# *Internet Appendix:*

# Excess Movement in Option-Implied Beliefs*

Ned Augenblick and Eben Lazarus

DECEMBER 2024

## Contents



*Contact: ned@haas.berkeley.edu and lazarus@berkeley.edu.

# Appendix B. Additional Derivations and Proofs of Theoretical Results

The proofs for Propositions 1–4 are provided in the main paper in Appendix A. We provide proofs for the remaining theoretical statements here.

## B.1 Additional Proofs for Sections 2–3

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

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

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

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

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

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

where the last line uses Assumption 1. Summing and applying the law of iterated expectations (LIE),

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

***Proof of Equation (13).*** This follows from a discrete-state application of Breeden and Litzenberger (1978), or see Brown and Ross (1991) for a general version. To review why the stated equation holds, the risk-neutral pricing equation for options can be written

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

This implies that for two adjacent return states $\theta_{j-1}$ and $\theta_j$,

$$q_{t,K_j} - q_{t,K_{j-1}} = \frac{1}{R^f_{t,T}} \left[ \sum_{j' \geqslant j} (K_{j'} - K_j)\, \mathbb{P}^*_t(S_T = K_{j'}) - \sum_{j' \geqslant j-1} (K_{j'} - K_{j-1})\, \mathbb{P}^*_t(S_T = K_{j'}) \right]$$

$$= \frac{1}{R^f_{t,T}} \left[ \sum_{j' \geqslant j} (K_{j-1} - K_j)\, \mathbb{P}^*_t(S_T = K_{j'}) \right] = \frac{1}{R^f_{t,T}} (K_{j-1} - K_j) \left[1 - \mathbb{P}^*_t(S_T < K_j)\right].$$

Rearranging,

$$R^f_{t,T} \frac{q_{t,K_j} - q_{t,K_{j-1}}}{K_j - K_{j-1}} = \mathbb{P}^*_t(S_T < K_j) - 1.$$

Repeating this for $\theta_j$ and $\theta_{j+1}$, we obtain $R^f_{t,T} \frac{q_{t,K_{j+1}} - q_{t,K_j}}{K_{j+1} - K_j} = \mathbb{P}^*_t(S_T < K_{j+1}) - 1$. Subtracting the preceding equation from this equation and using $\mathbb{P}^*_t(R_T = \theta_j) = \mathbb{P}^*_t(S_T = K_j)$ yields (13). □

## B.2 Proofs for Section 4

***Proof of Statements in Section 4.1.*** As in footnote 16 in the main text, statements 1–2 are immediate given the definition of CTI. We take the remaining statements in order:

3. The Gabaix (2012) economy features a representative agent with CRRA consumption utility, and log consumption and log dividends follow $c_{t+1} = c_t + g_c + \varepsilon^c_{t+1} + \log(B_{t+1})\mathfrak{D}_{t+1}$ and $d_{t+1} = d_t + g_d + \varepsilon^d_{t+1} + \log(F_{t+1})\mathfrak{D}_{t+1}$, respectively, where $\mathfrak{D}_{t+1} = \mathbb{1}\{\text{disaster}_{t+1}\}$; disasters in $t+1$ occur with probability $p_t$; $B_{t+1}$ and $F_{t+1}$ are possibly correlated variables with support $[0,1]$; and $(\varepsilon^c_{t+1}, \varepsilon^d_{t+1})'$ is i.i.d. bivariate normal (or a discretized approximation thereof) with mean zero and is independent of all disaster-related variables. *Resilience* is $\mathsf{H}_t = p_t \mathbb{E}_t[B^{-\gamma}_{t+1} F_{t+1} - 1 \mid \mathfrak{D}_{t+1}]$, and write $\mathsf{H}_t = \mathsf{H}_* + \widehat{\mathsf{H}}_t$. The dynamics of $p_t$ are governed by $\widehat{\mathsf{H}}_{t+1} = \frac{1+\mathsf{H}_*}{1+\mathsf{H}_t} e^{-\phi_\mathsf{H}} \widehat{\mathsf{H}}_t + \varepsilon^\mathsf{H}_{t+1}$, where $\varepsilon^\mathsf{H}_{t+1}$ is mean-zero and independent of all other shocks. Gabaix (2012, Theorem 1) shows that $S_t = \frac{D_t}{1-e^{-\beta_m}} \left(1 + \frac{e^{-\beta_m - h_*} \widehat{\mathsf{H}}_t}{1 - e^{-\beta_m - \phi_H}}\right)$, where $h_* \equiv \log(1 + \mathsf{H}_*)$ and $\beta_m \equiv -\log\beta + \gamma g_c - g_d - h_*$. Thus for any $\theta$ and $\mathsf{H}_0$, there exists some value $d_\theta$ and function $f(d_\theta, \widehat{\mathsf{H}}_T)$, which is strictly increasing in $d_\theta$ and strictly decreasing in $\widehat{\mathsf{H}}_T$, such that, by Bayes' rule,

$$\mathbb{P}_0\left( \left( \sum_{t=1}^T \mathfrak{D}_t \right) > 0 \,\middle|\, R_T \geqslant \theta \right) = \frac{\mathbb{P}_0\Big( R_T \geqslant \theta \mid \sum_{t=1}^T \mathfrak{D}_t > 0 \Big)\, \mathbb{P}_0\Big( \sum_{t=1}^T \mathfrak{D}_t > 0 \Big)}{\mathbb{P}_0(R_T \geqslant \theta)}$$
$$= \frac{\mathbb{P}_0\Big( D_T \geqslant f(d_\theta, \widehat{\mathsf{H}}_T) \,\Big|\, \sum_{t=1}^T \mathfrak{D}_t > 0 \Big)\, \mathbb{P}_0\Big( \sum_{t=1}^T \mathfrak{D}_t > 0 \Big)}{\mathbb{P}_0\Big( D_T \geqslant f(d_\theta, \widehat{\mathsf{H}}_T) \Big)}.$$

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

[1] To see why point (iii) holds, denote $\sigma_d \equiv \mathrm{Var}(\varepsilon^d_t)$, and then note that $\int_\epsilon^\infty \exp(-x^2/(2\sigma_d^2))/\sqrt{2\pi\sigma_d^2}\, dx < \int_\epsilon^\infty (x/\epsilon) \exp(-x^2/(2\sigma_d^2))/\sqrt{2\pi\sigma_d^2}\, dx = \sigma_d \exp(-\epsilon^2/(2\sigma_d^2))/(\sqrt{2\pi}\epsilon)$. A similar calculation can be used to derive a lower bound for the upper tail of the normal CDF. Then applying the previous upper-bound calculation to $\mathbb{P}_0(D_T \geqslant f(d_\theta, \widehat{\mathsf{H}}_T) \mid \sum_{t=1}^T \mathfrak{D}_t > 0)$ and the lower-bound calculation to $\mathbb{P}_0(D_T \geqslant f(d_\theta, \widehat{\mathsf{H}}_T))$, it follows that $\mathbb{P}_0(D_T \geqslant f(d_\theta, \widehat{\mathsf{H}}_T) \mid \sum_{t=1}^T \mathfrak{D}_t > 0)/\mathbb{P}_0(D_T \geqslant f(d_\theta, \widehat{\mathsf{H}}_T)) = o(1)$, as stated, since the distribution of the value in the denominator is shifted to the right relative to the distribution of the value in the numerator given (i)–(ii).

statement in footnote 18 follows. In particular, for any $\delta>0$, there exists a $\underline{\theta}$ such that $\forall\, \theta_j \geqslant \underline{\theta}$, $\mathbb{P}_0\big(\sum_{t=1}^{T} \mathbb{1}\{\text{disaster}_t\} > 0 \,\big|\, R_T \geqslant \underline{\theta}\big) < \delta$. Write $\delta = \delta_0$. It also follows immediately that for any $t > 0$ (with $t < T$), for any $\delta_t > 0$, there exists an $\underline{\theta}$ such that $\mathbb{P}_t(\sum_{\tau=1}^{T} \mathfrak{D}_t > 0 \,|\, R_T \geqslant \underline{\theta}) < \delta_t$ asymptotically $\mathbb{P}_0$-a.s. as $\delta_0 \to 0$. Given some $\delta_t > 0$, consider $\theta_j, \theta_{j+1}$ large enough that $\mathbb{P}_t(\sum_{\tau=1}^{T} \mathfrak{D}_t > 0 \,|\, R_T \in \{\theta_j, \theta_{j+1}\}) < \delta_t$. We then have from (14) that

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

The fraction in the last expression is constant given that $M_{t,T} = \beta^{T-t} e^{-\gamma g_c (T-t)}$ conditional on $\sum_{t=1}^{T} \mathfrak{D}_t = 0$, using eq. (2) of Gabaix (2012). Thus denoting $\phi_j \equiv \frac{\mathbb{E}_0[M_{t,T} \,|\, R_T = \theta_j, \sum_{t=1}^{T} \mathfrak{D}_\tau = 0]}{\mathbb{E}_0[M_{t,T} \,|\, R_T = \theta_{j+1}, \sum_{t=1}^{T} \mathfrak{D}_\tau = 0]}$, we have $\phi_{t,j} = \phi_j + \mathcal{O}(\delta_t)$. Since we can take $\delta_t \to 0$ asymptotically $\mathbb{P}_0$-a.s. as $\delta_0 \to 0$, we have $\phi_{t,j} = \phi_j + o_p(1)$ for any sequence of values $\delta = \delta_0 \to 0$. So CTI holds for the state pair $(\theta_j, \theta_{j+1})$ up to a negligible error, as stated.

4. The Epstein–Zin (1989) preference recursion is $U_t = \left[(1-\beta) C_t^{1-\psi^{-1}} + \beta(\mathbb{E}_t[U_{t+1}^{1-\gamma}])^{\frac{1-\psi^{-1}}{1-\gamma}}\right]^{\frac{1}{1-\psi^{-1}}}$, and it can be shown (e.g., Campbell, 2018, eq. (6.42)) that the SDF evolves in this case according to $M_{t,t+1} = \beta(C_{t+1}/C_t)^{-\vartheta/\psi}(1/R_{t,t+1})^{1-\vartheta}$, where $\vartheta \equiv (1-\gamma)/(1-\psi^{-1})$. In case (i) of the statement, $\gamma = 1$ and $M_{t,t+1} = \beta / R_{t,t+1}$, so $M_T$ depends only on the index return. Thus the numerator and denominator in equation (14) are constant, and CTI holds immediately. For case (ii), write $\Delta c_{t+1} = \mu_c + \rho \Delta c_t + \sigma \eta_{t+1}$, with $\eta_{t+1} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. Given $\psi = 1$, it follows from Hansen, Heaton, and Li (2008, eq. (3)) that the log SDF follows $m_{t,t+1} = -\Delta c_{t+1} + \frac{1-\gamma}{1-\beta\rho}\sigma\eta_{t+1}$ (up to a constant, as we ignore throughout). Further, the consumption-wealth ratio $C_t/S_t$ is a constant given $\psi = 1$, so $r_{t,t+1} = \Delta c_{t+1}$. Using this in the log SDF and summing from $t$ to $T$, $m_{t,T} = -r_{t,T} + \frac{1-\gamma}{1-\beta\rho}\sigma \sum_{\tau=t+1}^{T} \eta_\tau$. The first term is known conditional on $R_t$. In addition, recursive substitution and summation for $r_{t,t+1}$ gives that $r_{t,T} = \frac{\sigma}{1-\rho} \sum_{\tau=t+1}^{T} (1 - \rho^{T-\tau+1}) \eta_\tau$. Thus for the second term in $m_{t,T}$, conditioning on $R_T = \theta_j$ is equivalent to conditioning on $\sum_{\tau=t+1}^{T} (1-\rho^{T-\tau+1})\eta_\tau = \text{const} + \log\theta_j \equiv k_j$. Denoting $w_t \equiv (1 - \rho^{T-t+1})$, it can then be shown (e.g., Vrins, 2018, eq. (2)–(3)) that $(\sum_{\tau=t+1}^{T} \eta_\tau \mid \sum_{\tau=t+1}^{T} w_\tau \eta_\tau = k_j) \sim \mathcal{N}(\mu_{t,j}, \varsigma_t)$, where $\mu_{t,j} = k_j \frac{\sum_{\tau=t+1}^{T} w_\tau}{\sum_{\tau=t+1}^{T} w_\tau^2}$ and where $\varsigma_t$ does not depend on $k_j$. Therefore,

$$
\log \phi_{t,j} = \log \mathbb{E}_t\left[\sum_{\tau=t+1}^{T} \eta_\tau \mid R_T = \theta_j\right] - \log \mathbb{E}_t\left[\sum_{\tau=t+1}^{T} \eta_\tau \mid R_T = \theta_{j+1}\right] = \log\theta_j - \log\theta_{j+1},
$$

so CTI holds. Case (iii) follows immediately from eq. (17) of Kocherlakota (1990), which shows that $M_T \propto (R_T)^{-\gamma}$ in the i.i.d. case.

Next, we move to the statements made after the numbered remarks:

- **Models with stochastic volatility:** For this class of models, we provide just a brief (but clear) counterexample: the pricing kernel in equation (10) of Heston, Jacobs, and Kim (2024) immediately violates CTI.

