} \widetilde{\mathbb{P}}(\boldsymbol{\pi}_j^* = \mathfrak{b}_j), \tag{A.17}$$

and further $\widetilde{\mathbb{P}}_0^\diamond(R_T = \theta_j) = (\widetilde{\pi}_{0,j}^*/\widetilde{\pi}_{0,j})\,\widetilde{\mathbb{P}}_0(R_T = \theta_j)$ trivially. Combining these two equations yields $\widetilde{\mathbb{P}}_0^\diamond(\boldsymbol{\pi}_j^* = \mathfrak{b}_j \mid R_T = \theta_j) = \widetilde{\mathbb{P}}_0(\boldsymbol{\pi}_j^* = \mathfrak{b}_j \mid R_T = \theta_j)$. (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 $R_T = \theta_j$ are preserved, and similarly for $R_T = \theta_{j+1}$, as was the case for the simpler version in (A.1).) Thus

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

The same applies for $R_T = \theta_{j+1}$: for any $\mathfrak{b}_j$ such that $\widetilde{\pi}_{T,j}^* = 0$, (A.17) now becomes $\widetilde{\mathbb{P}}_0^\diamond(\boldsymbol{\pi}_j^* = \mathfrak{b}_j) = (1 - \widetilde{\pi}_{0,j}^*)/(1 - \widetilde{\pi}_{0,j})\,\widetilde{\mathbb{P}}(\boldsymbol{\pi}_j^* = \mathfrak{b}_j)$. Further, $\widetilde{\mathbb{P}}_0^\diamond(R_T = \theta_{j+1}) = (1 - \widetilde{\pi}_{0,j}^*)/(1 - \widetilde{\pi}_{0,j})\,\widetilde{\mathbb{P}}_0(R_T = \theta_{j+1})$, so again $\widetilde{\mathbb{P}}_0^\diamond(\boldsymbol{\pi}_j^* = \mathfrak{b}_j \mid R_T = \theta_{j+1}) = \widetilde{\mathbb{P}}_0(\boldsymbol{\pi}_j^* = \mathfrak{b}_j \mid R_T = \theta_{j+1})$. Thus $\widetilde{\mathbb{E}}_0^\diamond[\mathsf{m}_j^* \mid R_T = \theta_{j+1}] = \widetilde{\mathbb{E}}_0[\mathsf{m}_j^* \mid R_T = \theta_{j+1}]$. □

Note that definition (A.12) aligns with the measure defined in (A.1), so Lemmas A.3–A.4 prove the statements in the text connecting the RN measure in Section 2 to the general case in Section 3 (see after (A.1) and Lemma A.1(i)). Indeed, (A.13) is the precise analogue to Lemma A.1(i); (A.14) is the analogue to Lemma A.1(iii); and (A.16) gives immediately that $\widetilde{\mathbb{E}}_0^\diamond[X_j^* \mid R_T] = \widetilde{\mathbb{E}}_0[X_j^* \mid R_T]$, which was the main implication of Lemma A.1(i) used in deriving the results in Section 2. We will thus be able to apply those results in this case using the above two lemmas, as follows.

Proof of Proposition 3. No arbitrage gives the existence of a positive SDF for which (14) and

Assumption 3 are valid. We have

$$\widetilde{\pi}_{t,j} = \mathbb{E}_t[\widetilde{\pi}_{t+1,j}], \qquad \widetilde{\pi}^*_{t,j} = \widetilde{\mathbb{E}}^\diamond_t[\widetilde{\pi}^*_{t+1,j}],$$
$$\widetilde{\mathbb{E}}^\diamond_0[X^*_j] = 0, \qquad \widetilde{\mathbb{E}}^\diamond_0[X^*_j \mid R_T] = \widetilde{\mathbb{E}}_0[X^*_j \mid R_T],$$

where the first equality uses LIE and the rest use Lemmas A.3–A.4. The last equation implies, using the argument applied in Lemma A.2, that $\triangle_j \leqslant \widetilde{\pi}^*_{0,j}$. Further, (5)–(8) hold for $\widetilde{\pi}_{t,j}, \widetilde{\pi}^*_{t,j}, \phi_j$. We have thus obtained all the conditions used for Lemma 1, Propositions 1–2, and Corollary 1, and thus 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 $\triangle_j \equiv \widetilde{\mathbb{E}}_0[X^*_j \mid R_T = \theta_{j+1}] - \widetilde{\mathbb{E}}_0[X^*_j \mid R_T = \theta_j]$ replacing $\triangle$, as stated. □

Proof of Proposition 4. The result follows immediately from equation (7), with S_j and S_{j+1} replacing $C_{T,1}$ and $C_{T,0}$, respectively. □

References

AÏT-SAHALIA, Y., Y. WANG, AND F. YARED (2001): "Do Option Markets Correctly Price the Probabilities of Movement of the Underlying Asset?" *Journal of Econometrics*, 102, 67–110.

