Working Papers

FEDERAL RESERVE BANK OF BOSTON

No. 18-3

News-Driven Uncertainty Fluctuations

Dongho Song and Jenny Tang

Abstract:
We embed a news shock, a noisy indicator of the future state, in a two-state Markov-switching growth model. Our framework, combined with parameter learning, features rich history-dependent uncertainty dynamics. We show that bad news that arrives during a prolonged economic boom can trigger a "Minsky moment"—a sudden collapse in asset values. The effect is greatly amplified when agents have a preference for early resolution of uncertainty. We leverage survey recession probability forecasts to solve a sequential learning problem and estimate the full posterior distribution of model primitives. We identify historical periods in which uncertainty and risk premia were elevated because of news shocks.

JEL Classifications: C11, G12, E32, E37
Keywords: Bayesian learning, discrete environment, Minsky moment, news shocks, recursive utility, risk premium, survey forecasts, uncertainty

---

Dongho Song is an assistant professor of economics at Boston College. Jenny Tang is an economist in the research department of the Federal Reserve Bank of Boston. Their email addresses are dongho.song@bc.edu and jenny.tang@bos.frb.org, respectively.

The authors thank Andre Kurmann, Minchul Shin, Jonathan Wright, and the seminar and conference participants at Boston College, the 2017 Society for Economic Measurement Conference, the 2017 NBER Time-Series Conference at Northwestern University, the 2017 NBER DSGE Meeting at the Federal Reserve Bank of Philadelphia, the Zurich Workshop on Asset Pricing, and the 26th Symposium of the Society of Nonlinear Dynamics and Econometrics for helpful comments.

This version: January 2018

1 Introduction

Important news may come in the form of regular releases of macroeconomic data, policy announcements, or less frequent events (e.g., political elections, terrorist attacks, or other crises). The literature has developed with an overwhelming focus on identifying economically important news.[1] However, there have been relatively few attempts to understand the channels through which news influences subjective beliefs of economic agents, especially higher moments of beliefs.[2] The literature leaves several important questions unanswered. For example, does more information about the future always reduce subjective uncertainty? Should news that contradicts existing beliefs about the future state raise uncertainty? Should we expect a differential effect on subjective uncertainty when receiving good news in a good state compared with receiving good news in a bad state? Would bad news have a different effect depending on when it arrives? To what extent would the behavior of quantities such as subjective uncertainty change when news is believed to be of greater informative value?

In this paper, we incorporate news into an otherwise standard model in an attempt to answer these questions. In a parsimonious two-state Markov-switching growth model, we introduce an agent who receives news in every period that reveals the next period's state with some error. We consider cases in which the agent can rationally learn model parameters, including state transition probabilities or the accuracy of this news, using Bayes' rule as new data arrive. This model produces two novel properties of the influence of news on subjective uncertainty. First, we show that when news contradicts prior beliefs, uncertainty can increase as the agent's posterior beliefs are corrected to a more uniform distribution. For empirically plausible parameter values, the rise in uncertainty is more apparent when news is believed to be more accurate. This is in contrast to a Gaussian environment in which news always reduces uncertainty.[3] Thus,

[1] There is a growing literature that attempts to identify economically important news and its impact on financial markets (e.g., Boyd, Hu, and Jagannathan 2005; Gürkaynak, Sack, and Swanson 2005; Andersen et al. 2007; Faust et al. 2007; Savor and Wilson 2013; Lucca and Moench 2015; and Tang 2017).

[2] The existing literature studying the effect of news shocks (see Beaudry and Portier 2006, Jaimovich and Rebelo 2009, Barsky and Sims 2011, Schmitt-Grohe and Uribe 2012, Kurmann and Otrok 2013, Beaudry and Portier 2014, and Malkhozov and Tamoni 2015) focuses mainly on the transmission of these shocks to asset prices or the economy through their effects on agents' mean beliefs about future outcomes.

[3] Veronesi (1999) shows a similar relationship between dividend growth realizations and uncertainty in a model without news where agents must learn about a hidden dividend growth state. See the literature review below for a more extensive discussion.

even without stochastic volatility, the discrete-state environment allows news to drive large fluctuations in subjective uncertainty.

In addition to the above finding, we show that as the agent learns either the accuracy of news or the persistence of states, the effect of news on uncertainty is altered, thus giving rise to history-dependent behavior of uncertainty fluctuations. As a result, parameter learning can greatly amplify the response of uncertainty to news. To illustrate the intuition behind this result, suppose that the economy stays in a good state and that agents receive good news, which is correct ex-post, for an extended period of time. In this setting, the agent successively increases her estimate of the accuracy of news, and subjective uncertainty gradually falls as more good news arrives. When bad news arrives against this backdrop, the agent, now believing the news to be very accurate, sharply adjusts her beliefs toward a more uniform posterior distribution, which results in an upward jump in subjective uncertainty. In this example, the jump in uncertainty upon receiving bad news will be larger for longer runs of consecutive "good state, good news" realizations prior to the arrival of bad news. In general, parameter learning introduces additional state variables to the model that summarize information from past states and alter the effect of news on the agent's beliefs.

In light of these findings, we examine the extent to which news-driven uncertainty fluctuations can generate variation in asset prices. We consider an agent who has preference for early resolution of uncertainty in an endowment economy with a two-state Markov-switching growth process. We first isolate the role of news without parameter learning. It is interesting to highlight the implications for asset prices when an agent believes the probabilities of remaining in the same state and the news being correct to be identical and close to 1. In this setting, switching states, which occur when news contradicts existing beliefs about the future state, are ones in which the agent thinks there is equal chance of entering a nearly permanent good or bad state in the next period. Note that this cannot occur in a standard discrete-state Markov-switching model without news shocks, because probabilities of future state realizations are tightly linked to state persistence in that setting. News shocks break this link and thus create the potential for these episodes of extreme uncertainty and, as a result, highly elevated risk premia. Even for very moderate values of risk aversion and the intertemporal elasticity of substitution (both set to 1.5), the implied risk premium on a consumption claim can reach as much as 5 percent in these switching states, roughly 160 times greater than the counterfactual risk premium under log utility. In this highly uncertain state,

demand for the risky consumption claim is very low, and the agent requires a very high expected excess return on this asset in order to equilibrate shorting demand with the asset's zero net supply.

We then consider the theoretical implications of parameter learning in our model with news shocks. We first illustrate qualitative results using the anticipated utility approach to price assets.[4] We highlight the case of high prior means for the state persistence and news accuracy parameters in an experiment, similar to the one above, for forecast uncertainty in which bad news arrives after a prolonged economic boom. The excessive optimism of the agent with a preference for early resolution of uncertainty, i.e., the increase in her posterior belief of remaining in the good state, leads to a "Minsky moment" once she receives news that the future may not be as bright as expected.[5] The size of the risk premium is twice as large as the one that would have been obtained without parameter learning.

Next, we estimate the model. To obtain estimates of news parameters and realizations, we leverage a forward-looking variable that summarizes agents' information about the future state. Since agents in our model generate discrete belief distributions over future states based on both current fundamentals and potentially noisy news about the future state, we map their believed bad state probabilities to recession probability forecasts from the Survey of Professional Forecasters. We make an important technical innovation that allows us to use this variable to recover parameter estimates and filtered states, including news realizations. Specifically, we develop a novel filtering technique that uses both actual GDP growth and recession probability forecasts in the estimation. This technique can be interpreted as a sequential learning problem of an econometrician who observes only these two variables. We solve this problem and sample from the joint posterior of model parameters and states by augmenting the filtering technique with the particle learning algorithm developed by Carvalho et al. (2010). Using data from 1969:Q1 through 2016:Q3, we obtain the full posterior distribution of model primitives, including the filtered distribution of news, which characterizes the econometrician's full

[4] Johannes, Lochstoer, and Mou (2016) study parameter learning using an anticipated utility approach in a model with long-run risks but without news shocks and show that revisions in beliefs about consumption dynamics are quite volatile and that pricing assets using these beliefs improves the model's ability to fit moments as well as the historical path of equity prices. Collin-Dufresne, Johannes, and Lochstoer (2016) show that allowing fully rational pricing of parameter uncertainty further helps the model match asset pricing facts. See also Bansal and Yaron (2004).

[5] A "Minsky moment," which stems from the work of Hyman Minsky, is a sudden collapse of markets and economies after a long period of growth, sparked by debt or currency pressures.

learning problem.

This estimation yields three main empirical findings. First, we show that posterior beliefs vary significantly over time and, in particular, there is a pattern of greater forecast uncertainty driven by bad news during expansions relative to an estimation that does not allow for news shocks. In light of this property of our model, we expect news-driven uncertainty fluctuations to be important for generating sizable fluctuations in endogenous variables within growth regimes. Second, the estimated recession probability tightly identifies NBER recession dates. When only GDP growth is used in the estimation (without recession probability forecasts), the model's filtered recession probabilities tend to be biased upward, with bad states being less tightly identified. Third, we show that our model-implied probability of receiving bad news correlates strongly (around 0.6) with an index from the University of Michigan's Survey of Consumers that measures the prevalence of bad news. This serves as an external validation check. We also find that our news measure strongly negatively predicts one-step-ahead GDP growth after controlling for other factors as implied by the model.

Lastly, to assess our model's quantitative implications for asset prices, we incorporate our empirical findings into an asset pricing model in which the agent must learn the transition probabilities of growth states and this parameter uncertainty is rationally priced. To do so, we extend the policy function iteration method used by Collin-Dufresne, Johannes, and Lochstoer (2016) to accommodate a news shock and obtain fully nonlinear solutions for key variables. Adding news to this model changes the risks from parameter uncertainty by speeding up learning about transition probabilities. It also changes the risks coming from regime uncertainty through the mechanisms described above.

Using our estimated model parameters and filtered regime probabilities, we find that news shocks produce modest increases in both the average risk-free rate and equity risk premium while making a substantially larger impact on their volatility. This increase in volatility is particularly pronounced during economic expansions. Including news shocks in the model increases the standard deviation of equity premia within high growth states by a factor of more than 6.7. News shocks greatly improve the model's ability to match the conditional and unconditional volatilities of the risk-free rate and equity premium in the data.

This increased volatility is evident in the time series of the equity premia produced by our estimates. Furthermore, our estimates indicate that the arrival of bad news prior

to recessions often leads the equity premium to begin rising several quarters prior to the onset of recessions. In our model, bad news can also generate large spikes in the equity premium even when contemporaneous GDP growth remains high, such as during the Black Monday crash of 1987. This effect is amplified by the increase in subjective uncertainty when bad news contradicts the contemporaneous growth state. The equity premium generated by an identically calibrated model without news shocks does not have these features.

Literature Review. This paper is at the intersection of several literatures. First, it is related to papers that study news shocks theoretically and empirically in the context of asset pricing (such as Beaudry and Portier 2006, Kurmann and Otrok 2013, and Malkhozov and Tamoni 2015) or business cycles (such as Jaimovich and Rebelo 2009, Barsky and Sims 2011, Schmitt-Grohe and Uribe 2012, and the works surveyed in Beaudry and Portier 2014). Unlike most of this existing work, which focuses on the effects of news on mean forecasts of future economy activity, we focus on the additional effects that news shocks have on uncertainty. We also consider how news shocks affect parameter learning, and we look at their joint effects on uncertainty.

Our paper is also closely related to studies that examine the sources of uncertainty fluctuations. In particular, Kozeniauskas, Orlik, and Veldkamp (2016) also study uncertainty in a model that features parameter learning, but without news shocks. Stochastic volatility and disaster risks are crucial for generating uncertainty fluctuations in their environment. In contrast, we show that news shocks can generate uncertainty fluctuations even when the true data-generating process does not feature stochastic volatility. The recent work by Berger, Dew-Becker, and Giglio (2017) uses an estimated vector autoregression (VAR) to separate exogenous shocks to expected volatility from movements in realized volatility. In contrast, our setting does not allow for an exogenous shock to uncertainty. Instead, uncertainty fluctuates with realizations of actual economic data and news shocks about future states.

Our finding of a state-dependent relationship between news and uncertainty is related to that of Veronesi (1999), who also models asset prices in a setting with non-Gaussian random variables. That paper finds a similar relationship between observations of a dividend process and uncertainty in a model without news shocks where agents must infer a hidden Markov state that determines dividend growth. In contrast, we find this state-dependent relationship in a setting where the agent knows the current state but receives noisy news about the future state. We further explore the potential for rich

history-dependent behavior of uncertainty when the agent must also learn particular parameters of the model. Additionally, we estimate the model and empirically identify historical episodes in which uncertainty about future GDP growth was elevated because of news.

On the methodological side, Bianchi (2016) and Bianchi and Melosi (2016) provide analytical characterizations of uncertainty in Markov-switching models with settings where there is perfect information or where the agent must infer states and unknown parameter values through Bayesian learning, respectively. Our paper differs from these by focusing on the effect of news about the future (rather than signals about the current state) on uncertainty. We additionally provide an empirical estimation of the model as well as an analysis of the asset pricing implications of allowing for news.

Our empirical exercise is closely related to Milani and Rajrhandari (2012), Hirose and Kurozumi (2012), and Miyamoto and Nguyen (2015). These papers all incorporate forecast data into the estimation of news shocks. Nonetheless, two key differences remain. First, we develop new econometric methods that allow us to include recession probability forecasts in the estimation, while these papers focus only on mean forecasts. Furthermore, these studies estimate dynamic stochastic general equilibrium (DSGE) models with news shocks about future realizations of particular structural shocks (e.g., total factor productivity [TFP], demand, monetary policy, etc.). In contrast, we remain agnostic about the types of structural shocks that news may pertain to and instead estimate a latent news variable summarizing all information that is embedded in the recession probability forecasts and is not already captured by the current GDP growth state.

In addition to the above-mentioned asset pricing papers that consider parameter learning, our asset pricing application is related to other papers studying the effect of information quality on asset price moments. Veronesi (2000) focuses mainly on the interplay between information quality and risk aversion in determining moments of returns in a model similar to that of Veronesi (1999) but with agents also receiving another noisy signal about current dividend growth rates. Ai (2010) shows that, in a model with endogenous production and Kreps-Porteus preferences, including a dividend growth state inference problem similar to that of Veronesi (2000) helps to match moments of the wealth-consumption ratio and the return on wealth. Epstein and Schneider (2008) study a setting where ambiguity-averse investors behave according to a worst-case assessment of information quality that entails judging bad news to

be of high quality and good news to be of low quality. These preferences thus generate state-dependent effects of news, similarly to our model, but here the effects depend only on the news itself and not on the current growth regime. Matsumoto et al. (2011) and West (1988) both study the effect of news shocks on asset price volatility. West (1988) finds, in a partial equilibrium environment where asset prices are a discounted linear sum of future dividends and expectations are formed through linear projections, that asset price volatility must decrease when news is more accurate. In contrast, we find that a greater accuracy of news can increase the volatility of risk premia. The main reason for this opposing result is that expectations are not formed through linear projections in our setup, because we have Markov-switching growth regimes and discrete news realizations. Matsumoto et al. (2011) focus on news defined as information that reduces the ex-ante conditional variance of future fundamentals, which isn't always the case for our news shock, but find results opposite to those of West (1988) by allowing for general equilibrium effects of news on cash flows.

Lastly, this paper is also related to previous literature studying models in which agents can form beliefs from sources of data other than simply realized GDP growth rates. The works of Johannes, Lochstoer, and Mou (2016) and Constantinides and Ghosh (2016) allow agents to form beliefs by incorporating additional information from multiple sources of macroeconomic data—namely consumption, output, and/or labor market variables. However, in contrast to our paper, these studies model these additional variables as being informative only about the current state of the economy and not about future states. Furthermore, these papers do not use forecast data in their estimations, while we do. The use of forecast data allows us to be agnostic about the sources of additional information and instead capture a summary of information relevant for forecasting future output.

In Section 2, we describe our model of real output growth and news shocks. Section 3 presents the estimation of these exogenous driving processes. Section 4 presents stock market moments implied by these estimated values, and Section 5 concludes.

2 Modeling News

2.1 The environment

We begin with the following two-state Markov-switching model for real GDP growth:

$$\begin{aligned} y_t &= \mu_{S_t} + \sigma_{S_t}\epsilon_t, \quad \epsilon_t \sim N(0,1), \\ Pr(S_t = j | S_{t-1} = i) &= q_{ij} \quad \text{with} \quad \sum_j q_{ij} = 1. \end{aligned} \tag{1}$$

Here y_t represents the real GDP growth rates, and S_t is a discrete Markov state variable that takes on two values $S_t \in \{1, 2\}$. We assume $\mu_1 > \mu_2$ without loss of generality.

We introduce news that provides information about the next period's state,

$$n_t = S_{t+1}, \quad \text{w.p.} \quad \chi, \quad 0.5 \leq \chi \leq 1, \tag{2}$$

which is available in discrete form.[6] We assume that n_t can predict the future state with probability χ. The accuracy of the prediction increases with χ. The combination of two discrete Markov chains, one for the fundamental variable in (1) and the other for the news component in (2), results in the four-state Markov chain process. The model parameters are collected in

$$\theta = \left\{\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \chi, \sigma_z^2\right\}, \quad \Pi = \{q_{11}, q_{22}\}.$$

The parameter σ_z^2 is the variance of an error process that will be introduced in Section 3 when we estimate the model.

In terms of parameter learning, we will focus mainly on the case of learning about the transition probabilities, q_{11} and q_{22}, based on observing the true states and news realizations.[7] The Bayesian updating problem is detailed in Appendix A.

[6] Note that the label switching problem arises for χ values less than 0.5. Therefore, we restrict to $0.5 \leq \chi \leq 1$.

[7] Our notation will reflect that real GDP growth is also in the information set of agents, but growth state and news realizations are sufficient for Bayesian learning about q_{11} and q_{22}.

2.2 News implications

State prediction and parameter learning accuracy. We demonstrate how the accuracy of news affects the prediction for the future state and posterior beliefs about the transition probabilities. For ease of illustration, we work with conditional distributions and assume that a subset of the model parameters and the history of states may be known at time t when relevant.

We consider two boundary values of $\chi \in \{0.5, 1\}$. We start from the $\chi = 0.5$ case, in which news contains no information about the future state. Because it assigns equal probability to both states, the predictive distribution of the future state conditional on news is identical to the one without conditioning on news

$$p(S_{t+1} = 1|n_t, S^t, \Pi, \theta) = p(S_{t+1} = 1|S^t, \Pi, \theta).$$

Analogously, when agents must learn transition probabilities, we can deduce that news has no impact on the posterior mean of these parameters when $\chi = 0.5$

$$E(q_{ii}|n_t, y^t, S^t, \theta) = E(q_{ii}|y^t, S^t, \theta).$$

On the other hand, if news contains certain information about the future state, that is, $\chi = 1$, then knowledge about the current state no longer plays a role in predicting the future state. The state prediction becomes a degenerate distribution function

$$p(S_{t+1} = 1|n_t, S^t, \Pi, \theta) = p(S_{t+1} = 1|n_t, \theta) = \begin{cases} 1, & \text{if} \quad n_t = 1 \\ 0, & \text{otherwise.} \end{cases}$$

Because the future state is known with certainty, the posterior mean of transition probabilities is identical to the one that would have been obtained in time $t + 1$ in the absence of news

$$E(q_{ii}|n_t, y^t, S^t, \theta) = E(q_{ii}|y^{t+1}, S^{t+1}, \theta).$$

The above examples show that news can lead to improvement in prediction and parameter learning accuracy. Detailed derivations are provided in Appendix A and B.

Forecast uncertainty. The above examples illustrate that news improves ex-post

forecast accuracy, but we now consider the impact of news on ex-ante forecast uncertainty. In this model, one measure of forecast uncertainty is the conditional variance of one-period-ahead output growth defined as

$$\begin{aligned} Var(y_{t+1}|y^t, S^t, n^t, \Pi, \theta) &= \int (y_{t+1} - y_{t+1|t})^2 p(y_{t+1}|y^t, S^t, n^t, \Pi, \theta) dy_{t+1} \qquad (3) \\ &= \underbrace{\sigma_1^2 + (q_{3i}^B + q_{4i}^B)(\sigma_2^2 - \sigma_1^2)}_{\text{due to second moment}} + \underbrace{(q_{1i}^B + q_{2i}^B)(q_{3i}^B + q_{4i}^B)(\mu_1 - \mu_2)^2}_{\text{due to first moment}}, \end{aligned}$$

where $y_{t+1|t} = \int y_{t+1} p(y_{t+1}|y^t, S^t, n^t, \Pi, \theta) dy_{t+1}$, and $\sum_{j=1}^4 q_{ji}^B = 1$. Transition probabilities of an expanded state space that encompasses the realizations of both S_t and n_t are denoted by q_{ji}^B. We refer the reader to equation (A-15) in the Appendix for the relationship between q_{ji}, χ, and q_{ji}^B. Forecast uncertainty (3) comprises two components. The first component captures uncertainty with respect to innovation variances, and the second component arises from uncertainty about mean values. This is due to our discrete-state environment.