- **The Campbell and Cochrane (1999) model:** The Campbell and Cochrane (1999) economy features a representative agent with utility $\mathbb{E}_0\{\sum_{t=0}^{\infty} \beta^t [(C_t - \mathfrak{H}_t)^{1-\gamma} - 1]/(1-\gamma)\}$, where $\mathfrak{H}_t$ is the level of (exogenous) habit and other terms are standard. The *surplus-consumption ratio* is $S_t^c \equiv (C_t - \mathfrak{H}_t)/\mathfrak{H}_t$. Log dynamics are $s_{t+1}^c = (1-\phi)\bar{s}^c + \phi s_t^c + \lambda(s_t^c)\varepsilon_{t+1}$, $c_{t+1} = g + c_t + \varepsilon_{t+1}$, and $d_{t+1} = g + d_t + \eta_{t+1}$, where $\varepsilon_{t+1} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$, $\eta_{t+1} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\eta^2)$, $\mathrm{Corr}(\varepsilon_{t+1}, \eta_{t+1}) = \rho$, and the *sensitivity function* is $\lambda(s_t^c) = \left[\frac{1}{\bar{S}^c}\sqrt{1 - 2(s_t^c - \bar{s}^c)} - 1\right] \mathbb{1}\{s_t^c \leqslant s_{\max}^c\}$, with $\overline{S}^c = e^{\bar{s}^c} = \sigma_\varepsilon \sqrt{\frac{\gamma}{1-\phi}}$ and $s_{\max}^c = \bar{s}^c + (1 - \overline{S}^c)^2/2$. The SDF evolves according to $M_{t,t+1} = \beta \left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} \left(\frac{S_{t+1}^c}{S_t^c}\right)^{-\gamma}$, so

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

  For a counterexample to constant $\phi_t$, set $T = 2$ and $\rho = 1$ (so $\Delta c_t = \Delta d_t$, as in the simplest case in Campbell and Cochrane, 1999). A sufficient condition for non-constant $\phi_t$ is $\mathrm{Cov}_0(\phi_1, \mathbb{E}_1[M_{1,2} \mid R_2 = \theta_{j+1}]) \neq 0$, since then $\mathbb{E}_0[\phi_1] \neq \phi_0$. At $t = 0$, both $\varepsilon_1$ and $\varepsilon_2$ are relevant for $R_2$ and $M_{0,2}$: $\varepsilon_1$ determines $s_1^c$ and thus $\lambda(s_1^c)$. At $t = 1$, only $\varepsilon_2$ matters for uncertainty in $R_2$ and $M_{1,2}$: $s_2^c$ and $d_2$ determine $R_2$, and given $t = 1$ variables, these depend only on $\varepsilon_2$. Write $\varepsilon_j^1$ for the value of $\varepsilon_2$ needed to generate $R_2 = \theta_j$ given $\varepsilon_1$ (i.e., $\varepsilon_j^1 \equiv \{\varepsilon_2 : R_2 = \theta_j \mid \varepsilon_1\}$), and similarly $\varepsilon_{j+1}^1$ for $\theta_{j+1}$. Then $\mathbb{E}_1[M_{1,2} \mid R_2 = \theta_{j'}] = \exp(-\gamma(1 + \lambda(s_1^c))\varepsilon_{j'}^1)$ for $j' = j, j+1$, so $\phi_1 = \exp(-\gamma(1 + \lambda(s_1^c))(\varepsilon_j^1 - \varepsilon_{j+1}^1))$. Thus $\mathrm{Cov}_0(\phi_1, \mathbb{E}_1[M_{1,2} \mid R_2 = \theta_{j+1}]) = \mathrm{Cov}_0(\exp(-\gamma(1 + \lambda(s_1^c))(\varepsilon_j^1 - \varepsilon_{j+1}^1)), \exp(-\gamma(1 + \lambda(s_1^c))\varepsilon_{j+1}^1))$. For Gaussian $\varepsilon_1$, this value is generically non-zero.

- **The Basak (2000) model:** Take the two-agent CRRA case considered in Section 5 of Basak (2000), with notation adopted directly. Basak's Proposition 7 shows that when extraneous risk matters, state prices (and thus the SDF) depend on both the stochastic weighting process $\eta(t)$ and the aggregate endowment $\varepsilon(t)$. These two processes are driven respectively by independent shocks, $dW_z(t)$ (extraneous risk) and $dW_\varepsilon(t)$ (fundamental risk). Asset returns thus do not pin down the SDF realization, generating a generically path-dependent SDF and thus time-varying $\phi_t$ (see also the discussion in Atmaz and Basak, 2018, footnote 17). □

***Proof of Proposition 5.*** Given that $\phi_t$ can change, we explicitly allow it to depend on the signal history. RN beliefs are thus now denoted by $\pi_t^*(H_t) = \frac{\phi_t(H_t)\pi_t(H_t)}{(\phi_t(H_t)-1)\pi_t(H_t)+1}$, where we use the simpler notation from Section 2 for clarity throughout. Uncertainty about $\theta$ is again resolved by period $T$, and we again consider $X^*$ from 0 to $T$. Since $\pi_T \in \{0,1\}$ implies $\pi_T^* = \pi_T$, time variation in $\phi_t$ has no effect on $X^*$ for $t > T-1$.

Toward a contradiction, assume that there exists some DGP(s) in which $\phi_t$ changes such that $\mathbb{E}_t[\phi_{t+1}] \leqslant \phi_t$ and expected RN movement is higher than the bounds in Proposition 2 for some $T$. Consider a DGP from this set with the highest expected RN movement. We now consider the last meaningful movement of $\phi$ in this DGP. Specifically, given that $\phi_t$ is assumed to change at some point, but $\phi_t$ is constant when $t \geqslant T$, there must exist some history $H_t$ in which $\pi_t \in (0,1)$, $\phi_t$ can change between $t$ and $t+1$ (i.e., there exists a signal $s_{t+1}$ for which $\phi_{t+1}(H_t \cup s_{t+1}) \neq \phi_t(H_t)$, where $s_{t+1}$ includes the signal $s_{\phi_{t+1}}$) but for which $\phi_t$ is constant after $t+1$. Following any $H_t$, by assumption, $\phi_{t+1}(H_t \cup s_{t+1})$ can take two values: $\phi_{t+1}^H > \phi_t$ following signal $s_{t+1}^H$ with probability $q^H > 0$, and $\phi_{t+1}^L < \phi_t$ following signal $s_{t+1}^L$ with probability $q^L = 1 - q^H > 0$. We start by assuming that $\phi_t$ evolves as a martingale:

$$\sum_{i \in \{L,H\}} q^i \cdot \phi_{t+1}^i = \phi_t. \tag{B.1}$$

Given the maintained assumption that $\pi_t$ does not evolve in the same period as $\phi_t$ and therefore is constant immediately following history $H_t$, $\pi_t^*(H_t \cup s_{t+1})$ can take at most two values: $\pi_{t+1}^{*i} = \frac{\phi_{t+1}^i \cdot \pi_t}{(\phi_{t+1}^i - 1)\pi_t + 1}$ for $i \in \{L,H\}$. Now consider expected RN movement following $H_t$. From period $t$ to $t+1$, given signal $s_{t+1}^i$, RN beliefs move from $\pi_t^*$ to $\pi_{t+1}^{*i}$, leading to per-period RN movement

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

Given that the postulated $\phi_t$ process is constant after $t+1$, at that point our main bounds hold with $\pi_0^*$ replaced with $\pi_{t+1}^{*i}$ and $\phi$ replaced with $\phi_{t+1}^i$. Thus given signal $s_{t+1}^i$,

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

where the second line plugs in our bound for excess RN movement and uncertainty reduction given that uncertainty is zero at period $T$, and the third line states everything in terms of $\phi_t$ and $\pi_t$ and uses the assumption that $\pi_t = \pi_{t+1}$. Therefore, expected RN movement from period $t$ onward following history $H_t$ is bounded above by:

$$\mathbb{E}[\mathsf{m}_{t,T}^* | H_t] = \mathbb{E}[\mathsf{m}_{t,t+1}^* | H_t] + \mathbb{E}[\mathsf{m}_{t+1,T}^* | H_t]$$

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

We now show that this DGP will have higher RN movement if $\phi_t$ is constant from $H_t$ onward. To see this, consider the "worst-case" DGP in Proposition C.3 in which $\phi$ remains constant at $\phi_t$. In this case, RN movement is (arbitrarily close to) $\mathbb{E}_{maxDGP}[\mathsf{m}^*_{t,T}|H_t] = (1 - \pi_t) \cdot \frac{\phi_t \cdot \pi_t}{(\phi_t - 1)\pi_t + 1}$. We now subtract the expected RN movement given changing $\phi$ ($\mathbb{E}[\mathsf{m}^*_{t,T}|H_t]$) from the worst-case RN movement ($\mathbb{E}_{maxDGP}[\mathsf{m}^*_{t,T}|H_t]$) and show it is positive given the assumption that $\phi_t$ evolves as a martingale. The difference is positive if and only if

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

Using (B.1) in this inequality gives that $\mathbb{E}_{maxDGP}[\mathsf{m}^*_{t,T}|H_t] - \mathbb{E}[\mathsf{m}^*_{t,T}|H_t] > 0$ if and only if

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

It is straightforward to see that the expression on the left side of this inequality is positive: every parentheses contains a positive value as $\phi^H_{t+1} > \phi_t > \phi^L_{t+1} \geqslant 1$ and $\pi_t \in (0, 1)$. Therefore, we conclude that expected RN movement can be increased if $\phi_t$ remains constant following $H_t$ rather than changing. But this gives us a contradiction, as it violates the assumption that the DGP with $\phi_t$ moving following $H_t$ has the highest possible movement. Therefore, we conclude that there does not exist a DGP satisfying in which $\phi$ evolves as a martingale that produces more expected RN movement than the bound in Proposition 2.

We now extend this observation to DGPs in which movement in $\phi$ is a supermartingale rather than a martingale. We do so by showing that if there exists a DGP where $\phi$ evolves as super-martingale and leads to expected movement that is higher than our bound, there there must exist a martingale that leads to higher expected movement. Given the previous martingale result, this is impossible. Formally, assume that there exists a $DGP_{super}$ in which $\phi$ evolves as a supermartingale such that the expected movement of this DGP is higher than our bound for a given $T$. Consider the supermartingale DGP with the maximum expected movement, and consider a period $t$ (history $H_t$) with the last meaningful movement in $\phi$ in which $\phi$ is a strict supermartingale. If this period does not exist, the process is a martingale, and the previous results hold. Note that, following this movement, there cannot be further change in $\phi$. If there were and $\phi$ were a martingale, the previous result shows that no change in $\phi$ would produce more expected movement, contradicting the assumption that this DGP produces the highest expected movement in the class. If instead there was movement and the change in $\phi$ was a strict supermartingale, it would contradict the assumption that the previous movement was the last meaningful movement of that type.

Now, we show that it is possible to adjust $DGP_{super}$ following history $H_t$ to increase expected movement following $H_t$ by adjusting the change in $\phi$ from period $t$ to period $t + 1$ to be a martingale rather than a supermartingale. To do so, we first show that any upward movement from $\phi_t$ to

$\phi_{t+1} > \phi_t$ always leads to more total movement following $H_t$ than any downward movement from $\phi_t$ to $\phi_{t+1} < \phi_t$. Consider total expected movement from $H_t$ onward given a change from $\phi_t$ to $\phi_{t+1}$:

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

Our claim is that this is higher if $\phi_{t+1} > \phi_t$ than if $\phi_{t+1} < \phi_t$. To see this, compare the above with movement if $\phi_{t+1} = \phi_t$. In this case, $\mathbb{E}[\mathsf{m}^*_{t,T}|H_t, \phi_t = \phi_{t+1}] = (1 - \pi_t) \cdot \frac{\phi_t \cdot \pi_t}{(\phi_t - 1)\pi_t + 1}$. Subtracting from above and writing $\pi = \pi_t$ for simplicity yields:

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

As with the inequality in the martingale case, every component in this expression is weakly positive (as $0 < \pi < 1$ because the $\phi$ movement is meaningful and $\phi \geqslant 1$), except for $(\phi_t - \phi_{t+1})$. Therefore, this equation is positive if $\phi_{t+1} < \phi_t$ and negative if $\phi_{t+1} > \phi_t$. But then it must be that $\mathbb{E}[\mathsf{m}^*_{t,T}|H_t, \phi_t, \phi_{t+1}]$ is greater if $\phi_{t+1} > \phi_t$ than if $\phi_{t+1} < \phi_t$. In this case, we can adjust the evolution of $\phi$ following history $H_t$ — which was assumed to be a supermartingale — to be a martingale by taking a probability from downward change in $\phi$ and shifting it to an upward change in $\phi$. Specifically, if $\phi_t$ is a strict supermartingale at $H_t$, there must be at least some probability on a realization of $\phi_{t+1} < \phi$. Consider the lowest possible realization of $\phi^L_{t+1}$ with associated probability $q^L$. There are two possibilities. First, there is some value $\phi^H_{t+1} > \phi$ such shifting the probability $q^L$ from $\phi^L_{t+1}$ to $\phi^H_{t+1}$ makes $\phi$ a martingale. Second, there is some $q^H < q^L$ such that shifting $q^H$ from $\phi^L_{t+1}$ to $\phi^H_{t+1}$ makes $\phi$ a martingale. In either case, we are shifting probability from $\phi^L_{t+1} < \phi_t$ to $\phi^H_{t+1} > \phi_t$. But, as just proven above, it must be that $\mathbb{E}[\mathsf{m}^*_{t,T}|H_t, \phi_t, \phi_{t+1}]$ is greater if $\phi_{t+1} > \phi_t$ than if $\phi_{t+1} < \phi_t$. But then the total movement of the change from $\phi$ at $H_t$ must increase. This implies that there exists a martingale process for $\phi$ at $H_t$ that has higher expected movement than the strict supermartingale process for $\phi$ at $H_t$. This contradicts the assumption that the strict supermartingale process has the highest movement in the class of supermartingale processes (which includes martingales), completing the proof. □

***Proof of Proposition 6.*** In what follows, we often use $\mathbb{E}_i[\cdot]$ to make explicit that we are taking expectations over DGPs indexed by $i$, and we continue to use the notational simplifications used in the statement of the proposition. For (i), fixing $\pi^*_{0,i} = \pi^*_0$ across $i$ and applying Proposition 1,

$$\begin{aligned}
\mathbb{E}_i[\mathbb{E}[X^*_i]] &= \mathbb{E}_i[(\pi^*_0 - \pi_{0,i}) \cdot \triangle_i] = \pi^*_0 \cdot \mathbb{E}_i[\triangle_i] - \mathbb{E}_i[\pi_{0,i}] \cdot \mathbb{E}_i[\triangle_i] \\
&= (\pi^*_0 - \mathbb{E}_i[\pi_{0,i}]) \cdot \mathbb{E}_i[\triangle_i] = \mathbb{E}_i[\pi^*_0 - \pi_{0,i}] \cdot \mathbb{E}_i[\triangle_i] \\
&= \mathbb{E}_i\left[\pi^*_0 - \frac{\pi^*_0}{\phi_i + (1 - \phi_i)\pi^*_0}\right] \cdot \mathbb{E}_i[\triangle_i] \qquad \text{(B.2)}
\end{aligned}$$

where the last equality in the first line follows from the assumption that $\mathrm{Cov}(\pi_{0,i}, \triangle_i) = 0$.

