Learning with rare disasters*

Jessica A. Wachter[1,2] and Yicheng Zhu[1]

[1]University of Pennsylvania
[2]NBER

September 5, 2019

Abstract

Financial crises appear to have long-lasting effects, even after the crisis itself has past. This paper offers a simple explanation through Bayesian learning from rare events. Agents face a latent and time-varying probability of economic disaster. When a disaster occurs, learning results in greater effects on asset prices because agents update their probability of future disasters. Moreover, agents' belief that the disaster risk is high can rationally persist for years, even when it is in fact low. We generalize the model to allow for a noisy signal of the disaster probability. This generalized model explains excess stock market volatility together with negative skewness, effects that previous models in the literature struggle to explain.

*We thank Domenico Cuoco, Tyler Muir, Sergio Salgado Ibanez, and participants at the Wharton School and at the 2019 WFA meetings for helpful comments.

1 Introduction

Rare events are hard to learn about. While existing models use rare events to account for the equity premium, high equity volatility, option prices, and the cross-section of returns, they tend to abstract from the learning process. Yet the standard assumption of rational expectations equilibrium, in which the agent operates under full information, cannot be expected to hold for rare events. An economy with rare events and incomplete information can differ in fundamental ways from an economy in which information is (unrealistically) perfect.

In this paper, we consider the point of view of a Bayesian investor who learns about the probability of a rare event from realizations, and also from a noisy signal. The underlying probability of the rare event varies over time, so that the agent does not ever have full information about the rare event probability. Despite the complexity of the problem, we can derive analytical solutions. These analytical solutions show that learning can account for important aspects of the data that the full-information case cannot. First, learning generates a greater equity premium, because the realization of a disaster coincides teaches the agent that future disasters can occur. Embedding these expectations into stock prices leads these to fall by still more. This is important because a persistent criticism of rare events as an explanation of the equity premium is that rare events are simply not large enough to generate the needed effect.[1] Moreover, we show that a stock price decline can substantially exceed that of consumption during a disaster period, as shown in Muir (2017). Second, learning creates an extended recovery from disasters during which valuations are depressed, and precautionary savings remain high, possibly for years after the disaster has occurred. This effect is well-documented in the literature (Reinhart and Rogoff, 2009).

Our analytical approach also allows us to incorporate the effect of noisy signals concerning disaster. These noisy signals generate high volatility in stock prices, even when disasters do not occur. Moreover, the signals generate relatively small discontinuities in asset prices during normal times. The existence of such jumps is emphasized in a large empirical literature that specifies a reduced-form model of the pricing kernel (Broadie et al., 2007). They also explain a fact about aggregate market returns that have remained out of reach of benchmark asset pricing models: negative skewness in aggregate market returns.

[1]See Longstaff and Piazzesi (2004) and Mehra and Prescott (1988).

Our contribution relates to several recent strands of literature. The idea of a disaster that causes long-run effects through belief shifts is present also in Hennessy and Radnaev (2016), Kozlowski et al. (2018), and Moreira and Savov (2017). Hennessy and Radnaev (2016) focus on leverage effects in a production economy; Kozlowski et al. (2018) focus on the riskless rate. In their paper, learning is non-Bayesian. Moreira and Savov (2017) introduce learning about crash risk to a setting with intermediated assets; we show that some of their effects do not require an intermediary structure but are present in a frictionless economy. These papers do not present analytical results or discuss implications for skewness. The relation between learning and skewness is present in work of Schmalz and Zhuk (2018), who assume a partial equilibrium setting and focus on the cross-section. Finally, our work addresses the question of whether rationally anticipated learning has first-order effects. Collin-Dufresne et al. (2016) and Cogley and Sargent (2008) offer different perspectives on this issue. We show that learning can indeed generate an ex ante equity premium that is noticeably larger than otherwise, when it is rare events that are the subject of the learning.

2 The Model

2.1 Endowment and preferences

We assume an endowment economy with an infinitely-lived representative agent. The aggregate consumption (endowment) process is given by

$$\frac{dC_t}{C_{t^-}} = \mu_C dt + \sigma_C dB_{Ct} + (e^{-Z_t} - 1)dN_{1t}, \tag{1}$$

where B_{Ct} is the standard Brownian motion and N_{1t} is a Poisson process. The Brownian motion $\mu_C dt + \sigma_C dB_{Ct}$ summarizes normal time consumption growth, while the Poisson term $(e^{-Z_t} - 1)dN_{1t}$ captures disasters. The random variable Z_t is the change in log consumption when disasters occur. For tractability, we assume that Z_t follows an i.i.d process with distribution denoted by ν. We use the notation E_ν to denote the expectation taken with respect to the distribution of ν. The processes, B_{Ct}, N_{1t} and Z_t are assumed to be independent. The intensity of N_{1t} equals λ_{1t}. In what follows, we will assume that, while the agent perfectly observes consumption, λ_{1t} is latent. We model the learning problem about λ_{1t} in next section.

We assume the representative agent has recursive utility with EIS equal to 1, and use the continuous-time characterization of Epstein and Zin (1989) utility derived by Duffie and Epstein (1992):

$$V_t = \max E_t \int_t^\infty f(C_s, V_s) ds, \tag{2}$$

where

$$f(C_t, V_t) = \beta(1-\gamma)V_t \left(\log C_t - \frac{1}{1-\gamma} \log\left((1-\gamma)V_t\right) \right). \tag{3}$$

Here β represents agent's time-preference, and γ is the risk-aversion. When $\gamma > 1$, agents show preference for early resolution of uncertainty.

2.2 The processes for conditional jump intensity and learning

The conditional probability of disasters follows a Markovian regime switch model. We assume that the agent learns the probability of disasters from the realization of disasters. We also allow the agent to learn in another way: from signals. Because disasters occur rarely, these signals could be quite important. A signal might, for example, be a disaster realization elsewhere (if we consider this a model for the US economy, it could be a realized disaster in a foreign country).

We model learning from signals by assuming a second Poisson process, N_{2t}. We use λ_{2t} to denote the jump intensity of N_{2t} at time t.[2] Define

$$\lambda_t = [\lambda_{1t}, \lambda_{2t}]^\top.$$

Furthermore, assume two states, so that

$$\lambda^k = \left[\lambda_1^k, \lambda_2^k\right]^\top \qquad k \in \{L, H\}.$$

λ_t switches between λ^H and λ^L according to:

$$\begin{aligned} \Pr\left(\lambda_{t+dt} = \lambda^H | \lambda_t = \lambda^L\right) &= \phi_{L\to H} dt \\ \Pr\left(\lambda_{t+dt} = \lambda^L | \lambda_t = \lambda^H\right) &= \phi_{H\to L} dt. \end{aligned} \tag{4}$$

[2]This model assumes random arrival of a signal. Another interesting case is one in which the signal arrives at fixed times (pre-scheduled announcements). We explore this case in a companion paper (Wachter and Zhu, 2017).

The physical probability of a switch is independent of information other than λ_t. Moreover, the λ_1-state and the λ_2-state are perfectly correlated, an assumption that we make for simplicity. We assume that $\lambda_1^H > \lambda_1^L \geq 0$. [3] Note that state H is a high-risk state, because this is where the probability of economic disaster is highest. While nothing in our framework requires that $\lambda_2^H \geq \lambda_2^L \geq 0$ (namely that signals are more likely when disaster risk is high), we nonetheless make this intuitive assumption to fix ideas.

We use p_t to denote agent's *posterior* belief in the high-risk state, i.e.,

$$p_t \equiv \Pr\left(\lambda_t = \lambda^H | \mathcal{F}_t\right), \tag{5}$$

where $\mathcal{F}_t$ is agent's information set up to the observation at time t. We define notation for the agents' posterior jump intensities.

$$\begin{aligned} \bar{\lambda}_i(p_t) &\equiv \lambda_i^H p_t + \lambda_i^L (1 - p_t), \quad i = 1, 2 \\ \bar{\lambda}(p_t) &\equiv \lambda^H p_t + \lambda^L (1 - p_t). \end{aligned} \tag{6}$$

We assume that the agent leans about the regime through the realization of rare events, N_{1t} and N_{2t}. The following theorem gives provides the stochastic differential equation that characterizes the evolution of the posterior probability a Bayesian agent assigns on the high-risk state.

Theorem 1. The agent's posterior belief in the high-risk state, p_t, defined by (5), evolves according to

$$\begin{aligned} dp_t = & \underbrace{\left(\phi_{L\to H} - p_{t^-}(\phi_{H\to L} + \phi_{L\to H})\right) dt}_{(7.1)} + \underbrace{\iota^\top \left(\bar{\lambda}(p_{t^-}) - \lambda^H\right) p_{t^-} dt}_{(7.2)} \\ & + \underbrace{\left(\frac{\lambda_1^H - \bar{\lambda}_1(p_{t^-})}{\bar{\lambda}_1(p_{t^-})}\right) p_{t^-} dN_{1t} + \left(\frac{\lambda_2^H - \bar{\lambda}_2(p_{t^-})}{\bar{\lambda}_2(p_{t^-})}\right) p_{t^-} dN_{2t}}_{(7.3)}. \end{aligned} \tag{7}$$

where $\iota = [1, 1]^\top$.[4]

[3] Our 2-state Markov-switching framework follows assumptions of Benzoni et al. (2011). The literature on equilibrium models with learning about latent regimes include Lettau et al. (2008), David and Veronesi (2013), and Dergunov et al. (2018). The model is in contrast to that of Koulovatianos and Wieland (2011), who assume a transitory state for the disaster probability. Koulovatianos and Weiland also focus on the case with time-additive, as opposed to recursive utility.

[4] For simplicity, we assume that all types of rare events are equally easy to learn about. An

Proof. See Appendix A. □

What causes p_t to vary? The first term (7.1) reflects the physical drift in regime. Conditional on being in a low-risk state, the economy exists to the high state with probability $\phi_{L\to H}dt$; conditional on being in a high-risk state, the economy shifts to a low-risk state with probability $\phi_{H\to L}dt$ (this term appears with a negative sign because it represents a decrease in p_t). The shift in p_t then represents a weighted average:

$$(1 - p_{t^-})\phi_{L\to H}\, dt + p_{t^-}(-\phi_{H\to L})\, dt = (\phi_{L\to H} - p_{t^-}(\phi_{H\to L} + \phi_{L\to H}))\; dt$$

The terms (7.2) and (7.3) reflect learning from observations on the Poisson events. Note that

$$(\bar{\lambda}(p_t) - \lambda^H)p_t = (\lambda^L - \lambda^H)p_t(1 - p_t),$$

so these terms depend on the product of p_t with $1 - p_t$. Thus there is a nonlinear effect of p_t on learning. Learning is fastest when the agent is not certain, namely p_t is furthest away from 0 and from 1 (Veronesi, 1999).

The term (7.2) represents learning from the absence of Poisson shocks. This term is negative because $\bar{\lambda}_i(p_t) \le \lambda_i^H$. When nothing happens, the agent shifts his belief toward the low-risk state.[5] Though it appears that there is no news, the agent is still learning. In a precise sense, "no news is good news" (Campbell and Hentschel, 1992).

Finally, (7.3) captures the direct learning from Poisson arrivals. If either shock occurs, the agent updates the belief in favor of the high-risk state. Specializing to the case of $\lambda_j^L = 0$, note that the agent updates her probability from p_t to 1 should shock N_j occur.

---

interesting extension would be to allow some events to be easier to learn about than others. This might explain differential responses of risk premia to wars versus financial crises (Muir, 2017).

[5]Note that the process $N_{1t} + N_{2t}$ is itself a Poisson process with conditional jump intensity given by $\iota^\top \lambda_t$. When there is no Poisson arrival from N_{1t} or N_{2t}, the agent learns as if there is no Poisson arrival from $N_{1t} + N_{2t}$.

Note that, we can also rewrite (7) as

$$dp_t = \phi_{L\to H} - p_{t^-}\left(\phi_{H\to L} + \phi_{L\to H}\right)\right) dt + \left(\frac{\lambda_1^H - \lambda_1^L}{\bar{\lambda}_1(p_{t^-})}\right) p_{t^-}(1-p_{t^-})\left(dN_{1t} - \bar{\lambda}_1(p_{t^-})dt\right) + \left(\frac{\lambda_2^H - \lambda_2^L}{\bar{\lambda}_2(p_{t^-})}\right) p_{t^-}(1-p_{t^-})\left(dN_{2t} - \bar{\lambda}_2(p_{t^-})dt\right). \quad (8)$$

Equation 8 provides other characterizations of the learning process. First, when $p_{t^-} = 0$ or 1, the agent is certain about the current state, and there is no effect from learning. Meanwhile, the larger the difference between λ^H and λ^L, the stronger the effect of learning as the likelihood of a rare event in the high-risk state is higher compared to the low-risk state. Finally, as $\bar{\lambda}_1(p_{t^-})dt$ and $\bar{\lambda}_2(p_{t^-})dt$ are the agent's expected probability of Poisson jumps, the agent's expected change of p_t is $\phi_{L\to H} - p_{t^-}\left(\phi_{H\to L} + \phi_{L\to H}\right)) dt$, reflecting only the effect from physical dynamics of regime switch and implying that the learning should be unbiased.

2.3 The state-price density

We value streams of future cash flows using the state-price density process, which we will call π_t. The process is uniquely determined by the representative agent's utility and endowment process. We solve for the state-price density by characterizing the representative agent's value function first.

Proposition 1. The representative agent's continuation value V_t is given by

$$V_t = J(C_t, p_t),$$

where

$$J(C, p) = \frac{1}{1-\gamma} C^{1-\gamma} e^{(1-\gamma)j(p)}. \quad (9)$$

The function $j(p)$ is continuously differentiable and solves the ordinary differential equation (ODE):

$$j'(p) = \Bigg(\beta j(p) - \mu_C + \frac{1}{2}\gamma\sigma^2 - \frac{\bar{\lambda}_1(p)}{1-\gamma}\left(E_\nu\left[e^{(1-\gamma)Z}\right] e^{(1-\gamma)\left(j(p\lambda_1^H/\bar{\lambda}_1(p))-j(p)\right)} - 1\right)$$
$$- \frac{\bar{\lambda}_2(p)}{1-\gamma}\left(e^{(1-\gamma)\left(j(p\lambda_2^H/\bar{\lambda}_2(p))-j(p)\right)} - 1\right)\Bigg)$$
$$\times \left(\phi_{L\to H} - p(\phi_{H\to L} + \phi_{L\to H}) - p\iota^\top\left(\lambda^H - \bar{\lambda}(p)\right)\right)^{-1}, \quad (10)$$

with boundary condition

$$\beta j(p^*) - \mu_C + \frac{1}{2}\gamma\sigma^2$$
$$- \frac{\bar{\lambda}_1(p^*)}{1-\gamma}\left(E_\nu\left[e^{(1-\gamma)Z}\right] e^{(1-\gamma)\left(j(p^*\lambda_1^H/\bar{\lambda}_1(p^*))-j(p^*)\right)} - 1\right)$$
$$- \frac{\bar{\lambda}_2(p^*)}{1-\gamma}\left(e^{(1-\gamma)\left(j(p^*\lambda_2^H/\bar{\lambda}_2(p^*))-j(p^*)\right)} - 1\right) = 0, \quad (11)$$

for

$$p^* = \Big((\iota^\top(\lambda^H - \lambda^L) + \phi_{H\to L} + \phi_{L\to H})$$
$$- \sqrt{(\iota^\top(\lambda^H - \lambda^L) + \phi_{H\to L} + \phi_{L\to H})^2 - 4\iota^\top(\lambda^H - \lambda^L)\phi_{L\to H}}\Big)$$
$$\times \left(2\iota^\top(\lambda^H - \lambda^L)\right)^{-1}. \quad (12)$$

Proof. See Appendix B. □

The probability p^* has an economic interpretation. As time goes by without a Poisson realization N_t, the agent learns from the absence of Poisson realizations. At a certain point, however, belief update from learning will cancel with the belief update due to the knowledge of physical dynamics. This is the point at which the drift in (7) equals zero:

$$(\phi_{L\to H} - p^*(\phi_{H\to L} + \phi_{L\to H})) + \iota^\top\left(\bar{\lambda}(p^*) - \lambda^H\right)p^* = 0,$$

or equivalently:

$$\iota^\top(\lambda^H - \lambda^L)p^{*2} - (\iota^\top(\lambda^H - \lambda^L) + (\phi_{H\to L} + \phi_{L\to H}) +)p^* + \phi_{L\to H} = 0. \quad (13)$$

Note that (13) is a quadratic equation in p_t. The Equation 12 gives the unique root between 0 and 1. Furthermore, we can show that, if $\iota^\top \lambda^H \geq \iota^\top \lambda^L$

$$p^* \leq \frac{\phi_{L\to H}}{\phi_{H\to L} + \phi_{L\to H}}. \tag{14}$$

The right-hand side of (14) is the unconditional probability of the high-risk state. The physical dynamics of the Markov process pushes the agents' belief towards this probability. On the other hand, the absence of rare events pushes the agent's belief towards zero. When $p_t = p^*$, these forces exactly cancel each other out. Besides the economic intuition, p^* enables us to define a boundary condition for the ODE (10), which helps to accurately compute the function $j(p)$.

Given the representative agent's continuation value, we can then solve the unique state price density process, π_t, of the economy using the result, due to Duffie and Skiadas (1994) that

$$\pi_t = \exp\left\{\int_0^t \frac{\partial}{\partial V} f(C_s, V_s) ds\right\} \frac{\partial}{\partial C} f(C_t, V_t) \tag{15}$$

and therefore

$$\pi_t = \text{locally deterministic term } \times C_t^{-\gamma} e^{(1-\gamma)j(p_t)}. \tag{16}$$

To price assets, and understand risk premia, it is convenient to write down the stochastic process for π_t that (15) implies.

Theorem 2. The state-price density, π_t, solves the stochastic differential equation

$$\begin{aligned}\frac{d\pi_t}{\pi_{t-}} = \mu_{\pi t^-} dt + \sigma_{\pi t^-} dB_{Ct} &\\ + \underbrace{\left(e^{\gamma Z_t} e^{(1-\gamma)\left(j(p_{t^-}\lambda_1^H/\bar{\lambda}_1(p_{t^-})) - j(p_{t^-})\right)} - 1\right) dN_{1t}}_{(17.1)} &\\ + \underbrace{\left(e^{(1-\gamma)\left(j(p_{t^-}\lambda_2^H/\bar{\lambda}_2(p_{t^-})) - j(p_{t^-})\right)} - 1\right) dN_{2t}}_{(17.2)}, &\end{aligned} \tag{17}$$

for $\mu_{\pi t^-} = \mu_\pi(p_{t^-})$, $\sigma_{\pi t^-} = \sigma_\pi(p_{t^-})$, and

$$\mu_\pi(p) = -\left(\beta + \mu_C - \gamma\sigma_C^2\right) - \bar{\lambda}_1(p) E_\nu \left[e^{(\gamma-1)Z} e^{(1-\gamma)\left(j(p\lambda_1^H/\bar{\lambda}_1(p)) - j(p)\right)} - 1 \right] - \bar{\lambda}_2(p) \left(e^{(1-\gamma)\left(j(p\lambda_2^H/\bar{\lambda}_2(p)) - j(p)\right)} - 1 \right) \tag{18}$$

$$\sigma_\pi(p) = -\gamma\sigma_C, \tag{19}$$

where $j(p)$ is a continuous differentiable function characterized by ODE (10) and boundary condition (11).