We now explain the effect of news on the transition probabilities q_{ji}^B by which uncertainty is determined. For ease of illustration, we assume that persistence of each growth regime is identical $q = q_{11} = q_{22}$ and use $\mu_1 = 0.84, \mu_2 = -0.22, \sigma_1^2 = 0.47, \sigma_2^2 = 0.56$.[8] Figure 1 provides the value of forecast uncertainty as a function of $\chi \in [0.5, 1)$ for different values of $q \in [0.5, 1)$. We are considering only moderately persistent q values and excluding the end points ($q = 1, \chi = 1$) that would remove all state uncertainty. Note that because we consider only $q \geq 0.5$, the most likely outcome for next period's state is always the current state when news is not part of the information set.

Panel (A) of Figure 1 isolates the usual stochastic volatility component of forecast uncertainty. As the accuracy of news increases, forecast uncertainty converges to the variance of the state indicated by the news realization. It is interesting to consider panel (B), which isolates the part of uncertainty arising from different mean values (the final term in (3)). Note that this panel also illustrates the large fluctuations in uncertainty that can arise due to the presence of noisy news in a setting without stochastic volatility. If news suggests that the current regime will persist next period, then forecast uncertainty monotonically decreases as χ increases, as shown in the first and fourth columns. However, if news contradicts the implication of the current state alone and suggests switching into a different state, then forecast uncertainty increases in χ for $\forall \chi \leq q$ and

[8] These are the end-of-sample posterior median estimates presented in Section 3.

decreases in χ for $\forall\chi > q$. The intuition behind this result is that, with $q \in [0.5, 1)$, the current state is more likely than not to persist. News that suggests a switch in the state will shift the posterior distribution of next period's state to a more uniform one in which weight is placed more equally on the two possible realizations. Note that in order for this to occur, the news must be moderately informative. Completely uninformative news will not alter beliefs, while perfectly informative news will shift beliefs to the point where a switch will occur with certainty.

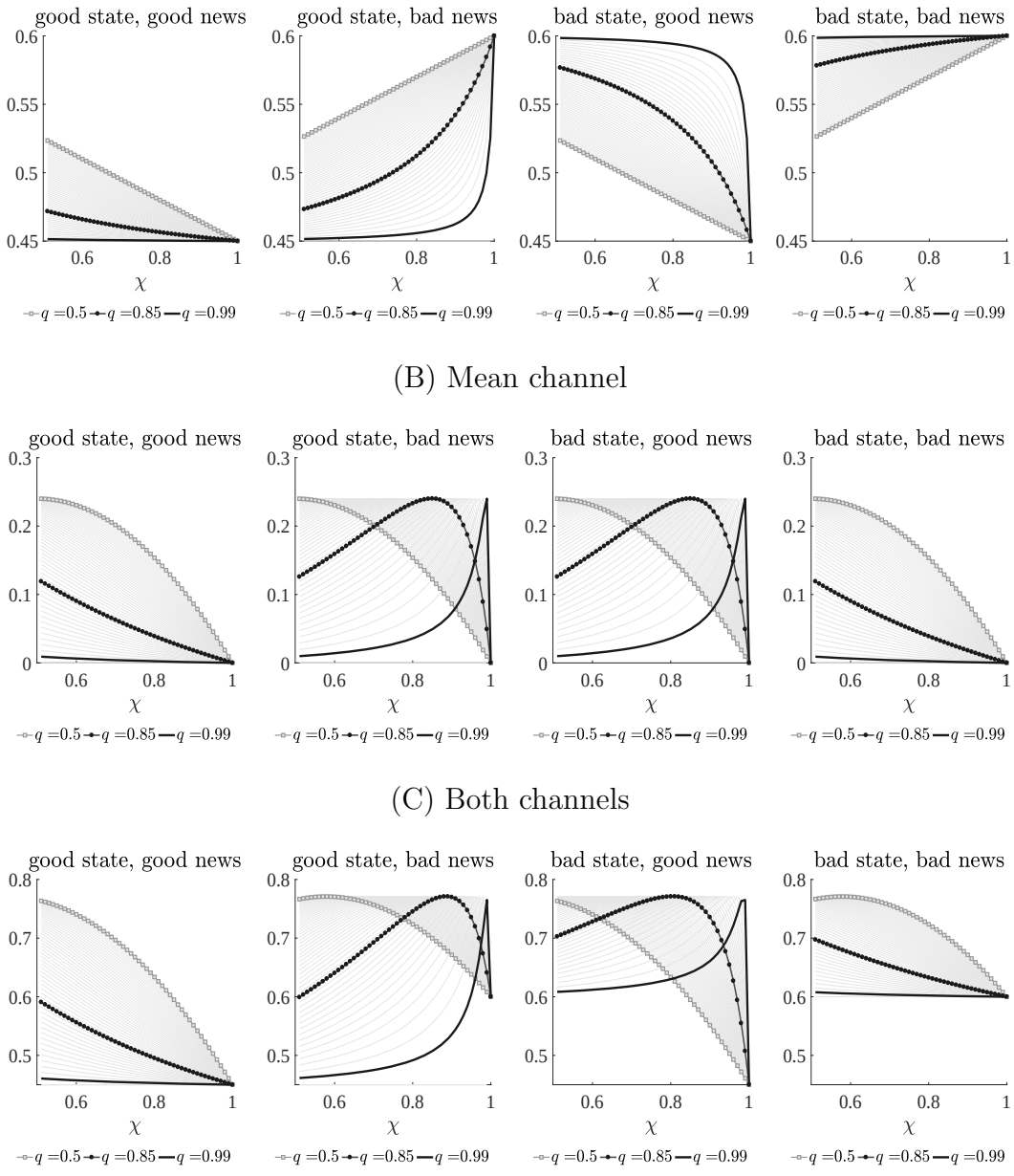
Panel (C) shows the forecast uncertainty when both channels are at play. It is important to note that the patterns are qualitatively similar to those shown in panel (B). The key takeaway is that, in contrast to standard Gaussian setups where more information always reduces uncertainty, our discrete-state environment allows moderately informative news to increase forecast uncertainty when it contradicts existing beliefs.

Having shown the role of news in forecast uncertainty, we discuss how parameter learning—in particular, learning about the accuracy of news and state persistence—can give rise to additional fluctuations in uncertainty and history-dependent behavior.[9] Consider the following experiment in which we are in a good growth state and receive good news for $t-1$ consecutive periods, but we receive bad news while remaining in the good growth state in the t^{th} period. We calculate the corresponding forecast uncertainty (3) when the agent learns about either state persistence or news accuracy and continue to use the assumption of $q = q_{11} = q_{22}$, along with $\mu_1 = 0.84, \mu_2 = -0.22, \sigma_1^2 = 0.47, \sigma_2^2 = 0.56$. For this exercise, learning begins with prior beliefs. We assume that the prior mean of the transition probability q is 0.8 and the prior mean for the news accuracy χ is 0.6. The length of the prior training sample is set to eight periods. To focus on learning only about these parameters, we assume that the remaining parameters are known.

Figure 2 shows the case of $t \in \{4, 9\}$. We start with the case of learning the accuracy of news χ while beliefs about q are never updated. Panel (A) of Figure 2 plots the posterior evolution of χ and the corresponding forecast uncertainty using circles that darken as time progresses. As news is revealed to be correct for t consecutive periods (see "good state, good news"), beliefs about news accuracy continually improve. Because of the improvement in news accuracy, forecast uncertainty gradually falls as more good

[9]The Bayesian learning process for news accuracy χ is analogous to those for state transition probabilities detailed in Appendix A. The main difference is that the posterior distribution of χ is updated in each period as past news is verified to be accurate or not.

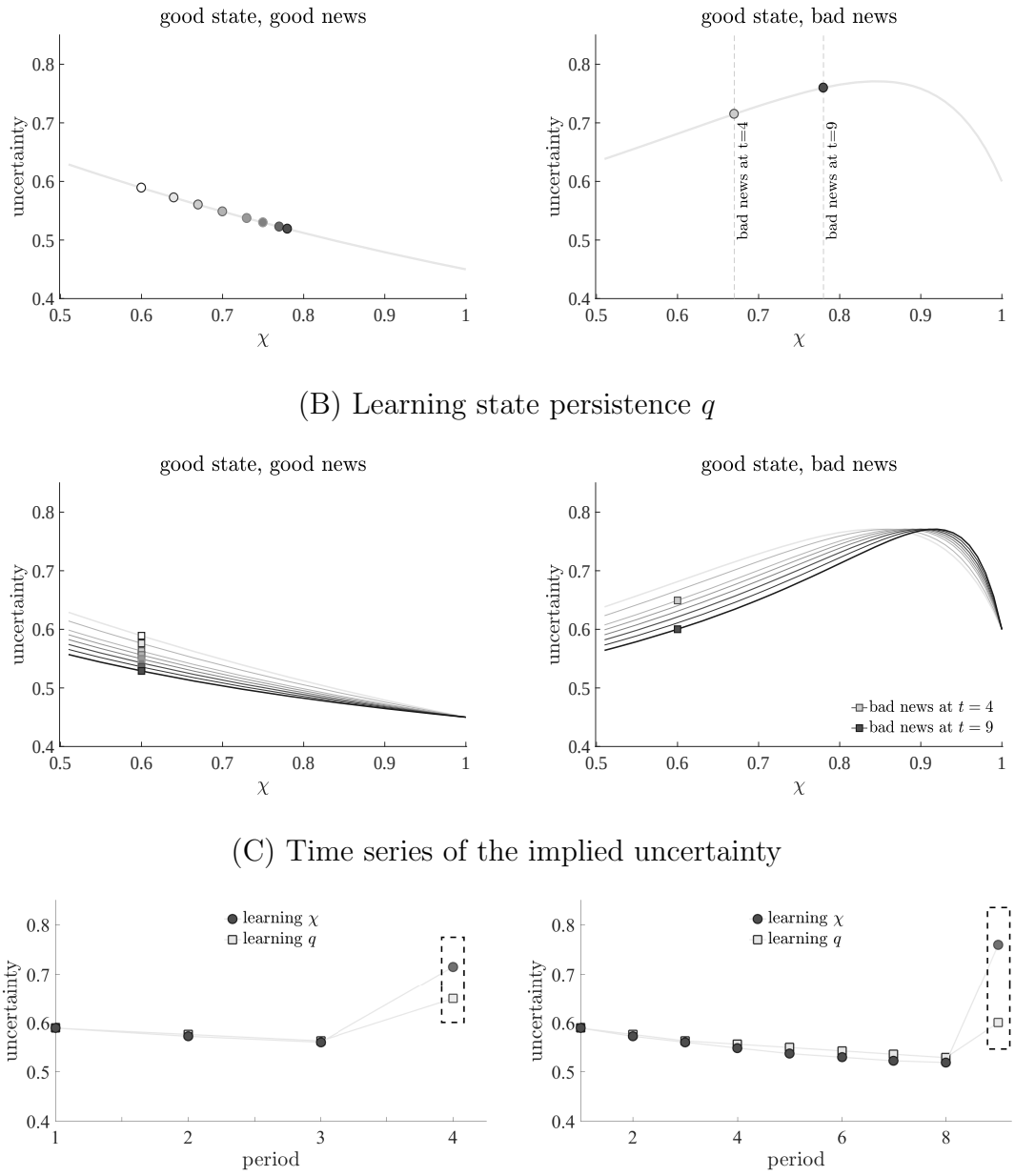
Figure 1: The role of news in forecast uncertainty



Notes: We assume $q = q_{11} = q_{22}$ and $\mu_1 = 0.84, \mu_2 = -0.22, \sigma_1^2 = 0.47, \sigma_2^2 = 0.56$. Squared lines indicate when $q = 0.5$, circled lines are when $q = 0.85$, and black solid lines are when $q = 0.99$.

news arrives. The circle corresponding to the arrival of bad news is plotted under "good state, bad news." Comparing the circles corresponding to $t = 4$ and $t = 9$ reveals that the longer the economy remains in "good state, good news" (the higher the t), the larger the jump in forecast uncertainty when bad news arrives. This is driven by two

Figure 2: The role of learning in forecast uncertainty



Notes: For illustrative purpose, we assume $\mu_1 = 0.84, \mu_2 = -0.22, \sigma_1^2 = 0.47, \sigma_2^2 = 0.56$. The prior mean of the transition probability is 0.8. The prior for the news accuracy is 0.6. The length of the prior training sample is set to eight periods. For $t-1$ consecutive periods, we are in a good state and receive good news. On t^{th} period, we are still in a good state, but receive bad news. We assume that bad news arrives at $t \in \{4, 9\}$. In the top and the middle panel, we use progressively darker shading to indicate the evolution of posterior beliefs. In the bottom panel, we use a dashed box to indicate the implied uncertainty from bad news.

forces, both of which stem from the fact that remaining in "good state, good news" for a longer period leads the agent to have a higher belief about news accuracy. This higher belief about χ has two implications. One is that, in the period just before the arrival of bad news, uncertainty is lower because the agent puts more stock in the current good news realization. Secondly, the agent also believes that the accuracy of the bad news about the next period is high, thus shifting beliefs about S_{t+1} more in the direction of a uniform distribution. We plot the time series of the implied uncertainty for these two cases as circles (compare the first column to the second column) in panel (C).

Panel (B) presents the case of learning only state persistence q, but with beliefs about χ remaining at the prior mean of 0.6. Note that the first point in this case is the same as that in panel (A) because we still start from the prior mean $q = 0.8$. Thus, the line in panel (A) also appears in panel (B). However, we no longer remain on this line now that posterior beliefs about q improve over time. This results in further declines in forecast uncertainty up to (and including) time $t-1$ as agents become more sure that they will remain in the good state. It is at time t, when news contradicts the implication of the current state and suggests switching into the bad state, that forecast uncertainty jumps up. However, both the level of uncertainty at time t and the jump from time $t-1$ are lower than those in the case of learning about news accuracy χ. We compare the time series of the implied uncertainty for these two cases as circles and squares in panel (C).

2.3 Embedding news in an asset pricing model

We have seen from the previous section that news shocks can generate significant variation in uncertainty, particularly when news contradicts existing beliefs. They can also produce rich history-dependent dynamics of uncertainty when combined with parameter learning. In light of these findings, we now examine the extent to which news-driven uncertainty fluctuations can generate variation in asset prices.

We consider an endowment economy with a representative agent that has Epstein and Zin (1989) preferences and maximizes lifetime utility,

$$V_t = \max_{C_t} \left[(1-\beta) C_t^{\frac{1-\gamma}{\alpha}} + \beta (E_t V_{t+1}^{1-\gamma})^{\frac{1}{\alpha}} \right]^{\frac{\alpha}{1-\gamma}}, \tag{4}$$

subject to budget constraint $W_{t+1} = W_t R_{c,t+1} - C_{t+1}$, where C_t is consumption of the agent, W_t is wealth, $R_{c,t+1}$ is the return on wealth, β is the discount rate, γ is risk

aversion, ψ is the intertemporal elasticity of substitution (IES), and $\alpha = \frac{1-\gamma}{1-1/\psi}$. The stochastic discount factor (SDF) is

$$M_{t+1} = \beta^{\alpha}\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}\left(\frac{PC_{t+1}+1}{PC_t}\right)^{\alpha-1}, \tag{5}$$

where PC_t is the wealth-consumption ratio.

To precisely understand the role of news, we start from the case of log utility preference and move to the case of the recursive utility with preference for early resolution of uncertainty (i.e., $\gamma > 1/\psi$). The former is a special (limiting) case of the latter, as $\gamma \to 1$ and $\psi \to 1$. We set $\beta = 0.994$ for both cases and use moderate values of $\gamma = 1.5$ and $\psi = 1.5$ for the latter. As with the previous subsection, we assume that persistence of each growth regime is identical, $q = q_{11} = q_{22}$, with values of $\mu_1 = 0.84, \mu_2 = -0.22, \sigma_1^2 = 0.47$, and $\sigma_2^2 = 0.56$ for illustrative purposes. We also assume that all assets are in zero net supply, so that $\Delta c_{t+1} = \Delta y_{t+1}$ and follows equation (1).

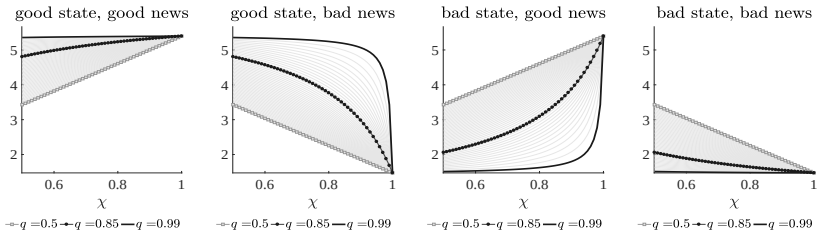
Figure 3 provides the corresponding values of the risk-free rate and the expected excess return (i.e., risk premium) on a consumption claim associated with each value of news accuracy χ and state persistence q. Panel (A) of Figure 3 shows the case of log utility preference in which the implied SDF is a function only of consumption growth (see equation (5) when $\alpha = 1$). We find that the model-implied risk-free rate is procyclical and risk premium is countercyclical. In the limiting case of completely uninformative news, $\chi = 0.5$, there exist only two values of the risk-free rate (risk premium), with the higher (lower) value realized in the good state. As the news accuracy increases toward the case of perfectly informative news, $\chi = 1$, the risk-free rate (risk premium) converges to a case in which there are again only two values, but they vary with news, with the higher (lower) value realized under good news. The intuition from Figure 1 carries over to the current risk premium graph: news alters the expected path of consumption growth and can give rise to higher risk premia when it suggests a switch in the state.

Panel (B) of Figure 3 provides the value of the risk-free rate and risk premium for the recursive utility case with preference for early resolution of uncertainty. For moderate values of q, we observe patterns of the risk-free rate and risk premium that are similar to the ones from the log utility case. It is interesting to note the case of $q = 0.99$ and $\chi = 0.99$. In this case, both switching news states are ones in which the agent thinks there is equal chance of entering a nearly permanent good or bad state in the next

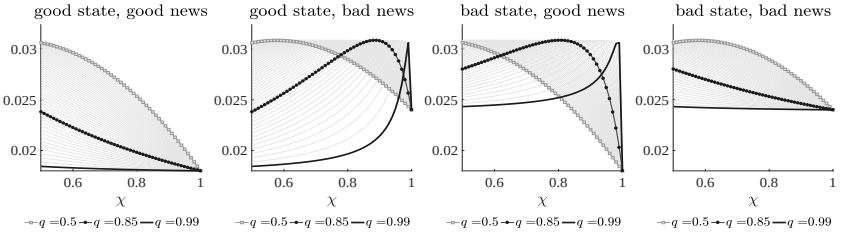
Figure 3: The role of news in asset prices

(A) Log utility

Risk-free rate

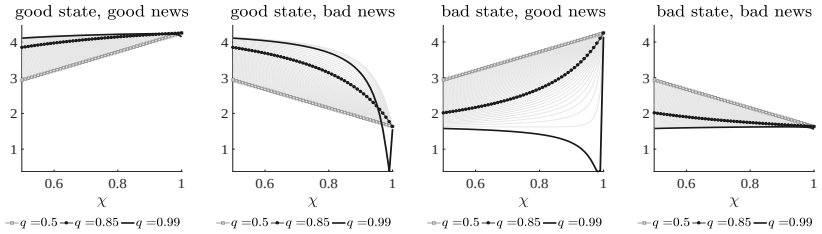


Expected excess return on consumption claim

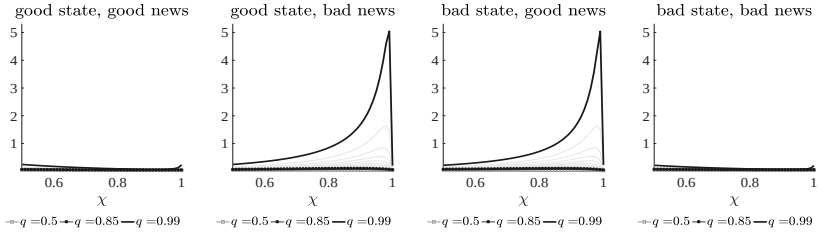


(B) Preference for early resolution of uncertainty: $\gamma = 1.5, \psi = 1.5$

Risk-free rate



Expected excess return on consumption claim



Notes: We assume $q = q_{11} = q_{22}$ and $\mu_1 = 0.84, \mu_2 = -0.22, \sigma_1^2 = 0.47, \sigma_2^2 = 0.56$. Squared lines indicate when $q = 0.5$, circled lines are when $q = 0.85$, and black solid lines are when $q = 0.99$. The numbers on the y-axis are in annualized percent terms.

period. We already learned from Figure 1 that the $q = \chi$ case maximizes the one-period-ahead uncertainty. When it comes to asset prices, the effect is greatly amplified with a preference for early resolution of uncertainty when the underlying growth states are nearly permanent. Note that the implied risk premium can be as much as 5 percent for the two switching news cases, which is roughly 170 times greater than the 0.03 percent value of the log utility case. The extremely high uncertainty about very persistent states produces a strong desire for the agent to save in a safe asset. In equilibrium, returns on the risk-free asset must be lower to equilibrate the agent's desire to save with the zero net supply of risk-free assets. This is reflected in a sharply lower risk-free rate for the values of $q = 0.99$ and $\chi = 0.99$ compared with other parameter values. Similarly, demand for the risky consumption claim is very low in this environment, and the agent requires a very high expected excess return on this asset in order to equilibrate shorting demand with the asset's zero net supply.