Now consider $\zeta_1(\phi_i, \pi_0^*) \equiv \pi_0^* - \frac{\pi_0^*}{\phi_i+(1-\phi_i)\pi_0^*}$. This function is concave in $\phi_i$: $\frac{\partial^2 \zeta_1}{\partial \phi_i^2} = \frac{-2\pi_0^*(1-\pi_0^*)^2}{(\pi_0^*+\phi(1-\pi_0^*))^3}$, which is weakly negative given $\pi_0^* \in [0,1]$ and $\phi \geqslant 1$. Thus by Jensen's inequality, the expectation of $\zeta_1$ over $\phi_i$ is less than $\zeta_1$ evaluated at $\underline{\phi} \equiv \mathbb{E}_i[\phi_i]$, so $\mathbb{E}_i\left[\pi_0^* - \frac{\pi_0^*}{\phi_i+(1-\phi_i)\pi_0^*}\right] \leqslant \pi_0^* - \frac{\pi_0^*}{\underline{\phi}+(1-\underline{\phi})\pi_0^*}$. Returning to (B.2), suppose that $\mathbb{E}_i[\triangle_i] > 0$. In this case,

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

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

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

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

For part (ii), first consider the situation in which $\pi_{0,i}^*$ is constant and equal to $\pi_0^*$. As above,

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

As above, given the concavity of $\zeta_2 \equiv \pi_0^* - \frac{\pi_0^*}{\phi_i+(1-\phi_i)\pi_0^*}$ with respect to $\phi_i$ and the fact that $\pi_0^* \geqslant 0$,

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

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

$$\mathbb{E}[X_i^*] \leqslant \mathbb{E}[\zeta_{2'}(\phi_i, \pi_{0,i}^*)] \leqslant \mathbb{E}\left[\zeta_{2'}\left(\mathbb{E}[\phi_i \mid \pi_{0,i}^*], \pi_{0,i}^*\right)\right] \leqslant \mathbb{E}\left[\zeta_{2'}(\overline{\phi}, \pi_{0,i}^*)\right], \tag{B.3}$$

where $\overline{\phi}$ is as in the proposition statement and where the last inequality uses $\partial\zeta_{2'}/\partial\phi_i \geqslant 0$. Substituting the definition of $\zeta_{2'}$ into this inequality yields equation (16).

For part (iii), as $(\pi_{0,i}^* - \frac{\pi_{0,i}^*}{\overline{\phi}+(1-\overline{\phi})\pi_{0,i}^*}) \leqslant \pi_{0,i}^*$ for any $\overline{\phi} \geqslant 1$, $\mathbb{E}[X_i^*] \leqslant \mathbb{E}[(\pi_{0,i}^* - 0)\pi_{0,i}^*] = \mathbb{E}[(\pi_{0,i}^*)^2]$, as stated. (Equivalently, one can use (B.3) and note again that $\partial\zeta_{2'}/\partial\overline{\phi} \geqslant 0$, so that the bound is most slack as $\overline{\phi} \to \infty$, giving the same bound.)

For part (iv), from Proposition C.1, if $\mathbb{E}[X^*|\theta = 0] \leqslant \mathbb{E}[X^*|\theta = 1]$, then $\mathbb{E}[X^*] \leqslant 0$. Therefore, if $\mathbb{E}[X_i^*|\theta = 0] \leqslant \mathbb{E}[X_i^*|\theta = 1]$ for all $i$, then $\mathbb{E}[X_i^*] \leqslant 0$ over all streams, completing the proof. □

***Proof of Proposition 7.*** Under the stated assumptions for $\epsilon_t$, observed RN movement satisfies

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

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

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

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

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

# Appendix C. Additional Technical Material

## C.1 Additional Theoretical Results Discussed in Section 2

Following the discussion in Section 2.3 in the main text, we now provide a number of additional theoretical results. First, when should we expect to see *negative* RN excess movement — as observed in the lowest dashed gray line in Figure 3 — even with risk aversion? If one is willing to make an assumption on the sign of $\triangle$ (discussed shortly), the following stronger bound applies effectively as a corollary of Proposition 1.

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

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

***Proof of Proposition C.1.*** As in (A.10), we have $\pi^*_0 - \pi_0 \geqslant 0$. Using this in the equality in (A.11) alongside the assumption that $\triangle = \mathbb{E}^*[\mathsf{m}^*|\theta = 0] - \mathbb{E}^*[\mathsf{m}^*|\theta = 1] \leqslant 0$ gives $\mathbb{E}[X^*] \leqslant 0$. $\square$

When $\triangle < 0$, $\mathbb{E}[X^*]$ is decreasing in $\phi$ and therefore $\mathbb{E}[X^*] < 0$ for any $\phi > 1$. Consequently, as formalized in Proposition C.1, the highest excess movement is $\mathbb{E}[X^*] = 0$. While most of our focus is on the conservative positive upper bounds for $\mathbb{E}[X^*]$, this corollary shows that asset-pricing settings with risk aversion do not necessarily entail positive excess movement in observed beliefs. We now further explore the statistical features of the DGP that are informative about the degree of RN excess movement to be expected under RE.

To explore the possibility of negative RN excess movement further, we now provide an additional result on how the DGP pins down $\mathbb{E}[X^*]$ precisely. Proposition 1 says that the deviation of $\mathbb{E}[X^*]$ from 0 depends on the product of $\pi^*_0 - \pi_0$ and $\triangle \equiv \mathbb{E}[X^*|\theta = 0] - \mathbb{E}[X^*|\theta = 1]$. The difference $\pi^*_0 - \pi_0$ is always positive and increases in $\phi$. But how are the sign and magnitude of $\triangle$

related to the DGP? The following result shows that given the arbitrary labeling of the two states, there is no reason to expect under RE that $\triangle$ should take a particular sign:

**PROPOSITION C.2.** *Fixing $\phi$, for every RN prior and DGP that leads to a given $\triangle$, there exists a different RN prior and DGP that leads to $-\triangle$.*

***Proof of Proposition C.2.*** Consider a given $\phi$, RN prior $\pi_0^*$, and signal DGPs $DGP(s_t|\theta=0,H_{t-1})$ and $DGP(s_t|\theta=1,H_{t-1})$ that lead to some $\mathbb{E}[X^*|\theta=0]$, $\mathbb{E}[X^*|\theta=1]$, and $\triangle$. Now consider the "reversed" DGP $\widehat{DGP}$ in which we modify the DGP by relabeling state 1 as state 0 and state 0 as state 1. That is, $\widehat{DGP}(s_t|\theta=0,H_{t-1}) \equiv DGP(s_t|\theta=1,H_{t-1})$ and $\widehat{DGP}(s_t|\theta=1,H_{t-1}) \equiv DGP(s_t|\theta=0,H_{t-1})$. Similarly, we consider the "reversed" RN prior $\widehat{\pi}_0^*=1-\pi_0^*$ implied by the physical prior $\widehat{\pi}_0=\frac{1-\pi_t^*}{\phi+(1-\phi)(1-\pi_t^*)}$.

With this relabeling, if the RN belief in the original DGP given history $H_t$ is $\pi_t^*(H_t)$, then the RN belief in the reversed $\widehat{DGP}$ with RN prior $1-\pi_0^*$ must be $\widehat{\pi}_t^*(H_t)=1-\pi_t^*(H_t)$. Thus $\mathbb{E}^*[\widehat{X}^*|\theta=0]=\mathbb{E}^*[X^*|\theta=1]$ and $\mathbb{E}^*[\widehat{X}^*|\theta=1]=\mathbb{E}^*[X^*|\theta=0]$. And since $\mathbb{E}^*[X^*|\theta]=\mathbb{E}[X^*|\theta]$ by Lemma A.1(i), $\mathbb{E}[\widehat{X}^*|\theta=0]=\mathbb{E}[X^*|\theta=1]$ and $\mathbb{E}[\widehat{X}^*|\theta=1]=\mathbb{E}[X^*|\theta=0]$. Thus for $\widehat{DGP}$, $\widehat{\triangle}\equiv\mathbb{E}[\widehat{X}^*|\theta=0]-\mathbb{E}[\widehat{X}^*|\theta=1]=-\triangle$. □

Intuitively, for any RN prior $\pi_0^*$ and DGP with some $\triangle$, the RN prior $1-\pi_0^*$ with the "reversed" DGP will necessarily lead to $-\triangle$. Consequently, there is no reason to assume that $\mathbb{E}[X^*]$ is more likely to be positive than negative given $\phi>1$.

In the next subsection of the Internet Appendix just below (Appendix C.2), we conduct numerical simulations to further explore how $\triangle$ and RN excess movement depend on the DGP. Those simulations suggest that extreme values of $\triangle$ only occur in highly asymmetric DGPs where movements in one direction are large and movements in the other direction are tiny. In fact, in the last additional result provided in this subsection, we can show that our upper bound in (10) is attainable asymptotically given the most asymmetric DGP possible:

**PROPOSITION C.3.** *There exists a sequence of DGPs, indexed by $T$, for which $\mathbb{E}[X^*]$ approaches the bound in Proposition 2 as $T\to\infty$. For each DGP in this sequence, downward movements ($\pi_{t+1}^*<\pi_t^*$) are resolving ($\pi_{t+1}^*=0$) and thus as large as possible, while upward movements are small ($\pi_{t+1}^*-\pi_t^*\to 0$ as $T\to\infty$). Meanwhile, the bound holds with strict inequality for any $T<\infty$ as long as $\phi>1$ and $\pi_0^*\in(0,1)$.*

***Proof of Proposition C.3.*** Consider a sequence of binary resolving DGPs indexed by $T$. There are two possible signals in each period, $l$ and $h$, and assume that for any history,

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

$$DGP(s_t=h|\theta=0)=\frac{\pi_{t-1}^*(1-\pi_{t-1}^*-\epsilon)}{(1-\pi_{t-1}^*)(\pi_{t-1}^*+\epsilon)}, \quad \text{with } \epsilon\equiv\frac{1-\pi_0^*}{T}. \tag{C.2}$$

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

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

Applying the transformation (8) to the $\pi_{t-1}$ values on the right side of this equation,

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

Now applying the transformation (5), we obtain that $\pi^*_t$ given an only-$h$ signal history (suppressing the dependence on this history for simplicity) is, after additional algebra,

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

Now using (C.2), we obtain after further algebra that $\pi^*_t - \pi^*_{t-1} = \epsilon$. Given the definition of $\epsilon$, this DGP is resolving for any $T$: given any $l$ signal at any $t$, beliefs resolve to 0, while given only $h$ signals, beliefs increase slowly ($\pi^*_t = \pi^*_0 + t\epsilon$) and resolve to 1 at period $T$. We thus have

$$\mathbb{E}[\mathsf{m}^*|\theta = 1] = T\epsilon^2 = T\left(\frac{1-\pi^*_0}{T}\right)^2 = \frac{(1-\pi^*_0)^2}{T} \xrightarrow{T\to\infty} 0.$$

Thus for such a sequence, using equation (A.8),

$$\triangle = \pi^*_0 - \frac{1}{1-\pi^*_0} \cdot \mathbb{E}[\mathsf{m}^*|\theta = 1] \xrightarrow{T\to\infty} \pi^*_0.$$

Using this in equation (A.9) gives $\mathbb{E}[X^*] \to (\pi^*_0 - \pi_0)\pi^*_0$ as $T \to \infty$, as stated. And as further stated, the sequence of DGPs is constructed such that any downward movement is resolving and any upward movement is small ($\pi^*_t - \pi^*_{t-1} = \epsilon \to 0$). We have thus proven the first two statements.

For the final statement, given $\phi > 1$ and $0 < \pi^*_0 < 1$, the inequality in (A.10) is strict, so that $\pi^*_0 - \pi_0 > 0$. Further, the only way to obtain $\mathsf{m}^* = 0$ for finite $T$ is if $\pi^*_0 = \pi^*_1 = \ldots = \pi^*_T$, which is ruled out by $0 < \pi^*_0 < 1$ since $\pi^*_T = 0$ or 1 with probability 1, so $\mathbb{E}[\mathsf{m}^*|\theta = 1] > 0$. Thus in (A.8), we have the strict inequality $\triangle < \pi^*_0$ for fixed $T < \infty$. Combining these in (A.9) gives $\mathbb{E}[X^*] < (\pi^*_0 - \pi_0)\pi^*_0$ for fixed $T$, as stated. □

One implication of this result is that the bound in Proposition 2 is approximately tight, as one can construct a DGP for which $\mathbb{E}[X^*]$ is close to the bound for large $T$. Perhaps more important, though, is that it points to the bound's conservatism: it holds under a somewhat perverse DGP that can be thought of as a "rare bonanzas" process, where with small probability the person receives news that the bad state ($\theta = 1$) will not be realized (so $\pi^*_{t+1} = 0$), and otherwise there is mostly uninformative bad news that increases $\pi^*_{t+1}$ slightly. More reasonable DGPs, or $T \ll \infty$, will give lower $\mathbb{E}[X^*]$. That said, the conservative bound has the advantage of being very simple and not requiring any estimation of $\triangle$. And as we show below, empirical excess movement is in fact so high that even these conservative bounds are often violated for reasonable values of $\phi$.