Proof. See Appendix B. □

The mean growth rate is (as usual) the riskfree rate, r_t, which we characterize in what follows. The term $-\gamma\sigma_C dB_{Ct}$ captures the effect of diffusive shocks to the consumption growth: a positive shock dB_{Ct} increases the agent's consumption level, and decreases the agent's marginal utility.

The term (17.1) captures the effect on the marginal utility from a disaster realization. Note that $e^{\gamma Z_t}$ gives the *direct* effect of a disaster: when a disaster hits, economy-wide consumption falls, directly raising marginal utility. It is this term that is responsible for the equity premium in models such as Barro (2006) and Rietz (1988). There is also an indirect effect on marginal utility, given in the term $e^{(1-\gamma)\left(j(p_{t^-}\lambda_1^H/\bar{\lambda}_1(p_{t^-})) - j(p_{t^-})\right)}$. This term enters multiplicatively, and thus amplifies the first. While superficially complicated, this term has a clear economic interpretation. First, note that if a disaster occurs, the agent updates her probability of the high-risk state from p to $p\lambda_1^H/\bar{\lambda}_1(p)$. We can see this directly from the law of motion (7). Given this change in p_t, the change to marginal utility follows from (15) and Proposition 1. The disaster increases marginal utility both directly, through its effect on consumption, and indirectly, through a slower resolution of uncertainty (due to the shift in probability) Note that when $\gamma = 1$, this term equals one, and the equation implies a rare-events form of the standard Breeden (1979) Consumption CAPM, namely only consumption risk is priced. The term (17.2) can be understood similarly, except in this case there is no direct effect on consumption.

The next theorem characterizes the equilibrium riskfree rate of the economy.

Theorem 3. The instantaneous riskfree rate r_{ft}, is given by $r_{ft} = r_f(p_t)$, where

$$r_f(p) = \beta + \mu_C - \gamma\sigma_C^2 + \bar{\lambda}_1(p) e^{(1-\gamma)\left(j(p\lambda_1^H/\bar{\lambda}_1(p)) - j(p)\right)} E_\nu \left[e^{\gamma Z} \left(e^{-Z} - 1 \right) \right]. \tag{20}$$

Proof. See Appendix B □

As in the work of Barro (2006), precautionary savings due to rare disasters lower the riskfree rate compared to what it would have been in a standard model. When the agent has a preference for early resolution of uncertainty, and when the agent must learn about the probability of a rare disaster, the riskfree rate is lower still. The agent fears disasters still more because of the change in her perception of the world that they bring about.[6]

2.4 Pricing equity

We model equity as the claim to a dividend process specified below:

$$\frac{dD_t}{D_{t^-}} = \mu_D dt + \varphi \sigma_C dB_{Ct} + \left(e^{-\varphi Z_t} - 1\right) dN_{1t}, \tag{21}$$

where φ is the dividend stream's leverage with respect to aggregate consumption growth. If we let S_t denote the time-t price of such claim, non-arbitrage implies that

$$S_t = \int_{s=0}^{\infty} E_t \left[\frac{\pi_{t+s}}{\pi_t} D_{t+s} \right] ds. \tag{22}$$

We price the equity claim by solving for the prices of individual future dividend payments, or dividend strips, first. We recursively solve for the price using an partial differential equation (PDE) which we show has a unique solution.

Theorem 4. The time-t price of an equity strip maturing at time $t+s$, scaled by current dividend D_t, is given by

$$E_t \left[\frac{\pi_{t+s}}{\pi_t} \frac{D_{t+s}}{D_t} \right] = G(p_t, s) = \exp(g(p_t, s)), \tag{23}$$

[6] Interestingly, the probability of a disaster enters into the riskfree rate. This does not usually happen for unitary EIS. The standard result is that, under unitary EIS, the riskfree rate depends solely on the distribution of consumption. In this case, however, the change in posterior parameters are perfectly correlated with the change in consumption, which is why the term appears.

where the continuously differentiable function $g(p,s)$ solves the following PDE:

$$\frac{\partial g}{\partial s} - \frac{\partial g}{\partial p}\left[\phi_{L\to H} - p(\phi_{L\to H}+\phi_{H\to L}) - p\left(\lambda^H - \bar{\lambda}(p)\right)\right]$$
$$= -\beta + \mu_D - \mu_C + \gamma(1-\varphi)\sigma_C^2$$
$$+ \underbrace{\bar{\lambda}_1(p)e^{(1-\gamma)\left(j(p\lambda_1^H/\bar{\lambda}_1(p))-j(p)\right)}E_\nu\left[e^{g(p\lambda_1^H/\bar{\lambda}_1(p),s)-g(p,s)+(\gamma-\varphi)Z} - e^{(\gamma-1)Z}\right]}_{(24.1)}$$
$$+ \underbrace{\bar{\lambda}_2(p)e^{(1-\gamma)\left(j(p\lambda_2^H/\bar{\lambda}_2(p))-j(p)\right)}\left(e^{g(p\lambda_2^H/\bar{\lambda}_2(p),s)-g(p,s)} - 1\right)}_{(24.2)} \quad (24)$$

with boundary condition

$$g(p,0) = 0, \qquad \forall p \in [0,1]. \quad (25)$$

Proof. See Appendix C □

To better understand the economic intuition of (24), it is helpful to consider the case when $p = p^*$ defined by (12). When $p = p^*$, p stops drifting without the realization of rare events, and the P.D.E (24) reduces to an O.D.E with respect to s:

$$\frac{\partial g}{\partial s}(p^*,s) = \underbrace{-\beta + \mu_D - \mu_C + \gamma(1-\varphi)\sigma_C^2}_{(26.1)}$$
$$\underbrace{+\bar{\lambda}_1(p^*)e^{(1-\gamma)\left(j(p^*\lambda_1^H/\bar{\lambda}_1(p^*))-j(p^*)\right)}E_\nu\left[e^{g(p^*\lambda_1^H/\bar{\lambda}_1(p^*),s)-g(p^*,s)+(\gamma-\varphi)Z} - e^{(\gamma-1)Z}\right]}_{(26.2)}$$
$$\underbrace{+\bar{\lambda}_2(p^*)e^{(1-\gamma)\left(j(p^*\lambda_2^H/\bar{\lambda}_2(p^*))-j(p^*)\right)}\left(e^{g(p^*\lambda_2^H/\bar{\lambda}_2(p^*),s)-g(p^*,s)} - 1\right)}_{(26.3)}, \quad (26)$$

with boundary condition $g(p^*,0) = 0$.

The function $g(p,s)$ is defined as the log price-dividend ratio of an equity strip. (26) summarizes the marginal effect of having one additional unit of maturity on the pricing of equity strip. Specifically, we can reorganize (26) and obtain the following equation:

$$
\begin{aligned}
&\frac{\partial g}{\partial s}(p^*, s)\\
&= -\beta - \mu_C + \gamma\sigma_C^2 - e^{(1-\gamma)\left(j(p^*\lambda_1^H/\bar{\lambda}_1(p^*)) - j(p^*)\right)} E_\nu\left[e^{(\gamma-1)Z} - e^{\gamma Z}\right]\\
&\qquad - \gamma\varphi\sigma_C^2 + \bar{\lambda}_1(p^*) E_\nu\left[\left(\frac{e^{-\varphi Z} e^{g(p^*\lambda_1^H/\bar{\lambda}_1(p^*), s)}}{e^{g(p^*, s)}} - 1\right) \frac{e^{\gamma Z} e^{(1-\gamma)j(p^*\lambda_1^H/\bar{\lambda}_1(p^*))}}{e^{(1-\gamma)j(p^*)}}\right]\\
&\qquad\qquad + \bar{\lambda}_2(p^*)\left(\frac{e^{g(p^*\lambda_2^H/\bar{\lambda}_2(p^*), s)}}{e^{g(p^*, s)}} - 1\right) \frac{e^{(1-\gamma)j(p^*\lambda_2^H/\bar{\lambda}_2(p^*))}}{e^{(1-\gamma)j(p^*)}} + \mu_D\\
&= -r_f(p^*)\\
&\qquad - \gamma\varphi\sigma_C^2 + \bar{\lambda}_1(p^*) E_\nu\left[\left(\frac{e^{-\varphi Z} e^{g(p^*\lambda_1^H/\bar{\lambda}_1(p^*), s)}}{e^{g(p^*, s)}} - 1\right)\left(\frac{e^{\gamma Z} e^{(1-\gamma)j(p^*\lambda_1^H/\bar{\lambda}_1(p^*))}}{e^{(1-\gamma)j(p^*)}} - 1\right)\right]\\
&\qquad\qquad + \bar{\lambda}_2(p^*)\left(\frac{e^{g(p^*\lambda_2^H/\bar{\lambda}_2(p^*), s)}}{e^{g(p^*, s)}} - 1\right)\left(\frac{e^{(1-\gamma)j(p^*\lambda_2^H/\bar{\lambda}_2(p^*))}}{e^{(1-\gamma)j(p^*)}} - 1\right)\\
&\qquad\qquad\quad + \underbrace{\mu_D + \bar{\lambda}_1(p^*) E_\nu\left[\frac{e^{-\varphi Z} e^{g(p^*\lambda_1^H/\bar{\lambda}_1(p^*), s)}}{e^{g(p^*, s)}} - 1\right] + \bar{\lambda}_2(p^*)\left(\frac{e^{g(p^*\lambda_2^H/\bar{\lambda}_2(p^*), s)}}{e^{g(p^*, s)}} - 1\right)}_{(27.1)}.
\end{aligned}
\tag{27}
$$

(27.1) is the expected growth rate of the price of equity strip, which captures the cash-flow effect. The remainder of Equation 27 summarizes the effect from discount rate, which is the sum of riskfree rate, and the risk premium (described in the theorem that follows).

The following theorem characterizes the risk-premium of the equity strips.

Theorem 5. The time-t instantaneous risk premium of a dividend strip with maturity s is given by $r_t(s) - r_{ft} = r(p_t; s) - r_f(p_t)$, where

$$
\begin{aligned}
r(p; s) - r_f(p) &= \gamma\varphi\sigma_C^2\\
&\quad - \bar{\lambda}_1(p) E_\nu\left[\left(\frac{e^{-\varphi Z} e^{g(p\lambda_1^H/\bar{\lambda}_1(p), s)}}{e^{g(p, s)}} - 1\right)\left(\frac{e^{\gamma Z} e^{(1-\gamma)j(p\lambda_1^H/\bar{\lambda}_1(p))}}{e^{(1-\gamma)j(p)}} - 1\right)\right]\\
&\qquad - \bar{\lambda}_2(p)\left(\left(\frac{e^{g(p\lambda_2^H/\bar{\lambda}_2(p), s)}}{e^{g(p, s)}} - 1\right)\left(\frac{e^{\gamma Z} e^{(1-\gamma)j(p\lambda_2^H/\bar{\lambda}_2(p))}}{e^{(1-\gamma)j(p)}} - 1\right)\right).
\end{aligned}
\tag{28}
$$

The following corollary summarizes the pricing of the equity.

Corollary 1. The time t price of the claim to the stream of dividend specified by (21) is given by

$$S(D_t, p_t) = \int_{s=0}^{\infty} E_t \left[\frac{\pi_{t+s}}{\pi_t} D_{t+s} \right] ds$$

where

$$S(D, p) = D \int_{s=0}^{\infty} G(p, s) ds. \tag{29}$$

Proof. The results directly follow from Theorem 4 and absence of arbitrage. □

Let r_t be the instantaneous expected return of the equity asset defined above. The following theorem characterizes equity asset's risk premium.

Theorem 6. The instantaneous risk premium for an equity asset as a claim to (21) is given by $r_t - r_{ft} = r(p_t) - r_f(p_t)$, where

$$\begin{aligned} r(p) - r_f(p) = & \underbrace{\gamma\varphi\sigma_C^2}_{(30.1)} \\ & \underbrace{- \bar{\lambda}_1(p) E_\nu \left[\left(\frac{e^{-\varphi Z} \int_{s=0}^{\infty} e^{g(p\lambda_1^H/\bar{\lambda}_1(p), s)} ds}{\int_{s=0}^{\infty} e^{g(p,s)} ds} - 1 \right) \left(\frac{e^{\gamma Z} e^{(1-\gamma) j(p\lambda_1^H/\bar{\lambda}_1(p))}}{e^{(1-\gamma) j(p)}} - 1 \right) \right]}_{(30.2)} \\ & \underbrace{- \bar{\lambda}_2(p) \left(\left(\frac{\int_{s=0}^{\infty} e^{g(p\lambda_2^H/\bar{\lambda}_2(p), s)} ds}{\int_{s=0}^{\infty} e^{g(p,s)} ds} - 1 \right) \left(\frac{e^{(1-\gamma) j(p\lambda_2^H/\bar{\lambda}_2(p))}}{e^{(1-\gamma) j(p)}} - 1 \right) \right)}_{(30.3)}. \end{aligned} \tag{30}$$

Equation 30 shows that the instantaneous equity premium can be decomposed into three parts. The first, (30.1), is the risk premium associated to the normal time consumption growth risk.

(30.2) and (30.3) are associated to the rare events, N_{1t} and N_{2t}, respectively. N_{1t} is associated with disaster realization. When a disaster realizes, the dividend of the equity jumps down; in addition, the price-dividend ratio of the equity asset also decreases because the agent revises his belief in the high state. The two effects jointly determines the return associated to disaster realization, and the covariance between the pricing kernel and the return conditioning on disaster realization determines the risk premium. N_{2t}, on the other hand, serves as a signal, affecting the price-dividend ratio only upon arrival.

This discussion has focused on incomplete information and the amplification of the risk premium. Incomplete information will have other consequences, namely to skewness. When information is incomplete, the occurrence of a disaster leads the agent to learn that disasters are more likely. This will raise the risk premium, decrease the interest rate, and also decrease expected cash flows. When $\phi > 1$, the net effect is negative (this does not require a preference for early resolution of uncertainty). The model thus embodies the partial equilibrium intuition of Campbell and Hentschel (1992), who show that negative skewness can arise when an increase in volatility leads to an increase in risk premia, which then lead prices to decline more than they would otherwise. The model also incorporates intuition of Hong and Stein (2003) that skewness captures a quicker release of negative information than positive information, though the mechanism is quite different than in that model (which assumes risk neutrality and short-sale constraints). The agent learns a lot, all at once, from the occurrence of a disasters or negative signal. In contrast, the agent learns gradually from the absence of events.

3 Calibration and Quantitative Results

In this section we focus on the quantitative performance of the model. We simulate 2,000 samples from the model, and calculate return moments in each samples. Tables and figures report means and the distribution across simulation samples. The simulation is performed at an intra-day frequency (to capture the dynamics of high-intensity Poisson processes), and returns are aggregated to an annual frequency. In what follows, we report moments on level (not log) annual returns and the differences (not log differences) between the return and the riskfree rate.

3.1 Calibration

We choose preference parameters and normal-times consumption parameters similar to the ones in Wachter (2013). However, the risk-aversion parameter is further lowered to 2. Similar to Wachter and Zhu (2017), we choose $\phi_{L\to H}$ and $\phi_{H\to L}$ such that the bad state is a rare event. The unconditional probability of the bad state is 8.26%. We then choose λ_1^H and λ_1^L such that the unconditional jump intensity of the disasters is 3.55% per annum. The disaster distribution is multinomial, as measured in the data by Barro and Ursúa (2008).

The market portfolio has a disaster sensitivity $\varphi = 3$. The jump intensity parameters of N_{2t} will govern the premium associated to the learning mechanism. We choose to pick different combinations of the pair to further explore the effect of learning on equity premium. Specifically, we choose $[\lambda_2^H, \lambda_2^L]$ such that $\lambda_2^H + \lambda_2^L$ and λ_2^H/λ_2^L are controlled. In addition, we also calibrate the models with 1) perfect information and 2) when $\lambda_2^H/\lambda_2^L = 1$. When $\lambda_2^H = \lambda_2^L$, the signal provides no information about the state at all.

The parameters except for $[\lambda_2^H, \lambda_2^L]$ in our calibration are reported in Table 1.

3.2 Learning versus full information

We first consider the case where there is no signal ($\lambda_2^H = \lambda_2^L$), and compare the case in which the agent learns from disasters relative to the full information case, which we solve in Appendix D. Table 2 reports moments from the full information case and the learning case. First note that the learning increases the equity premium. While the effect is modest (1 percentage point) it is noticeable. Relative eto full information, learning decreases the volatility of excess returns. Under full information, the volatility is 20%; it is 13% under learning. Interestingly, while measured risk is greater in the economy with full information, true risk is greater in the economy with learning, in that investors require a higher premium to hold equity.

Another qualitative difference between the learning and full-information benchmarks is return skewness. Return skewness equals -0.83 in annual data. This moment of returns is of interest because leading asset pricing models predict positive return skewness. For example, the model of Campbell and Cochrane (1999) implies positive skewness (Wachter, 2005), as does the model of Bansal and Yaron (2004) (Lorenz and Schumacher, 2018). Because these models imply a conditional lognormal distribution for returns, this is not necessarily surprising. More surprising is the fact that rare events models, which would seem to be a natural candidate for explaining negative skewness do not do so, as Panel I clearly shows. On the other hand, allowing for learning does imply negative skewness. The effect is about the same size as in the data. We explain the source of this result below.

Finally, and consistent with the results concerning the equity premium, the riskfree rate is lower under the model for learning. This is because of enhanced precautionary savings: the agent understands that disasters are worse, in the sense that they raise marginal utility for reasons other than the direct consumption response. The agent wishes to save more, pushing down the equilibrium interest rate in this consumption economy.

Figure 1 shows the evolution of valuation ratios, prices, risk premia, and riskfree rates following a disaster, and contrasts this with the full information case. The experiment in this figure is to consider a path in which the economy begins in a high-risk state, a disaster happens at the end of the first year, and then the economy reverts to the low-risk state in year 4. We set the agents beliefs at the unconditional probability of the high-risk state. The top-left corner shows the price-dividend ratio. In the full-information case, the dynamics are quite simple; the price-dividend ratio starts out depressed, and then switches immediately to its higher value when the state switches to low-risk. There is no effect of an economic disaster on the price-dividend ratio because the economic disaster contains no information about future dividend growth.

In the learning case, the price-dividend ratio begins below its level under full information. In the year preceding the disaster, it increases very slightly. Unlike in full-information case, the price-dividend ratio falls by nearly 50% in the event of a disaster. Following the disaster, it rises slowly, only reaching its pre-disaster level five years later. The initial (small) increase in the price-dividend ratio is due to learning: recall that the steady-state probability p^* lies below the unconditional probability. Thus, the agent does learn from the absence of disasters. However, this level of the price-dividend ratio lies below the full-information value, even for the high-risk state. Most importantly for our purposes, the price-dividend ratio falls when a disaster occurs because of an increase in risk and a decrease in expected cash flows (which together exceed the precautionary savings effect on the riskfree rate). As the agent observes months without disasters, the price-dividend ratio steadily rises.