We now consider the implications of parameter learning in our model that includes news shocks. Using a similar model without news shocks, Johannes, Lochstoer, and Mou (2016) are able to closely match several asset pricing moments.[10] One key feature of their model is that it generates strongly countercyclical return volatility and a high equity premium in recessions through parameter learning. This is because the infrequent nature of recessions implies greater parameter updating when the economy visits that state. Thus, long-run shocks to parameter beliefs will be largest during recessions, which contributes to the high average equity premium in recessions. It is interesting that, above and beyond the usual mechanics described in Johannes, Lochstoer, and Mou (2016), our model is able to generate large fluctuations in growth uncertainty when the growth regime itself remains unchanged through the addition of news shocks. On top of this, in our model, revisions in parameter beliefs will further interact with news shocks in a way that can produce richer dynamics for asset prices.

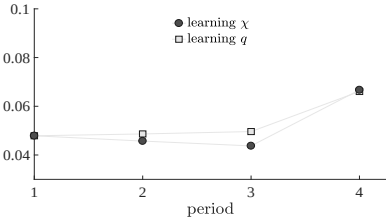
For illustrative purposes in this subsection, we rely on the anticipated utility approach to price assets, as in Johannes, Lochstoer, and Mou (2016).[11] We later examine the case of fully rational pricing in which the agent takes into account how future state realizations will impact future parameter beliefs. As we did in the previous section,

[10] We briefly summarize the mechanics of the asset pricing model as follows. If agents prefer an early resolution of uncertainty, changing beliefs are priced risks. Therefore, uncertainty about future revisions in beliefs leads to higher risk prices, equity premium, and return volatility.

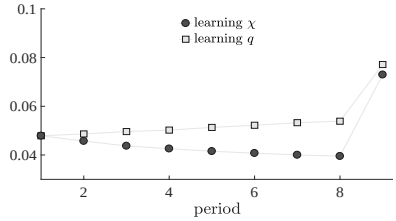
[11] Under this approach, the agent does not take into account future revisions in beliefs with respect to parameters, but she fully takes state uncertainty into account.

Figure 4: The role of learning in risk premium (under anticipated utility)

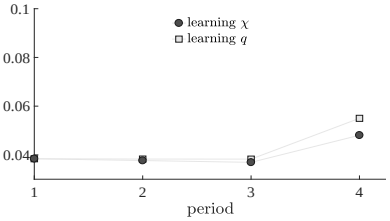
(A) RU, Prior: $E(\chi) = 0.6$, $E(q) = 0.8$



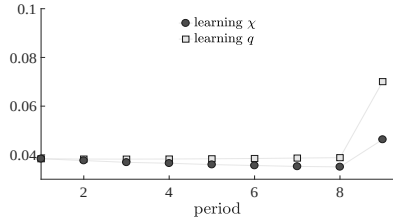
(B) RU, Prior: $E(\chi) = 0.6$, $E(q) = 0.8$



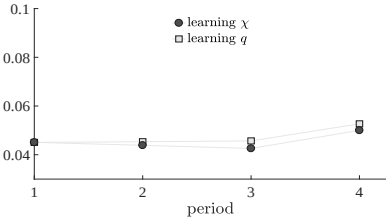
(C) RU, Prior: $E(\chi) = 0.8$, $E(q) = 0.6$



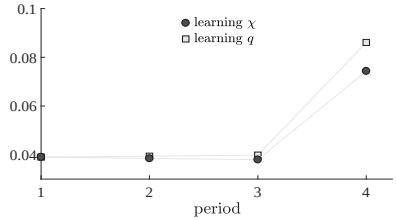
(D) RU, Prior: $E(\chi) = 0.8$, $E(q) = 0.6$



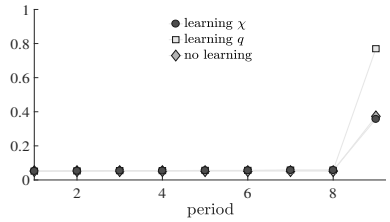
(E) RU, Prior: $E(\chi) = 0.6$, $E(q) = 0.6$



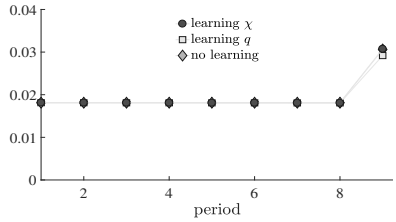
(F) RU, Prior: $E(\chi) = 0.8$, $E(q) = 0.8$



(G) RU, Prior: $E(\chi) = 0.95$, $E(q) = 0.95$



(H) LU, Prior: $E(\chi) = 0.95$, $E(q) = 0.95$



Notes: Risk premium is the expected excess return on consumption claim. RU indicates the case of preference for early resolution of uncertainty: $\gamma = 1.5, \psi = 1.5$. LU refers to the log utility case. We assume $\mu_1 = 0.84, \mu_2 = -0.22, \sigma_1^2 = 0.47$, and $\sigma_2^2 = 0.56$ to compute the risk premium, which is provided in annualized percent terms (y-axis). The length of the prior training sample is set to eight periods. For $t-1$ consecutive periods, we are in a good state and receive good news. On t^{th} period, we are still in a good state, but receive bad news. We consider $t \in \{4, 9\}$. We compare learning q and χ cases.

we consider the experiment in which for $t-1$ consecutive periods, the agent is in a

good growth state and receives good news. In the t^{th} period, she receives bad news while remaining in the good state. We calculate the corresponding risk premium on a consumption claim under parameter learning with the anticipated utility assumption. We provide various cases that have different prior means for q and χ and different values of t to understand the importance of prior beliefs and history dependence in this exercise.

Panels (A) and (B) of Figure 4 repeat the exercises shown in panel (C) of Figure 2 for the risk premium under our $\gamma = 1.5, \psi = 1.5$ recursive utility calibration. Consistent with the results for forecast uncertainty, we find that the risk premium jumps up when bad news arrives and that the jump is greater when the bad news arrives after the agent has spent a longer time in the good state receiving good news. Also as with the case of forecast uncertainty, the jump is larger when the agent learns about χ rather than about q.

Panels (C) and (D) show the same exercise under an alternate prior where $E(\chi) = 0.8$ and $E(q) = 0.6$. In this case, the jumps are greater when agents learn about q rather than χ. In these exercises, the prior can be interpreted as the initial belief held by agents at the start of the simulation. Therefore, the differences seen under different priors are another illustration of the history-dependent nature of the asset price response to news in this economy. Panels (E) and (F) further illustrate this fact, and we see that, overall, the effect of the bad news shock is greater for higher prior mean values of state persistence q and news accuracy χ.

Parameter learning also induces variations in the risk premium during periods prior to the arrival of bad news, but the effect is minimal when prior beliefs about q and χ are high. This can be seen, for example, with high prior means $E(\chi) = 0.95$ and $E(q) = 0.95$, a case that is presented in panel (G). The magnitude of the jump in the risk premium after the arrival of bad news is striking. For example, in the case of learning q, the risk premium starts at 0.05 percent in period 1 and jumps by a factor of nearly 15 to 0.77 percent with the arrival of bad news in period 9. It is important to emphasize that this is achieved with moderate values of $\gamma = 1.5, \psi = 1.5$. Without parameter learning, when beliefs about both q and χ remain at their prior values, the risk premium jumps to only 0.37 percent after the same sequence of state and news realizations. That's less than half the size of the risk premium with parameter learning. Also, when we repeat the exercise with log utility, displayed in panel (H), the size of the risk premium upon arrival of bad news reduces significantly to 0.03 percent. The size

of the risk premium at period 9 is roughly 1.6 times greater than the value at period 1, a stark contrast to the massive amplification (15 times greater) achieved when learning about q is combined with a preference for early resolution of uncertainty. Overall, the optimism built during a long period of economic expansion, i.e., the increase in posterior probability of remaining in the good state, can lead to a Minsky moment once the economic agent, endowed with a preference for early resolution of uncertainty, receives news indicating that the future is not as bright as expected. Parameter learning amplifies the effect of this news shock to generate a sudden increase in the risk premium despite no evidence of a recession in contemporaneous real growth rates.

Note that the risk premium associated with the case of learning q increases slightly over time as the economy remains in "good state, good news." This is because we keep $q = q_{11} = q_{22}$ identical across states: as the agent stays in the good state for longer, she learns that not only is the good state more persistent, but that the bad state is more persistent as well. In a more realistic setting in which the agent believes the two states differ in persistence and learns only about q_{11}, the risk premium decreases until period t as the agent remains in "good state, good news." Our simplifying assumption that $q_{11} = q_{22}$ also amplifies the magnitude of the jump in the risk premium when bad news arrives. If only q_{11} is updated in this scenario, then when the bad news comes, the agent places roughly equal weight on an almost permanent good state and a slightly less persistent bad state. The corresponding risk premium jumps up by a lesser extent. However, in the more general case of $q_{11} \neq q_{22}$ when the agent learns only about q_{11}, the jump can be still greater for higher prior values of χ, q_{11} and q_{22}.

3 Estimation

3.1 Finding an empirical proxy for the news component

Our model features news that arrives in discrete form, but there is no direct empirical proxy with this feature. We therefore look for a forward-looking variable that contains information about agents' beliefs regarding future states. Imagine that there are forecasters who know the true model parameters and current states, and they receive noisy news about the future state. Let $I_t = \{n^t, S^t, y^t, \Pi, \theta\}$ denote the information set for these forecasters. Here, we are assuming that these forecasters have full structural

knowledge of the economy, as in standard rational expectations models. Suppose that we can observe their probability forecast[12]

$$z_t = p(S_{t+1} = 1|I_t). \tag{6}$$

The Survey of Professional Forecasters' anxious index—the probability of a decline in real GDP in the quarter after a survey is taken—is a reasonable real-world proxy for (6). Professional forecasters have an information set that is closest to I_t, but in practice, their information set may not be perfect. Their forecasts could be biased or inefficient. We allow for this by modeling the anxious index as

$$z_{\text{SPF},t} = \Phi\left(b_z + \Phi^{-1}\Big(p(S_{t+1} = 1|I_t)\Big) + u_t\right), \quad u_t \sim N(0, \sigma_z^2), \tag{7}$$

where $\Phi(\cdot)$ is a cdf of the standard normal distribution.[13] Note that b_z and σ_z^2 capture potential bias and inefficiency in probability forecasts. If we set $b_z = \sigma_z^2 = 0$, then we recover (6). In the empirical analysis below, we proceed with the assumption that $b_z = 0$ but allow for errors with $\sigma_z^2 > 0$.

3.2 Information set

In Section 2, we assumed that some or all model parameters and the full history of states are known at time t to illustrate the implications of news for parameter learning, state prediction, forecast uncertainty, and asset returns. Here, in the econometrician's inference problem, this assumption is relaxed, with all model parameters and states being unknown. Consider an econometrician who observes past and current output growth and recession probability forecasts. She updates her beliefs about states and parameters sequentially using Bayes' rule as she obtains new data.

3.3 Solving the econometrician's sequential learning problem

Formally, we develop a novel filtering technique that uses both actual GDP growth y_t and recession probability forecasts $z_{\text{SPF},t}$ from the Survey of Professional Forecasters to

[12]Note that we could also use mean forecast, $E(y_{t+1}|I_t) = \sum_{i=1}^{2} \mu_i p(S_{t+1} = i|I_t)$, which is a function of $p(S_{t+1} = i|I_t)$. Instead, we rely on a direct measure of $p(S_{t+1} = i|I_t)$.

[13]The choice of the probit linking function is just for convenience.

solve the sequential state filtering and parameter learning problem. Both observables are collected in $x_t = [y_t, z_{\text{SPF},t}]$. Using data from 1969:Q1 through 2016:Q3, we obtain the full joint posterior distribution of the model primitives, including the filtered distribution of discrete states at each time. The joint posterior $p(\theta, \Pi, \kappa^t | x^t)$ summarizes subjective beliefs after observing x^t. We use κ_t to indicate a realization of the combined Markov state $\kappa_t = \{S_t, n_t\}$, which can take on four values.

As explained by Johannes, Lochstoer, and Mou (2016), the joint learning of states and parameters is a high-dimensional problem that incurs confounding effects arising from multiple sources of uncertainty. We tackle this problem by relying on particle methods to directly sample from the particle approximation to the joint posterior, which can be factorized into the product of the conditional posteriors

$$p(\theta, \Pi, \kappa^t | x^t) = \underbrace{p(\theta, \Pi | \kappa^t, x^t)}_{\text{(i) parameter learning}} \times \underbrace{p(\kappa^t | x^t)}_{\text{(ii) state filtering}} . \tag{8}$$

To sample from (i) and (ii) jointly, we extend the particle learning algorithm developed by Carvalho et al. (2010), which is a generalization of the mixture Kalman filter of Chen and Liu (2000). The idea is to utilize conditional sufficient statistics for parameters and states as particles. The details are provided in Appendix C and E.

Before moving on to the application to the U.S. data, we conduct a simulation exercise to check the performance of the particle learning algorithm and examine how news embedded in survey forecasts influences the econometrician's sequential learning problem in the presence of multiple confounding effects.

3.4 Simulation exercise

In this exercise, we run the particle learning algorithm over different sets of simulated data. In all cases, we fix parameters at their prior median values in Table 2 and set the length of simulated data to match the estimation sample. For all cases, we use the same set of simulated paths for output growth (y_t) and states (S_t) from equation (1). We simulate two different sets of time series of news (n_t): one in which news is informative, $\chi = 0.8$, and the other in which it is not, $\chi = 0.5$. Survey forecasts ($z_{\text{SPF},t}$) are simulated based on equation (7) for each of these two cases of χ using the simulated states and the corresponding news series.

Table 1: Root-mean-squared errors of parameter estimates and output growth forecasts from simulation exercises

	Case 1			Case 2			Case 3		
	$\chi = 0.5$ fix χ			$\chi = 0.8$ fix χ			$\chi = 0.8$ learn χ		
	5%	50%	95%	5%	50%	95%	5%	50%	95%
	Parameter learning								
μ_1	0.009	0.059	0.142	0.005	0.061	0.131	0.006	0.070	0.138
σ_1^2	0.002	0.048	0.114	0.003	0.046	0.118	0.002	0.043	0.122
μ_2	0.006	0.060	0.131	0.010	0.065	0.134	0.008	0.071	0.134
σ_2^2	0.003	0.037	0.108	0.002	0.038	0.110	0.002	0.037	0.114
q_{11}	0.003	0.018	0.049	0.002	0.016	0.040	0.002	0.016	0.045
q_{22}	0.002	0.017	0.046	0.002	0.015	0.039	0.002	0.015	0.042
	Output growth forecast								
y_t	0.656	0.727	0.783	0.628	0.694	0.750	0.629	0.693	0.752

Notes: We conduct three simulation exercises labeled Case 1, Case 2, and Case 3. In all data sets, output growth data are assumed to be identical. The survey forecasts, however, are generated based on uninformative news $\chi = 0.5$ (Case 1) and informative news $\chi = 0.8$ (Cases 2 and 3). In Cases 1 and 2, we keep the number of estimated parameters identical by fixing χ and σ_z^2 to their true values. In Case 3, we learn all parameters including χ and σ_z^2. In the top panel, we define RMSE $= \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\theta - \theta_{t|1:t})^2}$ where θ denotes the true parameter value and $\theta_{t|1:t}$ is the posterior median estimate of θ conditional on information at t. In the bottom panel, we define RMSE $= \sqrt{\frac{1}{T-1}\sum_{t=2}^{T}(y_t - y_{t|1:t-1})^2}$, where $y_{t|1:t-1}$ is the posterior mean one-step-ahead prediction of y_t conditional on information at $t-1$. In the table we report 5^{th}, 50^{th}, and 95^{th} percentiles of RMSE distributions based on 100 Monte Carlo simulations.

For the particle learning algorithm, the details of conjugate prior distributions are provided in Table 2. We set the length of the prior training sample (prior precision) to 10 years. Note that all simulation results are generated with unbiased priors.

To understand how news would influence the inference of other model parameters, we start by comparing two cases in which we keep the number of estimated parameters identical. In Case 1, news is uninformative, and the econometrician correctly fixes χ to 0.5 in the estimation. In Case 2, news is informative ($\chi = 0.8$), and the econometrician fixes χ and σ_z^2 to their true values in the estimation and estimates the same set of parameters as in Case 1. Comparing these first two cases allows us to see how the presence of news shocks affects the econometrician's learning problem while holding fixed the degree of parameter uncertainty. In Case 3, news is as informative as it is

in Case 2, but the econometrician now must also learn χ and σ_z^2. By comparing this case with Case 2, we aim to understand whether parameter uncertainty with respect to news makes learning about the other parameters more difficult.

In the top panel of Table 1, we provide 5^{th}, 50^{th}, and 95^{th} percentiles of root-mean-squared error (RMSE) distributions based on 100 Monte Carlo simulations. The RMSE of a parameter estimate is defined by $\sqrt{\frac{1}{T}\sum_{t=1}^{T}(\theta - \theta_{t|1:t})^2}$, where θ denotes the true parameter value and $\theta_{t|1:t}$ is the posterior median estimate of θ conditional on information at t. Two things are noteworthy. First, the small RMSEs across Monte Carlo simulations confirm that the particle learning algorithm performs well. Second, we find a roughly 10 percent improvement in parameter learning accuracy for q_{11} and q_{22} when news is informative. While the differences in absolute magnitudes are small, we find that RMSEs are uniformly smaller with informative news. The results for the other parameters are virtually identical across Cases 1 and 2. This is to be expected, because we impose the same degree of uncertainty in both cases and news in this model is informative about the realization of the future state and not the distribution of growth conditional on a state. It is only when we learn parameters associated with news, i.e., χ and σ_z^2 in addition to the others, that we find some marginal deterioration of the econometrician's learning accuracy for some parameters. The median RMSEs for μ_1 and μ_2 under Case 3 are larger than those in Case 2. Recognizing the confounding nature of parameter learning—when more uncertainty about one object makes learning about another more difficult—is key to understanding this result.

In the bottom panel of Table 1, we compute output growth forecast RMSEs. We find a 5 percent improvement in the median ex-post accuracy of output growth forecasts if the probability forecasts contain somewhat informative forward-looking news, regardless of whether we fix or estimate χ and σ_z^2. In sum, the simulation results are consistent with the implications of news that we derived in Section 2.

3.5 Application to U.S. data

We now apply the particle learning algorithm to U.S. data. We impose the same priors as in the simulation exercise.

Parameter estimates. Table 2 reports 5^{th}, 50^{th}, and 95^{th} percentiles of posterior

Table 2: Prior and posterior distributions of parameters

	Prior (1)			Posterior (2) With surveys Benchmark			(3) With surveys $\chi = 0.5$			(4) Without surveys		
	5%	50%	95%	5%	50%	95%	5%	50%	95%	5%	50%	95%
μ_1	-1.00	1.00	3.00	0.73	0.84	0.94	0.76	0.85	0.93	0.74	0.83	0.95
σ_1^2	0.12	0.40	1.50	0.34	0.47	0.63	0.37	0.44	0.53	0.22	0.29	0.36
μ_2	-2.00	0.00	2.00	-0.55	-0.22	0.12	-0.83	-0.55	-0.24	0.17	0.40	0.60
σ_2^2	0.12	0.40	1.50	0.26	0.56	1.10	0.36	0.53	0.98	0.76	1.02	1.42
q_{11}	0.64	0.80	0.93	0.84	0.89	0.94	0.90	0.94	0.96	0.80	0.85	0.89
q_{22}	0.64	0.80	0.93	0.51	0.65	0.78	0.58	0.68	0.78	0.74	0.80	0.86
χ	0.64	0.80	0.93	0.84	0.91	0.96	-	0.50	-	-	-	-
σ_z^2	0.06	0.24	0.96	0.39	0.71	1.28	0.28	0.35	0.42	-	-	-

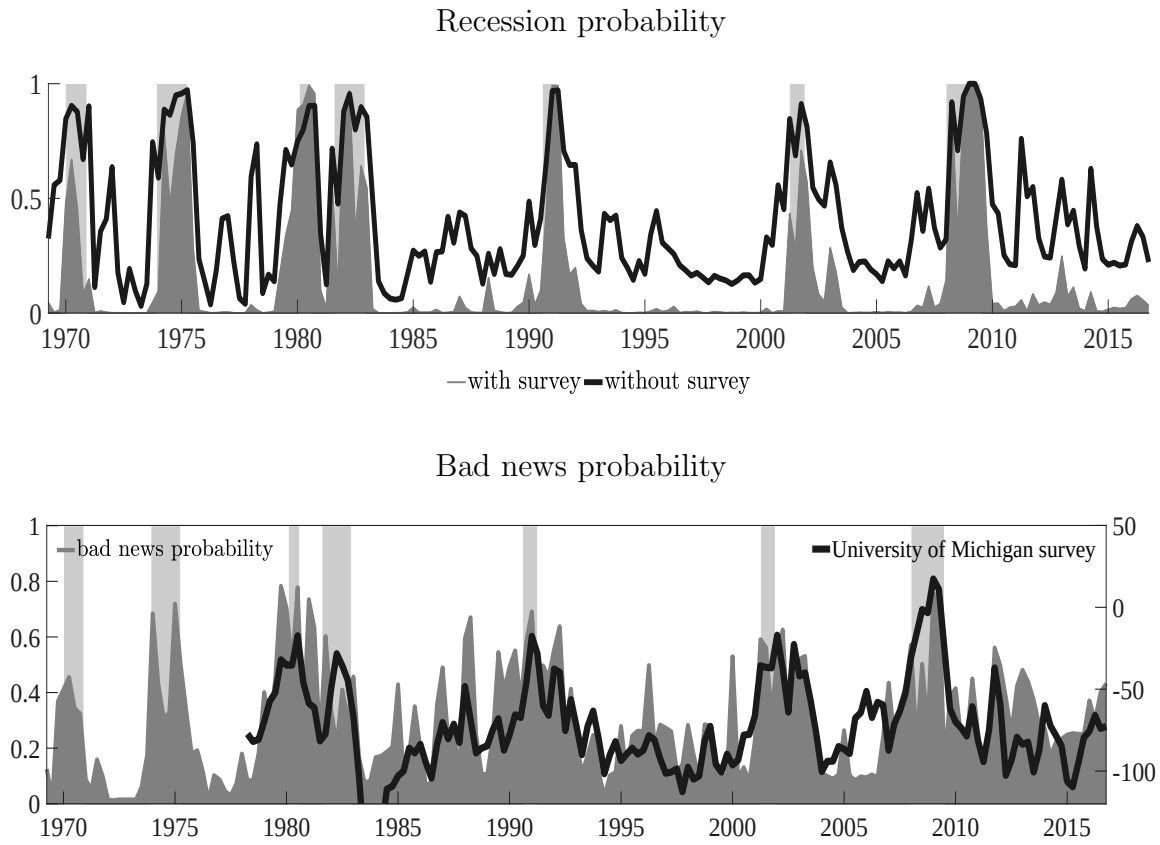
Notes: This table reports 5^{th}, 50^{th}, and 95^{th} percentiles of posterior distributions in Figure A-1. The details of prior choices are provided in Table A-1.

distributions.[14] The benchmark estimation results are provided in panel (2) of Table 2. First, the first growth regime ($S_t = 1$) is identified with a positive mean and the second growth regime ($S_t = 2$) with a negative mean. Note that we impose $\mu_1 > \mu_2$ only to deal with the label-switching problem in the estimation. We also find that the posterior median estimate of variance is larger in the second regime than in the first regime, which together with mean estimates, provides the natural interpretation that the first regime is the expansion regime and the second regime is the recession regime. Second, the posterior intervals associated with the expansion regime are much tighter than those associated with the recession regime. This is not surprising, because the average duration of expansions is much longer than the average duration of recessions, a property that is reflected in much lower estimates of q_{22} than q_{11}. Third, the posterior median estimate of χ is 0.91, which implies that news extracted from surveys is quite informative about the future growth regime.