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

As noted in Section 2.3 and discussed just above, we run numerical simulations of a large number of DGPs and priors in order to understand the precise impact of the RN prior and DGP on $\triangle$ (and therefore $\mathbb{E}[X^*]$). We consider the universe of history-independent binary-signal DGPs with a prior $\pi_0^*$ where $s_t \in \{l, h\}$ and $\mathbb{P}[s_t = h|\theta = 1]$ and (assumed lower) $\mathbb{P}[s_t = h|\theta = 0]$ are constant over $t$. These signal distributions imply likelihood ratios for the signals of $L_h \equiv \frac{\mathbb{P}[s_t=h|\theta=1]}{\mathbb{P}[s_t=h|\theta=0]} > 1$ and $L_l \equiv \frac{\mathbb{P}[s_t=l|\theta=0]}{\mathbb{P}[s_t=l|\theta=1]} > 1$. We use a fine grid to discretize $\pi_0^*$, $L_h$, and $L_l$, then conduct 1000 simulations with $T = 100$ and calculate $\triangle$ in all cases. We find:

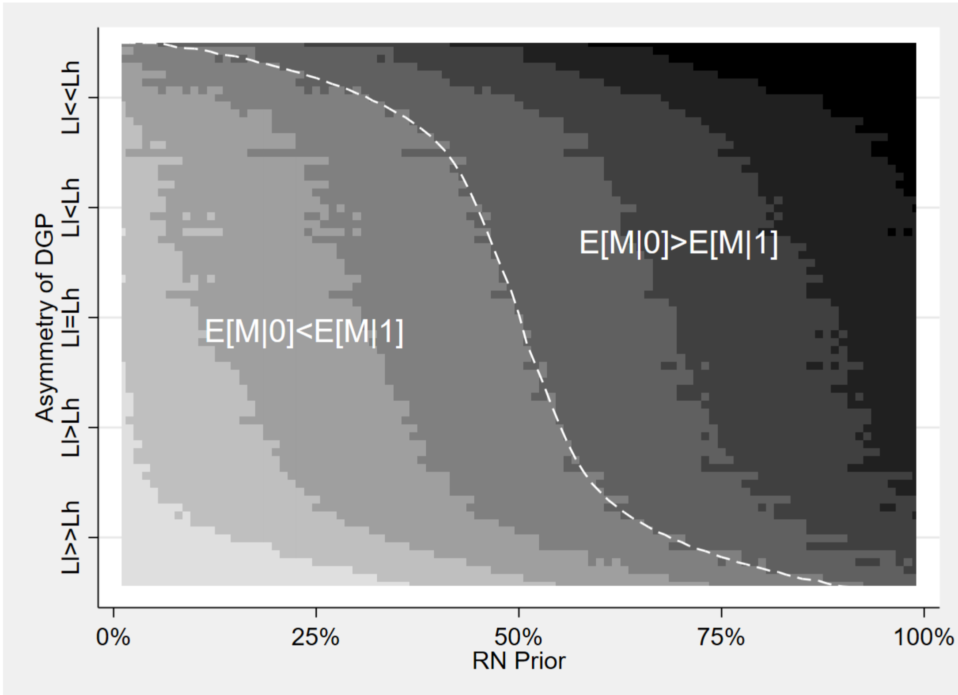
1. When $\pi_0^*$ is low, $\triangle > 0$ is very unlikely: the percentage of DGPs with positive $\triangle$ given a $\pi_0^* < .25$ is 2%. For $\pi_0^* < .5$, it is 11%.
2. When $\pi_0^*$ is low, the only DGPs in which $\triangle > 0$ are very asymmetric and extreme. For example, when $\pi_0^* = .25$, $\triangle > 0$ only occurs if $\mathbb{P}[s_t = h|\theta = 1] > .95$ and $L_l > 2 \cdot L_h$.
3. The converse is true when $\pi_0^*$ is high: $\triangle < 0$ is rare and only occurs given a very asymmetric and extreme DGP.
4. For symmetric DGPs ($L_h = L_l$), $\triangle \lesseqqgtr 0$ when $\pi_0^* \lesseqqgtr .5$.
5. Holding the DGP constant, $\triangle$ rises with $\pi_0^*$.
6. Holding all else constant, as $L_h$ rises and the size of upward updates rises, $\triangle$ falls. As $L_l$ rises and the size of upward-updates rises, $\triangle$ rises.

We present these results visually in Figure C.1. We reduce the dimensionality of the setting by focusing on the *likelihood ratio* $\frac{L_h}{L_l}$ rather than $L_h$ and $L_l$ individually. (While the impact of both $L_h$ and $L_l$ on $\triangle$ appears monotonic, the impact of $\frac{L_h}{L_l}$ is only monotonic on average, leading to a slightly messier graph.) The figure shows a contour plot with the RN prior on the $x$-axis, with the $y$-axis stacking all of the DGP combinations in order of the likelihood ratio, and the contour colors showing the approximate value of $\triangle$ (darker colors corresponding to higher values) for each prior and DGP (with the dotted line highlighting the points at which $\triangle = 0$). For example, drawing a vertical line at a prior of $\pi_0^* = 0.25$ suggests that a large portion of DGPs produce a $\triangle < 0$, and the only DGPs that produce $\triangle > 0$ have extreme likelihood ratios.

### C.3 Risk-Neutral Beliefs and Time-Varying Discount Rates

This section provides further context on the relationship between RN beliefs and discount rates, as discussed in Section 4.1 in the main paper. We again work in the setting in Section 2 here for simplicity of exposition. The price of the terminal consumption claim is given in equilibrium in

**Figure C.1: Contour Plot: Simulations for $\triangle$ by DGP and $\pi_0^*$**




*Note:* See text in Appendix C.2 for description of simulations and discussion of results.

by $P_t(C_T) = \mathbb{E}_t\left[\beta_t^{T-t}\frac{U'(C_T)}{U'(C_t)}C_T\right]$, where $\beta_t$ is now the agent's (possibly time-varying) time discount factor. Defining the gross return $R_{t,T}^C \equiv \frac{C_T}{P_t(C_T)}$, rearranging this equation for $P_t(C_T)$ yields

$$\mathbb{E}_t[R_{t,T}^C] = \frac{1 - \mathrm{Cov}_t\left(\beta_t^{T-t}\frac{U'(C_T)}{U'(C_T)}, C_T\right)}{\mathbb{E}_t\left[\beta_t^{T-t}\frac{U'(C_T)}{U'(C_t)}\right]} = \frac{\frac{U'(C_t)}{\beta_t^{T-t}} - \mathrm{Cov}_t(U'(C_T), C_T)}{\mathbb{E}_t[U'(C_T)]},$$

as usual. We can write $\mathbb{E}_t[U'(C_T)] = \pi_t U'(C_{\text{low}}) + (1-\pi_t)U'(C_{\text{high}})$ in our two-state setting, and $\mathrm{Cov}_t(U'(C_T), C_T)$ can be similarly rewritten as a function of $\pi_t$, $C_T$, and $U'(C_T)$. In this setting, discount-rate variation can arise from four sources:

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

Our framework allows for *any* discount-rate variation arising from the first three sources, but restricts the last one: under CTI, it must be the case that any changes to (expected) $U'(C_i)$ are proportional across states. (More generally, as in Section 4.1, permanent changes to the SDF are admissible, which by itself greatly generalizes this setting relative to one with constant discount rates.) With constant discount rates, meanwhile, none of the four changes are admissible, or any such changes must offset perfectly.

### C.4 Simulations with Time-Varying $\phi_t$

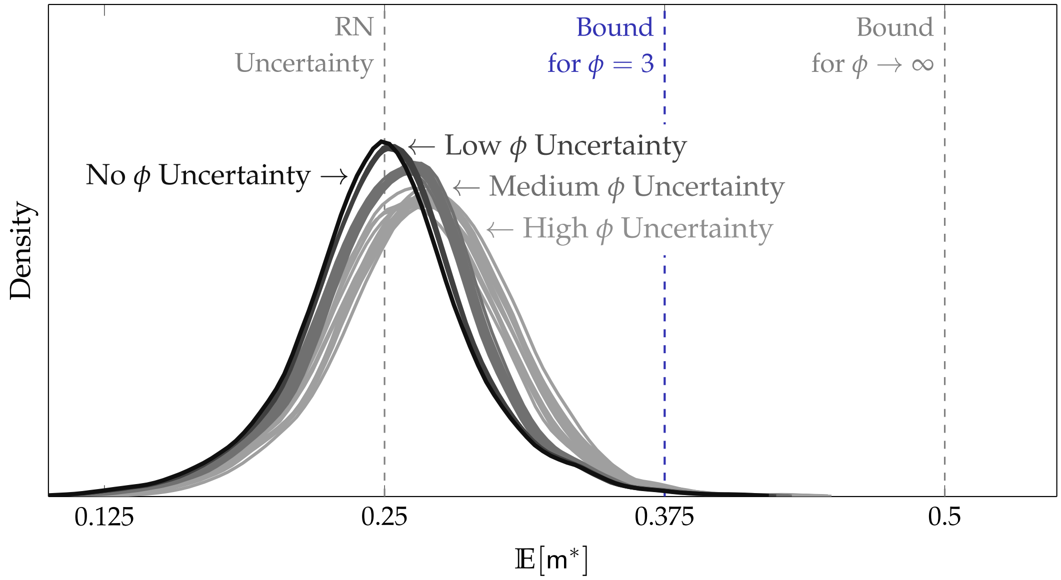
This section provides further detail for the setup and results of the simulations discussed in Section 4.2. Figure C.2 presents our main results, and we describe the setup and interpretation of these simulations step by step using this figure. The figure plots distributions for the estimated $\mathbb{E}[\mathsf{m}^*]$ (rather than $\mathbb{E}[X^*]$, as $\mathbb{E}[\mathsf{m}^*]$ is what changes with the DGP here) across simulations. Each DGP is simulated repeatedly to obtain an estimated $\mathbb{E}[\mathsf{m}^*]$ for that DGP. Each line represents a different $\mathbb{E}[\mathsf{m}^*]$ distribution given variation in the signal strengths for $\theta$, with the different lines showing different signal strengths for learning about the conditional values of $M_T$ (and thus $\phi$). In all cases, $\pi_0^* = 0.5$ and $\phi_0 = 3$.

The black line ("No $\phi$ Uncertainty") shows a baseline with $\phi_t = \phi = 3$ for all $t$. There is thus only uncertainty about $\theta$, and $\mathbb{E}[\mathsf{m}^*]$ varies depending on the signal DGP. To trace the distribution of $\mathbb{E}[\mathsf{m}^*]$ across DGPs, we attempt to cover the space of binary DGPs in which the signal strengths are constant over time. We start by looping over $\mathbb{P}[s_t = h|\theta = 1]$ from {1,.99,.98,...,.01}. Then we loop over $\mathbb{P}[s_t = l|\theta = 0]$ from {.01,.02,.03,....99} while constraining $\mathbb{P}[s_t = h|\theta = 1] > \mathbb{P}[s_t = l|\theta = 0]$ such that the $h$ signal leads to an upward movement. This process leads to 5052 DGPs. For each of these DGPs, we simulate 100 random streams of $T = 200$ periods, after which the state is perfectly observed. This number of periods allows beliefs to get very close to certainty prior to the resolving signal. We calculate $\mathsf{m}^*$ for each stream, from which we calculate the average $\mathsf{m}^*$ statistic as an estimate of $\mathbb{E}[\mathsf{m}^*]$ for each DGP. The distribution of $\mathbb{E}[\mathsf{m}^*]$ values across all such simulated DGPs is again shown in the dark line in Figure C.2. As in Section 2.3, when signals are symmetric, $\mathbb{E}[\mathsf{m}^*] = \mathsf{u}_0^* = 0.25$, and very asymmetric DGPs produce the tails. Up to smoothing noise, $\mathbb{E}[\mathsf{m}^*]$ never crosses the theoretical upper bound of 0.375 from (10).

Next, we allow additional uncertainty about the conditional realizations of the SDF $M_T$, so that $\phi_t$ also evolves over time. For each state ($j$ and $j+1$), we allow $M_T$ to take two possible values with equal probability, where we choose the values such that $\phi_0 = 3$. Here, we start to run into calculation timing constraints such that we limit the possible signal strengths. In particular, we allow signal strengths for the high signal of .55,.75,.95 and for the low signal of .05,.25,.45 for both states. Therefore we simulate nine DGPs for learning about $M_T$ in state $j$ and nine DGPs for learning about $j+1$, leading to 81 combined DGPs to learn about $M_T$. We combine each such DGP with each of the DGPs for $\theta$ discussed above, and we again simulate 100 random draws of movement of 200 periods.