The bottom left panel shows the conditional equity premium. The equity premium is higher in the full-information case when the agent knows that the economy is in the high-risk state. Interestingly, when the agent must learn the state, the equity premium, which starts out lower, briefly rises above the high-risk counterpart. This highlights the nonlinear effect of learning. Disasters cause a greater decline in the price under learning because they convey more information (top right panel). This is reflected in an equity premium that, in the worst state, exceeds that of full information. Finally, the bottom right panel shows the riskfree rate. The qualitative panel matches that of the equity premium; however, the riskfree rate in the learning case never falls below the value when there is full information about the high-risk state.[7]

To summarize: the equity premium is higher in the case with learning because the price impact of disasters on equities is greater (a disaster affects both valuations and cash flows), and also because of the greater impact on marginal utilities. The volatility, however, is lower, because, during times without disasters, there is very little release of information.

Figure 2 shows realized returns for the time path in Figure 1. The initial discontinuity in year 1 reflects the realization of disaster, which is substantially worse in case with learning. Qualitatively, though, learning and full-information both predict the same results for realized returns if a disaster occurs: namely, they are very negative. Something very different, however, happens right after a disaster. In the model with full information, returns simply revert to their previous level, reflecting the equity premium in the high-risk state (this figure abstracts from Brownian noise that would appear in real-world return observations). In the case with learning, the realized return is much higher, reflecting not just the risk premium of Figure 1, but also the fact that agents forecasted a second disaster, but that one did not occur. Hu et al. (2019) shows that this effect of extremely high returns occurring after extremely large market declines is a robust feature of the data; note that the model with learning can account for this effect, whereas the full information model cannot.

We now return to the question of negative skewness, and why it appears in the

[7]This subtle difference between the behavior of the riskfree rate and the behavior of the equity premium between learning and full information arises from the difference in the learning effects in (20) and (30). On the one hand, in our calibration, a disaster does not lead the agent to fully update in favor of the high-probability state because there is some chance, however small, that a disaster could occur in a low-probability state. Thus $\lambda_1(p) < \lambda^H$, even after a disaster has occurred. On the other hand, the effect of future learning enters marginal utility. For the riskfree rate, the latter effect is smaller than the former, leading the value to be very slightly less depressed in the learning case. In the case of the equity premium, the effect of learning (which enters both expected future cash flows as well as marginal utility) is larger than the difference in probabilities.

model with learning and not otherwise. Figure 2 shows that, while large negative returns characterize both the full-information model and the learning time series, the full-information time series also has very large positive returns. These large positive returns occur when the full information state is completely revealed to investors. The learning model has no counterpart to these very large upside surprises, nor does there appear to be a counterpart in the data. Rather, as in the actual time series, agents gradually learn from the absence of bad events.

The analysis in the section reveals important differences between the case of learning and the full-information case. The full-information case features high volatility, a lower equity premium, and positive skewness. The learning case has substantially lower volatility, combined with a higher equity premium, and negative skewness. Comparing these cases suggests a potential explanation for the lack of a observed relation between volatility and risk (Moreira and Muir, 2017). The question still remains, however, whether a model with learning can come close to explaining the level of volatility observed in the data.

3.3 Signals and volatility

Realistically, investors can learn from sources other than the observation (or lack thereof) of rare events. The model in the previous section, which allows for signals, is a particularly tractable way to model such learning. A signal is an event that reveals something about the state, without affecting cash flows directly. We conduct a two-way experiment in the model. We control the ratio of the signal intensity, λ_2^L/λ_2^H, and the frequency of the type-2 signals. The signal intensity ratio determines how informative the signals are (a lower ratio means that the signals are more informative). Thus, as the signals become more frequent, and increase in informativeness, the economy converges toward the case of full information. To maximally capture the effect of signals, we consider relatively high intensities.

Figure 3 shows the results for the aggregate market. The panels in this figure show the average moments, medians, interquartile ranges, and 95% confidence intervals. We also report the data estimate. We see that, as the signal intensity increases, the average return falls, the volatility rises, and the skewness goes from negative to positive. Increasing the signal intensity ratio (which reduces the informativeness of the signals) leads to higher equity premia, lower volatility, and more negative skewness. These figures show a tradeoff between explaining skewness and explaining volatility (lack of information helps explain skewness, but leads to lower volatility). Nonetheless, for many parameter configurations, the data fall within the interquartile range implied by the simulations.

Figure 4 performs a comparable exercise for the riskfree rate. As the figure shows, the less informative the economy (either because of a relatively low signal intensity, or a high intensity ratio), the lower the riskfree rate due to precautionary savings. Moreover, the less informative the economy, the lower the volatility of the interest rate.

4 Conclusion

In this paper, we assume that agents must learn about the probability of rare events, and focus on the effects of this learning for asset prices. Learning makes disasters more severe; once a disaster has occurred, agents rationally update their probability that a disaster might occur again. Agents' beliefs return to their baseline levels after about 5 years of data featuring no disasters; nonetheless, prices, valuation ratios, and interest rates remain depressed relative to the full-information benchmark. Agents rationally anticipate the effect of learning, leading to an equity premium that is higher than it would be in the full-information case. On the other hand, volatility is lower, because agents' beliefs about the probability change to a lesser degree.

Rare disasters imply an asymmetry in how agents learn about bad versus good news. Bad news is learned quickly, from the realization of a disaster, or of a signal. Agents, in contrast, must learn good news slowly, from the absence of rare events. Thus negative events in the economy are sharply negative, whereas positive ones are muted. Learning generates negative skewness, a fact present in the data that is difficult to explain in a full-information setting.

A Bayesian learning in a multiple regime switch model with multiple jumps

In this section, we show the proof for Theorem 1. We focus on a general case where there are finite number of states, and allow for finite numbers of Poisson and Itô processes as signals. Our result generalizes a related theorem for Brownian signals in Wonham (1965).

A.1 The setting

There are two types of processes: counting (Poisson) processes and Itô processes. Conditional moments of the processes depend a latent state. The agent learns about the state through the realization of the processes.

There are $I < \infty$ counting processes

$$N_{1,t}, N_{2,t}, \dots, N_{I,t},$$

with time-varying jump intensities

$$\lambda_{1,t}, \lambda_{2,t}, \dots, \lambda_{I,t},$$

respectively. In addition, there is a $J \times 1$ vector Itô process X_t evolving according to

$$dX_t = \mu_t dt + \sigma dB_t,$$

where μ_t is $J \times 1$, σ is $J \times J$, and B_t is a standard J-dimensional Brownian motion, independent of N_t. $J < \infty$. Let $\Sigma = \sigma\sigma^\top$. Furthermore, assume that μ_t itself is time-varying.

In what follows, let N_t and X_t denote vectors of the counting and Itô processes:

$$N_t \equiv \begin{bmatrix} N_{1,t} \\ N_{2,t} \\ \vdots \\ N_{I,t} \end{bmatrix}, \qquad X_t \equiv \begin{bmatrix} X_{1,t} \\ X_{2,t} \\ \vdots \\ X_{J,t} \end{bmatrix}.$$

The regime switch model. Let λ_t and μ_t denote the vectors of the conditional jump intensity and drift of the counting and Itô processes, respectively:

$$\lambda_t = \begin{bmatrix} \lambda_{1,t} \\ \lambda_{2,t} \\ \vdots \\ \lambda_{I,t} \end{bmatrix}, \qquad \mu_t = \begin{bmatrix} \mu_{1,t} \\ \mu_{2,t} \\ \vdots \\ \mu_{J,t} \end{bmatrix}.$$

In addition, let $S_t = [\lambda_t^\top, \mu_t^\top]^\top$ be the vector of the instantaneous jump intensity and drift processes. S_t follows a Markov regime switch model and switches among $K < \infty$ different regimes, $S^1, S^2, \ldots, S^K$.

Define the probability of a regime switch as follows

$$\Pr(S_{t+dt} = S^n | S_t = S^m) = \phi_{m \to n} dt, \qquad m \neq n. \tag{A.1}$$

A.2 Imperfect information and learning

The state S_t is not observable. We assume Bayesian agents who form a posterior probability of S_t by observing data on X_t and on N_t. In what follows, we describe the evolution of the agent's belief.

Define $\mathcal{F}_t$ to be the $\sigma-$algebra generated by $\{\{X_s\}_{s\in[0,t]}, \{N_s\}_{s\in[0,t]}\}$. Define the posterior probability of state k:

$$p_t^k \equiv \Pr(S_t = S^k | \mathcal{F}_t)$$

and the vector of posterior probabilities:

$$p_t = \begin{bmatrix} p_t^1 \\ p_t^2 \\ \vdots \\ p_t^K \end{bmatrix}.$$

The following theorem characterizes the evolution of p_t^k as a function of X_t and N_t:

Theorem A.1. With the hidden regime dynamics and signals given by (A.1), p_t^k, or the posterior probability a Bayesian agent assigns on state k, evolves according to

$$dp_t^k = -p_{t^-}^k \left(\sum_{m \neq k} \phi_{k\to m} \right) dt + \left(\sum_{m \neq k} p_{t^-}^m \phi_{m \to k} \right) dt$$
$$+ p_{t^-}^k \left(\mu^k - \bar{\mu}(p_{t^-}) \right) \Sigma^{-1} \left(dX_t - \bar{\mu}(p_{t^-}) \right)$$
$$+ p_{t^-}^k \sum_{i=1}^{I} \left(\frac{\lambda_i^k}{\bar{\lambda}_i(p_{t^-})} - 1 \right) \left(dN_{i,t} - \bar{\lambda}_i(p_{t^-}) dt \right), \quad \text{(A.2)}$$

where

$$\bar{\mu}(p) = \sum_{m=1}^{K} p^m \mu^m, \qquad \bar{\lambda}(p) = \sum_{m=1}^{K} p^m \lambda^m$$

and

$$\bar{\lambda}_i(p) = \sum_{m=1}^{K} p^m \lambda_i^m, \qquad \bar{\lambda}_i(p) = \sum_{m=1}^{K} p^m \lambda_i^m, \qquad i = 1, 2, \ldots, I,$$

are the posterior drifts and jump intensities.

Proving Theorem A.1 requires characterization of the evolution of the likelihood functions. The definition of conditional probability implies

$$\begin{aligned} p_t^k =& \Pr(S_t = S^k | \mathcal{F}_t) \\ =& \Pr\left(S_t = S^k | \mathcal{F}_0, \{X_u, N_u\}_{0 \le u \le t} \right) \\ =& \frac{\Pr\left(\{X_u, N_u\}_{0 \le u \le t}, S_t = S^k | \mathcal{F}_0 \right)}{\Pr\left(\{X_u, N_u\}_{0 \le u \le t} | \mathcal{F}_0 \right)} \\ =& \frac{\Pr\left(\{X_u, N_u\}_{0 \le u \le t}, S_t = S^k | \mathcal{F}_0 \right)}{\sum_{m=1}^{K} \Pr\left(\{X_u, N_u\}_{0 \le u \le t}, S_t = S^m | \mathcal{F}_0 \right)}. \end{aligned} \quad \text{(A.3)}$$

It suffices to show the evolution of conditional probability $\Pr\left(\{X_u, N_u\}_{0 \le u \le t}, S_t = S^k | \mathcal{F}_0 \right)$, or the likelihood. Then we can apply Itô 's Lemma to find the stochastic differential equation that characterizes the evolution of p_t^k.

However, since the path of states is latent, we do not have a closed-form solution to the likelihood. We resolve this issue by conditioning on a specific path of states, characterizing the evolution of likelihood, and then taking average across all sample paths.

In what follows, we characterize the evolution of the likelihood function. We begin the proof by defining a series of functions.

Define

$$p_{i,j}(t) \equiv \Pr(S_t = S^j | S_0 = S^i). \tag{A.4}$$

as the probability that $S_t = S^j$, conditioning on that $S_0 = S^i$. Markov property of the regime switch process implies that

$$p_{i,j}(t) = \Pr(S_{s+t} = S^j | S_s = S^i), \qquad \forall i, j, s, t. \tag{A.5}$$

Define function $\Phi(t_1, t_2, X_t, S_t)$ as

$$\Phi(t_1, t_2, \{X_u\}, \{N_u\}, \{S_u\}) = \exp\left(\int_{t_1}^{t_2} \mu_u^\top \Sigma^{-1} dX_u - \int_{t_1}^{t_2} \left(\frac{1}{2}\mu_u^\top \Sigma^{-1}\mu_u + \iota^\top \lambda_u\right) du\right) \times \prod_{i=1}^{I} \prod_{0<\tau_i(N,m)\leq t} \lambda_{i,\tau_i(N,m)^-}, \tag{A.6}$$

where $\tau_i(n,m) \equiv \min\{t : N_{i,t} \geq m\}$ is the arrival time of the mth jump of $N_{i,t}$ on the sample path $n = n_{i,t}$. $\Phi(t_1, t_2, X_t, N_t, S_t)$ is proportional to the likelihood of $\{X_u, N_u\}_{t_1 \leq u \leq t_2}$, conditioning on a specific path for state, $\{S_u\}_{t_1 \leq u \leq t_2}$.

Define a second process $[\tilde{\mu}_t, \tilde{\lambda}_t]$, identically distributed to $[\mu_t, \lambda_t]$, but independent of S_t and Y_t. In addition, define $h^k(t)$ as

$$h^k(t) \equiv \sum_{l=1}^{K} p^l(0) p_{l,k}(t) E\left[\Phi(0, t, \{X_u\}, \{N_u\}, \{\tilde{S}_u\}) | \tilde{S}_t = S^k, \tilde{S}_0 = S^l, \{X_u, N_u\}_{0\leq u\leq t}\right]. \tag{A.7}$$

The following provides a characterization of the likelihood function (A.6) for a small period of time.

Lemma A.1. Let $\delta > 0$. Then the following equation holds when $\lim_{\delta\to 0} S_{t-\delta} = S_t = S^k$ and $\lim_{\delta\to 0} N_{t-\delta} = N_t$:

$$\Phi(t-\delta, t, \{X_u\}, \{N_u\}, \{S_u\}) = 1 + {\mu^k}^\top \Sigma^{-1}(X_t - X_{t-\delta}) - \iota^\top \lambda^k \delta + o(\delta). \tag{A.8}$$

Proof. Note that

$$\int_{t_1}^{t_2} \mu_u^\top \Sigma^{-1} dX_u - \int_{t_1}^{t_2} \left(\frac{1}{2} \mu_u^\top \Sigma^{-1} \mu_u + \iota^\top \lambda_u \right) du \tag{A.9}$$

is a Itô process. When $\lim_{\delta \to 0} S_{t-\delta} = S_t$, the quadratic variation of (A.9) is given by

$$\mu^{k\top} \Sigma^{-1} \mu^k.$$

When $\lim_{\delta \to 0} N_{t-\delta} = N_t$, there is no jump realization at time t. Then we can apply Itô 's Lemma and obtain (A.8). □

It will turn out later that (A.8) will be driving the evolution of $h^k(t)$.

The following lemma connects $h^k(t)$ defined by Equation A.7 to the likelihood function.

Lemma A.2. The likelihood function of a sample path $\{X_u, N_u\}_{0 \le u \le t}$, with $S_t = S^k$, satisfies

$$\Pr\left(\{X_u, N_u\}_{0 \le u \le t}, S_t = S^k | \mathcal{F}_0 \right) \propto h^k(t). \tag{A.10}$$

Proof. The the probability density of a sample path, $\{x_u, n_u\}_{0 \le u \le t}$, conditioning on $S_t = s^k, S_0 = S^l$, is given by

$$\begin{aligned}
&\Pr\left(\{x_u, n_u\}_{0 \le u \le t} | S_t = S^k, S_0 = S^l \right) \\
=&E\left[\Pr\left(\{x_u, n_u\}_{0 \le u \le t} \middle| \{S_u\}_{0 \le u \le t} \right) \middle| S_t = S^k, S_0 = S^l \right] \\
\propto&E\left[\Phi(0, t, \{x_u\}, \{n_u\}, \{S_u\}) \middle| S_t = S^k, S_0 = S^l \right].
\end{aligned}$$

Note that $E\left[\Pr\left(\{x_u, n_u\}_{0 \le u \le t} \middle| \{S_u\}_{0 \le u \le t} \right) \middle| S_t = S^k, S_0 = S^l \right]$ is the density of the realized path $\{x_u, n_u\}_{0 \le u \le t}$, conditional on the state at time 0 and the state at time t (but not the intermediate states). The expectation operator integrates out over sample paths of $\{S_{u^-}\}_{0 < u \le t}$.

As a result, the probability density of the actual realization of $\{X_u, N_u\}_{0 \le u \le t}$ is given by

$$\Pr\left(\{x_u, n_u\}_{0\le u\le t} \,|S_t = S^k, S_0 = S^l\right)\Big|_{\{x_u,n_u\}_{0\le u\le t}=\{X_u,N_u\}_{0\le u\le t}}$$
$$\propto E\left[\Phi(0,t,\{x_u\},\{n_u\},\{S_u\})\Big|S_t = S^k, S_0 = S^l\right]\Big|_{\{x_u,n_u\}_{0\le u\le t}=\{X_u,N_u\}_{0\le u\le t}} \tag{A.11}$$

Note that, $\{S_u\}_{0\le u\le t}$ and $\{X_u, N_u\}_{0\le u\le t}$ are correlated, as a result (A.11) can not be re-written a conditional expectation, conditioning on $\{X_u, N_u\}_{0\le u\le t}$. Instead, we consider $[\tilde{\mu}_t, \tilde{\lambda}_t]$ process defined before. The assumptions on $[\tilde{\mu}_t, \tilde{\lambda}_t]$ imply the following equation:

$$E\left[\Phi(0,t,\{x_u\},\{n_u\},\{S_u\})\Big|S_t = S^k, S_0 = S^l\right]\Big|_{\{x_u,n_u\}_{0\le u\le t}=\{X_u,N_u\}_{0\le u\le t}}$$
$$=E\left[\Phi(0,t,\{X_u\},\{N_u\},\{\tilde{S}_u\})\Big|\tilde{S}_t = \tilde{S}^k, \tilde{S}_0 = S^l, \{X_u, N_u\}_{0\le u\le t}\right]$$
$$=h^k(t). \tag{A.12}$$

Then Equation A.10 follows. □

The following lemma shows that the posterior probability is a function of $h^m(t)$, $m = 1, 2, \ldots, K$.

Lemma A.3. The posterior probability of state k, p_t^k, satisfies

$$p_t^k = \frac{h^k(t)}{\sum\limits_{m=1}^{K} h^m(t)}. \tag{A.13}$$

Proof. The result follows immediately from Lemma A.2. □

The following lemma provides the stochastic differential equation that characterizes the evolution of $h^k(t)$.