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 E. LAZARUS (2022): "Restrictions on Asset-Price Movements Under Rational Expectations: Theory and Evidence," *Working Paper*.

——— (2023): "A New Test of Excess Movement in Asset Prices," *Working Paper*.

AUGENBLICK, N., E. LAZARUS, AND M. THALER (2025): "Overinference from Weak Signals and Underinference from Strong Signals," *Quarterly Journal of Economics*, 140, 335–401.

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

BACKUS, D., M. CHERNOV, AND I. MARTIN (2011): "Disasters Implied by Equity Index Options," *Journal of Finance*, 66, 1969–2012.

BANSAL, R. AND A. YARON (2004): "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59, 1481–1509.

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.

BASAK, S. (2000): "A Model of Dynamic Equilibrium Asset Pricing With Heterogeneous Beliefs and Extraneous Risk," *Journal of Economic Dynamics & Control*, 24, 63–95.

BEASON, T. AND D. SCHREINDORFER (2022): "Dissecting the Equity Premium," *Journal of Political Economy*, 130, 2203–2222.

BEKAERT, G., E. C. ENGSTROM, AND N. R. XU (2022): "The Time Variation in Risk Appetite and Uncertainty," *Management Science*, 68, 3975–4004.

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

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.

BROADIE, M., M. CHERNOV, AND M. JOHANNES (2007): "Model Specification and Risk Premia: Evidence from Futures Options," *Journal of Finance*, 62, 1453–1490.

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.

CARR, P. AND L. WU (2009): "Variance Risk Premiums," *Review of Financial Studies*, 22, 1311–1341.

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.

COLLIN-DUFRESNE, P. AND V. FOS (2016): "Insider Trading, Stochastic Liquidity, and Equilibrium Prices," *Econometrica*, 84, 1441–1475.

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

DRIESSEN, J., J. KOËTER, AND O. WILMS (2024): "Horizon Effects in the Pricing Kernel: How Investors Price Short-Term Versus Long-Term Risks," *Working Paper*.

DUFFIE, D., J. PAN, AND K. J. SINGLETON (2000): "Transform Analysis and Asset Pricing for Affine Jump-Diffusions," *Econometrica*, 68, 1343–1376.

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.

——— (1991): "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis," *Journal of Political Economy*, 99, 263–286.

ERAKER, B., M. JOHANNES, AND N. POLSON (2003): "The Impact of Jumps in Volatility and Returns," *Journal of Finance*, 58, 1269–1300.

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

FIGLEWSKI, S. (2010): "Estimating the Implied Risk-Neutral Density for the US Market Portfolio," in *Volatility and Time Series Econometrics: Essays in Honor of Robert Engle*, ed. by T. Bollerslev, J. R. Russell, and M. W. Watson, Oxford: Oxford University Press, chap. 15, 323–353.

FUSARI, N. (2017): *Matlab Toolbox for Option Pricing*.

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

GANDHI, M., N. J. GORMSEN, AND E. LAZARUS (2025): "Forward Return Expectations," *NBER Working Paper 31687*.

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.

HADDAD, V., A. MOREIRA, AND T. MUIR (2024): "Whatever It Takes? The Impact of Conditional Policy Promises," *Forthcoming, American Economic Review*.

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

JACKWERTH, J. C. (2000): "Recovering Risk Aversion from Option Prices and Realized Returns," *Review of Financial Studies*, 13, 433–451.

KELLY, B., S. MALAMUD, E. SIRIWARDANE, AND H. WU (2024): "Behavioral Impulse Responses," *Working Paper*.

KYLE, A. S. (1985): "Continuous Auctions and Insider Trading," *Econometrica*, 53, 1315–1335.

LAZARUS, E. (2022): "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 (2022): "A ReMeDI for Microstructure Noise," *Econometrica*, 90, 367–389.

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. (2013): "Consumption-Based Asset Pricing with Higher Cumulants," *Review of Economic Studies*, 80, 745–773.

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

NEUBERGER, A. (2012): "Realized Skewness," *Review of Financial Studies*, 25, 3424–3455.

NORDHAUS, W. D. (1987): "Forecasting Efficiency: Concepts and Applications," *Review of Economics and Statistics*, 69, 667–674.

POLKOVNICHENKO, V. AND F. ZHAO (2013): "Probability Weighting Functions Implied in Options Prices," *Journal of Financial Economics*, 107, 580–609.

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

SCHREINDORFER, D. AND T. SICHERT (2023): "Volatility and the Pricing Kernel," *Working Paper*.

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.

WEST, K. D. (1988): "Dividend Innovations and Stock Price Volatility," *Econometrica*, 56, 37–61.

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.