Each line in Figure C.2 represents a different $\mathbb{E}[\mathsf{m}^*]$ distribution given variation in the signal strengths for $\theta$, with the different lines showing different signal strengths for learning about the conditional values of $M_T$ (and thus $\phi$). In the dark gray lines ("Low $\phi$ Uncertainty"), $M_T$ in state $j$ can take the values 2.5 or 3.5 with equal probability and in state $j+1$ can take the values 0.833 or 1.167 with equal probability. Consequently, $\phi_0 = 3$, and $\phi_T$ can vary from 2.14 to 4.2 (with a coefficient of variation of 12%). The ex ante standard deviation of $\phi_T$ is $\sigma_\phi \equiv \mathrm{SD}_0(\phi_T) = 0.36$. Using

**Figure C.2: RN Belief Movement Distributions with Time-Varying $\phi_t$**




*Notes:* This figure shows the results of simulations studying the impact of time-varying $\phi_t$ on the distribution of $\mathbb{E}[\mathsf{m}^*]$ given different DGPs with possibly asymmetric signal strengths about $\theta$ and $\phi_T$ over time. The dark black line ("No $\phi$ Uncertainty") shows the distribution when there is no uncertainty about $\phi$. Each line in the slightly lighter dark gray set ("Low $\phi$ Uncertainty") represents the equivalent distribution for a DGP that also contains uncertainty about the SDF in both states. In this case, $\phi_0 = 3$, but $\phi_T$ can vary from 2.14 to 4.2. The set of gray lines ("Medium $\phi$ Uncertainty") allow $\phi_T$ to vary from 1.5 to 6, and the set of light-colored lines ("High $\phi$ Uncertainty") allow $\phi_T$ to vary from 1 to 9.

Proposition 4, if return states $\theta_j$ and $\theta_{j+1}$ differ by 5% (as in our empirical setting), this corresponds to relative risk aversion of $\gamma_0 = 40$ and standard deviation for $\gamma_T$ of $\sigma_\gamma = 7.2$. Changing $\phi$ has virtually no effect regardless of the signal structure: average $\mathbb{E}[\mathsf{m}^*]$ rises by 0.0012, and the number of DGPs for which $\mathbb{E}[\mathsf{m}^*]$ exceeds the bound rises by just 0.00007 percentage points (pp).

In the medium gray lines ("Medium $\phi$ Uncertainty"), $M_T$ in state $j$ can be 2 or 4 and in state $j+1$ can be 0.667 or 1.333, so that $\phi_T$ can vary from 1.5 to 6 (with a coefficient of variation of 54%, $\sigma_\phi = 1.62$, and $\sigma_\gamma = 32.4$). Even with such sizable variation, average $\mathbb{E}[\mathsf{m}^*]$ rises by 0.006, and DGPs above the bound by 0.0003 pp. Finally, in the light gray lines ("High $\phi$ Uncertainty"), $M_T$ in state $j$ can be 1.5 or 4.5 and in state $j+1$ can be 0.5 or 1.5, so that $\phi_T$ can vary from 1 to 9 (with a coefficient of variation of 100%, $\sigma_\phi = 3.0$, and $\sigma_\gamma = 60$). Average $\mathbb{E}[\mathsf{m}^*]$ still only increases by 0.015, and DGPs above the bound by 0.0012 pp. In all cases, the bound in Corollary 1 for $\phi \to \infty$ holds for 100% of the simulations.

## C.5 Additional Theoretical Results Discussed in Section 4

This subsection provides a set of robustness results discussed in Section 4.4 in the main text. We first consider Assumption 2, and show in the following proposition that incorrect updating is likely to be necessary for a bound violation. We continue to adopt the notation from Section 2 for clarity, but the following should be understood to apply for conditional beliefs for some state $j$.[2] Assume that Assumptions 3–4 continue to hold.

**PROPOSITION C.4.** *In place of Assumption 2, assume that the agent has an incorrect prior, $\pi_0 \neq \mathbb{P}_0(\theta)$, but updates correctly, in the sense that $\pi_t \propto \pi_{t-1} DGP(s_t \mid \theta, H_{t-1})$. Define $\check{\phi} \equiv \phi L$, where $L \equiv \frac{\pi_0/(1-\pi_0)}{\mathbb{P}_0(\theta)/(1-\mathbb{P}_0(\theta))}$ indexes the prior belief distortion, with $0 < L < \infty$. Then:*

*(i) For all $H_t$, the agent's RN beliefs $\pi_t^*$ are equivalent to the RN beliefs of a fictitious agent whose physical beliefs $\check{\pi}_t$ satisfy Assumption 2 but who has $\check{\phi}$ in place of $\phi$.*

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

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

***Proof of Proposition C.4.*** For part (i), first define the likelihood of a prior $\pi_0$ as

$$\mathcal{L}(\pi_0) \equiv \frac{\pi_0}{1-\pi_0}, \tag{C.3}$$

and the likelihood of a signal $s_t$ as

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

where the dependence of the latter on $H_{t-1}$ is left implicit for simplicity. The likelihood for any belief $\pi_t$ is defined as well following (C.3). The above likelihoods are well-defined for interior priors (as we assume given finite $L$ in the proposition) and for $DGP(s_t|\theta=0, H_{t-1}) > 0$ (we return to the situation in which $DGP(s_t|\theta=0, H_{t-1}) = 0$ shortly). From Bayes' rule, beliefs satisfy $\mathcal{L}(\pi_t) = \mathcal{L}(\pi_0) \cdot \mathcal{L}(s_1) \cdot \mathcal{L}(s_2) \cdots \mathcal{L}(s_t)$. Now note from (5) that $\mathcal{L}(\pi_0^*) \equiv \frac{\pi_0^*}{1-\pi_0^*} = \phi \frac{\pi_0}{1-\pi_0}$, from which it follows that under Bayesian updating,

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

For a fictitious agent with a rational prior, one could replace $\mathcal{L}(\pi_0)$ with $\mathcal{L}(\mathbb{P}_0(\theta=1))$. In our case, given the incorrect prior (but correct Bayesian updating), we have $\frac{\pi_t^*}{1-\pi_t^*} = \check{\phi} \frac{\mathbb{P}_0(\theta=1)}{1-\mathbb{P}_0(\theta=1)}$, where $\check{\phi} \equiv \phi L$, with $L$ defined as in the proposition. We can therefore write

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

As the likelihood ratios for the RN beliefs in this case are equal to those of a fictitious agent with a correct prior $\check{\pi}_0 = \mathbb{P}_0(\theta=1)$ and $\check{\phi}$ in place of $\phi$, we conclude that the RN beliefs are as well.

[2] For example, the incorrect prior is $\widetilde{\pi}_{0,j} \neq \mathbb{P}_0(R_T = \theta_j \mid R_T \in \{\theta_j, \theta_{j+1}\})$, and $L \equiv \frac{\widetilde{\pi}_{0,j}/(1-\widetilde{\pi}_{0,j})}{\widetilde{\mathbb{P}}_0(R_T=\theta_j)/(1-\widetilde{\mathbb{P}}_0(R_T=\theta_j))}$.

Finally, for the case in which $DGP(s_t|\theta=0,H_{t-1})=0$ and this signal $s_t$ is observed, the person will update to $\pi_t=1$, matching the belief of a rational agent again. We have thus shown part (i).

We can thus treat the agent with the incorrect prior as if she were rational (satisfying Assumption 2) but with $\check{\phi}$ in place of $\phi$. Further, $\check{\phi}$ satisfies Assumption 4, since $L$ is constant and $\phi$ is constant by that assumption as well. For part (ii) of the proposition, if $\check{\phi}\geqslant 1$, then Assumption 3 holds as well, so all three assumptions are satisfied, and the stated results carry through.

For part (iii), assuming $0<\check{\phi}<1$ (so Assumption 3 no longer holds for the fictitious rational agent), note first that the proof of Proposition 1 never employs Assumption 3 and therefore still holds straightforwardly, as we can write $\mathbb{E}[X^*]=(\pi_0^*-\check{\pi}_0)\triangle$ without use of this assumption. For Proposition 2, the result as stated for a rational agent requires that $\pi_0^*>\check{\pi}_0$, which is not true for $\check{\phi}<1$. But an alternative bound can be shown for this case, by obtaining a lower bound for $\triangle$ similar to the upper bound in Lemma A.2. Starting from (A.7) but solving now for $\mathbb{E}[\mathsf{m}^*|\theta=1]$, $\mathbb{E}[\mathsf{m}^*|\theta=1]=(1-\pi_0^*)-\frac{1-\pi_0^*}{\pi_0^*}\cdot\mathbb{E}[\mathsf{m}^*|\theta=0]$. Using this in (A.6),

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

Then, given that $\frac{1}{\pi_0^*}\geqslant 0$ and $\mathbb{E}[\mathsf{m}^*|\theta=0]\geqslant 0$, $\triangle$ must be bounded below by $-(1-\pi_0^*)$. Returning to the formula from Proposition 2, if $\check{\phi}<1$, then $\pi_0^*-\check{\pi}_0\leqslant 0$, which gives

$$\mathbb{E}[X^*]=(\pi_0^*-\check{\pi}_0)(\triangle)\leqslant(\check{\pi}_0-\pi_0^*)(1-\pi_0^*). \tag{C.4}$$

Further, as $\check{\pi}_0\leqslant 1$, $\mathbb{E}[X^*]\leqslant(\check{\pi}_0-\pi_0^*)(1-\pi_0^*)\leqslant(1-\pi_0^*)(1-\pi_0^*)=(1-\pi_0^*)^2$, as stated. And taking (ii) and (iii) together, we have that $\mathbb{E}[X^*]\leqslant\max(\pi_0^{*2},(1-\pi_0^*)^2)$. □

Part (i) formalizes that risk aversion is isomorphic to an incorrect prior, in that both have the same effect on $\pi_t^*$ relative to the objective $\mathbb{P}_t(\theta)$. Thus with a suitably altered value of $\phi$, the bounds generally cover the case of an incorrect prior, as in part (ii). The only case in which this argument requires slight amendment is when the prior is so downwardly distorted that $\pi_0^*<\mathbb{P}_0(\theta)$. Even in this case, though, a slightly altered version of Corollary 1 still applies, as in part (iii). An incorrect prior acts as a one-time belief distortion; while reverting to the correct belief in this case does require some excess movement, this is generally not sufficient for a full violation of the bound in Proposition 2. In general, then, incorrect updating behavior must be present in such a violation.

Next, turning to Assumption 3, we consider how our theoretical results change if $\phi<1$. We have the following result, which effectively holds as a corollary of Proposition C.4.

**Proposition C.5.** *Continue to maintain Assumptions 2 and 4. If $\phi<1$ rather than $\phi\geqslant 1$ in Assumption 3, then the bound from Proposition 2 becomes*

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

*and Corollary 1 becomes* $\mathbb{E}[X^*] \leqslant (1-\pi_0^*)^2$. *For any* $\phi$, *therefore,* $\mathbb{E}[X^*] \leqslant \max(\pi_0^{*2}, (1-\pi_0^*)^2)$.

***Proof of Proposition C.5.*** Case (iii) from the previous proof applies, with $\phi$ in place of $\check{\phi}$ and $\pi_0$ in place of $\check{\pi}_0$ (since the agent now has RE but $\phi < 1$). Thus (C.4) applies with these substitutions. The second expression for the bound given in the corollary then substitutes for $\pi_0$ (using (8)) and simplifies. Equivalently, by swapping the labels of states 0 and 1, the swapped RN beliefs become $1-\pi_t^*$ in place of $\pi_t^*$ and the swapped SDF ratio becomes $\frac{1}{\phi}$ in place of $\phi$. As $\phi < 1$, $\frac{1}{\phi} > 1$. Therefore, all of our results hold, with $\pi_t^*$ replaced by $1-\pi_t^*$ and $\phi$ replaced by $\phi^{-1} > 1$. □

The main bounds thus apply with minor modification, effectively flipping the role of the two states when $\phi < 1$. This entails replacing $\pi_0^*$ with $1-\pi_0^*$, and for Proposition 2, replacing $\phi$ with $\phi^{-1}$ as well. Thus it is *not* the case that anything goes when Assumption 3 is violated: excess movement is bounded no matter what, and its upper bound is a function of $\max(\phi, \phi^{-1})$, which in both cases indexes risk aversion across the two states. Bounds violations, meanwhile, retain their interpretation regardless of $\phi$.

### C.6 Data Cleaning and Measurement of Risk-Neutral Distribution

Before detailing measurement of the risk-neutral distribution, we note that we must collect additional data in order to follow the procedure below. In particular, in order to obtain the ex post return state for each option expiration date $T_i$ (and thereby assign probability 1 to that state on date $T_i$, so that our streams are resolving), we need S&P 500 index prices used as option settlement values. Our first step in this exercise is therefore to obtain end-of-day index prices (which we take as well from OptionMetrics). But the settlement value for many S&P 500 options in fact reflects the opening (rather than closing) price on the expiration date; for example, the payoff for the traditional monthly S&P 500 option contract expiring on the third Friday of each month depends on the opening S&P index value on that third Friday morning, while the payoff for the more recently introduced end-of-month option contract depends on the closing S&P index value on the last business day of the month.[3] To obtain the ex-post return state for A.M.-settled options, we hand-collect the option settlement values for these expiration dates from the Chicago Board Options Exchange (CBOE) website, which posts these values.

In addition, in order to measure the risk-neutral distribution *and* to measure realized excess index returns, we need risk-free zero-coupon yields $R^f_{t,T_i}$ for $t = 0_i, \ldots, T_i - 1$. To obtain these, we follow van Binsbergen, Diamond, and Grotteria (2022) and obtain the relevant yield directly from the cross-section of option prices by applying the put-call parity relationship. We apply their "Estimator 2," which obtains $R^f_{t,T_i} = \beta^{-1/T}$ from Theil–Sen (robust median) estimation of $q^{m,\text{put}}_{t,i,K} - q^{m,\text{call}}_{t,i,K} = \alpha + \beta K + \varepsilon_{t,i,K}$. This provides a very close fit to the option cross-sections (see van Binsbergen, Diamond, and Grotteria, 2022, for details) and thus produces a risk-free rate consistent with observed option prices, as is necessary to correctly back out the risk-neutral distribution.

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

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

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

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

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

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

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

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

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

6. Calculate the risk-neutral CDF for the date-$T_i$ index value at strike price $K$ using $\mathbb{P}_t^*(S_{T_i} < K) = 1 + R^f_{t,T_i}(\hat{q}_{t,i,K} - \hat{q}_{t,i,K-2})/2$. (See the proof of equation (13) in Appendix B.1 for a derivation of this result; the index-value distance between the two adjacent strikes is equal to 2 given that we evaluate the spline at intervals of two index points.)