Lemma A.4. The conditional probability $h^k(t)$, as defined by (A.7), follows the following stochastic differential equation:

$$dh^k(t) = h^k(t^-)\left(-\left(\sum_{m\neq k}\phi_{k\to m}\right)dt + \left(\mu^{k\top}\Sigma^{-1}dX_t - \iota^\top\lambda^k dt\right)\right) + \sum_{m\neq k} h^m(t^-)\phi_{m\to k}dt + h^k(t^-)\sum_{i=1}^{I}(\lambda_i^k - 1)dN_{i,t}. \quad \text{(A.14)}$$

Proof. Note that,

$$\begin{aligned}
& h^k(t) \\
=& \sum_{l=1}^{K} p^l(0)p_{l,k}(t)E\left[\Phi(0,t,\{X_u\},\{N_u\},\{\tilde{S}_u\})|\tilde{S}_t = S^k, \tilde{S}_0 = S^l, \{X_u, N_u\}_{0\le u\le t}\right] \\
\propto& \sum_{l=1}^{K} p^l(0)\Big(\sum_{k=1}^{K} p_{l,m}(t-\delta)p_{m,k}(\delta) \\
&\quad \times E\left[\Phi(0,t-\delta,\{X_u\},\{N_u\},\{\tilde{S}_u\})|\tilde{S}_t = S^k, \tilde{S}_{t-\delta} = S^m, \tilde{S}_0 = S^l, \{X_u, N_u\}_{0\le u\le t}\right] \\
&\quad\quad \times E\left[\Phi(t-\delta,t,\{X_u\},\{N_u\},\{\tilde{S}_u\})|\tilde{S}_t = S^k, \tilde{S}_{t-\delta} = S^m, \tilde{S}_0 = S^l, \{X_u, N_u\}_{0\le u\le t}\right]\Big) \\
=& \sum_{l=1}^{K} p^l(0)\Big(\sum_{m=1}^{K} p_{l,m}(t-\delta)p_{m,k}(\delta) \\
&\quad \times E\left[\Phi(0,t-\delta,\{X_u\},\{N_u\},\{\tilde{S}_u\})|\tilde{S}_{t-\delta} = S^m, \tilde{S}_0 = S^l, \{X_u, N_u\}_{0\le u\le t-\delta}\right] \\
&\quad\quad \times E\left[\Phi(t-\delta,t,\{X_u\},\{N_u\},\{\tilde{S}_u\})|\tilde{S}_t = S^k, \tilde{S}_{t-\delta} = S^m, \{X_u, N_u\}_{t-\delta\le u\le t}\right]\Big) \\
=& \sum_{m=1}^{K}\Big(\sum_{l=1}^{K} p^l(0)p_{l,m}(t-\delta) \\
&\quad \times E\left[\Phi(0,t-\delta,\{X_u\},\{N_u\},\{\tilde{S}_u\})|\tilde{S}_{t-\delta} = S^m, \tilde{S}_0 = S^l, \{X_u, N_u\}_{0\le u\le t-\delta}\right] \times \\
&\quad\quad \times p_{m,k}(\delta)E\left[\Phi(t-\delta,t,\{X_u\},\{N_u\},\{\tilde{S}_u\})|\tilde{S}_t = S^k, \tilde{S}_{t-\delta} = S^m, \{X_u, N_u\}_{t-\delta\le u\le t}\right]\Big) \\
=& \sum_{m=1}^{K}\left(h^m(t-\delta)p_{m,k}(\delta)E\left[\Phi(t-\delta,t,\{X_u\},\{N_u\},\{\tilde{S}_u\})|\tilde{S}_t = S^k, \tilde{S}_{t-\delta} = S^m, \{X_u, N_u\}_{t-\delta\le u\le t}\right]\right).
\end{aligned} \quad \text{(A.15)}$$

In addition, we know that, for a small and positive δ, the following equations hold.

$$p_{m,k} = \phi_{m\to k}\delta + o(\delta), \qquad m \neq k \tag{A.16}$$

$$p_{k,k} = 1 - \left(\sum_{m\neq k} \phi_{m\to k}\right)\delta + o(\delta). \tag{A.17}$$

In addition, when $m \neq k$ and $\lim_{\delta\to 0} N_{t-\delta} = N_t$, the following equation holds:

$$E\left[\Phi(t-\delta, t, \{X_u\}, \{N_u\}, \{\tilde{S}_u\})|\tilde{S}_t = S^k, \tilde{S}_{t-\delta} = S^m, \{X_u, N_u\}_{t-\delta\leq u\leq t}\right] = 1 + O(h). \tag{A.18}$$

Combining results from Lemma A.1 and Equations A.16, A.17 and A.18 implies the stochastic differential equation (A.14).

□

Lemma A.5. Let

$$\bar{h}(t) \equiv \sum_{l=1}^{K} h^k(t). \tag{A.19}$$

Then $\bar{h}_t$ is characterized by the following stochastic differential equation with jump:

$$d\bar{h}(t) = \bar{h}(t^-)\left(-\iota^\top\bar{\lambda}(p_{t^-})dt + \bar{\mu}(p_{t^-})^\top\Sigma^{-1}dX_t + \sum_{i=1}^{I}(\bar{\lambda}_i(p_{t^-}) - 1)dN_{i,t}\right) \tag{A.20}$$

Proof. By definition of $\bar{h}(t)$, we have

$$\begin{aligned}
d\bar{h}(t) =& d\left(\sum_{k=1}^{K} h^k(t)\right)\\
=& \sum_{k=1}^{K}\left(-h^k(t^-)\sum_{l\neq k}\phi_{k\to s} + \sum_{l\neq k} h^l(t^-)\phi_{l\to k}\right)dt - \bar{h}(t^-)\iota^\top\left(\sum_{k=1}^{K}\frac{h^k(t^-)}{\bar{h}(t^-)}\lambda^k\right)dt\\
&+ \bar{h}(t^-)\left(\sum_{k=1}^{K}\frac{h^k(t^-)}{\bar{h}(t^-)}\mu^k\right)^\top\Sigma^{-1}dX_t + \bar{h}(t^-)\sum_{i=1}^{I}\left(\sum_{k=1}^{K}\frac{h^k(t^-)}{\bar{h}(t^-)}\lambda_i^k - 1\right)dN_{i,t}\\
=& \bar{h}(t^-)\left(-\iota^\top\bar{\lambda}(p_{t^-})dt + \bar{\mu}(p_{t^-})^\top\Sigma^{-1}dX_t + \sum_{i=1}^{I}(\bar{\lambda}_i(p_{t^-}) - 1)dN_{i,t}\right).
\end{aligned}$$

□

Proof of Theorem A.1. From Lemma A.3 we know

$$p_t^k = \frac{h^k(t)}{\bar{h}(t)}. \tag{A.21}$$

Applying Itô 's Lemma, together with the results of Lemma A.4 and A.5 yields Equation A.2. □

Proof of Theorem 1. Let $K = 2$, $I = 2$, $J = 0$ for the general case in Theorem A.1. In addition, let state 2 represent the high-risk state. Then

$$p_t^1 = 1 - p_t^2. \tag{A.22}$$

Substituting (A.22) into (A.2) yields (7). □

B Solving the State Price Density

B.1 Representative agent's continuation value

Proof of Proposition 1. Conjecture that the representative agent's continuation value is given by

$$J(C,p) = \frac{1}{1-\gamma} C^{1-\gamma} e^{(1-\gamma)j(p)}, \tag{B.1}$$

where $j(p)$ is a continuously differentiable function of p.

Equation B.1 implies that

$$\frac{J\left(Ce^{-Z}, p\lambda_1^H/\bar{\lambda}_1(p)\right) - J(C,p)}{J(C,p)} = e^{(1-\gamma)\left(-Z + j(p\lambda_1^H/\bar{\lambda}_1(p)) - j(p)\right)} - 1 \tag{B.2}$$

$$\frac{J\left(C, p\lambda_2^H/\bar{\lambda}_2(p)\right) - J(C,p)}{J(C,p)} = e^{(1-\gamma)\left(j(p\lambda_2^H/\bar{\lambda}_2(p)) - j(p)\right)} - 1. \tag{B.3}$$

Optimality of the continuation value function implies the following Hamilton-Jacobian-Bellman Eqation:

$$\begin{aligned} f(C,J) &+ \frac{\partial J}{\partial C} C\mu_C + \frac{1}{2}\frac{\partial^2 J}{\partial C^2} C^2 \sigma^2 \\ &+ \frac{\partial J}{\partial p}\left[-p(\lambda^H - \bar{\lambda}(p)) + [\phi_{L\to H} - p(\phi_{H\to L} + \phi_{L\to H})]\right] \\ &+ \bar{\lambda}_1(p) E_\nu \left[J\left(Ce^{-Z}, p\lambda_1^H/\bar{\lambda}_1(p)\right) - J(C,p)\right] \\ &+ \bar{\lambda}_2(p) E_\nu \left[J\left(C, p\lambda_2^H/\bar{\lambda}_2(p)\right) - J(C,p)\right], \forall p \in [0,1]. \end{aligned} \tag{B.4}$$

Dividing both sides of (B.4) by $J(C,p)$ and then substituting (B.2) and (B.3) into the equation yields

$$\begin{aligned} &-\beta(1-\gamma)j(p) + (1-\gamma)\mu_C - \frac{1}{2}\gamma(1-\gamma)\sigma_C^2 \\ &+ (1-\gamma)j'(p)\left[\phi_{L\to H} - p(\phi_{L\to H} + \phi_{H\to L}) - p\left(\lambda^H - \bar{\lambda}(p_t)\right)\right] \\ &+ \bar{\lambda}_1(p)\left(E_\nu\left[e^{(\gamma-1)Z}\right] e^{(1-\gamma)\left(j(p\lambda_1^H/\bar{\lambda}_1(p)) - j(p)\right)} - 1\right) \\ &+ \bar{\lambda}_2(p)\left(e^{(1-\gamma)\left(j(p\lambda_2^H/\bar{\lambda}_2(p)) - j(p)\right)} - 1\right) = 0. \end{aligned} \tag{B.5}$$

When $-p\iota^\top(\lambda^H - \bar{\lambda}(p)) + [\phi_{L\to H} - p(\phi_{H\to L} + \phi_{L\to H})] \neq 0$, combining (B.4), (B.2) and (B.3) leads to Equation 10.

When $-p\iota^\top(\lambda^H - \bar{\lambda}(p)) + [\phi_{L\to H} - p(\phi_{H\to L} + \phi_{L\to H})] = 0$, the equation can be rewritten as the following quadratic function,

$$\iota^\top(\lambda^H - \lambda^L)p^2 - (\iota^\top(\lambda^H - \lambda^L) + \phi_{H\to L} + \phi_{L\to H})p + \phi_{L\to H} = 0, \tag{B.6}$$

which has the following roots,

$$p = \frac{(\iota^\top(\lambda^H - \lambda^L) + \phi_{H\to L} + \phi_{L\to H}) \pm \sqrt{(\iota^\top(\lambda^H - \lambda^L) + \phi_{H\to L} + \phi_{L\to H})^2 - 4\iota^\top(\lambda^H - \lambda^L)\phi_{L\to H}}}{2\iota^\top(\lambda^H - \lambda^L)}. \tag{B.7}$$

We have

$$\begin{aligned} &\left(\iota^\top(\lambda^H - \lambda^L) + \phi_{H\to L} + \phi_{L\to H}\right)^2 - 4\iota^\top(\lambda^H - \lambda^L)\phi_{L\to H} \\ =& \left(\iota^\top(\lambda^H - \lambda^L) - \phi_{L\to H}\right)^2 + 2\left(\iota^\top(\lambda^H - \lambda^L) + \phi_{L\to H}\right)\phi_{H\to L} + \phi_{H\to L}^2 > 0, \end{aligned}$$

As a results both of the root must be real.

In addition, define

$$f_p(p) = \iota^\top(\lambda^H - \lambda^L)p^2 - (\iota^\top(\lambda^H - \lambda^L) + \phi_{H\to L} + \phi_{L\to H})p + \phi_{L\to H}, \tag{B.8}$$

Then $f_p(0) = \phi_{L\to H} > 0$, $f_p(1) = -\phi_{H\to L} < 0$, and this implies that there must be at least one root between 0 and 1.

Finally, in fact we have

$$\begin{aligned} &\left(\iota^\top(\lambda^H - \lambda^L) + \phi_{H\to L} + \phi_{L\to H}\right)^2 - 4\iota^\top(\lambda^H - \lambda^L)\phi_{L\to H} \\ =& \left(\iota^\top(\lambda^H - \lambda^L) - \phi_{L\to H}\right)^2 + 2\left(\iota^\top(\lambda^H - \lambda^L) + \phi_{L\to H}\right)\phi_{H\to L} + \phi_{H\to L}^2 \\ >& \left(\iota^\top(\lambda^H - \lambda^L) - \phi_{L\to H}\right)^2, \end{aligned}$$

and as a result

$$
\begin{aligned}
&\left(\iota^\top(\lambda^H-\lambda^L)+\phi_{H\to L}+\phi_{L\to H}\right)+\sqrt{\left(\iota^\top(\lambda^H-\lambda^L)+\phi_{H\to L}+\phi_{L\to H}\right)^2-4\iota^\top(\lambda^H-\lambda^L)\phi_{L\to H}}\\
>&\left(\iota^\top(\lambda^H-\lambda^L)+\phi_{H\to L}+\phi_{L\to H}\right)+\left(\iota^\top(\lambda^H-\lambda^L)-\phi_{L\to H}\right)\\
=&2\iota^\top(\lambda^H-\lambda^L)+\phi_{H\to L}\\
>&2\iota^\top(\lambda^H-\lambda^L).
\end{aligned}
$$

This suggest that

$$
\frac{\left(\iota^\top(\lambda^H-\lambda^L)+\phi_{H\to L}+\phi_{L\to H}\right)+\sqrt{\left(\iota^\top(\lambda^H-\lambda^L)+\phi_{H\to L}+\phi_{L\to H}\right)^2-4\iota^\top(\lambda^H-\lambda^L)\phi_{L\to H}}}{2\iota^\top(\lambda^H-\lambda^L)}>1. \tag{B.9}
$$

Combining all results above, and we can conclude that p^* given by (12) is the unique solution in $[0,1]$ to Equation (B.6). Equation B.4 then reduces to Equation 11, uniquely pinning down the solution to (10). □

B.2 State price density

Proof of Theorem 2. Duffie and Skiadas (1994) show that

$$
\pi_t=\exp\left\{\int_0^t\frac{\partial}{\partial V}f(C_s,V_s)ds\right\}\frac{\partial}{\partial C}f(C_t,V_t). \tag{B.10}
$$

The functional form of f implies

$$
\begin{aligned}
\frac{\partial}{\partial C}f(C_t,V_t)&=\beta(1-\gamma)\frac{V_t}{C_t}\\
&=\beta C_t^{-\gamma}e^{(1-\gamma)j(p)}.
\end{aligned} \tag{B.11}
$$

Combining (B.10) and (B.11), we get

$$
\pi_t=\beta\exp\left\{\int_0^t\frac{\partial}{\partial V}f(C_s,V_s)ds\right\}C_t^{-\gamma}e^{(1-\gamma)j(p)}. \tag{B.12}
$$

(B.12) and Itô 's Lemma leads to

$$\frac{d\pi_t}{\pi_{t^-}} = \mu_{\pi t^-} dt + \sigma_{\pi t^-} dB_{C,t}$$
$$+\left(e^{\gamma Z_t} e^{(1-\gamma)\left(j(p_{t^-}\lambda_1^H/\bar{\lambda}_1(p_{t^-}))-j(p_{t^-})\right)} - 1\right) dN_{1t} + \left(e^{(1-\gamma)\left(j(p_{t^-}\lambda_2^H/\bar{\lambda}_2(p_{t^-}))-j(p_{t^-})\right)} - 1\right) dN_{2t}, \tag{B.13}$$

where

$$\mu_{\pi t^-} = -\beta[1 + (1-\gamma)j(p_{t^-})] - \gamma\mu_C$$
$$+ (1-\gamma)j'(p_{t^-})\left[\phi_{L\to H} - p_{t^-}(\phi_{H\to L} + \phi_{L\to H}) - p_{t^-}\left(\lambda^H - \bar{\lambda}(p_{t^-})\right)\right]$$
$$+ \frac{1}{2}\gamma(\gamma+1)\sigma_C^2, \tag{B.14}$$

and $\sigma_{\pi t}$ given by (19).

Substituting (B.5) into (B.14) yields

$$\mu_{\pi t^-} = -(\beta + \mu_C - \gamma\sigma_C^2)$$
$$- \bar{\lambda}_1(p_{t^-})\left(E_\nu\left[e^{(\gamma-1)Z_t}\right] e^{(1-\gamma)\left(j(p_{t^-}\lambda_1^H/\bar{\lambda}_1(p_{t^-}))-j(p_{t^-})\right)} - 1\right)$$
$$- \bar{\lambda}_2(p_{t^-})\left(e^{(1-\gamma)\left(j(p_{t^-}\lambda_2^H/\bar{\lambda}_2(p_{t^-}))-j(p_{t^-})\right)} - 1\right), \tag{B.15}$$

which verifies the functional form of $\mu_\pi(p)$ given by (18).

□

Proof of Theorem 3. Non-arbitrage implies that

$$E_{t^-}\left[\frac{d\pi_t}{\pi_{t^-}}\right] = -r_{ft^-} dt, \tag{B.16}$$

where r_{ft^-} is the riskfree rate. Substituting (B.13) and (18) into (B.16) yields

$$r_{ft^-} = \beta + \mu_C - \gamma\sigma_C^2 + \bar{\lambda}_1(p_{t^-})e^{(1-\gamma)\left(j(p_{t^-}\lambda_1^H/\bar{\lambda}_1(p_{t^-}))-j(p_{t^-})\right)} E_\nu\left[e^{\gamma Z_t}\left(e^{-Z_t} - 1\right)\right], \tag{B.17}$$

which and a function of p_{t^-}verifies the functional form of $r_f(p)$ in (20). □

C Pricing Equity

C.1 Pricing equity strips

We first characterize a necessary condition under which the price process of an equity strip with dividend process given by (21) is arbitrage-free.

Lemma C.1. Let

$$F(D_t, p_t, s) = D_t E_t \left[\frac{\pi_{t+s}}{\pi_t} \frac{D_{t+s}}{D_t} \right]. \tag{C.1}$$

Then $F_t = F(D_t, p_t, s)$ is the time-t price of an equity strip with maturity s, and

$$G(p_t, s) \equiv E_t \left[\frac{\pi_{t+s}}{\pi_t} \frac{D_{t+s}}{D_t} \right] \tag{C.2}$$

is the price-dividend ratio at time t. Moreover, H_t satisfies

$$\begin{aligned} \frac{dF_t}{F_{t^-}} &= \mu_{Ft^-} dt + \sigma_{Ft^-} dB_{Ct} \\ &\qquad + \left(\frac{e^{-\varphi Z_t} G(p_{t^-} \lambda_1^H / \bar{\lambda}_1(p_{t^-}), s)}{G(p_{t^-}, s)} - 1 \right) dN_{1t} \\ &\qquad\qquad + \left(\frac{G(p_{t^-} \lambda_2^H / \bar{\lambda}_2(p_{t^-}), s)}{G(p_{t^-}, s)} - 1 \right) dN_{2t}, \end{aligned} \tag{C.3}$$

with scalar processes $\mu_{Ft^-} = \mu_F(p_t^-, s)$ and $\sigma_{Ft^-} = \sigma_F(p_t^-, s)$ satisfying

$$\begin{aligned} &\mu_F(p, s) + \mu_\pi(p, s) + \sigma_F(p, s)\sigma_\pi(p, s) \\ &\qquad + \bar{\lambda}_1(p) E_\nu \left[\frac{G(p\lambda_1^H / \bar{\lambda}_1(p), s) e^{(\gamma - \varphi)Z} e^{(1-\gamma) j(p\lambda_1^H / \bar{\lambda}_1(p))}}{G(p, s) e^{(1-\gamma) j(p)}} - 1 \right] \\ &\qquad\qquad + \bar{\lambda}_2(p) \left(\frac{G(p\lambda_2^H / \bar{\lambda}_2(p), s) e^{(1-\gamma) j(p\lambda_2^H / \bar{\lambda}_2(p))}}{G(p, s) e^{(1-\gamma) j(p)}} - 1 \right) = 0, \forall t. \end{aligned} \tag{C.4}$$

Proof. Absence of arbitrage implies

$$\begin{aligned} F(D_t, p_t, s) &= E_t \left[\frac{\pi_{t+s} D_{t+s}}{\pi_t} \right] \\ &= D_t E_t \left[\frac{\pi_{t+s} D_{t+s}}{\pi_t D_t} \right] \end{aligned}$$

is the time-t price of the equity strip with maturity s, or (C.1). The Markovian property of the state variable implies (C.2). Itô 's Lemma leads to (C.3) and verifies that μ_{Ft^-} and σ_{Ft^-} are functions of p_{t^-} and s.