To understand how news would influence the inference of model parameters, we repeat the estimation by fixing $\chi = 0.5$. The corresponding results are provided in panel (3) of Table 2. We find that the posterior mean and variance estimates for the expan-

[14]The evolution of parameter learning is provided in Figure A-1. The credible intervals at time 0 correspond to the 90 percent prior intervals. As more information from observed data is incorporated into the posterior distributions over time, the 90 percent credible intervals shrink. Those at time T are posterior credible intervals one would obtain from the entire time series of data.

Figure 5: Filtered estimates of recession and bad news probabilities



Notes: In the top panel, we compare the two filtered recession probabilities obtained with (gray area) and without surveys (black solid line). In the bottom panel, the correlation of the model-implied $P(n_t = 2)$ (gray area) and the University of Michigan news index (black solid-line) is 0.60.

sion regime are similar to those reported in panel (2). However, the posterior mean for the recession regime becomes more negative, and both the mean and variance in the recession regime are estimated with more precision when we exgenously assume that there is no news. Similarly, both state transition probabilities are slightly higher and estimated with more precision. The width of the equal-tail probability 90 percent credible interval generally decreases for all parameters, highlighting that the confounding nature of learning news can alter uncertainty associated with parameters as well. This is consistent with the results from the simulation exercise of Case 3 in Table 1.

Finally, we remove survey forecasts from the estimation to understand how their presence affects the inference of model parameters. Panel (4) of Table 2 reports posterior distributions from the estimation in which only output growth is used. We find that

posterior median estimates associated with the recession regime are very different from those reported in panels (2) and (3). Specifically, the posterior median estimate of μ_2 is positive and σ_2^2 is estimated to be twice as large. Perhaps surprisingly, q_{22} is estimated to be much more persistent without surveys. The distinction can be seen clearly in Figure A-1, where the sequential parameter posterior distributions are plotted.

Regime probabilities. Another way to see the differences between the estimations with survey data and those without is to look at the filtered regime probabilities.[15] The top panel of Figure 5 combines the probabilities of "bad state, good news" and "bad state, bad news" and shows that this probability of the bad state roughly coincides with the NBER recession bars. However, when only output growth is used in the estimation (the "without survey" line), these recession probabilities tend to be biased upward and bad states are less tightly identified. This is reflected in the higher estimate of q_{22} in panel (4) of Table 2.

The bottom panel of Figure 5 combines the estimated probabilities of "good state, bad news" and "bad state, bad news" for our baseline case, giving the total probability of receiving bad news. Note that the correlation of the University of Michigan's Survey of Consumers news index (inverted) and our filtered probability of bad news realizations is around 0.6.[16] We also examine another measure of bad news that counts, for each quarter, the number of stories appearing in the *New York Times* and the *Washington Post* that include the word "recession."[17] We obtain a 0.47 correlation between our filtered probability of bad news and this index.

Forecasting GDP growth. We regress one-quarter-ahead GDP growth on various explanatory variables, including our estimated bad news probabilities (henceforth, bad news). We examine if the estimated news can predict one-period-ahead GDP growth

[15] Figure A-2 provides the probability of each regime in the benchmark estimation. Note that the economy spends most of the time in the "good state, good news" regime. However, the probability of "good state, bad news" is non-negligible. This is the regime in which the economy is currently in the good state, but news suggests that the economy will enter the bad state in the following period. On the other hand, the economy assigns very small probabilities to the "bad state, good news" and "bad state, bad news" regimes. Nevertheless, in periods when the probability of being in a bad state is high, the conditional probability of receiving good news can be substantial.

[16] This index is based on responses to the following question in the University of Michigan's Survey of Consumers: "During the last few months, have you heard of any favorable or unfavorable changes in business conditions?" The index is constructed as the percent replying "favorable" minus the percent replying "unfavorable" plus 100. We compare the negative of this index to our filtered bad news probability.

[17] This is a replication of the R-word index developed by *The Economist*. A description can be found at http://www.economist.com/node/566293.

Table 3: Predicting GDP growth

	(A)	(B)	(C)	(D)	(E)
Constant	0.45	0.67	0.54	0.60	0.40
	(0.08)	(0.05)	(0.08)	(0.08)	(0.10)
Current GDP growth	0.32		0.20	0.10	0.37
	(0.08)		(0.08)	(0.09)	(0.12)
Bad news		-0.34	-0.29		
		(0.05)	(0.07)		
UMich survey				-0.33	
				(0.07)	
R-word index					-0.04
					(0.05)
Adj. R^2	0.10	0.18	0.22	0.24	0.15

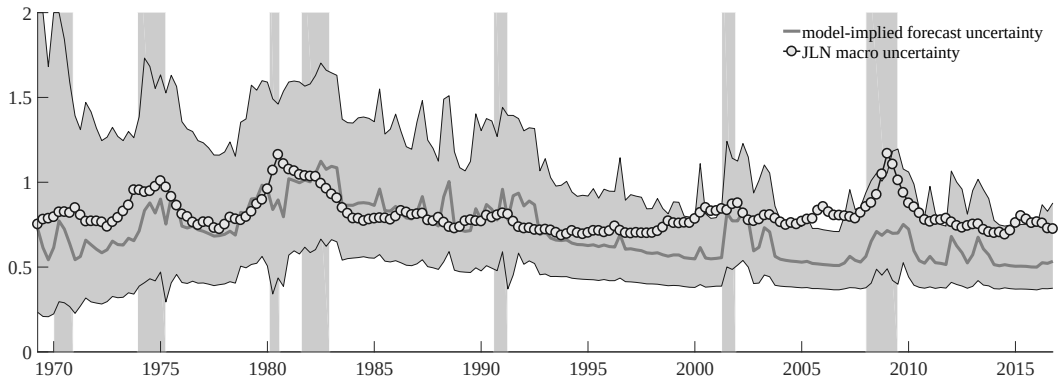
Notes: We regress one-quarter-ahead GDP growth on various explanatory variables. Bad news refers to our estimated bad news probability. UMich survey is the University of Michigan bad news index. The R-word index counts, for each quarter, the number of stories in the *New York Times* and the *Washington Post* that include the word "recession." The estimation sample for (A), (B), and (C) is from 1969:Q1 through 2016:Q3; the estimation sample for (D) is from 1978:Q1 through 2016:Q3; the estimation sample for (E) is from 1985:Q1 through 2016:Q3. The estimation results are qualitatively similar if we set the estimation sample identical across various specifications from 1985:Q1 through 2016:Q3. We report standard errors in parentheses. Since bad news is a generated regressor, asymptotic standard errors are constructed using generalized methods of moments.

as implied by the model. The regression results are provided in Table 3. As our model predicts, bad news significantly predicts lower future GDP growth and has strong forecasting power above the predictability contained in current GDP growth (compare across columns (A), (B), and (C)). We compare the forecasting power of our filtered bad news probability to other measures of bad news. We find that while the University of Michigan bad news index (column (D)) has the highest predictive power among all explanatory variables, the adjusted R^2 value is not too different from the value obtained from the benchmark prediction regression (column (C)). This is particularly notable because our bad news probabilities are filtered from professional forecasts collected during the middle month of a given quarter, while both the University of Michigan survey and R-word index contain information collected throughout the entire quarter. This difference should give these two alternate series a forecasting advantage over our bad news probabilities.

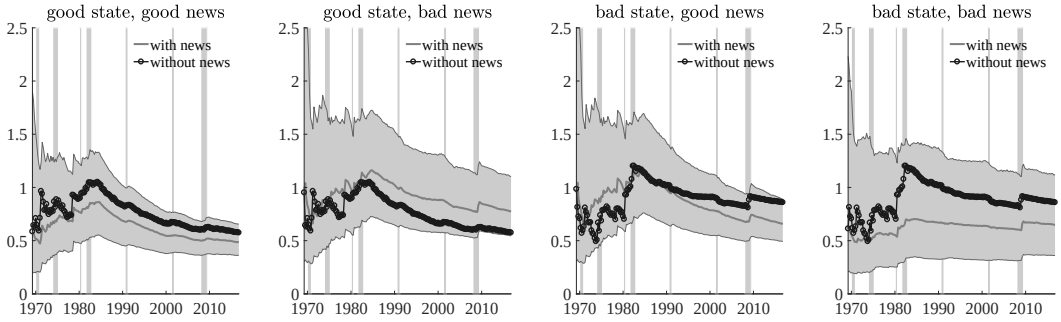
Forecast uncertainty. The top panel of Figure 6 plots the evolution of forecast error variance over time. The correlation between posterior median forecast error variance

Figure 6: Forecast uncertainty

Time series of the implied forecast uncertainty



Time series of the implied forecast uncertainty conditional on state



Notes: In the top panel, we plot posterior median forecast error variance with the 90 percent credible interval (shaded areas). We compare with the macroeconomic uncertainty (circled line) in Jurado, Ludvigson, and Ng (2015). The correlation between posterior median forecast error variance and JLN macro uncertainty is 0.49. In the bottom panel, we plot the conditional posterior median forecast error variance with the 90 percent credible interval (shaded areas). We compare with the conditional posterior median forecast error variance that assumes $\chi = 0.5$ (circled line).

and the macroeconomic uncertainty in Jurado, Ludvigson, and Ng (2015) is 0.49.[18] The bottom panel provides posterior forecast error variance conditional on each state, which is plotted against the median conditional posterior forecast error variance derived from the $\chi = 0.5$ (without news) case. Consistent with the implication in Figure 1, if news suggests that the current regime will persist next period, then forecast uncertainty monotonically decreases (first and fourth columns). However, if news contradicts the implication of the current state alone and suggests switching into a different state, then

[18]Note that Jurado, Ludvigson, and Ng (2015) provide an estimate of uncertainty that is based on ex-post forecast errors. In our model, the relationship between ex-ante forecast error variance and the size of ex-post forecast errors is not monotonic because, as discussed in Section 2.2, news always reduces ex-post forecast errors, while its relationship with ex-ante forecast variance is state dependent.

the weight is placed more equally on the two possible state outcomes. It turns out that for our full-sample posterior median estimates, forecast uncertainty increases with bad news in good states and decreases with good news in bad states (second and third columns).

4 Asset Pricing with Priced Parameter Uncertainty

Section 2.3 presented illustrative examples of the qualitative effects of news shocks on risk premia in the presence of parameter learning when using the anticipated utility approach. While this approach is highly tractable, taking it abstracts away from pricing risk due to parameter uncertainty. Instead, assets are priced as if the agent believes that her current mean estimates of parameter values are the true values. Collin-Dufresne, Johannes, and Lochstoer (2016) show that rationally pricing parameter uncertainty can lead to quantitatively large differences from the anticipated utility case. Therefore, in this section, we use the empirical estimates from Section 3 in a model where parameter uncertainty is rationally priced to assess the quantitative implications of our model for asset prices.

To be more precise, we solve an asset pricing problem that features fully rational sequential learning of unknown $\{q_{11}, q_{22}\}$, where the agent observes both the current true state as well as news about the one-period-ahead state.[19] To alleviate the computational burden, we assume the agent is fully aware of the rest of the model parameters, including the accuracy of news.[20] One reason to allow the agent to learn $\{q_{11}, q_{22}\}$ instead of the other parameters is that uncertainty about transition probabilities has the largest asset pricing impact. Recall that we compared the implied risk premium under the case of learning news accuracy χ with the case of learning state persistence q in Figure 4. For prior means of 0.8 or greater for both χ and q, we showed that learning q produced a larger jump in risk premia upon the arrival of bad news. From Collin-Dufresne, Johannes, and Lochstoer (2016), we also know that uncertainty about

[19]Note that in our model, equilibrium asset prices reflect only information already available to the agent. Therefore, it is sufficient for her to learn parameters only from observations of the true states and news without explicitly considering asset prices to also be in her information set.

[20]Each parameter that must be learned increases the number of state variables in the full-fledged parameter learning model. Therefore, the curse of dimensionality restricts us to considering learning over only a small number of parameters.

transition probabilities $\{q_{11}, q_{22}\}$ has a greater asset pricing impact than does uncertainty about conditional means or variances. Furthermore, as discussed in Section 3, since the news we are considering is about the future state and not about future growth rates, it provides information only about $\{q_{11}, q_{22}\}$ and not the other parameters governing the consumption or dividend processes in the model. Therefore, news in this model should interact more strongly with learning about $\{q_{11}, q_{22}\}$ than learning about $\{\mu_i, \sigma_i^2\}_{i=1,2}$.

4.1 The environment

Preferences. We continue to assume the same preferences represented by equation (4), which is repeated here for convenience,

$$V_t = \max_{C_t} \left[(1-\beta) C_t^{\frac{1-\gamma}{\alpha}} + \beta (E_t V_{t+1}^{1-\gamma})^{\frac{1}{\alpha}} \right]^{\frac{\alpha}{1-\gamma}}. \tag{9}$$

Solution. The equilibrium expression for the wealth-consumption ratio is

$$PC(\kappa_t, X_t)^\alpha = E\left[\beta^\alpha e^{(1-\gamma)(\mu_{\kappa_{t+1}} + \sigma_{\kappa_{t+1}} \epsilon_{t+1})} \big(PC(\kappa_{t+1}, X_{t+1}) + 1 \big)^\alpha | \kappa_t, X_t \right],$$

where X_t denotes summary statistics governing the parameter learning problem that evolve according to

$$X_{t+1} = f(\kappa_{t+1}, \kappa_t, X_t).$$

Transition probability estimates $E(q_{ii}|n^t, y^t, S^t, \theta)$ for $i \in \{1, 2\}$ are functions of $\{\kappa_t, X_t\}$, while $f(\cdot)$ is from Bayes' rule. The details are provided in Appendix A. Following Collin-Dufresne, Johannes, and Lochstoer (2016), we price an equity claim on an exogenous dividend whose growth is described by

$$\Delta d_{t+1} = \bar{\mu} + \rho(\Delta c_{t+1} - \bar{\mu}) + \sigma_d \epsilon_{d,t+1}, \quad \epsilon_{d,t+1} \sim N(0,1), \tag{10}$$

and we similarly solve for the price-dividend ratio of this claim. Here $\bar{\mu}$ is the unconditional mean of consumption growth. Its dependence on the state transition probabilities imply that the agent's beliefs about this quantity also evolve over time.

In this economy, the agent's beliefs about $\{q_{11}, q_{22}\}$ converge to the truth and her uncertainty about these parameters vanishes as time progresses. Therefore, the model

is solved by iterating the policy functions for the wealth-consumption and price-dividend ratios back from a distant endpoint that is approximated by an economy in which all parameters are known. Solutions for risk-free rates and risk premia are readily obtained from the solutions for the wealth-consumption and price-dividend ratios. Details on the numerical solution algorithm are provided in Appendix F.

4.2 Model-implied asset prices

We simulate asset prices using the calibrated values in Johannes, Lochstoer, and Mou (2016)

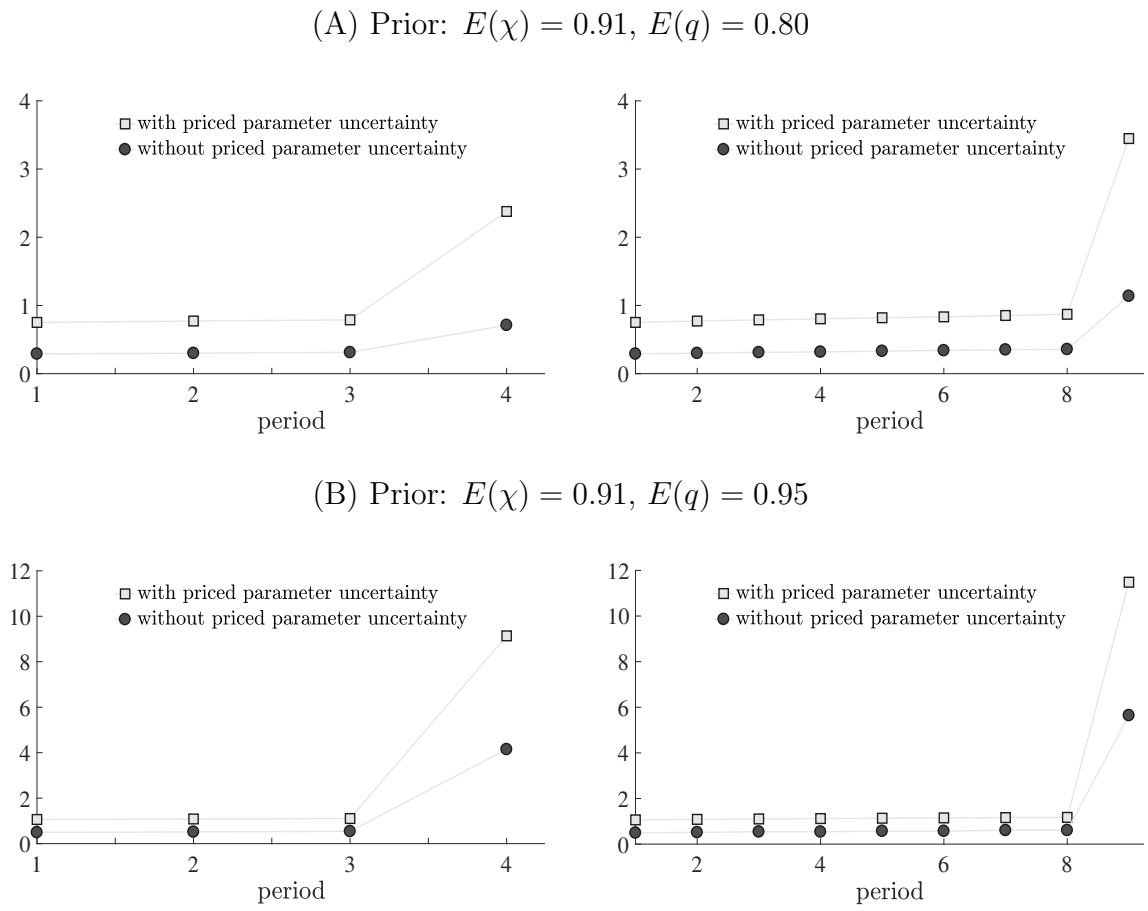
$$\beta = 0.994, \quad \gamma = 10, \quad \psi = 1.5, \quad \rho = 4.5, \quad \sigma_d = 4.81,$$

along with the "Benchmark" posterior median estimates of the growth and news process parameters in Table 2.