7. Defining $S_{i,j,\max}$ and $S_{i,j,\min}$ to be the date-$T_i$ index values corresponding to the upper and lower bounds, respectively, of the bin defining return state $\theta_j$,[5] we then calculate the risk-neutral probability for state $\theta_j$ will be realized at date $T_i$, referred to with slight notational abuse as $\mathbb{P}_t^*(\theta_j)$, as

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

   where the CDF values are taken from step 6 using linear interpolation between whichever two strike values $K \in \{200, 202, \ldots, 4000\}$ are nearest to $S_{i,j,\max}$ and $S_{i,j,\min}$, respectively.

Steps 1 and 2 represent the only point of distinction between our procedure and that of Malz, who assumes access to a single implied-volatility schedule without considering put or call prices directly; our procedure is accordingly essentially identical to his. Note that we transform the option prices into Black–Scholes implied volatilities simply for purposes of fitting the cubic spline and then transform these implied volatilities back into call prices before calculating risk-neutral beliefs, so this procedure does *not* require the Black–Scholes model to be correct.[6] The clamped cubic spline proposed by Malz (2014), and used in step 3 above, is chosen to ensure that the call-price schedule obtained in step 5 is decreasing and convex with respect to the strike price outside the range of observable strike prices, as required under the restriction of no arbitrage. Violations of these restrictions *inside* the range of observable strikes, as observed infrequently in the data, generate negative implied risk-neutral probabilities; in any case that this occurs, we set the associated risk-neutral probability to 0.

As noted in step 3, the clamped spline is an *interpolating* spline, as it is restricted to pass through all the observed data points so that the fitted-value set $\{\hat{q}_{t,i,K}\}$ contains the original values $\{q_{t,i,K}\}$. Some alternative methods for measuring risk-neutral beliefs use smoothing splines that are not constrained to exhibit such interpolating behavior. We discuss one leading alternative method in Appendix C.9 below.

## C.7 Noise Estimation and Matching to $X^*$ Observations

As introduced in Section 5.2, we first estimate $\text{Var}(\epsilon_t) = \text{Var}(\epsilon_{t,i,j})$ separately for each combination of trading day $t$, expiration date $T_i$, and return state pair $j$ in our intraday sample.[7] Our ReMeDI estimator for this noise variance follows the replication code provided by Li and Linton (2022):

[5] That is, formally, $S_{i,j,\min} = R^f_{0_i,T_i} S_{T_0} \exp(\theta_j - 0.05)$ and $S_{i,j,\max} = R^f_{0_i,T_i} S_{0_i} \exp(\theta_j)$. For example, for excess return state $\theta_2$, we have $S_{i,j,\min} = R^f_{0_i,T_i} S_{0_i} \exp(-0.2)$ and $S_{i,j,\max} = R^f_{0_i,T_i} S_{T_0} \exp(-0.15)$.

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

[7] For this exercise, to increase our available observations, we do not condition on the ex post state being $\theta_j$ or $\theta_{j+1}$.

$\widehat{\mathrm{Var}}(\epsilon_t) = \frac{1}{N_{\epsilon,n}} \sum_{i=2k_n}^{N_{\epsilon,n}-k_n} (\widehat{\pi}^*_{t_i} - \widehat{\pi}^*_{t_i-2k_n})(\widehat{\pi}^*_{t_i} - \widehat{\pi}^*_{t_i+k_n})$. We select $k_n$ for each return state using the algorithm in Section F.1 of the Online Appendix of Li and Linton (2022).[8]

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

### C.8 Details on Bootstrap Confidence Interval Construction

Our block-bootstrap resampling procedure is described in Section 5.4, and we provide further details on how we construct one-sided confidence intervals for Table 3 here. Fixing a given $\overline{\phi}$, denote the point estimate for $\overline{e_i^{\mathrm{main}}(\phi)}$ by $\widehat{e}(\overline{\phi})$. The null that $\overline{e_i^{\mathrm{main}}(\phi)} = 0$ is rejected at the 5% level if $2\widehat{e}(\overline{\phi}) - e^*_{(0.95)}(\overline{\phi}) > 0$, where $e^*_{(0.95)}(\overline{\phi})$ is the 95$^{\text{th}}$ percentile of the bootstrap distribution of $\overline{e_i^{\mathrm{main}}(\phi)}$ statistics (i.e., it is rejected if it is outside of the one-sided 95% basic bootstrap CI for $\overline{e_i^{\mathrm{main}}(\phi)}$). We conduct this procedure for all possible $\overline{\phi}$ values, and we obtain $\widehat{\phi}_{LB} = \min_{\overline{\phi}}$ s.t. $2\widehat{e}(\overline{\phi}) - e^*_{(0.95)}(\overline{\phi}) \leqslant 0$.

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

We note that our bootstrap procedure fully preserves the groupings of return-state pairs (indexed by $j = 1, \ldots, J-1$) for each set of observations indexed by $i$ (corresponding to the option expiration date) within each block, as we split the observations into blocks only by time and not by return states. We do so in order to obtain valid inference for the aggregate value $\overline{\boldsymbol{\phi}}$, which uses observations for state pairs $(\theta_2, \theta_3), \ldots, (\theta_{J-2}, \theta_{J-1})$, in the face of arbitrary dependence for the observations across those state pairs and a fixed number of return states $J$ (whereas we assume $N \to \infty$, and further the number of blocks $B \to \infty$ according to a sequence such that $(T_N + 1)/B \to \infty$). In

[8]The Online Appendix of the published version of their paper only contains the first three appendix subsections; for the full appendix (and Section F.1), see the supplementary materials for the working-paper version of their paper, provided in Li and Linton (2021).

this way our procedure is in fact a *panel* (or *cluster*) *block bootstrap*; see, for example, Palm, Smeekes, and Urbain (2011). Lahiri (2003, Theorem 3.2) provides a weak condition on the strong mixing coefficient of the relevant stochastic process — in our case, $\{(X^*_{i,j}, \tilde{\pi}^*_{0,i,j}, \{\widehat{\text{Var}}(\epsilon_{t,i,j})\})_{t,j}\}_i$ — under which the blocks are asymptotically independent and the bootstrap distribution estimator is consistent for the true distribution under the asymptotics above, so that our test-inversion confidence intervals have asymptotic coverage probability of at least 95% for the population parameters of interest in the presence of nearly arbitrary (stationary) autocorrelation and heteroskedasticity.[9] This coverage rate may in fact be greater than 95% given that we are estimating lower bounds for the parameters of interest rather than the parameters themselves, and this motivates our use of one-sided rather than two-sided confidence intervals, as in Section 5.4.

### C.9 Robustness Tests for Main Empirical Results

As described in Section 5.5, we conduct a number of robustness checks to probe the sensitivity of our main results to alternative measurement choices.

As in the main text, our first robustness check defines index-return states with 2-percentage-point return ranges, $\theta_1 = -10\%$, $\theta_2 = -8\%$, $\ldots$, $\theta_{11} = 10\%$. All other aspects of the RN excess movement measurement (steps 1–6 in Appendix C.6) are unchanged from the baseline. We also re-estimate the market microstructure noise variance for this case (as well as for all subsequent robustness tests), since changing the definition of the return states affects the degree of noise in the resulting RN beliefs. The ReMeDI estimation procedure (and tuning parameters) follow the same procedure as described in Appendix C.7 exactly.

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 (and associated terminal index realizations) for which the options' Black–Scholes delta values are between 1 and 0.9, state 2 between 0.9 and 0.8, and so on. All other measurement details and noise estimation procedures are unchanged.

Third, as introduced in Appendix C.6, we consider smoothing splines that do not necessarily pass through the observed option prices perfectly, whereas our baseline method uses an interpolating spline that passes through observed prices by construction. In particular, we follow the approach proposed by Bliss and Panigirtzoglou (2004). After steps 1–2 in Appendix C.6, we fit a smoothing spline in implied volatility–delta space, weighted by option vega, following equation (3) of Bliss and Panigirtzoglou (2004). We use their baseline smoothing parameter of $\lambda = 0.99$. We extrapolate beyond the range of observed strikes by extrapolating the last observed implied volatilities following the procedure described starting at the bottom of p. 416 of Bliss and Panigirtzoglou (2004). In step 4, we evaluate the spline at a denser set of strikes than in the baseline ($K = 200$,

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

**Table C.1: Robustness Checks: Bound Estimates for Different Specifications**

| | | *Conservative Lower Bound for:* | |
|---|---|---|---|
| | | SDF Slope $\overline{\phi}$ | RRA $\overline{\gamma}$ |
| (1) | Finer (2 pp) Return-State Bins | $\infty$ | $\infty$ |
| | | [$\infty$] | [$\infty$] |
| (2) | Return-State Bins Based on Ex Ante Option Delta | $\infty$ | N/A |
| | | [$\infty$] | [N/A] |
| (3) | Smoothing Spline for Option Prices | 26.1 | 501 |
| | | [7.7] | [134] |
| (4) | Smoothing Spline with GEV Tails | 25.4 | 488 |
| | | [7.9] | [138] |

*Notes:* This table reports estimation results for the bound for excess movement in Proposition 6(ii), across four alternative measurement strategies for RN beliefs as compared to the baseline. Confidence interval lower bounds, based on 10,000 bootstrap samples, are provided in brackets under each point estimate. See the text of Appendix C.9 for implementation details for each robustness check. See the notes to Table 3 for details on the estimation. For row (2), we cannot use Proposition 4 to translate a given $\overline{\phi}$ to a relative risk aversion value $\overline{\gamma}$, as the delta-based return states are no longer defined in terms of constant log differences in terminal index values. All estimates use conditional means of noise-adjusted excess movement for all interior state pairs, with noise variance re-estimated for each robustness check.

200.1, 200.2, . . ., 4000) to obtain a smoother RN distribution.[10] We then follow steps 5–7 with the same 5 pp excess return states as in the baseline.

Finally, we use the same Bliss and Panigirtzoglou (2004) smoothing spline over the range of observed strike prices, but we then extrapolate using tails from the generalized extreme value (GEV) distribution, rather than the lognormal distribution (as is implied by extrapolating constant IVs). This follows Section 5 of Figlewski (2010). For both the left and the right tail (using the 2% and 5% quantile of the distribution from observed prices for the left tail, and the 92% and 95% quantile for the right tail), we minimize the sum of squared deviations of (i) the GEV-implied CDF from the CDF implied by the observed prices at the less-extreme pasting point, (ii) the GEV-implied PDF from the observed PDF at the less-extreme pasting point, and (iii) the GEV-implied PDF from the observed PDF at the more-extreme pasting point. This follows Figlewski (2010) exactly, and see his Section 5 for formal details.

For all four robustness tests, to provide a summary of our results without excessive detail, we provide overall estimates across all interior state pairs (as in the first line of Table 3). We present the results of these robustness tests in Table C.1, with the rows in the order in which each measurement strategy was introduced above. In all four cases, the takeaways from our baseline estimation in Table 3 are qualitatively unchanged (or, in rows (1)–(2), strengthened slightly). The smoothing spline enforces some degree of smoothness in the resulting RN distribution, as is reflected in the very-slightly-lower estimated values in row (3) as compared to our baseline. In spite of this,

[10] We evaluate the spline at the option delta values corresponding to these strikes, since the spline is fit in IV–delta space, and then in step 5 we translate both dimensions to obtain a smoothed call price–strike price schedule.

we continue to need extremely high relative risk aversion for the data to be consistent with our bounds. The lowest estimated SDF slope is obtained for the last robustness test, in which we fit a smoothing spline following to the observed strikes, and then extrapolate by pasting GEV tails onto the resulting RN distribution following Figlewski (2010). The GEV extrapolation seems to very slightly decrease the amount of excess variation outside of the range of observed strikes. Under the baseline, the lognormal tails have some small amount of excess variation over time due to the implied volatility at the highest and lowest observed strikes changing over time (extrapolating these IVs is what generates the lognormal tails), and this is limited slightly when using GEV tails to extrapolate. That said, the estimates are effectively the same in row (4) as they are in row (3).

Finally, in addition to the results shown in the table, the normalized excess-movement statistics $\overline{X}^*/\overline{u}_0^*$ for the noise-adjusted $X^*$ across the four specifications are estimated to be: (1) 193% (i.e., $\overline{X}^*/\overline{u}_0^* = 1.93$), (2) 177%, (3) 117%, and (4) 116%, respectively, as compared to the value of 123% shown in the first row of Table 1. We conclude again that our main results are effectively unchanged across measurement strategies.

## C.10 Details on Simulations of Option Pricing Models

This subsection describes the model simulations in Section 6.1 in more detail. We start with the Christoffersen, Heston, and Jacobs (2013), or CHJ, model. The physical dynamics for the spot price and conditional variance, as well as the SDF, are respectively

$$\log(S_{t+1}) = \log(S_t) + r^f + \left(\mu - \frac{1}{2}\right) h_{t+1} + \sqrt{h_{t+1}} z_t,$$

$$h_{t+1} = \omega + \rho h_t + \alpha \left(z_t - \lambda \sqrt{h_t}\right)^2,$$

$$M_T = M_t \left(\frac{S_T}{S_t}\right)^{\kappa} \exp\left(\delta(T-t) + \eta \sum_{s=t+1}^{T} \left(h_s + \xi(h_{T+1} - h_{t+1})\right)\right),$$

where $S_t$ is again the spot index price and $z_t$ is a standard normal innovation, and where (from CHJ's Appendix B) $\kappa = -(\mu - 1/2 + \lambda)(1 - 2\alpha\xi) + \lambda - 1/2$. As shown in CHJ's Proposition 1, this specification generates the following risk-neutral dynamics:

$$\log(S_{t+1}) = \log(S_t) + r^f - \frac{1}{2} h_t^* + \sqrt{h_t^*} z_t^*,$$

$$h_{t+1}^* = \omega^* + \rho h_{t-1}^* + \alpha^* \left(z_t^* - \lambda^* \sqrt{h_t^*}\right)^2,$$

where $z_t^*$ is standard normal and where $h_t^* = h_t/(1-2\alpha\xi)$, $\omega^* = \omega/(1-2\alpha\xi)$, $\alpha^* = \alpha/(1-2\alpha\xi)^2$, and $\lambda^* = \lambda - \kappa$. We use the same daily parameter values as in CHJ's baseline estimates reported in column 3 of their Table 4, and we report these parameter values in Table C.2.