In addition, absence of arbitrage implies that the process $\pi_t F(D_t, \lambda_{1t}, s)$ must be a martingale. Consider a sufficiently small but positive Δt. It follows Itô 's Lemma that

$$\begin{aligned}\pi_{t+\Delta t}F_{t+\Delta t} = \pi_t F_t + \int_t^{t+\Delta t} \pi_u F_u \left(\mu_F(p_u) + \mu_\pi(p_u) + \sigma_F(p_u)\sigma_\pi(p_u)\right) du \\ + \int_t^{t+\Delta t} \pi_u F_u \left(\sigma_F(p_u) + \sigma_\pi(p_u)\right) dB_{Cu} \\ + \sum_{t<u_{1k}\leq t+\Delta t} (\pi_{u_{1k}} F_{u_{1k}} - \pi_{u_{1k}^-} F_{u_{1k}^-}) + \sum_{t<u_{2k}\leq t+\Delta t} (\pi_{u_{2k}} F_{u_{2k}} - \pi_{u_{2k}^-} F_{u_{2k}^-}). \quad \text{(C.5)}\end{aligned}$$

Here $u_{ik} = \min\{t : N_{it} \geq k\}, i = 1, 2$ is the arrival time of the i^{th} type-j Poisson jump. With (B.13) and (C.1), we know

$$\frac{\pi_{u_{1k}} F_{u_{1k}}}{\pi_{u_{1k}^-} F_{u_{1k}^-}} - 1 = \frac{G(p_{u_{1k}^-} \lambda_1^H / \bar{\lambda}_1(p_{u_{1k}^-}), s) e^{(\gamma-\varphi)Z_{u_{1k}}} e^{(1-\gamma)j(p_{u_{1k}^-} \lambda_1^H / \bar{\lambda}_1(p_{u_{1k}^-}))}}{G(p_{u_{1k}^-}, s) e^{(1-\gamma)j(p_{u_{1k}^-})}} - 1 \quad \text{(C.6)}$$

$$\frac{\pi_{u_{2k}} F_{u_{2k}}}{\pi_{u_{2k}^-} F_{u_{2k}^-}} - 1 = \frac{G(p_{u_{2k}^-} \lambda_2^H / \bar{\lambda}_2(p_{u_{2k}^-}), s) e^{(1-\gamma)j(p_{u_{2k}^-} \lambda_2^H / \bar{\lambda}_2(p_{u_{2k}^-}))}}{G(p_{u_{2k}^-}, s) e^{(1-\gamma)j(p_{u_{2k}^-})}} - 1. \quad \text{(C.7)}$$

Reordering and we have

$$
\begin{aligned}
\pi_{t+\Delta t}F_{t+\Delta t} = \pi_t F_t + \int_{t^+}^{t+\Delta t} \pi_u F_u \Bigg(& \mu_F(p_u,s) + \mu_\pi(p_u) + \sigma_F(p_u,s)\sigma_\pi(p_u) \\
& + \bar\lambda_1(p_u) E_\nu \left[\frac{G(p_u\lambda_1^H/\bar\lambda_1(p_u),s)e^{(\gamma-\varphi)Z_u}e^{(1-\gamma)j(p_u\lambda_1^H/\bar\lambda_1(p_u))}}{G(p_u,s)e^{(1-\gamma)j(p_u)}} - 1 \right] \\
& + \bar\lambda_2(p_u) \left(\frac{G(p_u\lambda_2^H/\bar\lambda_2(p_u),s)e^{(1-\gamma)j(p_u\lambda_2^H/\bar\lambda_2(p_u))}}{G(p_u,s)e^{(1-\gamma)j(p_u)}} - 1 \right) \Bigg) du \\
& + \underbrace{\int_{t^+}^{t+\Delta t} \pi_u F_u (\sigma_F(p_u,s) + \sigma_\pi(p_u)) dB_{Cu}}_{(C.8.1)} \\
& + \underbrace{\sum_{t<u_{1k}\le t+\Delta t} \left(\pi_{u_{1k}} F_{u_{1k}} - \pi_{u_{1k}^-} F_{u_{1k}^-} \right) - \int_{t^+}^{t+\Delta t} \bar\lambda_1(p_u) E_\nu \left[\frac{G(p_u\lambda_1^H/\bar\lambda_1(p_u),s)e^{(\gamma-\varphi)Z_u}e^{(1-\gamma)j(p_u\lambda_1^H/\bar\lambda_1(p_u))}}{G(p_u,s)e^{(1-\gamma)j(p_u)}} - 1 \right]}_{(C.8.2)} \\
& + \underbrace{\sum_{t<u_{2k}\le t+\Delta t} \left(\pi_{u_{2k}} F_{u_{2k}} - \pi_{u_{2k}^-} F_{u_{2k}^-} \right) - \int_{t^+}^{t+\Delta t} \bar\lambda_2(p_u) \left(\frac{G(p_u\lambda_2^H/\bar\lambda_2(p_u),s)e^{(1-\gamma)j(p_u\lambda_2^H/\bar\lambda_2(p_u))}}{G(p_u,s)e^{(1-\gamma)j(p_u)}} - 1 \right) du}_{(C.8.3)} .
\end{aligned}
\tag{C.8}
$$

As (C.8.1), (C.8.2) and (C.8.3) all equal zero in expectation, the first integral must equal zero in expectation as well, which implies (C.4). □

Proof of Theorem 4. Conjecture that $G(p,s) = e^{g(p,s)}$, where $g(p,s)$ is a continuously differentiable function with respect to p and s. As $F(D,p,0) = D$, $g(p,0) = 0$, $\forall p \in [0,1]$.

Itô 's Lemma implies that

$$
\mu_F(p,s) = \mu_D + \frac{\partial g}{\partial p}\left[\phi_{L\to H} - p(\phi_{H\to L} + \phi_{L\to H}) - p\iota^\top\left(\lambda^H - \bar\lambda(p)\right)\right] - \frac{\partial g}{\partial s} \tag{C.9}
$$

$$
\sigma_F(p,s) = \varphi\sigma_C. \tag{C.10}
$$

The conjecture and (B.12) yield that

$$
\begin{aligned}
& \frac{G(p\lambda_1^H/\bar\lambda_1(p),s)e^{(\gamma-\varphi)Z}e^{(1-\gamma)j(p\lambda_1^H/\bar\lambda_1(p))}}{G(p,s)e^{(1-\gamma)j(p)}} \\
& \qquad = e^{(\gamma-\varphi)Z} \times e^{(1-\gamma)\left(j(p\lambda_1^H/\bar\lambda_1(p)) - j(p)\right) + g(p\lambda_1^H/\bar\lambda_1(p),s) - g(p,s)}
\end{aligned}
\tag{C.11}
$$

$$\frac{G(p\lambda_2^H/\bar{\lambda}_2(p), s)e^{(1-\gamma)j(p\lambda_2^H/\bar{\lambda}_2(p))}}{G(p,s)e^{(1-\gamma)j(p)}}$$
$$= e^{(1-\gamma)\left(j(p\lambda_2^H p/\bar{\lambda}_2(p))-j(p)\right)+g(p\lambda_2^H/\bar{\lambda}_2(p),s)-g(p,s)} \quad \text{(C.12)}$$

Substituting (C.9), (C.10), (C.11), (C.12) into (C.4) yields Equation 24.

□

Proof of Theorem 5. The instantaneous expected return of a dividend strip with maturity s is given by

$$\begin{aligned} r_{t^-}(s) - r_{ft^-} &= E_{t^-}\left[\frac{dF_{t^-}}{F_{t^-}dt}\right] - r_f(p_{t^-}) \\ &= \mu_F(p_{t^-}, s) + \lambda_1(p_{t^-})E_\nu\left[\frac{e^{-\varphi Z}G(\lambda_1^H p_{t^-}/\bar{\lambda}_1(p_{t^-}), s)}{G(p_{t^-}, s)}\right] \\ &\quad + \lambda_2(p_{t^-})E_\nu\left[\frac{G(\lambda_2^H p_{t^-}/\bar{\lambda}_2(p_{t^-}), s)}{G(p_{t^-}, s)}\right] - r_f(p_{t^-}) \end{aligned} \quad \text{(C.13)}$$

Substituting (C.9) , (24) and (20) into (C.13) yields

$$
\begin{aligned}
r_{t^-}(s) - r_{ft^-} &= E_{t^-}\left[\frac{dF_{t^-}}{F_{t^-}dt}\right] - r_f(p_{t^-}) \\
&= \mu_D + \frac{\partial g}{\partial p_{t^-}}\left[\phi_{L\to H} - p_{t^-}(\phi_{H\to L} + \phi_{L\to H}) - p_{t^-}\iota^\top\left(\lambda^H - \bar{\lambda}(p_{t^-})\right)\right] - \frac{\partial g}{\partial s} \\
&\quad + \bar{\lambda}_1(p_{t^-})E_\nu\left[\frac{e^{-\varphi Z}G(p_{t^-}\lambda_1^H/\bar{\lambda}_1(p_{t^-}), s)}{G(p_{t^-}, s)} - 1\right] \\
&\quad + \bar{\lambda}_2(p_{t^-})E_\nu\left[\frac{G(p_{t^-}\lambda_2^H/\bar{\lambda}_2(p_{t^-}), s)}{G(p_{t^-}, s)} - 1\right] \\
&\quad - \beta - \mu_C + \gamma\sigma_C^2 - \bar{\lambda}_1(p_{t^-})e^{(1-\gamma)\left(j(p_{t^-}\lambda_1^H/\bar{\lambda}_1(p_{t^-})) - j(p_{t^-})\right)}E_\nu\left[e^{\gamma Z}\left(e^{-Z} - 1\right)\right] \\
&= \gamma\varphi\sigma_C^2 + \bar{\lambda}_1(p_{t^-})E_\nu\left[\frac{e^{-\varphi Z}G(p_{t^-}\lambda_1^H/\bar{\lambda}_1(p_{t^-}), s)}{G(p_{t^-}, s)} - 1\right] \\
&\quad - \bar{\lambda}_1(p_{t^-})e^{(1-\gamma)\left(j(p_{t^-}\lambda_1^H/\bar{\lambda}_1(p_{t^-})) - j(p_{t^-})\right)}E_\nu\left[e^{\gamma Z}\left(e^{-Z} - 1\right)\right] \\
&\quad - \bar{\lambda}_1(p_{t^-})e^{(1-\gamma)\left(j(p_{t^-}\lambda_1^H/\bar{\lambda}_1(p_{t^-})) - j(p_{t^-})\right)}E_\nu\left[\frac{G(p_{t^-}\lambda_2^H/\bar{\lambda}_2(p_{t^-}), s)}{G(p_{t^-}, s)}e^{(\gamma-\varphi)Z} - e^{(\gamma-1)Z}\right] \\
&\quad + \bar{\lambda}_2(p_{t^-})\left(\frac{G(p_{t^-}\lambda_2^H/\bar{\lambda}_2(p_{t^-}), s)}{G(p_{t^-}, s)} - 1\right) \\
&\quad - \bar{\lambda}_2(p_{t^-})e^{(1-\gamma)\left(j(p_{t^-}\lambda_2^H/\bar{\lambda}_2(p_{t^-})) - j(p_{t^-})\right)}\left(\frac{G(p_{t^-}\lambda_2^H/\bar{\lambda}_2(p_{t^-}), s)}{G(p_{t^-}, s)} - 1\right) \\
&= \gamma\varphi\sigma_C^2 \\
&\quad - \bar{\lambda}_1(p_{t^-})E_\nu\left[\left(\frac{e^{-\varphi Z}e^{g(p_{t^-}\lambda_1^H/\bar{\lambda}_1(p_{t^-}), s)}}{e^{g(p_{t^-}, s)}} - 1\right)\left(e^{\gamma Z + (1-\gamma)\left(j\left(p_{t^-}\lambda_1^H/\bar{\lambda}_1(p_{t^-})\right) - j(p_{t^-})\right)}\right)\right] \\
&\quad - \bar{\lambda}_2(p_{t^-})\left(\left(\frac{e^{g(p_{t^-}\lambda_2^H/\bar{\lambda}_2(p_{t^-}), s)}}{e^{g(p_{t^-}, s)}} - 1\right)\left(e^{(1-\gamma)\left(j(p_{t^-}\lambda_2^H/\bar{\lambda}_2(p_{t^-})) - j(p_{t^-})\right)}\right)\right),
\end{aligned}
$$

which verifies that the instantaneous risk premium is a function of p_{t^-}, and the functional form given by (28).zen □

C.2 Pricing equity

We hereby present a lemma parallel to Lemma C.1.

Lemma C.2. Let $\{D_{t+s}\}_{s>0}$ be a dividend stream with dynamics given by 21, $\forall s$.

Then

$$S(D_t, p_t) = \int\limits_{s=0}^{\infty} F(D_t, p_t, s) ds$$
$$= D_t \int\limits_{s=0}^{\infty} G(p_t, s) ds \tag{C.14}$$

is the time-t price of the dividend stream, or equity. Moreover, there exist processes μ_{Ft} and σ_{Ft}, such that

$$\frac{dS_t}{S_{t^-}} = \mu_{St^-} dt + \sigma_{St^-} dB_{C,t} + \left(\frac{S(D_{t^-} e^{-\varphi Z_t}, p_{t^-} \lambda_1^H / \lambda_1(p_{t^-}))}{S(D_{t^-}, p_{t^-})} - 1 \right) dN_{1t} + \left(\frac{S(D_{t^-}, p_{t^-} \lambda_2^H / \lambda_2(p_{t^-}))}{S(D_{t^-}, p_{t^-})} - 1 \right) dN_{2t}, \tag{C.15}$$

with $\mu_{St^-} = \mu_S(p_{t^-})$ and $\sigma_{St^-} = \sigma_S(p_{t^-})$ satisfying

$$\mu_S(p) + \mu_\pi(p) + \frac{D}{S(p)} + \bar{\lambda}_1(p) E_\nu \left[\frac{S(De^{-\varphi Z}, p\lambda_1^H / \bar{\lambda}_1(p)) e^{\gamma Z} e^{(1-\gamma) j(p\lambda_1^H / \bar{\lambda}_1(p))}}{S(D, p) e^{(1-\gamma) j(p)}} - 1 \right] + \bar{\lambda}_2(p) \left(\frac{S(D, p\lambda_2^H / \bar{\lambda}_2(p)) e^{(1-\gamma) j(p\lambda_2^H / \bar{\lambda}_2(p))}}{S(D, p) e^{(1-\gamma) j(p)}} - 1 \right) = 0. \forall p. \tag{C.16}$$

Proof. Equation C.14 follows the absence of arbitrage. For simplicity, denote $F_t(s) = F(D_t, p_t, s)$.

Apply Itô 's lemma on both sides of Equation C.14, and we get

$$\sigma_{St} = \int_0^\infty \frac{G(p_t, s)}{\int_0^\infty G(p_t, u) du} \sigma_H(p_t, s) ds, \tag{C.17}$$

which is a function of p_t and verifies $\sigma_{St} = \sigma_S(p_t)$.

In addition, we have

$$\begin{aligned}\pi_t S(D_t, p_t) - \pi_{t^-} S(D_{t^-}, p_{t^-}) &= \pi_t \int_0^\infty F(D_t, p_t, s) ds - \pi_{t^-} \int_0^\infty F(D_{t^-}, p_{t^-}, s) ds \\ &= \int_0^\infty \big(\pi_t F(D_t, p_t, s) - \pi_{t^-} F(D_{t^-}, p_{t^-}, s)\big) ds.\end{aligned} \tag{C.18}$$

Finally, by Itô 's Lemma, we can see

$$S(D_t, p_t)\mu_S(D_t, p_t) = \int_0^\infty F(D_t, p_t, s)\mu_F(p_t, s) ds - D_t, \tag{C.19}$$

D_t term shows up as $F(D_t, p_t, 0) = D_t$.

Then we have

$$
\begin{aligned}
&\mu_S(p) + \frac{D}{S(D,p)} + \sigma_\pi(p)\sigma_S(p) \\
&\qquad + \bar{\lambda}_1(p) E_\nu \left[\frac{S(De^{-\varphi Z}, \lambda_1^H p/\bar{\lambda}_1(p)) e^{\gamma Z} e^{(1-\gamma) j(p\lambda_1^H/\bar{\lambda}_1(p))}}{S(D,p) e^{(1-\gamma) j(p)}} - 1 \right] \\
&\qquad\qquad + \bar{\lambda}_2(p) \left(\frac{S(D, p\lambda_2^H/\bar{\lambda}_2(p)) e^{(1-\gamma) j(p\lambda_2^H/\bar{\lambda}_2(p))}}{S(D,p) e^{(1-\gamma) j(p)}} - 1 \right) \\
=& \frac{1}{S(D,p)} \left(\int_0^\infty F(D,p,s) \mu_F(p,s) ds \right) + \frac{1}{S(D,p)} \sigma_\pi(p) \int_0^\infty F(D,p,s) \sigma_F(p,s) ds \\
&\qquad + \bar{\lambda}_1(p) \frac{1}{S(D,p)} \int_0^\infty E_\nu \left[\frac{F(De^{-\varphi Z}, p\lambda_1^H/\bar{\lambda}_1(p), s) e^{\gamma Z} e^{(1-\gamma) j(p\lambda_1^H/\bar{\lambda}_1(p))}}{e^{(1-\gamma) j(p)}} - 1 \right] ds \\
&\qquad\qquad + \bar{\lambda}_2(p) \frac{1}{S(D,p)} \int_0^\infty \left(\frac{F(D, p\lambda_2^H/\bar{\lambda}_2(p), s) e^{(1-\gamma) j(p\lambda_2^H/\bar{\lambda}_2(p))}}{e^{(1-\gamma) j(p)}} - 1 \right) ds \\
=& \frac{1}{S(D,p)} \int_0^\infty F(D,p,s) \Bigg(\mu_F(p,s) + \sigma_\pi(p) \sigma_F(p,s) \\
&\qquad + \bar{\lambda}_1(p) E_\nu \left[\frac{e^{-\varphi Z} G(p\lambda_1^H/\bar{\lambda}_1(p), s) e^{\gamma Z} e^{(1-\gamma) j(p\lambda_1^H/\bar{\lambda}_1(p))}}{G(p,s) e^{(1-\gamma) j(p)}} - 1 \right] \\
&\qquad\qquad + \bar{\lambda}_2(p) \left(\frac{G(p\lambda_2^H/\bar{\lambda}_2(p), s) e^{(1-\gamma) j(p\lambda_2^H/\bar{\lambda}_2(p))}}{G(p,s) e^{(1-\gamma) j(p)}} - 1 \right) \Bigg) ds \\
=& \frac{1}{S(D,p)} \int_0^\infty F(D,p,s)(-\mu_\pi(p)) ds \\
=& -\mu_\pi(p) \frac{1}{S(D,p)} \int_0^\infty F(D,p,s) ds \\
=& -\mu_\pi(p).
\end{aligned}
\tag{C.20}
$$

Replace t^- with t, and we can get Equation C.16. □

C.3 Equity premium

We first proposes a lemma that characterizes the equity premium for a general equity asset.