Quantitative effect of pricing parameter uncertainty. We first attempt to understand the quantitative effect of pricing parameter uncertainty on the risk premium of an agent compared to the anticipated utility case discussed in Section 2.3. For ease of comparison with Figure 4, we set $q = q_{11} = q_{22}$ and compute the expected excess return on a consumption claim.[21] Figure 7 compares this consumption claim risk premium with parameter uncertainty to the one under anticipated utility, as in Figure 4. Here, we set the prior value for news accuracy $\chi = 0.91$ and consider two prior values, 0.80 and 0.95, for state persistence q. As shown in the figure, we find a significant increase in the magnitude of risk premium jump upon the arrival of bad news when parameter uncertainty is priced. This is consistent with the finding in Collin-Dufresne, Johannes, and Lochstoer (2016) that parameter uncertainty gives rise to a quantitatively significant effect on asset prices. We find that, compared with the value obtained under the anticipated utility approach, the risk premium value is at least twice as large with priced parameter uncertainty. Consistent with the findings discussed in the previous section, the quantitative difference is more apparent for a higher prior mean for q. As

[21] Note that for this example we use the solution of the more general model described above, where the agent believes that q_{11} and q_{22} can differ and her beliefs about these parameters follow separate Bayesian updating processes. However, to facilitate comparison with Figure 4, we exogenously set the path of q_{22} beliefs to be equal to that for q_{11}, even though in our example the agent should learn only q_{11}. In a model where the agent believes that a single parameter governs both transition probabilities, $q = q_{11} = q_{22}$, the parameter uncertainty and corresponding risk premia would likely be lower, though our qualitative results should still hold.

Figure 7: The role of priced parameter uncertainty in the behavior of the risk premium on a consumption claim



Notes: Risk premium is the expected excess return on consumption claim, which is derived under the case of preference for early resolution of uncertainty: $\gamma = 10$, $\psi = 1.5$. We solve an asset pricing problem that features fully rational sequential learning of unknown $q = q_{11} = q_{22}$. We compare this with the risk premium when parameter uncertainty is not priced. The numbers on the y-axis are in annualized percent terms.

in the anticipated utility case, we continue to see that bad news causes a larger jump in risk premia when it arrives after a longer expansion.

Simulated asset price moments. For the rest of the paper, we revert to the more general case of learning separate transition probabilities $q_{11} \neq q_{22}$. We simulate asset prices based on the posterior parameter estimates in Table 2 and the filtered states underlying Figure 5.[22] In Table 4, we present moments from the data and the 5^{th}, 50^{th},

[22]The particle filter produces a set of draws from the posterior distribution at each point in time. It does not yield draws of complete time series from the posterior distribution. Therefore, these draws can be used to produce moments for ex-ante returns that depend only on current observations, but they

and 95$^{\text{th}}$ percentiles of moments based on our model simulations. The table contains three cases. The first is our benchmark case and the second is a case without news shocks, which is equivalent to fixing $\chi = 0.5$ while keeping all other parameters the same. That is, the parameters used in both the "Benchmark" and $\chi = 0.5$ cases here come from the "Benchmark" posterior median estimates in Table 2. Both cases also use the same set of filtered states. Comparing these two cases isolates the effect of news shocks in this model conditional on our other benchmark parameters. In the third case, there are no news shocks and all parameters governing the data-generating process of consumption are estimated without the use of survey data. That is, we use the "Without surveys" posterior median estimates from Table 2. We include this case because it features an estimation method commonly used to obtain parameter and regime estimates in many macroeconomic and asset pricing contexts. Comparing this case with the benchmark illustrates the role of news shocks while also highlighting how the change in inference from the addition of survey data affects asset prices.

Focusing first on the average risk-free rate and ex-ante equity premium, we see that adding news has an effect, though not a large one. The greatest difference, in terms of these first two moments, comes from including survey data in the estimation. We see from the "Without surveys" case that applying the same set of asset pricing parameters while estimating the data-generating process of output using only real GDP growth data produces a higher average risk-free rate and a substantially lower average equity premium, particularly in the bad state. Given these asset pricing parameters, it is clear that the data-generating process estimated by using both actual real GDP growth and recession-probability forecasts does a better job of matching average risk-free rates and equity premia. Another interpretation of these results is that a researcher using a data-generating process estimated from only real GDP growth data would infer that an alternate set of asset pricing parameters is needed to match average risk-free rates and equity premia.

Next, we turn to volatilities of risk-free rates and equity premia. Now it becomes clear that the main effect of allowing for moderately informative news in this asset pricing model is the amplification of these volatilities. The within-state volatility of the risk-free rate grows by 5.4 and 4.5 times in the good and bad states, respectively, relative to the case where news is irrelevant, thus bringing the model much closer to matching the data in this dimension. In terms of the equity premium, we see a similar

cannot be used to compute ex-post returns that are a function of both current and lagged observations.

Table 4: Asset pricing moments

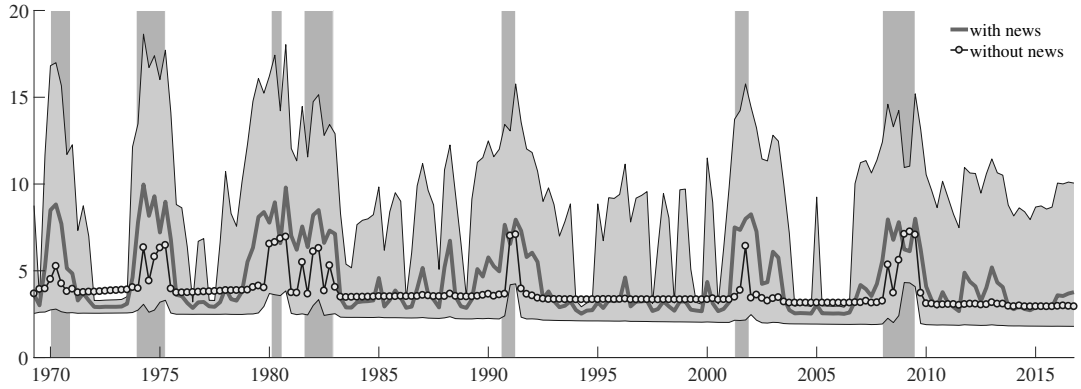
	Data	With surveys						Without surveys		
		Benchmark			$\chi = 0.5$					
		5%	50%	95%	5%	50%	95%	5%	50%	95%
					Average risk-free rate					
Full sample	0.93	3.22	3.57	3.78	3.25	3.36	3.47	3.79	3.84	3.89
$S_t = 1$	1.04	3.49	3.79	3.96	3.60	3.67	3.71	4.19	4.20	4.22
$S_t = 2$	0.44	1.41	2.33	2.94	1.31	1.64	2.02	3.29	3.30	3.33
					Average ex-ante equity premium					
Full sample	7.70	3.16	4.21	6.40	3.33	4.00	4.96	1.76	1.82	1.87
$S_t = 1$	7.18	2.75	3.48	5.17	2.91	3.41	4.25	1.40	1.44	1.46
$S_t = 2$	9.37	5.36	8.28	14.09	5.61	7.15	9.67	2.33	2.38	2.44
					Standard deviation of risk-free rate					
Full sample	1.96	0.93	1.08	1.31	0.60	0.74	0.86	0.43	0.44	0.46
$S_t = 1$	1.81	0.81	0.97	1.19	0.15	0.18	0.21	0.04	0.05	0.06
$S_t = 2$	2.36	0.39	0.72	0.96	0.05	0.16	0.30	0.04	0.05	0.06
					Standard deviation of ex-ante equity premium					
Full sample	2.35	1.67	2.85	5.17	0.99	1.42	2.06	0.46	0.48	0.51
$S_t = 1$	2.00	1.28	2.16	3.98	0.15	0.32	0.56	0.06	0.07	0.10
$S_t = 2$	2.59	1.05	2.61	5.76	0.39	0.81	1.86	0.12	0.15	0.19

Notes: We report the conditional averages of the log risk-free rate, log price-dividend ratio, and equity premium. We use the model-implied equity premium from Schorfheide, Song, and Yaron (2016). In the data, the conditional averages are computed based on expansion and recession states.

amplification of volatility, particularly in the good state where the within-state standard deviation grows by more than 6.7 times relative to the case without news. The model-implied conditional and unconditional equity premium standard deviations come close to matching the corresponding data moments.

This increase in volatilities of ex-ante returns follows from the effect of news on uncertainty that was highlighted in Section 2.2. In that section, for a simplified case with $q_{11} = q_{22}$, we show that news can increase uncertainty in switching states. Consistent with this intuition, we see that equity premia are higher in switching states, thus increasing equity premia volatility conditional on the growth state. The results from this simplified case also suggest that when news contradicts the current growth state, equity premia are at their highest when $\chi = q$. Since our posterior estimates feature a χ that is much closer to q_{11} than q_{22}, this suggests that news shocks should increase equity

Figure 8: Time series of the equity premium



Notes: We provide the model-implied median equity premium with the 90 percent credible intervals (shaded areas). We compare it to the median equity premium computed without news (circled line). We assume $\beta = 0.994$, $\gamma = 10$, $\psi = 1.5$, $\rho = 4.5$, and $\sigma_d = 4.81$. The numbers on the y-axis are in annualized percent terms.

premia volatility by more in the $S_t = 1$ growth state than in the $S_t = 2$ state. This is consistent with the results in Table 4.

Model-implied equity premium. Figure 8 provides the time series of the model-implied equity premium for our benchmark model and the case without news shocks. We find that allowing for news shocks produces an equity premium that fluctuates substantially more during recession and expansion periods. The additional volatility generated by news shocks is particularly prominent during the expansion of the late 1980s. Note also that our estimates indicate the arrival of bad news prior to recessions often leads the model-implied equity premium to begin rising several quarters prior to the onset of recessions. For example, the equity premium soared from (roughly) 3 percent in 1978:Q2 to 8 percent in 1979:Q2 and remained there until the recession started in 1980:Q2. The benchmark model with news also generates a spike in the equity premium in 1987:Q4 and 1988:Q1 that corresponds to the Black Monday crash. Even though we do not include asset prices directly in the estimation, the survey recession probabilities in these periods reflect an economic outlook darkened by the stock market crash. Our estimation interprets this as a heightened filtered probability of bad news about future GDP growth (as seen in Figure 5) that produces a high equity premium during this period despite high contemporaneous GDP growth. The arrival of this bad news at a time when the economy had been expanding for quite a while amplified

its effect on the equity premium, creating a Minsky moment.[23] The equity premium generated in the absence of news does not have these features.

5 Conclusion

Economic agents consider news as well as current fundamentals when forming beliefs about the future. In this paper, we study the channels through which news influences subjective beliefs of economic agents, with particular focus on their subjective uncertainty. In a two-state Markov-switching economy, we consider what happens when an agent receives news each period that reveals the next period's growth state with some error. We show that when news contradicts the agent's existing beliefs, the beliefs are revised toward a uniform distribution, raising subjective uncertainty. When the agent rationally learns model parameters, the response of uncertainty to news depends on past states as well as the current state. In this environment, parameter learning introduces additional state variables to the model that summarize information from past states, altering the effects of news on the agent's beliefs.

Given these properties of our model, we examine the extent to which news-driven uncertainty fluctuations can generate variation in asset prices. We show that the optimism built during a long period of economic boom, i.e., the increase in the believed probability of remaining in the good state, can trigger a Minsky moment when the agent receives news that the future may not be as bright as expected. The effect is greater when the expansion has been longer lasting, the prior beliefs about state persistence or news accuracy have been stronger, and the agent prefers early resolution of uncertainty.

We employ a novel filtering technique and estimate the model using recession probability forecasts from the Survey of Professional Forecasters. Our filtered probability of bad news correlates strongly with survey- and media-based measures and strongly negatively predicts one-step-ahead GDP growth after controlling for current GDP growth as implied by the model. Posterior beliefs vary significantly over time, with bad news frequently raising uncertainty during expansions.

Based on model estimates, we assess our model's quantitative implications for asset prices in a setting where the agent prefers an early resolution of uncertainty. We show

[23]Relative to a model with only Gaussian growth and news shocks, this increase in the equity premium is also amplified by the increase in uncertainty when bad news contradicts the current high-growth state.

that the main effect of allowing for informative news in this asset pricing model is the amplification of asset return volatilities. This is because news can increase uncertainty when it contradicts the agent's existing beliefs. In sum, news shocks greatly improve the model's ability to match both conditional and unconditional moments in the data. Finally, we identify historical periods when equity premia were elevated due to news.

References

Ai, H. 2010. "Information Quality and Long-Run Risk: Asset Pricing Implications." *The Journal of Finance* 65(4): 1333–1367.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and C. Vega. 2007. "Real-Time Price Discovery in Global Stock, Bond and Foreign Exchange Markets." *Journal of International Economics* 73(2): 251–277.

Bansal, R., and A. Yaron. 2004. "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles." *Journal of Finance* 59: 1481–1509.

Barsky, R., and E. Sims. 2011. "News Shocks and Business Cycles." *Journal of Monetary Economics* 58(3): 273–289.

Beaudry, P., and F. Portier. 2006. "Stock Prices, News, and Economic Fluctuations." *American Economic Review* 96(4): 1293–1307.

Beaudry, P., and F. Portier. 2014. "News-Driven Business Cycles: Insights and Challenges." *Journal of Economic Literature* 52(4): 993–1074.

Berger, D., I. Dew-Becker, and S. Giglio. 2017. "Uncertainty Shocks as Second-moment News Shocks." Manuscript.

Bianchi, F. 2016. "Methods for Measuring Expectations and Uncertainty in Markov-Switching Models." *Journal of Econometrics* 190(1): 79–99.

Bianchi, F., and L. Melosi. 2016. "Modeling the Evolution of Expectations and Uncertainty in General Equilibrium." *International Economic Review* 57(2): 717–756.

Boyd, J., J. Hu, and R. Jagannathan. 2005. "The Stock Market's Reaction to Unemployment News: Why Bad News Is Usually Good for Stocks." *Journal of Finance* 60: 649–672.

Carvalho, C., M. Johannes, H. Lopes, and N. Polson. 2010. "Particle Learning and Smoothing." *Statistical Science* 25(1): 88–106.

Chen, R., and J. Liu. 2000. "Mixture Kalman Filters." *Journal of the Royal Statistical Society Series B* 62: 493–508.

Collin-Dufresne, P., M. Johannes, and L. Lochstoer. 2016. "Parameter Learning in General Equilibrium: The Asset Pricing Implications." *American Economic Review* 106(3): 664–698.

Constantinides, G., and A. Ghosh. 2016. "What Information Drives Asset Prices?" Manuscript.

Epstein, L., and S.E. Zin. 1989. "Substitution, Risk Aversion and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework." *Econometrica* 57: 937–969.

Epstein, L. G., and M. Schneider. 2008. "Ambiguity, Information Quality, and Asset Pricing." *The Journal of Finance* 63(1): 197–228.

Faust, J., J.H. Rogers, S.Y.B. Wang, and J.H. Wright. 2007. "The High-Frequency Response of Exchange Rates and Interest Rates to Macroeconomic Announcements." *Journal of Monetary Economics* 54(4): 1051–1068.

Gürkaynak, R.S., B. Sack, and E. Swanson. 2005. "The Excess Sensitivity of Long-Term Interest Rates: Evidence and Implications for Macroeconomic Models." *American Economic Review* 95(1): 425–436.

Hirose, Y., and T. Kurozumi. 2012. "Identifying News Shocks with Forecast Data." CAMA Working Paper 1/2012.

Jaimovich, N., and S. Rebelo. 2009. "Can News about the Future Drive the Business Cycle?" *American Economic Review* 99(4): 1097–1118.

Johannes, M., L. Lochstoer, and Y. Mou. 2016. "Learning about Consumption Dynamics." *Journal of Finance* 71: 551–600.

Jurado, K., S. Ludvigson, and S. Ng. 2015. "Measuring Uncertainty." *American Economic Review* 105(3): 1177–1216.

Kozeniauskas, N., A. Orlik, and L. Veldkamp. 2016. "The Common Origin of Uncertainty Shocks." Working Paper 22384. National Bureau of Economic Research.

Kurmann, A., and C. Otrok. 2013. "News Shocks and the Slope of the Term Structure of Interest Rates." *American Economic Review* 103(6): 2612–2632.

Lucca, D., and E. Moench. 2015. "The Pre-FOMC Announcement Drift." *Journal of Finance* 70: 329–371.

Malkhozov, A., and A. Tamoni. 2015. "News Shocks and Asset Prices." LSE Research Online Documents on Economics 62004. London School of Economics and Political Science.

Matsumoto, A., P. Cova, M. Pisani, and A. Rebucci. 2011. "News Shocks and Asset Price Volatility in General Equilibrium." *Journal of Economic Dynamics and Control* 35(12): 2132–2149.

Milani, F., and A. Rajrhandari. 2012. "Observed Expectations, News Shocks, and the Business Cycle." Manuscript.

Miyamoto, W., and T.L. Nguyen. 2015. "News Shocks and Business Cycles: Evidence from Forecast Data." SSRN Scholarly Paper ID 2496739. Rochester, NY: Social Science Research Network.

Savor, P., and M. Wilson. 2013. "How Much Do Investors Care about Macroeconomic Risk? Evidence from Scheduled Economic Announcements." *Journal of Financial and Quantitative Analysis* 48: 343–375.

Schmitt-Grohe, S., and M. Uribe. 2012. "What's News in Business Cycles." *Econometrica* 80(6): 2733–2764.

Schorfheide, F., D. Song, and A. Yaron. 2016. "Identifying Long-Run Risks: A Bayesian Mixed-Frequency Approach." Manuscript.

Tang, J. 2017. "FOMC Communication and Interest Rate Sensitivity to News." Federal Reserve Bank of Boston Research Department Working Paper 17-12.

Veronesi, P. 1999. "Stock Market Overreaction to Bad News in Good Times: A Rational Expectations Equilibrium Model." *Review of Financial Studies* 12(5): 975–1007.

Veronesi, P. 2000. "How Does Information Quality Affect Stock Returns?" *The Journal of Finance* 55(2): 807–837.

West, K. D. 1988. "Dividend Innovations and Stock Price Volatility." *Econometrica* 56(1): 37–61.

Online Appendix:
News-Driven Uncertainty Fluctuations

Dongho Song and Jenny Tang

A Posterior of the Markov-Transition Probabilities

A.1 Without news

Prior. At $t = 0$, the agent is given an initial (potentially truncated) beta-distributed prior over each of these parameters, and thereafter she updates her beliefs sequentially upon observing the time series of realized regimes, S_t.

$$\begin{aligned} p(\Pi|\theta) &= p(q_{11})p(q_{22}) \\ &\propto q_{11}^{a_{1,0}-1}(1-q_{11})^{b_{1,0}-1}q_{22}^{a_{2,0}-1}(1-q_{22})^{b_{2,0}-1}, \end{aligned} \tag{A-1}$$

where $\{a_{1,0}, b_{1,0}, a_{2,0}, b_{2,0}\}$ can be interpreted as observations from a training sample of $a_{1,0} + b_{1,0} + a_{2,0} + b_{2,0} - 4$ periods that contained $a_{i,0} - 1$ observations of state i to state i transitions and $b_{i,0} - 1$ observations of state i to state $j \neq i$ transitions.