Following the main text, we conduct 10,000 simulations of daily data. In each simulation, we initialize $S_0 = 100$ and set $h_0$ to its long-run mean of $(\omega + \alpha)/(1 - \rho - \alpha\lambda^2)$. We then draw standard

**Table C.2: CHJ Model Parameters Used in Simulations**

| $\omega$ | $\alpha$ | $\rho$ | $\lambda$ | $\mu$ | $(1-2\alpha\xi)^{-1}$ | $\omega^*$ | $\alpha^*$ | $\lambda^*$ | $r^f$ |
|---|---|---|---|---|---|---|---|---|---|
| 0 | $8.887 \times 10^{-7}$ | 0.756 | 515.57 | 1.594 | 1.2638 | 0 | $1.419 \times 10^{-6}$ | 409.32 | 5% |

*Notes:* Parameter values for CHJ model simulations are taken from CHJ's baseline estimates in column 3 of their Table 4. All relevant parameters are daily aside from the risk-free rate $r^f$, which is annualized.

normal $z_t$ for $t = 1, \ldots, 60{,}480$ days to obtain price and variance realizations. The first 10 years (2,520 trading days) are the burn-in sample. In the remaining days, we set option expiration dates to be spaced evenly every 3 months (63 trading days). The 230-year span of a given simulation is much greater than our empirical sample span because in the empirical data, option panels are generally overlapping (i.e., there are options for many expiration dates trading on the same date), while in the simulation, we consider non-overlapping panels. Using this 230-year span ensures that we have roughly the same number of $X^*$ observations — 1840, or two for each of the 920 option expiration dates (for the two state pairs bracketing the realized outcome) — as in our empirical sample. The first option panel (i.e., daily stream of option cross-sections with a fixed expiration date) starts on $t_{0,1} = 2,521$ (where the subscript "1" refers to the first stream), and these options' expiration date is $T_1 = t_{0,1} + 63$. On that expiration date, the next panel begins, expiring 63 days thereafter, and so on.

At the beginning of each option panel (on day $t_{0,i}$), given the spot price $S_{0,i}$, we determine the return space by partitioning the set of strike prices in terms of 5 pp intervals of log excess returns, exactly as in the main text:

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

with $R_{0_i,T_i} = \exp(5\% \times 0.25)$ (since the contracts run for three months, or one quarter). We then keep this same binning of the strike space until expiration, as in the main text.

On each trading date $t$ with expiration date $T$, we take the simulated spot price $S_t$ and volatility $h_t$; then, for each of the nine strike prices corresponding to the edges of the return-state bins,[11] we calculate the RN CDF as

$$\pi_t^*(R_T \leqslant K) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \mathrm{Re}\left(\frac{K^{-i\varphi} g^*_{t,T}(i\varphi)}{i\varphi}\right) d\varphi, \tag{C.5}$$

where $g^*_{t,T}(\cdot)$ is the conditional RN moment-generating function (MGF). (Note in this equation that $i$ now represents the imaginary unit rather than the option panel index, so we omit the option panel index subscript when referencing the expiration date $T$.) We use the solution for the conditional MGF provided in Appendix D of CHJ; see that appendix for the formulas. We calculate the integral in (C.5) using numerical quadrature.[12] The above representation of the RN CDF is from Heston and

[11] For example, if $S_{0,i} = 100$, then the set of strikes for which we calculate the RN CDF would be $K = 82.9$ (corresponding to a log excess return of -0.2), 87.2, and so on up through 123.7.

[12] Our code is adapted from the Christoffersen, Jacobs, and Jeon (2014) "GARCH Options Toolbox," and we thank the

Nandi (2000), equation (A12), and it allows us to compute RN probabilities directly without first solving for vanilla option prices. Given these RN CDF values, we then calculate RN probabilities for bin $j$ as $\pi^*_{t,j} = \pi^*_t(R_T \leqslant K_j) - \pi^*_t(R_T \leqslant K_{j-1})$, where $K_j$ is the strike price corresponding to the right edge of return state $j$. (For $j = 1$, we use $\pi^*_t(R_T \leqslant K_1)$; for $j = 10$, we use $1 - \pi^*_t(R_T \leqslant K_9)$.) On $t = T$, we place probability 1 in the realized return bin, and then we move to the next contract.

We then calculate RN binary beliefs, $X^*$, and $\mathsf{u}^*_0$ statistics for each expiration date (keeping only the two state pairs bracketing the realized return outcome, as in the main text). Using these, we calculate the simulation-sample means $\overline{\mathsf{X}}^*$ and $\overline{\mathsf{u}}^*_0$ over all expiration dates and interior state pairs, and we estimate our overall bound $\overline{\phi}$ using our conservative lower bound, exactly as in the empirical estimation. We then translate this to an RRA value $\overline{\gamma}$ as in the empirical estimation. We do this for each simulation and report averages across simulations in Table 4.

To calculate the true value for $\overline{\phi}_0$ reported in that table, we first calculate the average binary physical and RN probabilities $\widetilde{\pi}_{0,j}$ and $\widetilde{\pi}^*_{0,j}$. To do so, we partition the state space for volatility $h_t$ into 100 percentile bins (i.e., bin 1 contains the lowest 1% of volatility values $h_t$ for the whole simulation sample, and so on); for each volatility bin, we calculate the average starting value $\widetilde{\pi}^*_{0,j}$ (for the two state pairs bracketing that contract's realized outcome), and for the physical probability we calculate the percentage of panels for which return state $j$ is realized (conditional on $j$ or $j+1$, and again for each volatility bin). We then go contract by contract, and based on which volatility bin the starting volatility value $h_0$ falls in, we take that bin's average $\widetilde{\pi}_{0,j}$ and $\widetilde{\pi}^*_{0,j}$ values and calculate

$$\phi_{0,j} = \frac{\frac{\widetilde{\pi}^*_{0,j}}{1-\widetilde{\pi}^*_{0,j}}}{\frac{\widetilde{\pi}_{0,j}}{1-\widetilde{\pi}_{0,j}}},$$

as derived after equation (14). We then calculate $\overline{\phi}_{0,j} = \max_{\pi^*_{0,j}} \mathbb{E}[\phi_{0,j} \mid \pi^*_{0,j}]$ (from Proposition 6), where the maximum is taken over all contracts within this simulation and all interior state pairs $j$. The true value for $\overline{\phi}_0$ reported in the table is then the average of this value across all simulations. We convert that to an effective RRA value as above.

We move next to the stochastic volatility with correlated jumps (SVCJ) model originally formulated by Duffie, Pan, and Singleton (2000), which is cast in continuous time (and then simulated at a daily interval given the continuous-time dynamics). We follow the version of the model in Broadie, Chernov, and Johannes (2007). The physical dynamics for the spot price and variance are respectively

$$dS_t = S_t(r^f + \zeta_t)dt + S_t\sqrt{V_t}dW^s_t + d\left(\sum_{n=1}^{N_t} S_{\tau_{n-}}\left[e^{Z^s_n} - 1\right]\right) - S_t\lambda\overline{\mu}_s dt, \tag{C.6}$$

$$dV_t = \kappa_v(\varsigma_v - V_t)dt + \sigma_v\sqrt{V_t}dW^v_t + d\left(\sum_{n=1}^{N_t} Z^v_n\right), \tag{C.7}$$

authors for posting their code.

**Table C.3: SVCJ Model Parameters Used in Simulations**

| $\kappa_v$ | $\varsigma_v$ | $\sigma_v$ | $\rho_w$ | $\lambda$ | $\mu_s$ | $\sigma_s$ | $\mu_v$ | $\eta_v$ | $\mu_s^*$ | $\sigma_s^*$ | $\mu_v^*$ | $r^f$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.026 | 0.54 | 0.08 | −0.48 | 0.006 | −2.63% | 2.89% | 1.48 | 0 | −5.01% | 7.51% | 3.71 | 5% |

*Notes:* Physical parameter values (columns $\kappa_v$ through $\mu_v$) are taken from the estimates of Eraker, Johannes, and Polson (2003), reported in Table I (row "SVCJ EJP") of Broadie, Chernov, and Johannes (2007). Risk-neutral parameter values (the remaining columns) are taken from the preferred estimates of Broadie, Chernov, and Johannes (2007), reported in the last row of their Table IV (and discussed on p. 1478 of the text as "more reasonable" than other parameter combinations). For the risk-free rate, we use the same value as in the CHJ simulations. All relevant parameters are daily aside from the risk-free rate $r^f$, which is annualized.

where $W_t^s$ and $W_t^v$ are correlated brownian motions with $\mathbb{E}[W_t^s W_t^v] = \rho_w t$, $N_t$ is a Poisson process with arrival intensity $\lambda$, $Z_n^s | Z_n^v \sim \mathcal{N}(\mu_s + \rho_s Z_n^v, \sigma_s^2)$ are spot-price jumps, $Z_n^v \sim \exp(\mu_v)$ are volatility jumps, $\zeta_t = \eta_s V_t + \lambda \overline{\mu}_s - \lambda^* \overline{\mu}_s^*$ is the equity premium, $\overline{\mu}_s = \exp(\mu_s + \sigma_s^2/2) - 1$, and finally $\overline{\mu}_s^* = \exp(\mu_s^* + (\sigma_s^*)^2/2) - 1$, where $\mu_s^*$ and $\sigma_s^*$ are risk-neutral parameters.

We denote the risk-neutral diffusion and jump processes using $\mathbb{Q}$ for clarity (and risk-neutral parameters again using an asterisk). The risk-neutral dynamics for the spot price and variance are

$$dS_t = S_t r^f dt + S_t \sqrt{V_t} dW_t^s(\mathbb{Q}) + d\left( \sum_{n=1}^{N_t(\mathbb{Q})} S_{\tau_{n-}} \left[ e^{Z_n^s(\mathbb{Q})} - 1 \right] \right) - S_t \lambda^* \overline{\mu}_s^* dt,$$

$$dV_t = \kappa_v^* (\varsigma_v - V_t) dt + \sigma_v \sqrt{V_t} dW_t^v(\mathbb{Q}) + d\left( \sum_{n=1}^{N_t(\mathbb{Q})} Z_n^v(\mathbb{Q}) \right),$$

where $W_t^s(\mathbb{Q})$ and $W_t^v(\mathbb{Q})$ are $\mathbb{Q}$ Brownian motions with correlation $\rho_w$ as above, $N_t(\mathbb{Q})$ is a Poisson process with arrival intensity $\lambda^*$ set to $\lambda^* = \lambda$, $Z_n^s(\mathbb{Q}) | Z_n^v(\mathbb{Q}) \sim \mathcal{N}(\mu_s^*, (\sigma_s^*)^2)$ are $\mathbb{Q}$ price jumps, $Z_n^v(\mathbb{Q}) \sim \exp(\mu_v^*)$ are $\mathbb{Q}$ volatility jumps, and $\kappa_v^* = \kappa_v + \eta_v$.

We use the same 5% risk-free rate as used in the CHJ simulations. For the main parameters, we use estimates from Eraker, Johannes, and Polson (2003) for daily parameter values under the physical distribution, and we use the preferred estimates of Broadie, Chernov, and Johannes (2007) for parameters under the RN distribution. The Eraker, Johannes, and Polson (2003) physical parameters are in Table I (row "SVCJ EJP") of Broadie, Chernov, and Johannes (2007), and the RN parameters are in the last row of Table IV of Broadie, Chernov, and Johannes (2007). We report these parameter values in Table C.3.

For the simulations, we follow almost exactly the same procedure as described for the CHJ simulations (from the paragraph beginning "Following the main text, we conduct. . .," through the paragraph beginning "To calculate the true value. . ."). The only distinction is that we adapt the code to handle the more complex continuous-time setting efficiently. We build on the Fusari (2017) "Matlab Toolbox for Option Pricing," and we thank Nicola Fusari for sharing his code. We modify the toolbox to simulate daily observations of the physical continuous-time price and volatility processes described in (C.6)–(C.7). For option prices, the toolbox solves for prices efficiently using the Fourier cosine series expansion as proposed by Fang and Oosterlee (2009). We modify this part of the toolbox code in two ways. First, we solve for digital option prices (and thereby the RN CDF) directly using the expression at the top of p. 833 of Fang and Oosterlee (2009). Second, we rewrite the code to be compatible with MATLAB's GPU functionality, and we run our code on a GPU to decrease the simulations' run time to a manageable length. All other simulation details are the same as for the CHJ model described above, and results for both are reported in the main text.

### C.11 Details and Results of RN Excess Movement Regressions

As discussed in Section 6.2, we consider reduced-form evidence on the macroeconomic and financial correlates of RN excess movement, with results in Table C.4. The dependent variable in all cases is the monthly average noise-adjusted RN excess movement $X^*_{t,t+1,i,j}$, with the average calculated over all available expiration dates and interior state pairs for all trading days $t$ in a month.