Lemma C.3. For an asset with claim to a stream of dividends with time-t price

$S(D_t, p_t)$, its instantaneous premium is given by $r_t - r_{ft} = r(p_t) - r_f(p_t)$, where

$$
\begin{aligned}
r(p) - r_f(p) &= -\sigma_\pi(p)\sigma_S(p) \\
&\quad - \bar{\lambda}_1(p) E_\nu \left[\left(\frac{S(De^{-\varphi Z}, p\lambda_1^H/\bar{\lambda}_1(p))}{S(D,p)} - 1 \right) \left(e^{\gamma Z + (1-\gamma)\left(j(p\lambda_1^H/\bar{\lambda}_1(p)) - j(p)\right)} - 1 \right) \right] \\
&\quad - \bar{\lambda}_2(p) \left(\left(\frac{S(D, p\lambda_2^H/\bar{\lambda}_2(p))}{S(D,p_t)} - 1 \right) \left(e^{(1-\gamma)\left(j(p\lambda_2^H/\bar{\lambda}_2(p)) - j(p)\right)} - 1 \right) \right) \quad \text{(C.21)}
\end{aligned}
$$

Proof.

$$
\begin{aligned}
r_{t^-} - r_{ft^-} &= E_t \left[\frac{(dS_t + D_{t^-} dt)}{S_{t^-} dt} \right] - r(p_{t^-}) \\
&= \mu_S(p_{t^-}) + \frac{D_{t^-}}{S_{t^-}} + \bar{\lambda}_1(p_{t^-}) E_\nu \left[\frac{S(D_{t^-} e^{-\varphi Z_t}, p_{t^-}\lambda_1^H/\bar{\lambda}_1(p_{t^-}))}{S(D_{t^-}, p_{t^-})} - 1 \right] \quad \text{(C.22)} \\
&\quad + \bar{\lambda}_2(p_{t^-}) E_\nu \left[\frac{S(D_{t^-}, p_{t^-}\lambda_2^H/\bar{\lambda}_2(p_{t^-}))}{S(D_{t^-}, p_{t^-})} - 1 \right] - r(p_{t^-})
\end{aligned}
$$

Substituting (C.16) and (18) into (C.22) yields The RHS of (C.21), with $p = p_t$. As $S(D,p)$ is homothetic in D, we can show that in fact the RHS of (C.21) is a function of p, which implies that $r_t - r_{ft}$ is a function of p. Finally, as $r_{ft} = r_f(p_t)$, r_t must be a function of p_t, and this verifies the functional form of $r(p_t)$. □

Proof of Theorem 6. Note that

$$
\begin{aligned}
\sigma_\pi(p) &= -\gamma\sigma_C \\
\sigma_S(p) &= \varphi\sigma_C.
\end{aligned}
$$

In addition,

$$E_\nu\left[\left(\frac{S(De^{-\varphi Z}, p\lambda_1^H/\bar{\lambda}_1(p))}{S(D,p)}-1\right)\left(e^{\gamma Z+(1-\gamma)\left(j(p\lambda_1^H/\bar{\lambda}_1(p))-j(p)\right)}-1\right)\right]$$
$$=E_\nu\left[\left(\frac{e^{-\varphi Z}\int\limits_{s=0}^{\infty}e^{g(p\lambda_1^H/\bar{\lambda}_1(p),s)}ds}{\int\limits_{s=0}^{\infty}e^{g(p,s)}ds}-1\right)\left(\frac{e^{\gamma Z}e^{(1-\gamma)j\left(p\lambda_1^H/\bar{\lambda}_1(p)\right)}}{e^{(1-\gamma)j(p)}}-1\right)\right]$$
$$\left(\frac{S(D, p\lambda_2^H/\bar{\lambda}_2(p))}{S(D,p)}-1\right)\left(e^{(1-\gamma)\left(j(p\lambda_2^H/\bar{\lambda}_2(p))-j(p)\right)}-1\right)$$
$$=\left(\frac{\int\limits_{s=0}^{\infty}e^{g(p\lambda_2^H/\bar{\lambda}_2(p),s)}ds}{\int\limits_{s=0}^{\infty}e^{g(p,s)}ds}-1\right)\left(\frac{e^{(1-\gamma)j(p\lambda_2^H/\bar{\lambda}_2(p))}}{e^{(1-\gamma)j(p)}}-1\right).$$

Combining the results above and the results from Lemma C.3 yields Equation 30. □

D The model with full information

In this section we solve the model with complete information. The endowment process and the representative agent's preference are the same as is in the main model. However the representative agent knows exactly the value of λ_{1t}. In this case N_{2t} does not provide any further information and we can focus on a model with N_{1t} only.

D.1 The representative agent's value function

For simplicity, we define a Poisson process N_t^S to capture the regime switch. The process of the regime switch can then be characterized by the following equation:

$$d\lambda_{1t} = (\lambda_1^H + \lambda_1^L - 2\lambda_{1t})dN_t^S. \tag{D.1}$$

When $dN_t^S = 1$, the λ_{1t} changes value and the economy switches the regime it is in.

The (conditional) jump intensity for N_t^S, or the probability of regime switch, is given by:

$$\phi_t = \phi(\lambda_{1t}) = \phi_{H\to L}\mathbf{1}_{\lambda_{1t}=\lambda_1^H} + \phi_{L\to H}\mathbf{1}_{\lambda_{1t}=\lambda_1^L}, \tag{D.2}$$

which is a function of λ_{1t}.

The representative agent's value function is characterized by the following proposition.

Proposition D.1. The representative agent's continuation value V_t is given by

$$V_t = J(C_t, p_t),$$

where

$$J(C, \lambda_1) = \frac{1}{1-\gamma}C^{1-\gamma}e^{(1-\gamma)j(\lambda_1)}, \tag{D.3}$$

where $j(\lambda_1)$, which is defined on $\{\lambda^L, \lambda_1^H\}$, is the solution to the following equations:

$$\beta(1-\gamma)j(\lambda_1^L) = (1-\gamma)\mu_C - \frac{1}{2}\gamma(1-\gamma)\sigma^2 \tag{D.4}$$

$$+ \lambda_1^L E_\nu\left[e^{(\gamma-1)Z} - 1\right] + \phi_{L\to H}\left(e^{(1-\gamma)\left(j(\lambda_1^H)-j(\lambda_1^L)\right)} - 1\right) \tag{D.5}$$

$$\beta(1-\gamma)j(\lambda_1^H) = (1-\gamma)\mu_C - \frac{1}{2}\gamma(1-\gamma)\sigma^2 \tag{D.6}$$

$$+ \lambda_1^H E_\nu\left[e^{(\gamma-1)Z} - 1\right] + \phi_{H\to L}\left(e^{(1-\gamma)\left(j(\lambda_1^L)-j(\lambda_1^H)\right)} - 1\right). \tag{D.7}$$

Proof. Conjecture that the representative agent's continuation value is given by (D.3).

Conditioning on λ_{1t}, the value function must satisfy the following Hamilton-Jacobian-Bellman Equation:

$$f(C,J) + \frac{\partial J}{\partial C}C\mu_C + \frac{\partial^2 J}{\partial C^2}C^2\sigma^2 + \lambda_1 \mathbf{E}_\nu\left[J(Ce^{-Z},\lambda_1) - J(C,\lambda_1)\right] + \phi(\lambda_1)\left(J(C,\lambda^H + \lambda^L - \lambda_1) - J(C,\lambda_1)\right) = 0. \tag{D.8}$$

Given the conjecture in (D.3),

$$\frac{1}{J}\left(J(Ce^{-Z},\lambda_1) - J(C,\lambda_1)\right) = e^{(\gamma-1)Z} - 1. \tag{D.9}$$

Dividing both sides of (D.8) by $J(C,\lambda_1)$, and substituting (D.9) and (3) into the equation yields

$$\beta(1-\gamma)g(\lambda_1) = (1-\gamma)\mu_C - \frac{1}{2}\gamma(1-\gamma)\sigma^2 + \lambda_1 E_\nu\left[e^{(\gamma-1)Z} - 1\right] + \phi(\lambda_1)\left[e^{(1-\gamma)\left(j(\lambda_1^H+\lambda_1^L-\lambda_1)-j(\lambda_1)\right)} - 1\right]. \tag{D.10}$$

As $\lambda_1 \in \{\lambda_1^H, \lambda_1^L\}$, we can explicitly write down the Hamilton-Jacobian-Bellman Equations when $\lambda_1 = \lambda_1^H$ or $\lambda_{1t} = \lambda_1^L$, respectively. This yields (D.4) and (D.6). □

D.2 The state price density

The following theorem characterizes the state price density process of the model.

Theorem D.1. The state-price density of the economy is given by

$$\frac{d\pi_t}{\pi_{t^-}} = \mu_{\pi t^-} dt + \sigma_{\pi t^-} dB_{Ct} + \left(e^{-\gamma Z_t} - 1\right) dN_{1t} + \left(e^{(1-\gamma)\left(j(\lambda_1^H + \lambda_1^L - \lambda_{1t}) - j(\lambda_{1t})\right)} - 1\right) dN_t^S, \quad \text{(D.11)}$$

where $\mu_{\pi t^-} = \mu_\pi(\lambda_{1t^-})$, $\sigma_{\pi t^-} = \sigma_\pi(\lambda_{1t^-})$and

$$\mu_\pi(\lambda_1) = -(\beta + \mu_C - \gamma\sigma_C^2) - \lambda_1 E_\nu\left[e^{(\gamma-1)Z} - 1\right] - \phi(\lambda_1)\left(e^{(1-\gamma)\left(j(\lambda_1^H + \lambda_1^L - \lambda_1) - j(\lambda_1)\right)} - 1\right) \quad \text{(D.12)}$$

$$\sigma_\pi(\lambda_1) = -\gamma\sigma_C. \quad \text{(D.13)}$$

Proof. Duffie and Skiadas (1994) show that

$$\pi_t = \exp\left\{\int_0^t \frac{\partial}{\partial V} f(C_s, V_s)\right\} \frac{\partial}{\partial C} f(C_t, V_t). \quad \text{(D.14)}$$

The functional form of f implies

$$\begin{aligned}\frac{\partial}{\partial C} f(C_t, V_t) &= \beta(1-\gamma)\frac{V_t}{C_t} \\ &= \beta C_t^{-\gamma} e^{(1-\gamma)j(\lambda_{1t})}.\end{aligned} \quad \text{(D.15)}$$

Combining (D.14) and (D.15), we get

$$\pi_t = \beta \exp\left\{\int_0^t \frac{\partial}{\partial V} f(C_s, V_s)\right\} C_t^{-\gamma} e^{(1-\gamma)j(\lambda_{1t})}. \quad \text{(D.16)}$$

With Itô 's Lemma, we have

$$\frac{d\pi_t}{\pi_{t^-}} = \mu_{\pi t^-} dt + \sigma_{\pi t^-} dB_{Ct} + \left(e^{\gamma Z_t} - 1\right) dN_{1t} + \left(e^{(1-\gamma)\left(j(\lambda_1^H + \lambda_1^L - \lambda_{1t}) - j(\lambda_{1t})\right)} - 1\right) dN_t^S, \quad \text{(D.17)}$$

where

$$\mu_{\pi t} = \frac{\partial}{\partial V} f(C_t, V_t) - \gamma \mu_C + \frac{1}{2}\gamma(\gamma+1)\sigma_C^2 \tag{D.18}$$

$$\sigma_{\pi t} = -\gamma \sigma_C \tag{D.19}$$

Combining (3), (D.3) and (D.10), and we have

$$\begin{aligned} \frac{\partial}{\partial V} f(C_t, V_t) &= -\beta - \beta(1-\gamma) j(\lambda_{1t}) \\ &= -\beta - (1-\gamma)\mu_C + \frac{1}{2}\gamma(1-\gamma)\sigma^2 \\ &\quad - \lambda_{1t} E_\nu \left[e^{(\gamma-1)Z_t} - 1 \right] - \phi(\lambda_{1t}) \left[e^{(1-\gamma)\left(j(\lambda_1^H + \lambda_1^L - \lambda_{1t}) - j(\lambda_{1t}) \right)} - 1 \right]. \end{aligned} \tag{D.20}$$

Substituting (D.20) into (D.18), and we get

$$\begin{aligned} \mu_{\pi t} &= -(\beta + \mu_C - \gamma \sigma_C^2) \\ &\quad - \lambda_{1t} E_\nu \left[e^{(\gamma-1)Z} - 1 \right] - \phi(\lambda_{1t}) \left(e^{(1-\gamma)\left(j(\lambda_1^H + \lambda_1^L - \lambda_{1t}) - j(\lambda_{1t}) \right)} - 1 \right), \end{aligned} \tag{D.21}$$

which is a function of λ_{1t} and confirms (D.12). □

Theorem D.2. The riskfree rate, r_{ft}, is given by

$$r_{ft} = r_f(\lambda_{1t}) = \beta + \mu_C - \gamma\sigma^2 + \lambda_{1t} E_\nu \left[e^{\gamma Z_t} \left(e^{-Z_t} - 1 \right) \right]. \tag{D.22}$$

Proof. Non-arbitrage implies that

$$\mathbf{E}_t \left[\frac{d\pi_t}{\pi_t} \right] = -r_t, \tag{D.23}$$

where r_t is the instantaneous riskfree rate. Combining (D.17) and (D.23) yields

$$\mu_{\pi t} = -\left(r_{ft} + \lambda_{1t} E_\nu \left[e^{\gamma Z_t} - 1 \right] + \phi(\lambda_{1t}) e^{(1-\gamma)\left[j(\lambda_1^H + \lambda_1^L - \lambda_{1t}) - j(\lambda_{1t}) \right]} \right). \tag{D.24}$$

Substituting (D.21) into (D.24), and we get (D.22). □

D.3 Pricing an equity strip

The dividend process is given by (21). Again the following lemma characterizes the non-arbitrage condition for the process of equity strips.

Lemma D.1. Define the function

$$F(D_t, \lambda_{1t}, s) = D_t E_t \left[\frac{\pi_{t^*}}{\pi_t} \frac{D_{t^*}}{D_t} \right]. \tag{D.25}$$

Then $F_t = F(D_t, \lambda_{1t}, s)$ is the time-t price of the dividend with maturity s ,and

$$G(\lambda_{1t}, s) \equiv E_t \left[\frac{\pi_{t^*}}{\pi_t} \frac{D_{t^*}}{D_t} \right] \tag{D.26}$$

is the price-dividend ratio at time t. Moreover, F_t satisfies

$$\begin{aligned} \frac{dF_t}{F_{t^-}} &= \mu_{Ft^-} dt + \sigma_{Ft^-} dB_{Ct} \\ &\quad + \left(e^{-\varphi Z_t} - 1 \right) dN_{1t} + \left(\frac{G(\lambda^H + \lambda^L - \lambda_{1t^-}, s)}{G(\lambda_{1t^-}, s)} - 1 \right) dN_t^S, \end{aligned} \tag{D.27}$$

with scalar processes $\mu_{Ht} = \mu_H(\lambda_{1t}, s)$ and $\sigma_{Ht} = \sigma_H(\lambda_{1t} s)$, satisfying

$$\begin{aligned} &\mu_H(\lambda_1, s) + \mu_\pi(\lambda_1) + \sigma_H(\lambda_1, s)\sigma_\pi(\lambda_1) + \lambda_1 E_\nu \left[e^{(\gamma-\varphi)Z_t} - 1 \right] \\ &\quad + \phi(\lambda_1) \left(\frac{G(\lambda^H + \lambda^L - \lambda_1, s) e^{(1-\gamma)j(\lambda^H + \lambda^L - \lambda_1)}}{G(\lambda_1, s) e^{(1-\gamma)j(\lambda_1)}} - 1 \right) = 0, \end{aligned} \tag{D.28}$$

with $\mu_\pi(\lambda_1)$ and $\sigma_\pi(\lambda_1)$ given by (D.12) and (D.13), and $\phi(\lambda_1)$ given by

$$\begin{aligned} \phi(\lambda_1^H) &= \phi_{H \to L} \\ \phi(\lambda_1^L) &= \phi_{L \to H}. \end{aligned}$$

Proof. Non-arbitrage condition implies (D.25), and Markovian property of the state variables implies (D.26).

Itô 's Lemma leads to (D.27) and the functional forms of μ_H and σ_H.

Absence of arbitrage implies that the process $\pi_t F(D_t, \lambda_{1t}, s)$ must be a martingale.

Consider a sufficiently small but positive Δt. It follows Itô 's Lemma that

$$\pi_{t+\Delta t}F_{t+\Delta t} = \int_t^{t+\Delta t} \pi_u F_u(\mu_{Fu} + \mu_{\pi u} + \sigma_{Fu}\sigma_{\pi u})du + \int_t^{t+\Delta t} \pi_u F_u(\sigma_{Fu} + \sigma_{\pi u})dB_{Cu} + \sum_{t<u_{1k}\leq t+\Delta t} \left(\pi_{u_{1k}}F_{u_{1k}} - \pi_{u_{1k}^-}F_{u_{1k}^-}\right) + \sum_{t<u_k^S\leq t+\Delta t} \left(\pi_{u_k^S}F_{u_k^S} - \pi_{u_k^S-}F_{u_k^S-}\right), \quad \text{(D.29)}$$

where $u_{1k} \equiv \min\{t : N_{1t} \geq k\}$ and $u_k^S \equiv \min\{t : N_t^S \geq k\}$. With (D.17) and (D.25), we know

$$\pi_{u_{1k}}H_{u_{1k}} - \pi_{u_{1k}^-}H_{u_{1k}^-} = \frac{G(\lambda_{1u_{1k}}, s)e^{(\gamma-\varphi)Z_{u_{1k}}}}{G(\lambda_{1u_{1k}}, s)} - 1 = e^{(\gamma-\varphi)Z_{u_{1k}}} - 1 \quad \text{(D.30)}$$

$$\pi_{u_k^S}H_{u_k^S} - \pi_{u_k^S-}H_{u_k^S-} = \frac{G(\lambda_{1u_k^S}, s)e^{(1-\gamma)j(\lambda_{1u_k^S})}}{G(\lambda_{1u_k^S-}, s)e^{(1-\gamma)j(\lambda_{1u_k^S-})}} - 1 \quad \text{(D.31)}$$

Reordering and we have

$$\begin{aligned}\pi_{t+\Delta t}H_{t+\Delta t} = \pi_t H_t + \int_{t^+}^{t+\Delta t} \pi_u H_u \Bigg(& \mu_H(\lambda_{1u}, s) + \mu_\pi(\lambda_{1u}) + \sigma_H(\lambda_{1u}, s)\sigma_\pi(\lambda_{1u}) \\ & + \lambda_{1u}E_\nu\left[e^{(\gamma-\varphi)Z_u} - 1\right] \\ & + \phi(\lambda_{1u})\left(\frac{G(\lambda^H + \lambda^L - \lambda_{1u}, s)e^{(1-\gamma)g(\lambda^H+\lambda^L-\lambda_{1u})}}{G(\lambda_{1u}, s)e^{(1-\gamma)g(\lambda_{1u})}} - 1\right)\Bigg)du \\ & + \underbrace{\int_{t^+}^{t+\Delta t} \pi_u H_u(\sigma_{Hu} + \sigma_{\pi u})dB_{Cu}}_{(D.32.1)} \\ & + \underbrace{\sum_{t<u_{1k}\leq t+\Delta t}\left(\pi_{u_{1k}}H_{u_{1k}} - \pi_{u_{1k}^-}H_{u_{1k}^-}\right) - \int_{t^+}^{t+\Delta t}\lambda_{1u}E_\nu\left[e^{(\gamma-\varphi)Z_u} - 1\right]}_{(D.32.2)} \\ & + \underbrace{\sum_{t<u_k^S\leq t+\Delta t}\left(\pi_{u_k^S}H_{u_k^S} - \pi_{u_k^S-}H_{u_k^S-}\right) - \int_{t^+}^{t+\Delta t}\phi(\lambda_{1u})\left(\frac{G(\lambda^H + \lambda^L - \lambda_{1u}, s)e^{(1-\gamma)g(\lambda^H+\lambda^L-\lambda_{1u})}}{G(\lambda_{1u}, s)e^{(1-\gamma)g(\lambda_{1u})}} - 1\right)du}_{(D.32.3)}.\end{aligned} \quad \text{(D.32)}$$

As (D.32.1), (D.32.2) and (D.32.3) all equal zero in expectation, the first integral must equal zero in expectation as well, which implies (D.28). □

The following theorem gives the functional form of the price of an equity strip.