Likelihood. The binomial likelihood is

$$p(S^t|\Pi, \theta) = q_{11}^{a_{1,t}-a_{1,0}-1}(1-q_{11})^{b_{1,t}-b_{1,0}-1}q_{22}^{a_{2,t}-a_{2,0}-1}(1-q_{22})^{b_{2,t}-b_{2,0}-1}. \tag{A-2}$$

Observations. The standard Bayes' rule shows that the updating equations count the number of times state i has been followed by state i versus the number of times state i has been followed by state j.

$$\begin{aligned} a_{i,t} &= a_{i,0} + \#\text{ (state } i \text{ has been followed by state } i), \\ b_{i,t} &= b_{i,0} + \#\text{ (state } i \text{ has been followed by state } j). \end{aligned} \tag{A-3}$$

The law of motions for $a_{i,t}$ and $b_{i,t}$ are

$$\begin{aligned} a_{i,t+1} &= a_{i,t} + \mathbb{I}_{\{S_{t+1}=i\}}\mathbb{I}_{\{S_t=i\}} \\ b_{i,t+1} &= b_{i,t} + (1 - \mathbb{I}_{\{S_{t+1}=i\}})\mathbb{I}_{\{S_t=i\}}. \end{aligned} \tag{A-4}$$

Posterior. The prior beta distribution, coupled with the realization of regimes, leads to a conjugate prior, and so posterior beliefs are also beta distributed. Using this prior (A-1) and the binomial likelihood (A-2) associated with observations $\{a_{1,t}, b_{1,t}, a_{2,t}, b_{2,t}\}_{t=1}^{T}$, the posterior is

$$\begin{aligned} p(\Pi|y^t, S^t, \theta) &= \frac{p(y^t, S^t|\Pi, \theta)p(\Pi|\theta)}{p(y^t, S^t|\theta)} \\ &= \frac{p(y^t|S^t, \Pi, \theta)p(S^t|\Pi, \theta)p(\Pi|\theta)}{p(y^t|S^t, \theta)p(S^t|\theta)} \\ &= \frac{p(S^t|\Pi, \theta)p(\Pi|\theta)}{p(S^t|\theta)} \quad \text{since} \quad p(y^t|S^t, \Pi, \theta) = p(y^t|S^t, \theta) \\ &= \frac{q_{11}^{a_{1,t}-1}(1-q_{11})^{b_{1,t}-1}q_{22}^{a_{2,t}-1}(1-q_{22})^{b_{2,t}-1}}{B(a_{1,t}, b_{1,t})B(a_{2,t}, b_{2,t})}. \end{aligned} \tag{A-5}$$

From (A-1) and (A-2)

$$\begin{aligned} p(S^t|\theta) &= \int p(S^t|\Pi, \theta)p(\Pi|\theta)d\Pi \\ &= \int q_{11}^{a_{1,t}-1}(1-q_{11})^{b_{1,t}-1}dq_{11} \int q_{22}^{a_{2,t}-1}(1-q_{22})^{b_{2,t}-1}dq_{22} \\ &= B(a_{1,t}, b_{1,t})B(a_{2,t}, b_{2,t}). \end{aligned} \tag{A-6}$$

Posterior (A-5) is independent and equal to

$$\begin{aligned} p(\Pi|y^t, S^t, \theta) &= p(q_{11}|y^t, S^t, \theta) \cdot p(q_{22}|y^t, S^t, \theta) \\ &= p(q_{11}|S^t, \theta) \cdot p(q_{22}|S^t, \theta) \\ &= \frac{q_{11}^{a_{1,t}-1}(1-q_{11})^{b_{1,t}-1}}{B(a_{1,t}, b_{1,t})} \cdot \frac{q_{22}^{a_{2,t}-1}(1-q_{22})^{b_{2,t}-1}}{B(a_{2,t}, b_{2,t})}, \end{aligned} \tag{A-7}$$

and the posterior means are

$$\begin{aligned} E(q_{11}|y^t, S^t, \theta) &= \int q_{11} p(\Pi|y^t, S^t, \theta) d\Pi \\ &= \int \frac{q_{11}^{a_{1,t}}(1-q_{11})^{b_{1,t}-1}}{B(a_{1,t}, b_{1,t})} dq_{11} \int \frac{q_{22}^{a_{2,t}-1}(1-q_{22})^{b_{2,t}-1}}{B(a_{2,t}, b_{2,t})} dq_{22} \\ &= \frac{B(a_{1,t}+1, b_{1,t})}{B(a_{1,t}, b_{1,t})} \\ &= \frac{a_{1,t}}{a_{1,t}+b_{1,t}} \\ E(q_{22}|y^t, S^t, \theta) &= \frac{a_{2,t}}{a_{2,t}+b_{2,t}}, \end{aligned} \tag{A-8}$$

where the last two lines follow from the definition of beta and gamma distribution

- $p(y) = \frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)}$ for $0 \leq y \leq 1$
- $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$
- $\Gamma(\alpha+n) = \frac{(\alpha+n-1)!}{(\alpha-1)!}\Gamma(\alpha)$
- $E(y^n) = \frac{B(\alpha+n,\beta)}{B(\alpha,\beta)} = \frac{\Gamma(\alpha+n)\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+n)\Gamma(\alpha)}$.

A.2 With news

We can compute

$$p(S^t|\Pi, \theta)p(\Pi|\theta) = q_{11}^{a_{1,t}-1}(1-q_{11})^{b_{1,t}-1} q_{22}^{a_{2,t}-1}(1-q_{22})^{b_{2,t}-1} \tag{A-9}$$

and

$$\begin{aligned} p(n_t|S^t, \Pi, \theta) &= \sum_{S_{t+1}\in\{1,2\}} p(n_t|S_{t+1}, S^t, \Pi, \theta) p(S_{t+1}|S^t, \Pi, \theta) \\ &= \begin{cases} \chi q_{11} + (1-\chi)(1-q_{11}) = 1-\chi+(2\chi-1)q_{11}, & \text{if} \quad n_t = 1, S_t = 1 \\ (1-\chi)q_{11} + \chi(1-q_{11}) = \chi-(2\chi-1)q_{11}, & \text{if} \quad n_t = 2, S_t = 1 \end{cases} \end{aligned} \tag{A-10}$$

and from (A-9), (A-10),

$$p(n_t, S^t|\theta) = \int p(n_t|S^t, \Pi, \theta)p(S^t|\Pi, \theta)p(\Pi|\theta)d\Pi \tag{A-11}$$
$$= \begin{cases} \left(1-\chi+(2\chi-1)\frac{a_{1,t}}{a_{1,t}+b_{1,t}}\right)B(a_{1,t},b_{1,t})B(a_{2,t},b_{2,t}), & \text{if} \quad n_t=1, S_t=1 \\ \left(1-\chi+(2\chi-1)\frac{b_{1,t}}{a_{1,t}+b_{1,t}}\right)B(a_{1,t},b_{1,t})B(a_{2,t},b_{2,t}), & \text{if} \quad n_t=2, S_t=1. \end{cases}$$

From (A-11), we can deduce that the conditional posterior distribution of the Markov-switching transition probability matrix is

$$p(\Pi|n_t, y^t, S^t, \theta) = p(\Pi|n_t, S^t, \theta) = p(q_{11}|n_t, S^t, \theta)p(q_{22}|n_t, S^t, \theta) \tag{A-12}$$
$$= \begin{cases} \frac{\left(1-\chi+(2\chi-1)q_{11}\right)}{\left(1-\chi+(2\chi-1)\frac{a_{1,t}}{a_{1,t}+b_{1,t}}\right)} \cdot \frac{q_{11}^{a_{1,t}-1}(1-q_{11})^{b_{1,t}-1}}{B(a_{1,t},b_{1,t})} \cdot \frac{q_{22}^{a_{2,t}-1}(1-q_{22})^{b_{2,t}-1}}{B(a_{2,t},b_{2,t})}, & \text{if} \quad n_t^1=1, S_t=1 \\ \frac{q_{11}^{a_{1,t}-1}(1-q_{11})^{b_{1,t}-1}}{B(a_{1,t},b_{1,t})} \cdot \frac{\left(1-\chi+(2\chi-1)q_{22}\right)}{\left(1-\chi+(2\chi-1)\frac{a_{2,t}}{a_{2,t}+b_{2,t}}\right)} \cdot \frac{q_{22}^{a_{2,t}-1}(1-q_{22})^{b_{2,t}-1}}{B(a_{2,t},b_{2,t})}, & \text{if} \quad n_t^1=2, S_t=2 \\ \frac{\left(1-\chi+(2\chi-1)(1-q_{11})\right)}{\left(1-\chi+(2\chi-1)\frac{b_{1,t}}{a_{1,t}+b_{1,t}}\right)} \cdot \frac{q_{11}^{a_{1,t}-1}(1-q_{11})^{b_{1,t}-1}}{B(a_{1,t},b_{1,t})} \cdot \frac{q_{22}^{a_{2,t}-1}(1-q_{22})^{b_{2,t}-1}}{B(a_{2,t},b_{2,t})}, & \text{if} \quad n_t^1=2, S_t=1 \\ \frac{q_{11}^{a_{1,t}-1}(1-q_{11})^{b_{1,t}-1}}{B(a_{1,t},b_{1,t})} \cdot \frac{\left(1-\chi+(2\chi-1)(1-q_{22})\right)}{\left(1-\chi+(2\chi-1)\frac{b_{2,t}}{a_{2,t}+b_{2,t}}\right)} \cdot \frac{q_{22}^{a_{2,t}-1}(1-q_{22})^{b_{2,t}-1}}{B(a_{2,t},b_{2,t})}, & \text{if} \quad n_t^1=1, S_t=2. \end{cases}$$

The posterior means are

$$E(q_{11}|n_t^1, y^t, S^t, \theta) = \begin{cases} \frac{\chi(a_{1,t}+1)+(1-\chi)b_{1,t}}{\chi a_{1,t}+(1-\chi)b_{1,t}} \cdot \frac{a_{1,t}}{a_{1,t}+b_{1,t}+1}, & \text{if} \quad n_t^1=1, S_t=1 \\ \frac{(1-\chi)(a_{1,t}+1)+\chi b_{1,t}}{(1-\chi)a_{1,t}+\chi b_{1,t}} \cdot \frac{a_{1,t}}{a_{1,t}+b_{1,t}+1}, & \text{if} \quad n_t^1=2, S_t=1 \\ \frac{a_{1,t}}{a_{1,t}+b_{1,t}}, & \text{if} \quad n_t^1 \in \{1,2\}, S_t=2 \end{cases}$$

and

$$E(q_{22}|n_t^1, y^t, S^t, \theta) = \begin{cases} \frac{\chi(a_{2,t}+1)+(1-\chi)b_{2,t}}{\chi a_{2,t}+(1-\chi)b_{2,t}} \frac{a_{2,t}}{a_{2,t}+b_{2,t}+1}, & \text{if} \quad n_t^1=2, S_t=2 \\ \frac{(1-\chi)(a_{2,t}+1)+\chi b_{2,t}}{(1-\chi)a_{2,t}+\chi b_{2,t}} \frac{a_{2,t}}{a_{2,t}+b_{2,t}+1}, & \text{if} \quad n_t^1=1, S_t=2 \\ \frac{a_{2,t}}{a_{2,t}+b_{2,t}}, & \text{if} \quad n_t^1 \in \{1,2\}, S_t=1. \end{cases}$$

To calculate the posterior variances, we need $E(q_{ii}^2|n_t^1, y^t, S^t, \theta) - (E(q_{ii}|n_t^1, y^t, S^t, \theta))^2$

$$E(q_{11}^2|n_t^1, y^t, S^t, \theta) = \begin{cases} \frac{\chi(a_{1,t}+2)+(1-\chi)b_{1,t}}{\chi a_{1,t}+(1-\chi)b_{1,t}} \frac{a_{1,t}}{a_{1,t}+b_{1,t}+1} \frac{a_{1,t}+1}{a_{1,t}+b_{1,t}+2}, & \text{if} \quad n_t^1 = 1, S_t = 1 \\ \frac{(1-\chi)(a_{1,t}+2)+\chi b_{1,t}}{(1-\chi)a_{1,t}+\chi b_{1,t}} \frac{a_{1,t}}{a_{1,t}+b_{1,t}+1} \frac{a_{1,t}+1}{a_{1,t}+b_{1,t}+2}, & \text{if} \quad n_t^1 = 2, S_t = 1 \\ \frac{a_{1,t}}{a_{1,t}+b_{1,t}} \frac{a_{1,t}+1}{a_{1,t}+b_{1,t}+1}, & \text{if} \quad n_t^1 \in \{1,2\}, S_t = 2 \end{cases}$$

and

$$E(q_{22}^2|n_t^1, y^t, S^t, \theta) = \begin{cases} \frac{\chi(a_{2,t}+2)+(1-\chi)b_{2,t}}{\chi a_{2,t}+(1-\chi)b_{2,t}} \frac{a_{2,t}}{a_{2,t}+b_{2,t}+1} \frac{a_{2,t}+1}{a_{2,t}+b_{2,t}+2}, & \text{if} \quad n_t^1 = 2, S_t = 2 \\ \frac{(1-\chi)(a_{2,t}+2)+\chi b_{2,t}}{(1-\chi)a_{2,t}+\chi b_{2,t}} \frac{a_{2,t}}{a_{2,t}+b_{2,t}+1} \frac{a_{2,t}+1}{a_{2,t}+b_{2,t}+2}, & \text{if} \quad n_t^1 = 1, S_t = 2 \\ \frac{a_{2,t}}{a_{2,t}+b_{2,t}} \frac{a_{2,t}+1}{a_{2,t}+b_{2,t}+1}, & \text{if} \quad n_t^1 \in \{1,2\}, S_t = 1. \end{cases}$$

B The State Transition Probability

We compute the state transition probabilities

$$p(S_{t+1}, n_{t+1}|S_t, n_t, y_t, \Pi, \theta) = p(n_{t+1}|S_{t+1}, S_t, n_t, y_t, \Pi, \theta) p(S_{t+1}|S_t, n_t, y_t, \Pi, \theta). \quad \text{(A-13)}$$

The first component of (A-13) can be expressed by

$$\begin{aligned} p(n_{t+1}|S_{t+1}, S_t, n_t, y_t, \Pi, \theta) &= \sum_{S_{t+2}} p(n_{t+1}|S_{t+2}, S_{t+1}, S_t, n_t, y_t, \Pi, \theta) p(S_{t+2}|S_{t+1}, S_t, n_t, y_t, \Pi, \theta) \\ &= \sum_{S_{t+2}} p(n_{t+1}|S_{t+2}, \Pi, \theta) p(S_{t+2}|S_{t+1}, \Pi, \theta) \\ &= \begin{cases} (1-\chi) + (2\chi-1)q_{11}, & \text{if} \quad S_{t+1} = 1, n_{t+1} = 1 \\ \chi - (2\chi-1)q_{11}, & \text{if} \quad S_{t+1} = 1, n_{t+1} = 2 \\ \chi - (2\chi-1)q_{22}, & \text{if} \quad S_{t+1} = 2, n_{t+1} = 1 \\ (1-\chi) + (2\chi-1)q_{22}, & \text{if} \quad S_{t+1} = 2, n_{t+1} = 2. \end{cases} \end{aligned}$$

The second component of (A-13) can be expressed by

$$p(S_{t+1}|n_t, S^t, \Pi, \theta) = \frac{p(n_t|S_{t+1}, S^t, \Pi, \theta)p(S_{t+1}|S^t, \Pi, \theta)}{p(n_t|S_t, \Pi, \theta)} \tag{A-14}$$

$$= \begin{cases} \frac{\chi q_{11}}{(1-\chi)+(2\chi-1)q_{11}}, & \text{if} \quad n_t = 1, S_t = 1, S_{t+1} = 1 \\ \frac{(1-\chi)(1-q_{11})}{(1-\chi)+(2\chi-1)q_{11}}, & \text{if} \quad n_t = 1, S_t = 1, S_{t+1} = 2 \\ \frac{(1-\chi)q_{11}}{\chi-(2\chi-1)q_{11}}, & \text{if} \quad n_t = 2, S_t = 1, S_{t+1} = 1 \\ \frac{\chi(1-q_{11})}{\chi-(2\chi-1)q_{11}}, & \text{if} \quad n_t = 2, S_t = 1, S_{t+1} = 2, \end{cases}$$

$$= \begin{cases} \frac{\chi(1-q_{22})}{\chi-(2\chi-1)q_{22}}, & \text{if} \quad n_t = 1, S_t = 2, S_{t+1} = 1 \\ \frac{(1-\chi)q_{22}}{\chi-(2\chi-1)q_{22}}, & \text{if} \quad n_t = 1, S_t = 2, S_{t+1} = 2 \\ \frac{(1-\chi)(1-q_{22})}{1-\chi+(2\chi-1)q_{22}}, & \text{if} \quad n_t = 2, S_t = 2, S_{t+1} = 1 \\ \frac{\chi q_{22}}{1-\chi+(2\chi-1)q_{22}}, & \text{if} \quad n_t = 2, S_t = 2, S_{t+1} = 2. \end{cases}$$

Putting these two back into (A-13) gives a four-state transition matrix

$$p(S_{t+1}, n_{t+1}|S_t, n_t, y_t, \Pi, \theta) = \Pi'_B =$$

$$\begin{bmatrix} \frac{\chi q_{11}(1-\chi+(2\chi-1)q_{11})}{1-\chi+(2\chi-1)q_{11}} & \frac{\chi q_{11}(\chi-(2\chi-1)q_{11})}{1-\chi+(2\chi-1)q_{11}} & \frac{(1-\chi)(1-q_{11})(\chi-(2\chi-1)q_{22})}{1-\chi+(2\chi-1)q_{11}} & \frac{(1-\chi)(1-q_{11})(1-\chi+(2\chi-1)q_{22})}{1-\chi+(2\chi-1)q_{11}} \\ \frac{(1-\chi)q_{11}(1-\chi+(2\chi-1)q_{11})}{\chi-(2\chi-1)q_{11}} & \frac{(1-\chi)q_{11}(\chi-(2\chi-1)q_{11})}{\chi-(2\chi-1)q_{11}} & \frac{\chi(1-q_{11})(\chi-(2\chi-1)q_{22})}{\chi-(2\chi-1)q_{11}} & \frac{\chi(1-q_{11})(1-\chi+(2\chi-1)q_{22})}{\chi-(2\chi-1)q_{11}} \\ \frac{\chi(1-q_{22})(1-\chi+(2\chi-1)q_{11})}{\chi-(2\chi-1)q_{22}} & \frac{\chi(1-q_{22})(\chi-(2\chi-1)q_{11})}{\chi-(2\chi-1)q_{22}} & \frac{(1-\chi)q_{22}(\chi-(2\chi-1)q_{22})}{\chi-(2\chi-1)q_{22}} & \frac{(1-\chi)q_{22}(1-\chi+(2\chi-1)q_{22})}{\chi-(2\chi-1)q_{22}} \\ \frac{(1-\chi)(1-q_{22})(1-\chi+(2\chi-1)q_{11})}{1-\chi+(2\chi-1)q_{22}} & \frac{(1-\chi)(1-q_{22})(\chi-(2\chi-1)q_{11})}{1-\chi+(2\chi-1)q_{22}} & \frac{\chi q_{22}(\chi-(2\chi-1)q_{22})}{1-\chi+(2\chi-1)q_{22}} & \frac{\chi q_{22}(1-\chi+(2\chi-1)q_{22})}{1-\chi+(2\chi-1)q_{22}} \end{bmatrix}. \tag{A-15}$$

The rows and columns of the transition matrix correspond to $\kappa_t \in \{1, 2, 3, 4\}$, where

$$\kappa_t = \begin{cases} 1 & \text{if} \quad S_t = 1, n_t = 1 \\ 2 & \text{if} \quad S_t = 1, n_t = 2 \\ 3 & \text{if} \quad S_t = 2, n_t = 1 \\ 4 & \text{if} \quad S_t = 2, n_t = 2. \end{cases} \tag{A-16}$$

The rows are time t and columns are time $t + 1$, so that one can compute conditional means in time t by pre-multiplying outcomes with this matrix .

C Hamilton Filter

Define α_t as the 4×1 vector with i^{th} element equal to 1 when $\kappa_t = i$ and all other elements equal to 0 (refer to A-16). This implies that $E(\alpha_t|\alpha_{t-1}) = \Pi_B \alpha_{t-1}$, or

$$\alpha_t = \Pi_B \alpha_{t-1} + \xi_t, \tag{A-17}$$

where ξ_t is a disturbance vector uncorrelated with α_{t-1}. Note that the transition probabilities satisfy the condition $\sum_i \pi_{ij} = 1$ for $j = 1, ..., 4$. The disturbance term ξ_t can take one of a possible set of 4^2 discrete values and so is not normally distributed.

Note that $p(S_{t+1}|S_t, n_t, \Pi, \theta) = \Psi \Pi_B \alpha_t$ where $\Psi = \begin{bmatrix} 1 & 1 & 0 & 0 \end{bmatrix}$. We define

$$z_t = \Phi(w_t), \quad \text{where} \quad w_t \sim N(\xi_t, \sigma_z^2), \tag{A-18}$$

where

$$\xi_t = b_z + \Phi^{-1}\big(p(S_{t+1} = 1|S_t, n_t, \Pi, \theta)\big) = b_z + \Phi^{-1}\big(\Psi \Pi_B \alpha_t\big). \tag{A-19}$$

Then,

$$p(z_t|S_t, n_t, y_t, \Pi, \theta) = f_w(\Phi^{-1}(z_t); \xi_t, \sigma_z^2) \times \Big|\frac{1}{\phi(\Phi^{-1}(z_t))}\Big|, \tag{A-20}$$

where $f_w(\cdot)$ is a density function of $\mathcal{N}(\xi_t, \sigma_z^2)$, and $\phi(\cdot)$ is a pdf of the standard normal distribution.[24]

Define $x_t = [y_t, z_t]$ and $\mathcal{I}_t = x^t$. The Hamilton filter is an iterative algorithm for calculating the distribution of the state variable α_t.

$$\begin{aligned} \alpha_{t|t} &= E(\alpha_t|\mathcal{I}_t) \\ \alpha_{t|t-1} &= E(\alpha_t|\mathcal{I}_{t-1}), \end{aligned}$$

with the i^{th} element given by $Pr(\kappa_t = i|\mathcal{I}_t)$ and $Pr(\kappa_t = i|\mathcal{I}_{t-1})$, respectively. The

[24]Note that $\Phi(\Phi^{-1}(z_t)) = z_t$, thus

$$\begin{aligned} \frac{\partial \Phi(\Phi^{-1}(z_t))}{\partial z_t} = \phi(\Phi^{-1}(z_t)) \frac{\partial(\Phi^{-1}(z_t))}{\partial z_t} &= 1 \\ \frac{\partial(\Phi^{-1}(z_t))}{\partial z_t} &= \frac{1}{\phi(\Phi^{-1}(z_t))}. \end{aligned}$$

Hamilton filter comprises two recursive equations: (1) the prediction equation, defining $\alpha_{t|t-1}$, and (2) the updating equation, defining $\alpha_{t|t}$.