For the dependent variables, from top to bottom in the table, option bid-ask spread is the volume-weighted average bid-ask spread for all S&P 500 options with less than a year to maturity and positive bid prices in the given month. Option volume is total monthly dollar trading volume in that same sample, detrended using an estimated exponential trend given the steady growth in option volume over the sample. The negative of net public OTM put buys is obtained from Constantinides and Lian (2021). (We thank these and subsequent authors for making the relevant data available.) We start with their net public OTM put buys. A line of work (including both their paper and Chen, Joslin, and Ni, 2019) argues that this is a measure of the tightness of intermediary constraints in the option market, with net public put buys decreasing at times of significantly heightened intermediary constraints. We therefore take the negative of this net buy variable, so that it is intuitively higher at times when there may be higher excess movement. We also take its 3-month moving average before using it as a regressor to eliminate apparent high-frequency noise, which otherwise renders the estimated coefficient close to zero.[13]

Continuing down the table, RN belief stream length is the average full-stream length $\overline{T_i}$ over all contracts $i$ active in that month. VIX$^2$ is calculated using the average VIX in the given month. The variance risk premium, following Bollerslev, Tauchen, and Zhou (2009), is VIX$^2$ minus realized variance, and we use the data provided by Lochstoer and Muir (2022) for this VRP. For the risk-aversion proxy and its volatility, we start by obtaining the proxy $ra_t^{BEX}$ from Bekaert, Engstrom, and Xu (2022) via Nancy Xu's website (`https://www.nancyxu.net/risk-aversion-index`). [14] We then take the sum of squared daily changes in $ra_t^{BEX}$ in a given month (winsorized at the 5th and 95th percentiles) to measure the volatility of this proxy. We obtain the monthly repurchase-adjusted log price-dividend ratio $pd_t$ from Nagel and Xu (2022), and we calculate the absolute value of its deviation from its sample mean $\overline{pd}$. The 12-month S&P 500 change is the log change in the S&P

[13]The resulting smoothed public buying pressure series looks quite similar to the net purchase variable plotted in Chen, Joslin, and Ni (2019). Note that the buying pressure–volatility relation was different in an earlier sample (as documented in Bollen and Whaley, 2004, and discussed by Chen, Joslin, and Ni, 2019), but for our sample, negative buying pressure appears positively related to intermediary constraints and option-market volatility.

[14]Using daily data, they estimate time-varying relative risk aversion $ra_t^{BEX}$ for a representative agent with habit-like preferences and preference shocks.

**Table C.4: Regressions for Monthly Average of RN Excess Movement**

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Option Bid-Ask Spread | 0.24 | | | | | | -0.01 |
| | [0.15] | | | | | | [0.11] |
| Option Volume | 0.07 | | | | | | -0.05 |
| | [0.09] | | | | | | [0.11] |
| Negative of Net Public OTM Put Buys | | 0.09 | | | | | 0.04 |
| | | [0.08] | | | | | [0.04] |
| RN Belief Stream Length | | | 0.28 | | | 0.16 | 0.17 |
| | | | [0.15] | | | [0.05] | [0.07] |
| $\text{VIX}^2$ | | | | 0.33 | | 0.57 | 0.59 |
| | | | | [0.16] | | [0.31] | [0.35] |
| Variance Risk Premium | | | | 0.40 | | | |
| | | | | [0.25] | | | |
| Volatility of Risk-Aversion Proxy | | | | 0.06 | | | |
| | | | | [0.10] | | | |
| Repurchase-Adjusted $\lvert pd_t - \overline{pd} \rvert$ | | | | | 0.38 | 0.19 | 0.19 |
| | | | | | [0.13] | [0.06] | [0.06] |
| 12-Month S&P 500 Return | | | | | 0.32 | 0.55 | 0.55 |
| | | | | | [0.17] | [0.23] | [0.22] |
| $R^2$ | 0.08 | 0.01 | 0.08 | 0.28 | 0.15 | 0.37 | 0.36 |
| Obs. | 264 | 246 | 264 | 264 | 264 | 264 | 246 |

*Notes:* Heteroskedasticity- and autocorrelation-robust standard errors are in brackets, calculated using the equal-weighted periodogram estimator with $0.4\,\text{Obs.}^{2/3} = 16$ degrees of freedom following Lazarus et al. (2018) and Lazarus, Lewis, and Stock (2021). Dependent variable in all regressions is the mean noise-adjusted $X^*_{t,t+1,i,j}$, for all available expiration dates and interior state pairs, over trading dates within a given month. All variables are signed so that they should intuitively comove positively with excess movement, and all variables are normalized to have unit standard deviation. All regressions include a constant. See Appendix C.11 for variable construction details.

price from month $t-12$ to $t$, using data from Robert Shiller's website (`http://www.econ.yale.edu/~shiller/data.htm`). All variables (both dependent and dependent) are normalized to have zero mean and standard deviation of 1, and all regressions include a constant.

The first column shows that proxies for option illiquidity and trading activity — namely, volume-weighted average monthly bid-ask spread in our options sample, and exponentially detrended option trading volume — are insignificant as predictors of $X^*$. Column (2) shows that net put buying pressure (a proxy for intermediary constraints) is also insignificant. These results provide evidence that option-market frictions are unlikely to be the main drivers of our main empirical results. Column (3) shows, however, that one economically meaningful factor specific to the option market *does* robustly predict excess movement: the average length of RN belief streams (i.e., $\overline{T_i}$ for contracts $i$ traded in the given month). As discussed in the main text, excess volatility seems to be concentrated at longer horizons.

Column (4) considers volatility-related predictors. Excess movement has a significant positive relationship with the (squared) VIX; a weak positive relationship with the variance risk premium, calculated as $\text{VIX}^2$ minus realized variance following Bollerslev, Tauchen, and Zhou (2009); and essentially no relationship with the volatility of Bekaert, Engstrom, and Xu's (2022) high-frequency risk-aversion proxy. This suggests, as in the text, that $X^*$ comoves strongly with the quantity of market uncertainty; slightly less strongly with the price of this uncertainty; and not at all with the volatility of risk aversion, which can be thought of as a proxy for $\text{Var}(\phi_t)$. We thus find no evidence, at least with this set of predictors, for meaningful comovement between variation in the price of risk and our measured RN excess movement.

Column (5) considers proxies for (mis)valuation and return reversals, in the form of the absolute deviation of the log repurchase-adjusted price-dividend ratio (from Nagel and Xu, 2022) and the trailing 12-month S&P return. Both are significantly positively related to $X^*$. As noted by Greenwood and Shleifer (2014), the trailing 12-month return predicts Gallup survey-based return expectations well, suggesting a plausible role here for similar survey expectations to predict excess movement. Column (6) considers all four predictors from (1)–(5) that are significant separately at the 10% level and shows that they remain significant jointly, and explain 37% of the variation in $X^*$. Column (7) adds back the predictors related to option-market frictions; they remain insignificant, while the other predictors from (6) retain their significance. Taken together, these results suggest that RN excess movement is a real phenomenon, as discussed further in the main text.

### C.12 Details of Quantification for Overall Index Volatility

This subsection describes the index-variance quantification exercise in Section 6.3 in more detail. We start with the RN histograms over 2 pp excess-return bins. These are the same bins as described in the first robustness test in Appendix C.9, but we use the full histogram (without noise adjustment) rather than the binarized noise-adjusted beliefs.[15] As usual, we define return bins as of date 0 (in terms of realized excess returns from date 0 to expiration date $T$) and then keep the bins the same for every trading date up to $T$. For example, if the starting index value is $S_0 = 1$ and the net risk-free rate is 0, outcome $R_T \in [-10\%, -8\%]$ is realized if $S_T$ is between 0.9 and 0.92.[16]

We then conduct the following calculations for each option expiration date $T$, given excess-return tail cutoffs of $[\underline{\theta}, \overline{\theta}]$. We will use the example of $[\underline{\theta}, \overline{\theta}] = [-10\%, 10\%]$ for concreteness in what follows, but equivalent calculations are done for all four cutoffs.[17] First, we assume all the mass of the distribution is concentrated at the midpoint of each bin. For example, if there is a date-0 RN probability of 8% that the excess return will be between -10% and -8%, we assume there is an 8% probability of the excess return being exactly -9%. Second, normalize $S_{0,\text{alt}} = 1/R^f_{0,T}$. This means that a net excess return of -9% will result in a final index value of $S_{T,\text{alt}} = 0.91$, for example; this is the midpoint of the index outcome for the excess-return bin $[-10\%, -8\%]$. And setting $[\underline{\theta}, \overline{\theta}] = [-10\%, 10\%]$ is equivalent to setting upper and lower limits for the terminal index value of $[\underline{S}, \overline{S}] = [0.9, 1.1]$.

[15] We remove any trading date containing an RN probability below the first percentile of the full distributions of RN probabilities in this data, to remove very negative observations (this cuts off anything below a -2% RN probability). For any remaining negative probabilities, we set them to 0 and renormalize the distribution we work with to sum to one.

[16] This is approximate, since we in fact define bins in terms of log excess returns, but we maintain this approximation throughout.

[17] Note that we abuse notation slightly in going back and forth between net and gross returns when defining the return state, but this does not substantively affect any of the analysis.

Next, for each trading date from 0 to $T$, we calculate $S_{t,\text{alt}} = \sum_{\underline{S} \leqslant s \leqslant \overline{S}} P_t^*(S_T = s | \underline{S} \leqslant S_T \leqslant \overline{S}) \cdot s$ using the risk-neutral histogram. For the $[\underline{\theta}, \overline{\theta}] = [-10\%, 10\%]$ example, the terminal states $s$ are given by $s = 0.91, 0.93, \ldots, 1.09$. We also, as of date 0, calculate the ex ante RN implied variance as $\mathrm{u}_0(S_{\text{alt}}) = \sum_{\underline{S} \leqslant s \leqslant \overline{S}} P_0^*(S_T = s | \underline{S} \leqslant S_T \leqslant \overline{S}) \cdot (s - S_{0,\text{alt}})^2$. We then compute tail-free index movement and excess movement exactly as in the text.

For the full distribution without tail cutoffs (the case of $S_{t,\text{alt}} = S_t$), we must make an assumption about where the expected value given a tail realization is. We set it somewhat arbitrarily to be at an excess return of $-15\%$ for the left tail, and 15% for the right tail, since we do not find that alternative choices make a meaningful difference. So we assume that all of the probability mass in the left tail, for example, tells us the RN probability that the excess return will be exactly equal to $-15\%$ (and $S_{T,\text{alt}}$ will be equal to 0.85). Given this, all calculations are the same as the case with tail cutoffs as above.

As in the main analysis, we keep streams with $T > 4$ to avoid very short panels. We then measure the empirical average movement and initial uncertainty and calculate our statistic $\overline{X}(S_{\text{alt}})/\overline{\mathrm{u}}_0(S_{\text{alt}})$ (and, for the analysis in the last paragraph of Section 6.3, $\overline{X}/\overline{\mathrm{m}}$). For both statistics, we calculate bootstrap standard errors using the same block bootstrap as used for Table 1. We classify observations of movement, initial uncertainty, and excess movement based on the month in which they expire, and then we redraw blocks of data with block size of one month (with replacement) until our bootstrap sample contains as many observations as our empirical sample. We then recalculate each statistic ($\overline{X}/\overline{\mathrm{u}}_0$ and $\overline{X}/\overline{\mathrm{m}}$) in that bootstrap sample. We redo this exercise for 10,000 bootstrap samples, and we report standard errors as the standard deviation of each statistic over the bootstrap draws. The table and text then report the results of this analysis.

## Appendix References


ATMAZ, A. AND S. BASAK (2018): "Belief Dispersion in the Stock Market," *Journal of Finance*, 73, 1225–1279.

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

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

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

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

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

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

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

BOLLERSLEV, T., G. TAUCHEN, AND H. ZHOU (2009): "Expected Stock Returns and Variance Risk Premia," *Review of Financial Studies*, 22, 4463–4492.

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.

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

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

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

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

CHEN, H., S. JOSLIN, AND S. X. NI (2019): "Demand for Crash Insurance, Intermediary Constraints, and Risk Premia in Financial Markets," *Review of Financial Studies*, 32, 228–265.

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

CHRISTOFFERSEN, P., K. JACOBS, AND Y. JEON (2014): *GARCH Options Toolbox*.

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

CONSTANTINIDES, G. M. AND L. LIAN (2021): "The Supply and Demand of S&P 500 Put Options," *Critical Finance Review*, 10, 1–20.

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.

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

FANG, F. AND C. W. OOSTERLEE (2009): "A Novel Pricing Method for European Options Based on Fourier-Cosine Series Expansions," *SIAM Journal on Scientific Computing*, 31, 826–848.

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.

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

HANSEN, L. P., J. C. HEATON, AND N. LI (2008): "Consumption Strikes Back? Measuring Long-Run Risk," *Journal of Political Economy*, 116, 260–302.

HESTON, S. L., K. JACOBS, AND H. J. KIM (2024): "A New Closed-Form Discrete-Time Option Pricing Model with Stochastic Volatility," *Working Paper*.

HESTON, S. L. AND S. NANDI (2000): "A Closed-Form GARCH Option Valuation Model," *Review of Financial Studies*, 13, 585–625.

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

KOCHERLAKOTA, N. R. (1990): "Disentangling the Coefficient of Relative Risk Aversion from the Elasticity of Intertemporal Substitution: An Irrelevance Result," *Journal of Finance*, 45, 175–190.

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

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

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

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

LI, Z. M. AND O. B. LINTON (2021): "Supplementary Materials for 'A ReMeDI for Microstructure Noise'," *Working Paper*.

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

LOCHSTOER, L. A. AND T. MUIR (2022): "Volatility Expectations and Returns," *Journal of Finance*, 77, 1055–1096.

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

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

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

NAGEL, S. AND Z. XU (2022): "Asset Pricing with Fading Memory," *Review of Financial Studies*, 35, 2190–2245.

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

VRINS, F. (2018): "Sampling the Multivariate Standard Normal Distribution Under a Weighted Sum Constraint," *Risks*, 6, 1–13.