Theorem D.3. The time-t price of an equity strip maturing at time $t+s$, D_{t+s}, is given by

$$F_t = F(D_t, \lambda_t, s) = D_t G(s; \lambda_{1t}) = D_t e^{g(s;\lambda_{1t})}, \tag{D.33}$$

where the continuously differentiable function $g(s; \lambda_{1t})$ is the solution to the following system of ordinary differentiable equations:

$$\frac{d}{ds} g(s; \lambda_1^L) = -\beta + \mu_D - \mu_C + \gamma(1-\varphi)\sigma_C^2 + \lambda_1^L E_\nu \left[e^{(\gamma-\varphi)Z} - e^{(\gamma-1)Z} \right] + \phi_{L\to H} e^{(1-\gamma)\left[j(\lambda_1^H) - j(\lambda_1^L)\right]} \left(e^{g(s;\lambda_1^H) - g(s;\lambda_1^L)} - 1 \right) \tag{D.34}$$

$$\frac{d}{ds} g(s; \lambda_1^H) = -\beta + \mu_D - \mu_C + \gamma(1-\varphi)\sigma_C^2 + \lambda_1^H E_\nu \left[e^{(\gamma-\varphi)Z} - e^{(\gamma-1)Z} \right] + \phi_{L\to H} e^{(1-\gamma)\left[j(\lambda_1^L) - j(\lambda_1^H)\right]} \left(e^{g(s;\lambda_1^L) - g(s;\lambda_1^H)} - 1 \right). \tag{D.35}$$

$$\tag{D.36}$$

with boundary condition

$$g(0, \lambda_{1t}) = 0, \;\; \lambda_{1t} = \lambda_1^L, \lambda_1^H. \tag{D.37}$$

Proof. We proof this by conjecture and verify.

Conjecture that the time-t price of an equity strip with maturity s is given by (D.33). First $D_t = H(D_t, \lambda_{1t}, s)$ implies (D.37).

Itô 's Lemma suggest that

$$\mu_H(\lambda_1, s) = \mu_D - \frac{d}{ds} G(s; \lambda_1) \tag{D.38}$$

$$\sigma_H(\lambda_1, s) = \varphi \sigma_C. \tag{D.39}$$

In addition,

$$\frac{G(\lambda^H + \lambda^L - \lambda_1, s) e^{(1-\gamma) j(\lambda^H + \lambda^L - \lambda_1)}}{G(\lambda_1, s) e^{(1-\gamma) j(\lambda_1)}} - 1 = e^{(1-\gamma)\left(j(\lambda^H + \lambda^L - \lambda_1) - j(\lambda_1)\right) + g(s;\lambda^H + \lambda^L - \lambda_1) - g(s;\lambda_1)} - 1. \tag{D.40}$$

Substituting (D.12), (D.13), (D.38), (D.39) and (D.40) into (D.28) yields

$$\frac{d}{ds}g(s;\lambda_1) = -\beta + \mu_D - \mu_C + \gamma(1-\varphi)\sigma_C^2 + \lambda_1 E_\nu \left[e^{(\gamma-\varphi)Z} - e^{(\gamma-1)Z}\right] + \phi(\lambda_1)e^{(1-\gamma)\left[j(\lambda^H+\lambda^L-\lambda_1)-j(\lambda_1)\right]}\left(e^{g(s;\lambda^H+\lambda^L-\lambda_1)-g(s;\lambda_1)} - 1\right). \quad \text{(D.41)}$$

As $\lambda_{1t} \in \{\lambda_1^H, \lambda_1^L\}$, we then can specifically write down (D.34) and (D.35) conditioning on the value of λ_1.

□

D.4 Equity premium

In what follows we present a lemma parallel to Lemma D.1.

Lemma D.2. Let $\{D_{t+s}\}_{s>0}$ be a dividend stream with dynamics given by 21, $\forall s$. Then

$$\begin{aligned} S(D_t, \lambda_{1t}) &= \int_{s=0}^{\infty} F(D_t, \lambda_{1t}, s)ds \\ &= D_t \int_{s=0}^{\infty} G(\lambda_{1t}, s)ds \end{aligned} \quad \text{(D.42)}$$

is the time-t price of the dividend stream, or equity. Moreover, there exist processes μ_{Ft} and σ_{Ft}, such that

$$\frac{dS_t}{S_{t^-}} = \mu_{St^-}dt + \sigma_{St^-}dB_{C,t} + \left(e^{-\varphi Z_t} - 1\right)dN_{1t} + \left(\frac{S(D_{t^-}, \lambda_1^H + \lambda_1^L - \lambda_{1t^-})}{S(D_{t^-}, \lambda_{1t^-})} - 1\right)dN_{2t}, \quad \text{(D.43)}$$

with $\mu_{St^-} = \mu_S(p_{t^-})$ and $\sigma_{St^-} = \sigma_S(p_{t^-})$ satisfying

$$\mu_S(p) + \mu_\pi(p) + \frac{D}{S(p)} + \bar{\lambda}_1(p) E_\nu \left[\frac{S(De^{-\varphi Z}, \lambda_1) e^{\gamma Z}}{S(D, \lambda_1)} - 1 \right] + \bar{\lambda}_2(p) \left(\frac{S(D, \lambda_1^H + \lambda_1^L - \lambda_1) e^{(1-\gamma) j(\lambda_1^H + \lambda_1^L - \lambda_1)}}{S(D, \lambda_1) e^{(1-\gamma) j(\lambda_1)}} - 1 \right) = 0. \forall p. \quad \text{(D.44)}$$

Proof. Equation D.42 follows the absence of arbitrage. For simplicity, denote $F_t(s) = F(D_t, \lambda_{1t}, s)$.

Apply Itô 's lemma on both sides of Equation D.42, and we get

$$\sigma_{St} = \int_0^\infty \frac{G(\lambda_{1t}, s)}{\int_0^\infty G(\lambda_{1t}, u) du} \sigma_H(\lambda_{1t}, s) ds, \quad \text{(D.45)}$$

which is a function of λ_{1t} and verifies $\sigma_{St} = \sigma_S(\lambda_{1t})$.

In addition, we have

$$\begin{aligned} \pi_t S(D_t, \lambda_{1t}) - \pi_{t^-} S(D_{t^-}, \lambda_{1t^-}) &= \pi_t \int_0^\infty F(D_t, \lambda_{1t}, s) ds - \pi_{t^-} \int_0^\infty F(D_{t^-}, \lambda_{1t^-}, s) ds \\ &= \int_0^\infty \big(\pi_t F(D_t, \lambda_{1t}, s) - \pi_{t^-} F(D_{t^-}, \lambda_{1t^-}, s) \big) ds. \end{aligned} \quad \text{(D.46)}$$

Finally, by Itô 's Lemma, we can see

$$S(D_t, \lambda_{1t}) \mu_S(D_t, \lambda_{1t}) = \int_0^\infty F(D_t, \lambda_{1t}, s) \mu_F(\lambda_{1t}, s) ds - D_t, \quad \text{(D.47)}$$

D_t term shows up as $F(D_t, \lambda_{1t}, 0) = D_t$.

Then we have

$$
\begin{aligned}
&\mu_S(\lambda_1) + \frac{D}{S(D,\lambda_1)} + \sigma_\pi(\lambda_1)\sigma_S(\lambda_1) \\
&\qquad + \lambda_1 E_\nu \left[\frac{S(De^{-\varphi Z},\lambda_1)e^{\gamma Z}}{S(D,\lambda_1)} - 1 \right] \\
&\qquad\qquad + \phi(\lambda_1)\left(\frac{S(D,\lambda_1^H+\lambda_1^L-\lambda_1)e^{(1-\gamma)j(\lambda_1^H+\lambda_1^L-\lambda_1)}}{S(D,\lambda_1)e^{(1-\gamma)j(\lambda_1)}} - 1 \right) \\
=&\frac{1}{S(D,\lambda_1)}\left(\int_0^\infty F(D,\lambda_1,s)\mu_F(\lambda_1,s)ds \right) + \frac{1}{S(D,\lambda_1)}\sigma_\pi(\lambda_1)\int_0^\infty F(D,\lambda_1,s)\sigma_F(\lambda_1,s)ds \\
&\qquad + \lambda_1 \frac{1}{S(D,\lambda_1)}\int_0^\infty E_\nu\left[F(De^{-\varphi Z},\lambda_1,s)e^{\gamma Z} - 1 \right] ds \\
&\qquad\qquad + \phi(\lambda_1)\frac{1}{S(D,\lambda_1)}\int_0^\infty \left(\frac{F(D,\lambda_1^H+\lambda_1^L-\lambda_1,s)e^{(1-\gamma)j(\lambda_1^H+\lambda_1^L-\lambda_1)}}{e^{(1-\gamma)j(\lambda_1)}} - 1 \right) ds \\
=&\frac{1}{S(D,\lambda_1)}\int_0^\infty F(D,\lambda_1,s)\Bigg(\mu_F(\lambda_1,s) + \sigma_\pi(\lambda_1)\sigma_F(\lambda_1,s) \\
&\qquad + \lambda_1 E_\nu\left[\frac{G(\lambda_1,s)e^{(\gamma-\varphi)Z}}{G(\lambda_1,s)} - 1 \right] \\
&\qquad\qquad + \phi(\lambda_1)\left(\frac{G(\lambda_1^H+\lambda_1^L-\lambda_1 s)e^{(1-\gamma)j(\lambda_1^H+\lambda_1^L-\lambda_1)}}{G(\lambda_1,s)e^{(1-\gamma)j(\lambda_1)}} - 1 \right) \Bigg) ds \\
=&\frac{1}{S(D,\lambda_1)}\int_0^\infty F(D,\lambda_1,s)(-\mu_\pi(\lambda_1))ds \\
=&-\mu_\pi(\lambda_1)\frac{1}{S(D,\lambda_1)}\int_0^\infty F(D,\lambda_1,s)ds \\
=&-\mu_\pi(\lambda_1).
\end{aligned}
\tag{D.48}
$$

Replace t^- with t, and we can get Equation D.44. □

In what follows, we proposes a lemma that characterizes the equity premium for a general equity asset.

Lemma D.3. For an asset with claim to a stream of dividends with time-t price

$S(D_t, \lambda_{1t})$, its instantaneous premium is given by $r_t - r_{ft} = r(\lambda_{1t}) - r_f(\lambda_{1t})$, where

$$\begin{aligned} r(\lambda_1) - r_f(\lambda_1) = & -\sigma_\pi(\lambda_1)\sigma_S(\lambda_1) \\ & - \lambda_1 E_\nu\left[\left(\frac{S(De^{-\varphi Z}, \lambda_1)}{S(D, \lambda_1)} - 1\right)\left(e^{\gamma Z} - 1\right)\right] \\ & - \phi(\lambda_1)\left(\left(\frac{S(D, \lambda_1^H + \lambda_1^L - \lambda_1)}{S(D, \lambda_1)} - 1\right)\left(e^{(1-\gamma)\left(j(\lambda_1^H + \lambda_1^L - \lambda_1) - j(\lambda_1)\right)} - 1\right)\right) \end{aligned} \quad \text{(D.49)}$$

Proof.

$$\begin{aligned} r_{t^-} - r_{ft^-} &= E_{t^-}\left[\frac{(dS_t + D_{t^-}dt)}{dt}\right] - r(\lambda_{1t^-}) \\ &= \mu_S(\lambda_{1t^-}) + \frac{D_{t^-}}{S_{t^-}} + \lambda_{1t^-}E_\nu\left[\frac{S(D_{t^-}e^{-\varphi Z_t}, \lambda_{1t^-})}{S(D_{t^-}, \lambda_{1t^-})} - 1\right] \\ &\quad + \phi(\lambda_{1t^-})E_\nu\left[\frac{S(D_{t^-}, \lambda_1^H + \lambda_1^L - \lambda_{1t^-})}{S(D_{t^-}, \lambda_{1t^-})} - 1\right] - r(\lambda_{1t^-}) \end{aligned} \quad \text{(D.50)}$$

Substituting (D.44) and (D.12) into (D.50) yields The RHS of (D.49), with $\lambda_1 = \lambda_{1t}$. As $S(D, \lambda_1)$ is homothetic in D, we can show that in fact the RHS of (D.49) is a function of λ_1, which implies that $r_t - r_{ft}$ is a function of λ_1. Finally, as $r_{ft} = r_f(\lambda_1)$, r_t must be a function of λ_{1t}, and this verifies the functional form of $r(\lambda_{1t})$. □

The following theorem characterizes equity premium in the case with full information.

Theorem D.4. The instantaneous risk premium for an equity asset as a claim to (21) is given by

$$\begin{aligned} r(\lambda_1) - r_f(\lambda_1) = & \gamma\varphi\sigma_C^2 \\ & - \bar{\lambda}_1 E_\nu\left[\left(e^{-\varphi Z} - 1\right)\left(e^{\gamma Z} - 1\right)\right] \\ & - \phi(\lambda_1)\left(\left(\frac{S(D, \lambda_1^H + \lambda_1^L - \lambda_1))}{S(D, \lambda_1)} - 1\right)\left(e^{(1-\gamma)\left(j(\lambda_1^H + \lambda_1^L - \lambda_1) - j(\lambda_1)\right)} - 1\right)\right). \end{aligned} \quad \text{(D.51)}$$

Proof of Theorem D.4. Note that

$$\begin{aligned} \sigma_\pi(p) &= -\gamma\sigma_C \\ \sigma_S(p) &= \varphi\sigma_C. \end{aligned}$$

In addition,

$$
\begin{aligned}
& E_\nu\left[\left(\frac{S(De^{-\varphi Z},\lambda_1)}{S(D,\lambda_1)}-1\right)\left(e^{\gamma Z}-1\right)\right] \\
= & E_\nu\left[\left(e^{-\varphi Z}-1\right)\left(e^{\gamma Z}-1\right)\right] \\
& \left(\frac{S(D,\lambda_1^H+\lambda_1^L-\lambda_1)}{S(D,\lambda_1)}-1\right)\left(e^{(1-\gamma)\left(j(\lambda_1^H+\lambda_1^L-\lambda_1)-j(\lambda_1)\right)}-1\right) \\
= & \left(\frac{S(D,\lambda_1^H+\lambda_1^L-\lambda_1))}{S(D,\lambda_1)}-1\right)\left(e^{(1-\gamma)\left(j(\lambda_1^H+\lambda_1^L-\lambda_1)-j(\lambda_1)\right)}-1\right).
\end{aligned}
$$

Combining the results above and the results from Lemma C.3 yields Equation 30. □

References

Bansal, R. and Yaron, A. (2004). Risks for the long run: a potential resolution of asset pricing puzzles. *Journal of Finance*, 59(4):1481–1509.

Barro, R. J. (2006). Rare disasters and asset markets in the twentieth century. *Quarterly Journal of Economics*, 121(3):823–866.

Barro, R. J. and Ursúa, J. F. (2008). Macroeconomic crises since 1870. *Brookings Papers on Economic Activity*, 39(1 (Spring)):255–350.

Benzoni, L., Collin-Dufresne, P., and Goldstein, R. S. (2011). Explaining asset pricing puzzles associated with the 1987 market crash. *Journal of Financial Economics*, 101(3):552 – 573.

Breeden, D. T. (1979). An intertemporal asset pricing model with stochastic consumption and investment opportunities. *Journal of Financial Economics*, 7:265–296.

Broadie, M., Chernov, M., and Johannes, M. (2007). Model specification and risk premia: Evidence from futures options. *Journal of Finance*, 62(3):1453–1490.

Campbell, J. Y. and Cochrane, J. H. (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy*, 107:205–251.

Campbell, J. Y. and Hentschel, L. (1992). No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of Financial Economics*, 31(3):281–318.

Cogley, T. and Sargent, T. J. (2008). Anticipated utility and rational expectations as approximations of bayesian decision making. *International Economic Review*, 49(1):185–221.

Collin-Dufresne, P., Johannes, M., and Lochstoer, L. A. (2016). Parameter learning in general equilibrium: The asset pricing implications. *American Economic Review*, 106(3):664–698.

David, A. and Veronesi, P. (2013). What ties return volatilities to price valuations and fundamentals? *Journal of Political Economy*, 121(4):682–746.

Dergunov, I., Meinerding, C., and Schlag, C. (2018). Extreme inflation and time-varying disaster risk. Working paper, Goethe University Frankfurt and Research Center SAFE.

Duffie, D. and Epstein, L. G. (1992). Stochastic differential utility. *Econometrica*, 60(2):353–394.

Duffie, D. and Skiadas, C. (1994). Continuous-time asset pricing: A utility gradient approach. *Journal of Mathematical Economics*, 23:107–132.

Epstein, L. and Zin, S. (1989). Substitution, risk aversion and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica*, 57:937–969.

Hennessy, C. A. and Radnaev, B. (2016). Learning and leverage cycles in general equilibrium: Theory and evidence. *Review of Finance*, page rfw058.

Hong, H. and Stein, J. C. (2003). Differences of opinion, short-sales constraints, and market crashes. *The Review of Financial Studies*, 16(2):487–525.

Hu, G. X., Pan, J., Wang, J., and Zhu, H. (2019). Premium for heightened uncertainty: Solving the FOMC puzzle. Working Paper 25817, National Bureau of Economic Research.

Koulovatianos, C. and Wieland, V. (2011). Asset Pricing under Rational Learning about Rare Disasters. IMFS Working Paper Series No. 46.

Kozlowski, J., Veldkamp, L., and Venkateswaran, V. (2018). The tail that keeps the riskless rate low. National Bureau of Economic Research Working Paper 24362.

Lettau, M., Ludvigson, S. C., and Wachter, J. A. (2008). The declining equity premium: What role does macroeconomic risk play? *Review of Financial Studies*, 21:1653–1687.