Prediction Equation. The Hamilton filter prediction equation is

$$\alpha_{t|t-1} = \Pi_B \alpha_{t-1}. \tag{A-21}$$

Updating Equation. The log-likelihood function of the observations x_t is given by

$$L = \sum_{t=1}^{T} \log p(x_t|\kappa_t, \mathcal{I}_{t-1}, \Pi, \theta), \tag{A-22}$$

where

$$\begin{aligned} p(x_t|\kappa_t, \mathcal{I}_{t-1}) &= p(y_t|\kappa_t, \mathcal{I}_{t-1}, \Pi, \theta) p(z_t|y_t, \kappa_t, \mathcal{I}_{t-1}, \Pi, \theta) \\ &= N\Big([\mu_1, \mu_1, \mu_2, \mu_2]\alpha_t, [\sigma_1^2, \sigma_1^2, \sigma_2^2, \sigma_2^2]\alpha_t\Big) f_w(\Phi^{-1}(z_t); \xi_t, \sigma_z^2) \Big| \frac{1}{\phi(\Phi^{-1}(z_t))} \Big|. \end{aligned}$$

The ratio of the two represents the optimal inference on κ_t based on $\mathcal{I}_t$:

$$Pr(\kappa_t = i|\mathcal{I}_t, \Pi, \theta) = \frac{p(x_t, \kappa_t = i|\mathcal{I}_{t-1}, \Pi, \theta)}{p(x_t|\mathcal{I}_{t-1}, \Pi, \theta)}. \tag{A-23}$$

Define v_t to be the $n \times 1$ vector with i^{th} element given by $p(x_t|\kappa_t = i, \mathcal{I}_{t-1}, \Pi, \theta)$. Then the marginal distribution is

$$p(x_t|\mathcal{I}_{t-1}, \Pi, \theta) = v_t' \alpha_{t|t-1}, \tag{A-24}$$

and $p(x_t, \kappa_t = i|\mathcal{I}_{t-1}, \Pi, \theta)$ is the i^{th} element of the $n \times 1$ vector

$$v_t \odot \alpha_{t|t-1}, \tag{A-25}$$

where $\odot$ is the element-wise multiplication operator. Thus, equation (A-23) can be written as the (nonlinear) updating equation for $\alpha_{t|t}$

$$\alpha_{t|t} = \frac{v_t \odot \alpha_{t|t-1}}{v_t' \alpha_{t|t-1}}. \tag{A-26}$$

D Parameter Learning (without News)

Parameters. The joint prior over the mean μ_i and the variance σ_i^2 is normal-inverse-gamma, where

$$\begin{aligned} p(\sigma_i^2|y^t, S^t, \Pi) &= IG(\frac{v_{i,t}}{2}, \frac{K_{i,t}}{2}) \\ p(\mu_i|\sigma_i^2, y^t, S^t, \Pi) &= N(m_{i,t}, \sigma_i^2 C_{i,t}). \end{aligned} \tag{A-27}$$

Since they are independent,

$$\begin{aligned} p(\mu|\sigma^2, y^t, S^t, \Pi) &= p(\mu_i|\sigma^2, y^t, S^t, \Pi)p(\mu_j|\sigma^2, y^t, S^t, \Pi) \\ p(\sigma^2|y^t, S^t, \Pi) &= p(\sigma_i^2|y^t, S^t, \Pi)p(\sigma_j^2|y^t, S^t, \Pi). \end{aligned} \tag{A-28}$$

These prior beliefs lead to posterior beliefs that are of the same form. The joint posterior distribution of the mean μ and the variance σ^2 can be factorized as

$$p(\mu, \sigma^2|y^{t+1}, S^{t+1}, \Pi) = p(\mu|\sigma^2, y^{t+1}, S^{t+1}, \Pi)p(\sigma^2|y^{t+1}, S^{t+1}, \Pi). \tag{A-29}$$

Note that

$$\begin{aligned} p(\mu|\sigma^2, y^{t+1}, S^{t+1}, \Pi) &\propto p(y_{t+1}, S_{t+1}|\mu, \sigma^2, y^t, S^t, \Pi)p(\mu|\sigma^2, y^t, S^t, \Pi) \\ &= p(y_{t+1}|\mu, \sigma^2, y^t, S^{t+1}, \Pi)p(S_{t+1}|\mu, \sigma^2, y^t, S^t, \Pi)p(\mu|\sigma^2, y^t, S^t, \Pi) \end{aligned} \tag{A-30}$$

and

$$\begin{aligned} p(\sigma^2|y^{t+1}, S^{t+1}, \Pi) &\propto p(y_{t+1}, S_{t+1}|\sigma^2, y^t, S^t, \Pi)p(\sigma^2|y^t, S^t, \Pi) \\ &= p(y_{t+1}|S_{t+1}, \sigma^2, y^t, S^t, \Pi)p(S_{t+1}|\sigma^2, y^t, S^t, \Pi)p(\sigma^2|y^t, S^t, \Pi). \end{aligned} \tag{A-31}$$

Assume that $S_{t+1} = i$. We re-express (A-30) as

$$p(\mu|\sigma^2, y^{t+1}, S^{t+1}) \propto p(y_{t+1}|S_{t+1}, \mu, \sigma^2, y^t, S^t, \Pi)p(S_{t+1}|\mu, \sigma^2, y^t, S^t, \Pi)p(\mu|\sigma^2, y^t, S^t, \Pi) \tag{A-32}$$

$$\begin{aligned}
&\propto p(y_{t+1}|S_{t+1}, \mu, \sigma^2, y^t, S^t, \Pi)p(\mu|\sigma^2, y^t, S^t, \Pi) \\
&= \Big(\sum_{i=1}^{2} \mathbb{I}_{\{S_{t+1}=i\}} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\Big\{ -\frac{1}{2\sigma_i^2}(y_{t+1} - \mu_i)^2 \Big\} \Big) \\
&\times \frac{1}{\sqrt{2\pi\sigma_i^2 C_{i,t}}} \exp\Big\{ -\frac{(\mu_i - m_{i,t})^2}{2\sigma^2 C_{i,t}} \Big\} \times \frac{1}{\sqrt{2\pi\sigma_j^2 C_{j,t}}} \exp\Big\{ -\frac{(\mu_j - m_{j,t})^2}{2\sigma^2 C_{j,t}} \Big\} \\
&\propto \frac{1}{\sqrt{2\pi\big(1 + \frac{1}{C_{i,t}}\big)^{-1}\sigma_i^2}} \exp\Big\{ -\frac{1}{2\sigma_i^2}\Big(1 + \frac{1}{C_{i,t}}\Big)\Big(\mu_i - \frac{y_{t+1} + \frac{m_{i,t}}{C_{i,t}}}{1 + \frac{1}{C_{i,t}}}\Big)^2 \Big\} \\
&= N(m_{i,t+1}, C_{i,t+1}\sigma_i^2).
\end{aligned}$$

We can deduce that, $\forall i = \{1, 2\}$,

$$\begin{aligned}
C_{i,t+1} &= \big(\mathbb{I}_{\{S_{t+1}=i\}} + \frac{1}{C_{i,t}}\big)^{-1} \\
\frac{1}{C_{i,t+1}} &= \frac{1}{C_{i,t}} + \mathbb{I}_{\{S_{t+1}=i\}} \\
m_{i,t+1} &= \big(\mathbb{I}_{\{S_{t+1}=i\}} + \frac{1}{C_{i,t}}\big)^{-1}\big(\mathbb{I}_{\{S_{t+1}=i\}} y_{t+1} + \frac{m_{i,t}}{C_{i,t}}\big) \\
&= C_{i,t+1}\big(y_{t+1}\mathbb{I}_{\{S_{t+1}=i\}} + \frac{m_{i,t}}{C_{i,t}}\big) \\
\frac{m_{i,t+1}}{C_{i,t+1}} &= \frac{m_{i,t}}{C_{i,t}} + y_{t+1}\mathbb{I}_{\{S_{t+1}=i\}}.
\end{aligned} \tag{A-33}$$

Analogously for σ^2,

$$\begin{aligned}
p(\sigma^2|y^{t+1}, S^{t+1}, \Pi) &\propto p(y_{t+1}|\sigma^2, y^t, S^{t+1}, \Pi)p(S_{t+1}|\sigma^2, y^t, S^t, \Pi)p(\sigma^2|y^t, S^t, \Pi) \quad \text{(A-34)} \\
&\propto \Big(\sum_{i=1}^{2} \mathbb{I}_{\{S_{t+1}=i\}} \frac{1}{\sqrt{2\pi(C_{i,t}+1)\sigma_i^2}} \exp\Big\{ -\frac{1}{2\sigma_i^2}\frac{(y_{t+1} - m_{i,t})^2}{(C_{i,t}+1)} \Big\} \Big) \\
&\times \frac{\big(\frac{K_{i,t}}{2}\big)^{(\frac{v_{i,t}}{2})}}{\Gamma(\frac{v_{i,t}}{2})}(\sigma_i^2)^{-\frac{v_{i,t}}{2}-1} e^{-\frac{K_{j,t}}{2\sigma_j^2}} \times \frac{\big(\frac{K_{j,t}}{2}\big)^{(\frac{v_{j,t}}{2})}}{\Gamma(\frac{v_{j,t}}{2})}(\sigma_j^2)^{-\frac{v_{j,t}}{2}-1} e^{-\frac{K_{j,t}}{2\sigma_j^2}} \\
&\propto (\sigma_i^2)^{-\frac{v_{i,t}+1}{2}-1} \exp\Big\{ -\frac{1}{\sigma_i^2}\frac{\big(K_{i,t} + \frac{(y_{t+1}-m_{i,t})^2}{(C_{i,t}+1)}\big)}{2} \Big\},
\end{aligned}$$

from which we can deduce that

$$\begin{aligned} v_{i,t+1} &= v_{i,t} + \mathbb{I}_{\{S_{t+1}=i\}} \\ K_{i,t+1} &= K_{i,t} + \frac{(y_{t+1} - m_{i,t})^2}{(C_{i,t}+1)} \mathbb{I}_{\{S_{t+1}=i\}}. \end{aligned} \tag{A-35}$$

Transition Probabilities. At $t = 0$, the agent is given an initial (potentially truncated) beta-distributed prior over each of these parameters, and thereafter she updates beliefs sequentially upon observing the time series of realized regimes, S_t. The prior beta distribution, coupled with the realization of regimes, leads to a conjugate prior, and so posterior beliefs are also beta distributed. The probability density function of the beta distribution is

$$p(\pi|a,b) = \frac{\pi^{a-1}(1-\pi)^{b-1}}{B(a,b)}, \tag{A-36}$$

where $B(a,b)$ is the beta function (a normalization constant). The parameters a and b govern the shape of the distribution. The expected value is

$$E(\pi|a,b) = \frac{a}{a+b}. \tag{A-37}$$

The standard Bayes' rule shows that the updating equations count the number of times state i has been followed by state i versus the number of times state i has been followed by state j. Given this sequential updating, we let the a and b parameters have a subscript for the relevant state (1 or 2) and a time subscript

$$\begin{aligned} a_{i,t} &= a_{i,0} + \#\text{ (state } i \text{ has been followed by state } i), \\ b_{i,t} &= b_{i,0} + \#\text{ (state } i \text{ has been followed by state } j). \end{aligned} \tag{A-38}$$

The law of motions for $a_{i,t}$ and $b_{i,t}$ are

$$\begin{aligned} a_{i,t+1} &= a_{i,t} + \mathbb{I}_{\{S_{t+1}=i\}}\mathbb{I}_{\{S_t=i\}} \\ b_{i,t+1} &= b_{i,t} + (1 - \mathbb{I}_{\{S_{t+1}=i\}})\mathbb{I}_{\{S_t=i\}}. \end{aligned} \tag{A-39}$$

We can deduce that posterior distribution of Π is

$$p(\Pi|y^{t+1}, S^{t+1}, \theta) = B(a_{1,t+1}, b_{1,t+1})B(a_{2,t+1}, b_{2,t+1}). \tag{A-40}$$

E Particle Learning

We collect two observations in

$$x_t = [y_t, z_t],$$

and the model parameters are collected in

$$\theta = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2), \quad \Pi = (q_{11}, q_{22}), \quad \Lambda = (\chi, \sigma_z^2).$$

Denote sufficient statistics for θ, Π, and Λ by $F_{\theta,t}$, $F_{\Pi,t}$, and $F_{\Lambda,t}$ respectively. Specifically,

$$F_{\theta,t} = \{m_{i,t}, C_{i,t}, v_{i,t}, K_{i,t}\}_{i=1}^{2}, \quad F_{\Pi,t} = \{a_{i,t}, b_{i,t}\}_{i=1}^{2}, \quad F_{\Lambda,t} = \{a_{z,t}, b_{z,t}, v_{z,t}, K_{z,t}\}. \tag{A-41}$$

Sufficient statistics imply that the full posterior distribution of the parameters conditional on the entire history of latent states and data takes a known functional form conditional on a vector of sufficient statistics

$$p(\theta, \Pi, \Lambda | x^t, \kappa^t) = p(\theta, \Pi, \Lambda | F_{\theta,t}, F_{\Pi,t}, F_{\Lambda,t}) = p(\theta | F_{\theta,t}) p(\Pi | F_{\Pi,t}) p(\Lambda | F_{\Lambda,t}). \tag{A-42}$$

Ultimately, we are interested in

$$p(\theta, \Pi, \Lambda, \kappa^t | x^t) = p(\theta, \Pi, \Lambda | \kappa^t, x^t) p(\kappa^t | x^t). \tag{A-43}$$

The idea of particle learning is to sample from $p(\theta, \Pi, \Lambda, F_{\theta,t}, F_{\Pi,t}, F_{\Lambda,t}, \kappa^t | x^t)$ and then from $p(\theta, \Pi, \Lambda, \kappa^t | x^t)$.

$$p(\theta, \Pi, \Lambda, F_{\theta,t}, F_{\Pi,t}, F_{\Lambda,t}, \kappa^t | x^t) = \underbrace{p(\theta, \Pi, \Lambda | F_{\theta,t}, F_{\Pi,t}, F_{\Lambda,t})}_{\text{(4) Drawing Parameters}} \times \underbrace{p(F_{\theta,t}, F_{\Pi,t}, F_{\Lambda,t}, \kappa^t | x^t)}_{\substack{\text{Propagating (2) State,} \\ \text{(3) Sufficient Statistics}}}. \tag{A-44}$$

The particle learning algorithm can be described through the following steps.

E.1 Algorithm

Assume at time t, we have particles $\left\{\kappa_t^{(k)}, \theta^{(k)}, \Pi^{(k)}, \Lambda^{(k)}, F_{\theta,t}^{(k)}, F_{\Pi,t}^{(k)}, F_{\Lambda,t}^{(k)}\right\}_{k=1}^{N}$.

1. **Resample Particles:**

   Resample $\left\{\kappa_t^{(k)}, \theta^{(k)}, \Pi^{(k)}, \Lambda^{(k)}, F_{\theta,t}^{(k)}, F_{\Pi,t}^{(k)}, F_{\Lambda,t}^{(k)}\right\}$ with weights $w_t^{(k)}$,

$$\begin{aligned} w_{t+1}^{(k)} \quad \propto \quad & \sum_{i=1}^{2} p\Bigg(x_{t+1}|\kappa_{t+1}=i, \left\{\kappa_t^{(k)}, \theta^{(k)}, \Pi^{(k)}, \Lambda^{(k)}, F_{\theta,t}^{(k)}, F_{\Pi,t}^{(k)}, F_{\Lambda,t}^{(k)}\right\} \Bigg) \\ & \times p\Bigg(\kappa_{t+1}=i| \left\{\kappa_t^{(k)}, \theta^{(k)}, \Pi^{(k)}, \Lambda^{(k)}, F_{\theta,t}^{(k)}, F_{\Pi,t}^{(k)}, F_{\Lambda,t}^{(k)}\right\} \Bigg). \end{aligned} \tag{A-45}$$

   Denote them by $\left\{\tilde{\kappa}_t^{(k)}, \tilde{\theta}^{(k)}, \tilde{\Pi}^{(k)}, \tilde{\Lambda}^{(k)}, \tilde{F}_{\theta,t}^{(k)}, \tilde{F}_{\Pi,t}^{(k)}, \tilde{F}_{\Lambda,t}^{(k)}\right\}_{k=1}^{N}$.

2. **Propagate State:**

$$\kappa_{t+1}^{(k)} \sim p\Bigg(\kappa_{t+1}|x_{t+1}, \left\{\tilde{\kappa}_t^{(k)}, \tilde{\theta}^{(k)}, \tilde{\Pi}^{(k)}, \tilde{\Lambda}^{(k)}, \tilde{F}_{\theta,t}^{(k)}, \tilde{F}_{\Pi,t}^{(k)}, \tilde{F}_{\chi,t}^{(k)}\right\} \Bigg).$$

3. **Propagate Sufficient Statistics:**

   (a) $F_{\theta,t+1} \sim \mathcal{F}(\tilde{F}_{\theta,t}^{(k)}, S_{t+1}^{(k)}, x_{t+1})$.

$$\begin{aligned} \frac{m_{i,t+1}^{(k)}}{C_{i,t+1}^{(k)}} &= \frac{\tilde{m}_{i,t}^{(k)}}{\tilde{C}_{i,t}^{(k)}} + y_{t+1}\mathbb{I}_{\left\{S_{t+1}^{(k)}=i\right\}} \\ \frac{1}{C_{i,t+1}^{(k)}} &= \frac{1}{\tilde{C}_{i,t}^{(k)}} + \mathbb{I}_{\left\{S_{t+1}^{(k)}=i\right\}} \\ v_{i,t+1}^{(k)} &= \tilde{v}_{i,t}^{(k)} + \mathbb{I}_{\left\{S_{t+1}^{(k)}=i\right\}} \\ K_{i,t+1}^{(k)} &= \tilde{K}_{i,t}^{(k)} + \frac{(y_{t+1} - \tilde{m}_{i,t}^{(k)})^2}{(\tilde{C}_{i,t}^{(k)} + 1)}\mathbb{I}_{\left\{S_{t+1}^{(k)}=i\right\}}. \end{aligned} \tag{A-46}$$

   (b) $F_{\Pi,t+1} \sim \mathcal{F}(\tilde{F}_{\Pi,t}^{(k)}, S_{t+1}^{(k)}, x_{t+1})$.

$$\begin{aligned} a_{i,t+1}^{(k)} &= \tilde{a}_{i,t}^{(k)} + \mathbb{I}_{\left\{S_{t+1}^{(k)}=i\right\}}\mathbb{I}_{\left\{S_t^{(k)}=i\right\}} \\ b_{i,t+1}^{(k)} &= \tilde{b}_{i,t}^{(k)} + (1 - \mathbb{I}_{\left\{S_{t+1}^{(k)}=i\right\}})\mathbb{I}_{\left\{S_t^{(k)}=i\right\}}. \end{aligned} \tag{A-47}$$

(c) $F_{\Lambda,t+1} \sim \mathcal{F}(\tilde{F}^{(k)}_{\Lambda,t}, S^{(k)}_{t+1}, n^{(k)}_{t+1}, x_{t+1})$.

$$\begin{aligned} a^{(k)}_{z,t+1} &= \tilde{a}^{(k)}_{z,t} + \mathbb{I}_{\left\{S^{(k)}_{t+1}=i\right\}} \mathbb{I}_{\left\{n^{(k)}_{t}=i\right\}} \\ b^{(k)}_{z,t+1} &= \tilde{b}^{(k)}_{z,t} + (1 - \mathbb{I}_{\left\{S^{(k)}_{t+1}=i\right\}}) \mathbb{I}_{\left\{n^{(k)}_{t}=i\right\}} \\ v^{(k)}_{z,t+1} &= \tilde{v}^{(k)}_{z,t} + 1 \\ K^{(k)}_{z,t+1} &= \tilde{K}^{(k)}_{z,t} + (\Phi^{-1}(z_t) - \tilde{\xi}_t)^2 \end{aligned} \tag{A-48}$$

where $\tilde{\xi}_t$ is provided in (A-19).

Note that $\mathcal{F}$s are analytically known.

4. **Draw Parameters:**

(a) $\theta^{(k)} \sim p(\theta|F_{\theta,t+1})$.

$$\begin{aligned} \sigma^{2,(k)}_i &\sim IG(\frac{v^{(k)}_{i,t+1}}{2}, \frac{K^{(k)}_{i,t+1}}{2}) \\ \mu^{(k)}_i &\sim N(m^{(k)}_{i,t+1}, \sigma^{2,(k)}_i C^{(k)}_{i,t+1}). \end{aligned} \tag{A-49}$$

(b) $\Pi^{(k)} \sim p(\Pi|F_{\Pi,t+1})$.

$$\begin{aligned} q^{(k)}_{11} &\sim B(a^{(k)}_{1,t+1}, b^{(k)}_{1,t+1}) \\ q^{(k)}_{22} &\sim B(a^{(k)}_{2,t+1}, b^{(k)}_{2,t+1}). \end{aligned} \tag{A-50}$$

(c) $\Lambda^{(k)} \sim p(\Lambda|F_{\Lambda,t+1})$.

$$\chi^{(k)} \sim B(a^{(k)}_{z,t+1}, b^{(k)}_{z,t+1}) \tag{A-51}$$

$$\sigma^{2,(k)}_z \sim IG(\frac{v^{(k)}_{z,t+1}}{2}, \frac{K^{(k)}_{z,t+1}}{2}). \tag{A-52}$$

To initialize the algorithm, we provide the priors in Table A-1. The length of the prior training sample (prior precision), T^{prior}, is set to 10 years.