Longstaff, F. A. and Piazzesi, M. (2004). Corporate earnings and the equity premium. *Journal of Financial Economics*, 74:401–421.

Lorenz, F. and Schumacher, M. D. (2018). Downside risks and the price of variance uncertainty. Working paper, University of Zurich.

Mehra, R. and Prescott, E. (1988). The equity risk premium: A solution? *Journal of Monetary Economics*, 22:133–136.

Moreira, A. and Muir, T. (2017). Volatility-managed portfolios. *The Journal of Finance*, 72(4):1611–1644.

Moreira, A. and Savov, A. (2017). The Macroeconomics of Shadow Banking. *Journal of Finance*, 72(6):2381–2432.

Muir, T. (2017). Financial crises and risk premia. *The Quarterly Journal of Economics*, 132(2):765–809.

Reinhart, C. M. and Rogoff, K. S. (2009). *This Time is Different: Eight Centuries of Financial Folly*. Princeton University Press, Princeton, NJ.

Rietz, T. A. (1988). The equity risk premium: A solution. *Journal of Monetary Economics*, 22:117–131.

Schmalz, M. C. and Zhuk, S. (2018). Revealing downturns. forthcoming, *Review of Financial Studies*.

Veronesi, P. (1999). Stock market overreaction to bad news in good times: A rational expectations equilibrium model. *The Review of Financial Studies*, 12(5):33.

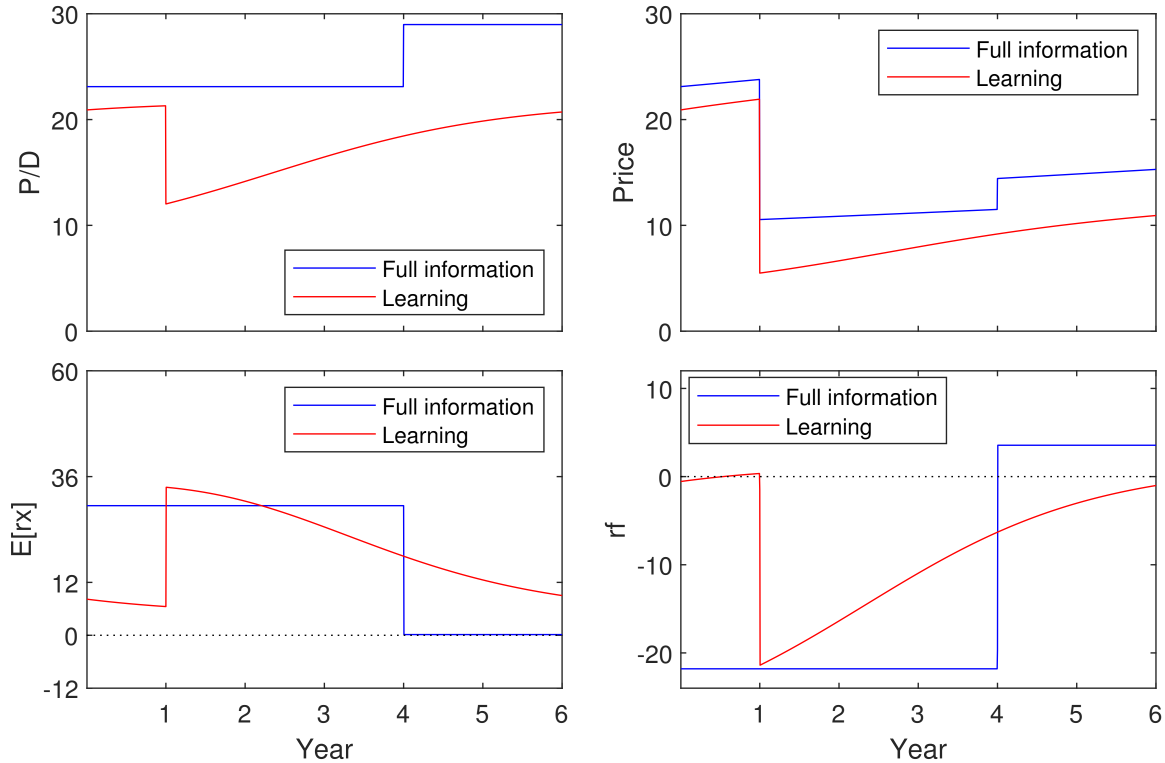
Wachter, J. A. (2005). Solving models with external habit. *Finance Research Letters*, 2:210–226.

Wachter, J. A. (2013). Can time-varying risk of rare disasters explain aggregate stock market volatility? *Journal of Finance*, 68(3):987–1035.

Wachter, J. A. and Zhu, Y. (2017). The macroeconomic announcement premium. Rodney L. White Center Working Paper.

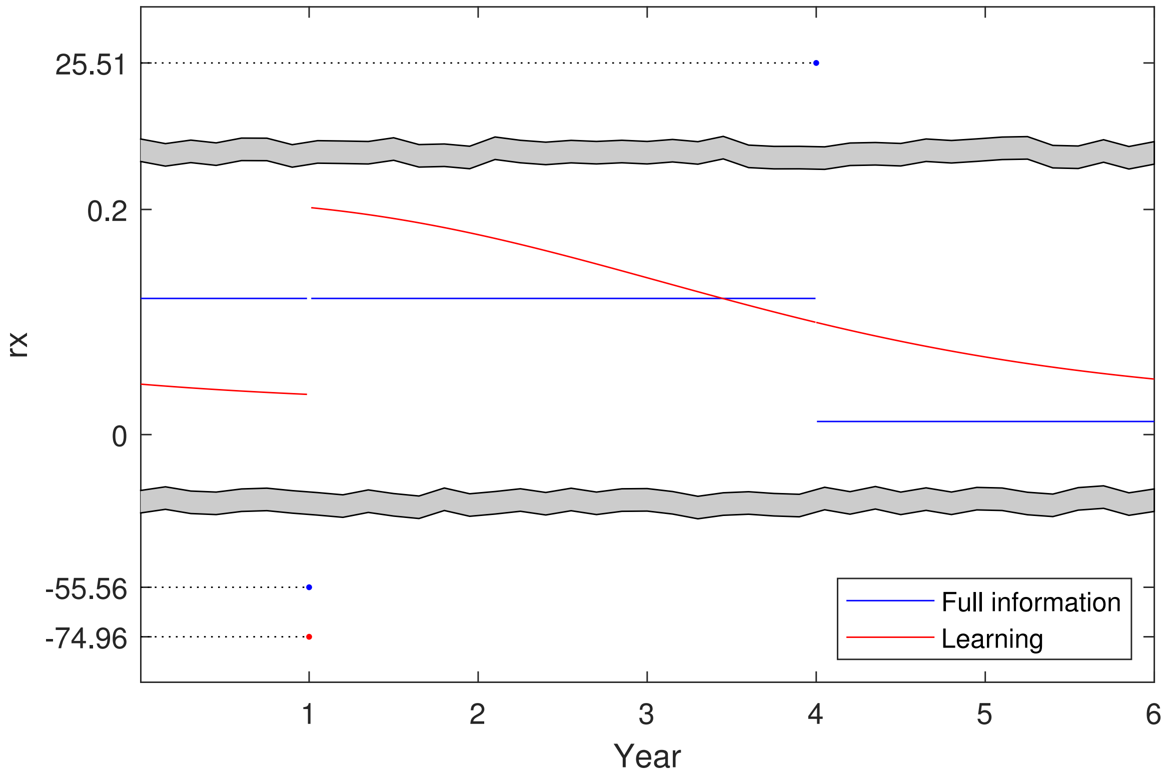
Wonham, W. M. (1965). Some applications of stochastic differential equations to optimal nonlinear filtering. *Journal of the Society for Industrial and Applied Mathematics Series A Control*, 2:347–369.

Figure 1: Evolution of financial moments given realization of one disaster and one regime switch



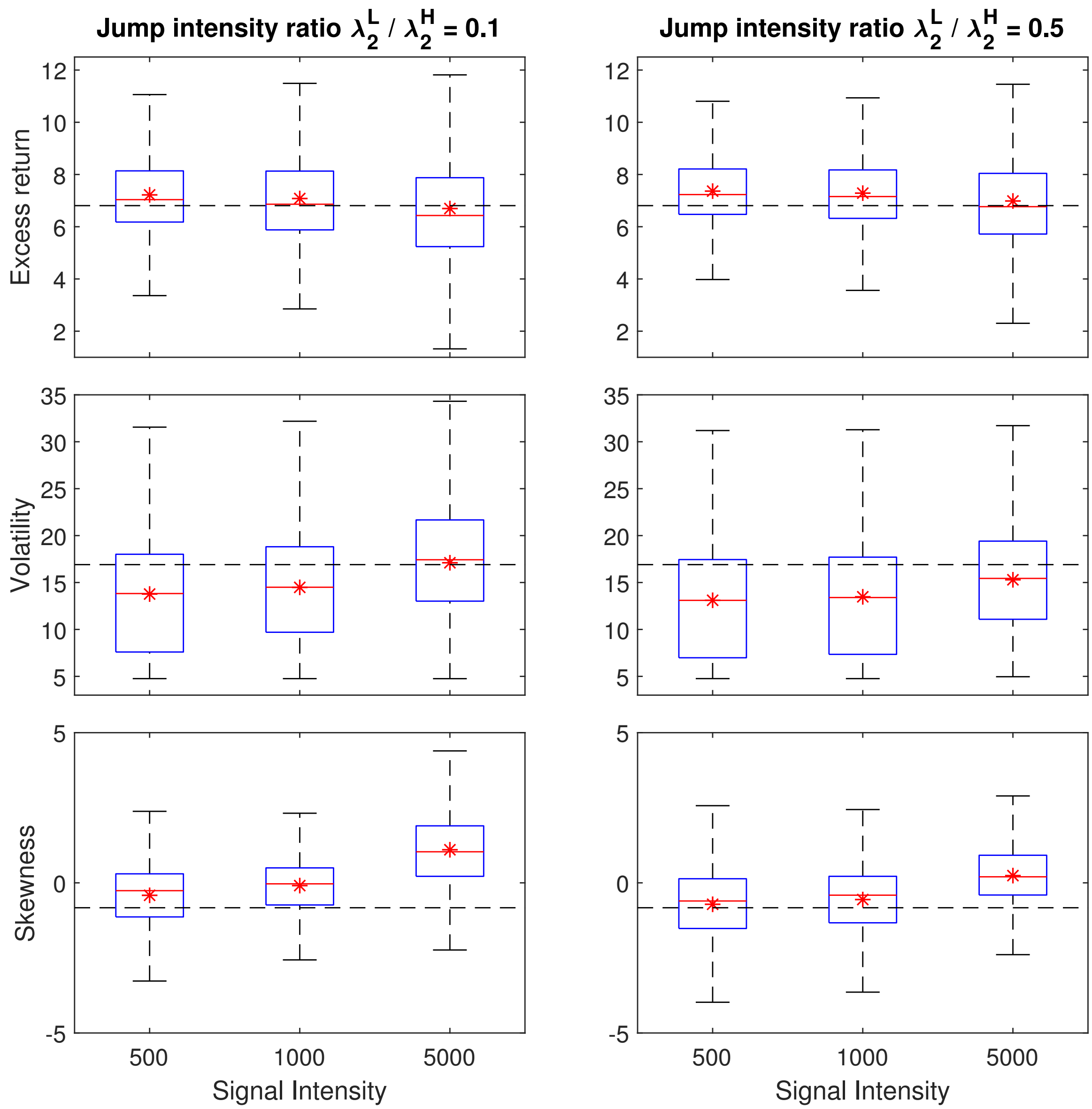
Note: Simulated moments of the equity asset in the case of a disaster realization and a regime switch. The sample length is 1440 trading days (6 years, with each year having 240 trading days). The realization of disaster is on the 240^{th} day of the simulation sample, while the regime switch happens on the 960^{th} day of the simulation sample. The physical regime starts with the high-risk state, and switches to the one with low risk. The belief of the agent in the learning case starts with the unconditional probability of the high-risk state ($\phi_{L\to H}/(\phi_{H\to L} + \phi_{L\to H})$). There are no diffusion shocks ($B_{Ct} = 0$) throughout the simulation. rx denotes the excess returns, and rf denotes the riskfree returns. The dividend at the beginning of the simulation sample is 1. The unit is percentage per annum.

Figure 2: Realized daily excess return given realization of one disaster and one regime switch



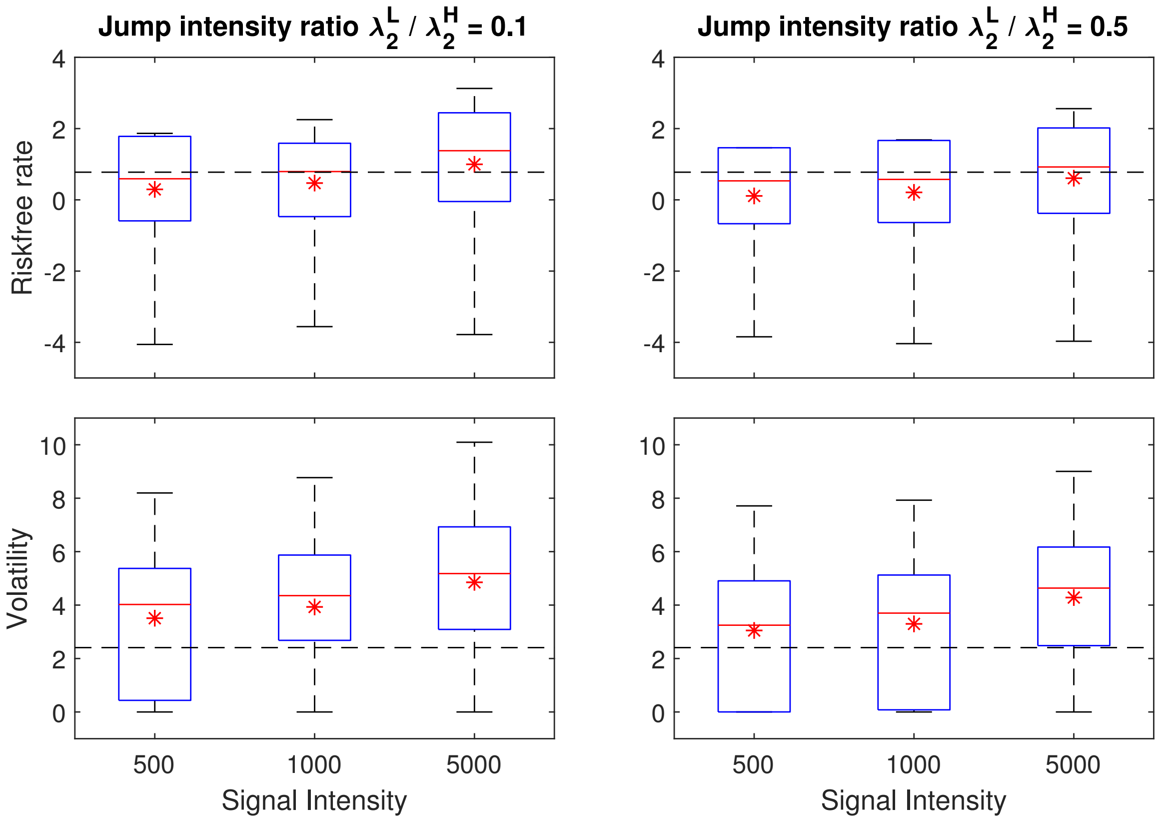
Note: Simulated realized excess returns of the equity asset in the case of a disaster realization and a regime switch. The sample length is 1440 trading days (6 years, with each year having 240 trading days). The realization of disaster is on the 240^{th} day of the simulation sample, while the regime switch happens on the 960^{th} day of the simulation sample. The physical regime starts with the high-risk state, and switches to the one with low risk. The belief of the agent in the learning case starts with the unconditional probability of the high-risk state $(\phi_{L\to H}/(\phi_{H\to L}+\phi_{L\to H}))$. There are no diffusion shocks $(B_{Ct}=0)$ throughout the simulation. The unit is percentage per day.

Figure 3: Simulation sample moments of market portfolio excess return with different signal characteristics



Note: Distribution of simulated moments of market excess returns. There are 2,000 parallel simulation samples, and each sample has length of 58 years. The figures on the left are with unconditional signal intensity ratio $\lambda_2^L/\lambda_2^H = 0.1$, while the figures on the right are with unconditional signal intensity ratio $\lambda_2^L/\lambda_2^H = 0.5$. The red lines stand for sample median, the black dashed lines stand for empirical estimates, and the red stars stand for sample mean. The unit is percentage per annum.

Figure 4: Simulation sample moments of riskfree rates with different signal characteristics



Note: Distribution of simulated moments of riskfree rates. There are 2,000 parallel simulation samples, and each sample has length of 58 years. The figures on the left are with unconditional signal intensity ratio $\lambda_2^L/\lambda_2^H = 0.1$, while the figures on the right are with unconditional signal intensity ratio $\lambda_2^L/\lambda_2^H = 0.5$. The red lines stand for sample median, the black dashed lines stand for empirical estimates, and the red stars stand for sample mean. The unit is percentage per annum.

Table 1: Calibration and simulation parameters

Panel A: Basic parameters	
Average log growth in consumption $\mu_C(\%)$	2.50
Average log growth in dividend $\mu_D(\%)$	2.90
Volatility of consumption growth $\sigma_C(\%)$	2.00
Leverage of equity asset φ	3.00
Rate of time preference β	0.012
Relative risk aversion γ	2.6
Panel B: The process for λ_{1t}	
Probability of switching to the high state $\phi_{L\to H}(\%)$	3.33
Probability of switching to the low state $\phi_{H\to L}(\%)$	33.33
Probability of disaster in the low state $\lambda_1^L(\%)$	0.07
Probability of disaster in the high state $\lambda_1^H(\%)$	30.75

Note: Parameter values for the main calibration, expressed in annual terms.

Table 2: Summary statistic of the market portfolio & riskfree rates with infrequent signals

Panel I: Full Information			
Statistics	Estimate	Mean	90% CI
$\bar{RX}_t$	6.81	6.29	$[3.01, 10.73]$
σ_{RX}	16.91	19.58	$[6.22, 31.87]$
Skewness	-0.83	2.05	$[-0.96, 4.80]$
$\bar{R}_{ft}$	0.78	1.49	$[-2.45, 3.62]$
σ_{R_f}	2.41	5.18	$[0.00, 9.88]$

Panel II: Learning Case			
Statistics	Estimate	Mean	90% CI
$\bar{RX}_t$	6.81	7.40	$[5.46, 9.66]$
σ_{RX}	16.91	12.92	$[5.80, 22.74]$
Skewness	-0.83	-0.80	$[-2.66, 0.70]$
$\bar{R}_{ft}$	0.78	0.06	$[-2.33, 1.34]$
σ_{R_f}	2.41	2.92	$[0.00, 6.14]$

Note: This table compares the simulation results with the cases of learning and full information. Panel I reports the simulated moments when the economy has perfect information, while Panel II reports the results when λ_{1t} is latent and agent learns the value through disasters. We simulate 2,000 parallel samples, and each sample has length 58 years. The empirical moments are computed using annual weighted average returns of NYSE/AMEX/NASDAQ and gross returns of one-month treasury in each calendar year. The sample period is 1961-2018.