Table A-1: Priors

Parameter	5%	50%	95%	Sufficient Statistics	Distribution
μ_1	-1.0	1.0	3.0	$m_{1,0}, C_{1,0}$	$N(1, 0.5)$, $IG(1, 2/3)$
σ_1^2	0.12	0.40	1.50	$v_{1,0}, K_{1,0}$	$G(5, 5)$, $G(5, 2)$
μ_2	-2.0	0.0	2.0	$m_{2,0}, C_{2,0}$	$N(0, 0.5)$, $IG(1, 2/3)$
σ_2^2	0.12	0.40	1.50	$v_{2,0}, K_{2,0}$	$G(5, 5)$, $G(5, 2)$
q_{11}	0.64	0.80	0.93	$a_{1,0}, b_{1,0}$	$Mult(T^{\text{prior}}, 0.8, 0.2)$
q_{22}	0.64	0.80	0.93	$a_{2,0}, b_{2,0}$	$Mult(T^{\text{prior}}, 0.8, 0.2)$
χ	0.64	0.80	0.93	$a_{z,0}, b_{z,0}$	$Mult(T^{\text{prior}}, 0.8, 0.2)$
σ_z^2	0.06	0.24	0.96	$v_{z,0}, K_{z,0}$	$G(5, 5)$, $G(3, 2)$

Notes: N, G, IG, $Mult$ are normal distribution, gamma distribution, inverse gamma distribution, and multinomial distribution, respectively.

F Asset Pricing Solution

F.1 Sufficient statistics

In our asset pricing model, we assume that the agent's information sets at time t are $I_t = \{\kappa^t, y^t, \theta\}$ and she rationally learns the transition probabilities Π using observed data. To conserve notation, we now add to the parameter vector θ parameters governing the prior beliefs about Π as well as preference and dividend process parameters from the asset pricing model. We follow Collin-Dufresne, Johannes, and Lochstoer (2016) in mapping the sufficient statistics of the Bayesian learning problem in Appendix A to an alternative set of statistics that yields a more convenient solution method. More specifically, we map $\{a_{1,t}, b_{1,t}, a_{2,t}, b_{2,t}\}$ to $X_t \equiv \{\lambda_{1,t}, \tau_{1,t}, \lambda_{2,t}, \tau_{2,t}\}$ as follows:

$$\begin{aligned}
\tau_{1,t} &= a_{1,t} - a_{1,0} + b_{1,t} - b_{1,0} \\
\lambda_{1,t} &= \frac{a_{1,t}}{a_{1,t} + b_{1,t}} \\
\tau_{2,t} &= a_{2,t} - a_{2,0} + b_{2,t} - b_{2,0} \\
\lambda_{2,t} &= \frac{a_{2,t}}{a_{2,t} + b_{2,t}}.
\end{aligned}$$

Note that the expressions for the posterior estimates of Π given in Appendix A can be rewritten in terms of $\{\lambda_{1,t}, \tau_{1,t}, \lambda_{2,t}, \tau_{2,t}, \theta\}$. Further, note that X_t follows a process that

depends only on its own lag as well as $\{\kappa_t, \kappa_{t-1}\}$:

$$\begin{aligned}
\lambda_{1,t} &= \lambda_1(\kappa_t, \kappa_{t-1}, X_{t-1}, \theta) = \lambda_{1,t-1} + \frac{\mathbb{I}\{S_t = 1\} - \lambda_{1,t-1}}{a_{1,0} + b_{1,0} + \tau_{1,t-1} + 1}, \\
\tau_{1,t} &= \tau_1(\kappa_t, \kappa_{t-1}, X_{t-1}, \theta) = \tau_{1,t-1} + \mathbb{I}\{S_{t-1} = 1\} \\
\lambda_{2,t} &= \lambda_2(\kappa_t, \kappa_{t-1}, X_{t-1}, \theta) = \lambda_{2,t-1} + \frac{\mathbb{I}\{S_t = 2\} - \lambda_{2,t-1}}{a_{2,0} + b_{2,0} + \tau_{2,t-1} + 1} \\
\tau_{2,t} &= \tau_1(\kappa_t, \kappa_{t-1}, X_{t-1}\theta) = \tau_{2,t-1} + \mathbb{I}\{S_{t-1} = 2\}.
\end{aligned} \tag{A-53}$$

Lastly, this set of sufficient statistics simplifies the solution method, because now only two of the sufficient statistics grow without bound ($\tau_{1,t}$ and $\tau_{2,t}$) while $\lambda_{1,t}, \lambda_{2,t} \in [0, 1]$.

For the solution below, the agent's belief about the probability of κ_{t+1} conditional on I_t is given by:

$$\begin{aligned}
p(\kappa_{t+1}|I_t) &= p(\kappa_{t+1}|\kappa_t, y_t, X_t, \theta) \\
&= \int p(\kappa_{t+1}|\kappa_t, X_t, y_t, \Pi, \theta) p(\Pi|\kappa_t, y_t, X_t, \theta) d\Pi,
\end{aligned}$$

where $p(\kappa_{t+1}|\kappa_t, X_t, y_t, \Pi, \theta) = \Pi_B^T$ as defined in (A-15), and $p(\Pi|\kappa_t, y_t, X_t, \theta) = p(\Pi|n_t, S^t, \theta)$ as given in (A-12). Solving this integration yields:

$$\Pi_{B,t}^T \equiv p(\kappa_{t+1}|\kappa_t, X_t, y_t, \Pi, \theta) = \tag{A-54}$$

$$\begin{bmatrix}
\frac{\lambda_1\chi[\chi\lambda_1' + (1-\chi)(1-\lambda_1')]}{\lambda_1\chi + (1-\lambda_1)(1-\chi)} & \frac{\lambda_1\chi[(1-\chi)\lambda_1' + \chi(1-\lambda_1')]}{\lambda_1\chi + (1-\lambda_1)(1-\chi)} & \frac{(1-\lambda_1)(1-\chi)[\chi(1-\lambda_2') + (1-\chi)\lambda_2']}{\lambda_1\chi + (1-\lambda_1)(1-\chi)} & \frac{(1-\lambda_1)(1-\chi)[(1-\chi)(1-\lambda_2') + \chi\lambda_2']}{\lambda_1\chi + (1-\lambda_1)(1-\chi)} \\
\frac{\lambda_1(1-\chi)[\chi\lambda_1' + (1-\chi)(1-\lambda_1')]}{\lambda_1(1-\chi) + (1-\lambda_1)\chi} & \frac{\lambda_1(1-\chi)[(1-\chi)\lambda_1' + \chi(1-\lambda_1')]}{\lambda_1(1-\chi) + (1-\lambda_1)\chi} & \frac{(1-\lambda_1)\chi[\chi(1-\lambda_2') + (1-\chi)\lambda_2']}{\lambda_1\chi + (1-\lambda_1)(1-\chi)} & \frac{(1-\lambda_1)\chi[(1-\chi)(1-\lambda_2') + \chi\lambda_2']}{\lambda_1\chi + (1-\lambda_1)(1-\chi)} \\
\frac{(1-\lambda_2)\chi[\chi\lambda_1' + (1-\chi)(1-\lambda_1')]}{(1-\lambda_2)\chi + \lambda_2(1-\chi)} & \frac{(1-\lambda_2)\chi[(1-\chi)\lambda_1' + \chi(1-\lambda_1')]}{(1-\lambda_2)\chi + \lambda_2(1-\chi)} & \frac{\lambda_2(1-\chi)[\chi(1-\lambda_2') + (1-\chi)\lambda_2']}{(1-\lambda_2)\chi + \lambda_2(1-\chi)} & \frac{\lambda_2(1-\chi)[(1-\chi)(1-\lambda_2') + \chi\lambda_2']}{(1-\lambda_2)\chi + \lambda_2(1-\chi)} \\
\frac{(1-\lambda_2)(1-\chi)[\chi\lambda_1' + (1-\chi)(1-\lambda_1')]}{(1-\lambda_2)(1-\chi) + \lambda_2\chi} & \frac{(1-\lambda_2)(1-\chi)[(1-\chi)\lambda_1' + \chi(1-\lambda_1')]}{(1-\lambda_2)(1-\chi) + \lambda_2\chi} & \frac{\lambda_2\chi[\chi(1-\lambda_2') + (1-\chi)\lambda_2']}{(1-\lambda_2)(1-\chi) + \lambda_2\chi} & \frac{\lambda_2\chi[(1-\chi)(1-\lambda_2') + \chi\lambda_2']}{(1-\lambda_2)(1-\chi) + \lambda_2\chi}
\end{bmatrix},$$

where we use the shorthand notation $\lambda_i \equiv \lambda_{i,t}$ and $\lambda_i' \equiv \lambda_{i,t+1} = \lambda_i(\kappa_{t+1}, \kappa_t, X_t, \theta)$ for $i \in \{1, 2\}$. Note that in this section, the superscript ^{T} will be used to denote matrix transposition, while $'$ will be used to denote next-period variables.

F.2 Wealth-consumption and price-dividend ratios

Agents have Epstein-Zin preferences given by

$$V_t = \left\{ (1-\beta)\, C_t^{1-\frac{1}{\psi}} + \beta \left[E_t V_{t+1}^{1-\gamma} \right]^{\frac{1-\frac{1}{\psi}}{1-\gamma}} \right\}^{\frac{1}{1-\frac{1}{\psi}}}.$$

The consumption process is governed by the exogenous process for output as follows:

$$\Delta c_t = \Delta y_t = \mu_{s_t} + \sigma_{s_t} \varepsilon_t.$$

Lastly, we assume that the dividend process is given by

$$\Delta d_{t+1} = \bar{\mu} + \rho \left(\Delta c_{t+1} - \bar{\mu} \right) + \sigma_d \eta_{t+1}.$$

The solution for the wealth-consumption ratio in this setting is

$$\begin{aligned} PC_t^{\alpha} &= E[\beta^{\alpha} e^{(1-\gamma)\Delta c_{t+1}} (PC_{t+1}+1)^{\alpha} | I_t] \\ &= E[\beta^{\alpha} E\left[e^{(1-\gamma)\Delta c_{t+1}} | I_t, S_{t+1} \right] (PC_{t+1}+1)^{\alpha} | I_t] \\ &= E[\beta^{\alpha} e^{(1-\gamma)\tilde{\mu}_{\kappa_{t+1}} + \frac{1}{2}(1-\gamma)^2 \tilde{\sigma}^2_{\kappa_{t+1}}} (PC_{t+1}+1)^{\alpha} | I_t], \end{aligned}$$

where $\alpha \equiv \frac{1-\gamma}{1-1/\psi}$ and

$$\begin{aligned} &\tilde{\mu}_1 = \tilde{\mu}_2 = \mu_1,\ \tilde{\mu}_3 = \tilde{\mu}_4 = \mu_2, \\ &\tilde{\sigma}_1 = \tilde{\sigma}_2 = \mu_1,\ \tilde{\sigma}_3 = \tilde{\sigma}_4 = \mu_2. \end{aligned}$$

Similarly, the price-dividend ratio is

$$PD_t = E\left[\beta^{\alpha} e^{(\rho-\gamma)\tilde{\mu}_{\kappa_{t+1}} + \frac{1}{2}(\rho-\gamma)^2 \tilde{\sigma}^2_{\kappa_{t+1}} + (1-\rho)\bar{\mu}(q_{11}, q_{22})} \left(\frac{PC_{t+1}+1}{PC_t} \right)^{\alpha-1} (PD_{t+1}+1) \middle| I_t \right],$$

where $\bar{\mu}\left(q_{11}, q_{22}\right)$ is the long-run mean of consumption growth as a function of $\{q_{11}, q_{22}\}$.

Since the exogenous states are Markov and the sufficient statistics of the Bayesian learning problem satisfy $X_{t+1} = F\left(S_{t+1}, S_t, X_t\right)$ for a function F that summarizes (A-53), the equilibrium wealth-consumption and price-dividend ratios can be written as functions of state variables and known parameters $\{\kappa_t, X_t, \theta\}$ that satisfy the following

recursions:

$$\begin{aligned} PC\left(\kappa_t, X_t, \theta\right)^{\alpha} &= E\left[\beta^{\alpha} e^{(1-\gamma)\tilde{\mu}_{\kappa_{t+1}} + \frac{1}{2}(1-\gamma)^2 \tilde{\sigma}^2_{\kappa_{t+1}}} \left(PC\left(\kappa_{t+1}, X_{t+1}, \theta\right)+1\right)^{\alpha} | I_t\right] \\ PD\left(\kappa_t, X_t, \theta\right) &= E\left[\beta^{\alpha} e^{(\rho-\gamma)\tilde{\mu}_{\kappa_{t+1}} + \frac{1}{2}(\rho-\gamma)^2 \tilde{\sigma}^2_{\kappa_{t+1}} + (1-\rho)\bar{\mu}(q_{11}, q_{22})}\right. \\ &\quad \left. \times \left(\frac{PC\left(\kappa_{t+1}, X_{t+1}, \theta\right)+1}{PC\left(\kappa_t, X_t, \theta\right)}\right)^{\alpha-1} \left(PD\left(\kappa_{t+1}, X_{t+1}, \theta\right)+1\right) \middle| I_t \right]. \end{aligned}$$

We follow Collin-Dufresne, Johannes, and Lochstoer (2016) in solving this recursion by using the property that beliefs about $\{q_{11}, q_{22}\}$ converge to the true values under Bayesian learning as $\{\tau_{1,t}, \tau_{2,t}\}$ grow. This allows us to iterate back from the known parameters solution. Details on this procedure are provided below:

F.3 Full information case (both q_{11}, q_{22} known)

Although news has no impact on parameter learning in this case, n_t still serves as a signal about S_{t+1}. The relevant state transition matrix is the one given in (A-15). In this case, we solve for the equilibrium wealth-consumption and price-dividend ratios on a three-dimensional grid of values for $\{\kappa, q_{11}, q_{22}\}$.

F.3.1 Wealth-consumption ratio

The equilibrium condition for the wealth-consumption ratio when $\{q_{11}, q_{22}\}$ are known is

$$\mathbf{PC}^{\circ\alpha} = \Pi_B^T \left(\beta^{\alpha} e^{(1-\gamma)\tilde{\mu}' + \frac{1}{2}(1-\gamma)^2 (\tilde{\sigma}')^{\circ 2}} \odot (\mathbf{PC}' + 1)^{\circ\alpha}\right),$$

where bolded variables now denote 4×1 vectors indexed by κ and dependence on $\{\Pi, \theta\}$ is suppressed for brevity. The symbol $\circ$ is used to denote element-wise exponentiation. For fixed values of $\{\Pi, \theta\}$, this is simply a system of four nonlinear equations that can be solved numerically.

F.3.2 Price-dividend ratio

Similarly, the price-dividend ratio can be obtained by solving the following system of equations:

$$\mathbf{PD} = \Pi_B^T \left(\beta^\alpha e^{(\rho-\gamma)\tilde{\mu}' + \frac{1}{2}(\rho-\gamma)^2 (\tilde{\sigma}')^{\circ 2} + (1-\rho)\bar{\mu}} \odot \left((\mathbf{PC}' + 1) \oslash \mathbf{PC} \right)^{\circ(\alpha-1)} \odot (\mathbf{PD}' + 1) \right),$$

where $\oslash$ denotes element-wise division.

F.4 Case with q_{ii} known and q_{jj} unknown

In this case, we need to carry a subset of the sufficient statistics $X_{j,t} \equiv \{\lambda_{j,t}, \tau_{j,t}\}$. The relevant state transition matrix is denoted by $\Pi^T_{B,t,ii} \equiv p(\kappa_{t+1}|\kappa_t, X_{j,t}, y_t, q_{ii}, \theta)$. One can show that this matrix is equal to (A-54) with the substitution $\lambda_i = q_{ii}$ in all time periods. For both sets of $\{i, j\}$, we solve for the equilibrium wealth-consumption and price-dividend ratios on a four-dimensional grid of values for $\{\kappa, \lambda_j, \tau_j, q_{ii}\}$, where $\tau_{j,t}$ is truncated at a large value $\tau_{j,T}$.

F.4.1 Wealth-consumption ratio

Now, the wealth-consumption ratio is no longer stationary, because $\tau_{j,t}$ grows without bound. It must satisfy the recursion

$$PC\left(\kappa_t, X_{j,t}, q_{ii}, \theta\right)^\alpha = E\left[\beta^\alpha e^{(1-\gamma)\tilde{\mu}_{\kappa_{t+1}} + \frac{1}{2}(1-\gamma)^2 \tilde{\sigma}^2_{\kappa_{t+1}}} \left(PC\left(\kappa_{t+1}, X_{j,t+1}, q_{ii}, \theta\right) + 1 \right)^\alpha | I_t \right].$$

To obtain the PC solution for a given q_{ii}, we start with the approximation that $PC\left(\kappa_T, X_{j,T}, q_{ii}, \theta\right) = PC\left(\kappa_T, \Pi, \theta\right)$ for some large $\tau_{j,T}$, because this relationship becomes exact as $\tau_{j,T} \to \infty$. We then iterate back by repeating the following two steps:

1. If $S_t = j$, $\lambda_{j,t+1}$ and $\tau_{j,t+1}$ will update as above for both $S_{t+1} = 1$ and $S_{t+1} = 2$, and any observation n_t would be informative, so the values of PC for κ_t corresponding to $S_t = j$ are a function of the values of PC with τ_j incremented by one. That is,

we iterate τ_j back one step as follows:

$$\begin{bmatrix} PC(\kappa_t = 2j-1, \lambda_{j,t}, \tau_{j,t})^\alpha \\ PC(\kappa_t = 2j, \lambda_{j,t}, \tau_{j,t})^\alpha \end{bmatrix} = \beta^\alpha \Pi^T_{B,t,ii,[2j-1:2j]} \times \left(e^{(1-\gamma)\tilde{\mu}' + \frac{1}{2}(1-\gamma)^2 (\tilde{\sigma}')^{\circ 2}} \right.$$
$$\left. \odot \left(\begin{bmatrix} PC(\kappa_t = 1, \lambda_{j,t+1}, \tau_{j,t}+1) \\ PC(\kappa_t = 2, \lambda_{j,t+1}, \tau_{j,t}+1) \\ PC(\kappa_t = 3, \lambda_{j,t+1}, \tau_{j,t}+1) \\ PC(\kappa_t = 4, \lambda_{j,t+1}, \tau_{j,t}+1) \end{bmatrix} + 1 \right)^{\circ\alpha} \right),$$

where we use the fact that $\kappa_t = 2(S_t - 1) + n_t$, and $\Pi^T_{B,t,ii,[2j-1:2j]}$ is the 2×4 submatrix formed by rows $2j-1$ and $2j$ of matrix $\Pi^T_{B,t,ii}$. The dependence on $\{q_{ii}, \theta\}$ is suppressed here for brevity.

2. If $S_t = i$, nothing can be learned about q_{jj} in the next period regardless of the value of S_{t+1} or n_t. In other words, $\lambda_{j,t+1} = \lambda_{j,t}$ and $\tau_{j,t+1} = \tau_{j,t}$ for $S_t = i$. Thus, given the values from the left-hand side of the previous step, we can directly solve for PC for the remaining values of κ_t using this system of equations:

$$\begin{bmatrix} PC(\kappa_t = 2i-1, \lambda_{j,t}, \tau_{j,t})^\alpha \\ PC(\kappa_t = 2i, \lambda_{j,t}, \tau_{j,t})^\alpha \end{bmatrix} = \beta^\alpha \Pi^T_{B,t,ii,[2i-1:2i]} \left(e^{(1-\gamma)\tilde{\mu}' + \frac{1}{2}(1-\gamma)^2 (\tilde{\sigma}')^{\circ 2}} \odot (\mathbf{PC}' + 1)^{\circ\alpha} \right),$$

where $\mathbf{PC}'$ is formed by stacking

$$\begin{bmatrix} PC(\kappa_t = 2i-1, \lambda_{j,t}, \tau_{j,t}) \\ PC(\kappa_t = 2i, \lambda_{j,t}, \tau_{j,t}) \end{bmatrix}, \quad \begin{bmatrix} PC(\kappa_t = 2j-1, \lambda_{j,t}, \tau_{j,t}) \\ PC(\kappa_t = 2j, \lambda_{j,t}, \tau_{j,t}) \end{bmatrix}$$

appropriately given the values of i and j.

F.4.2 Price-dividend ratio

Once we have the solution for PC, the solution for the price-dividend ratio procedes analogously based on the recursion

$$PD(\kappa_t, X_{j,t}, q_{ii}, \theta) = E\left[\beta^\alpha e^{(\rho-\gamma)\tilde{\mu}_{\kappa_{t+1}} + \frac{1}{2}(\rho-\gamma)^2 \tilde{\sigma}^2_{\kappa_{t+1}} + (1-\rho)\bar{\mu}(q_{ii}, q_{jj})} \right.$$
$$\left. \times \left(\frac{PC(\kappa_{t+1}, X_{j,t+1}, q_{ii}, \theta) + 1}{PC(\kappa_t, X_{j,t+1}, q_{ii}, \theta)} \right)^{\alpha-1} (PD(\kappa_{t+1}, X_{j,t+1}, q_{ii}, \theta) + 1) \middle| I_t \right].$$

We start with an analogous approximation that $PD\left(\kappa_T, X_{j,T}, q_{ii}, \theta\right) = PD\left(\kappa_T, \Pi, \theta\right)$ for some large $\tau_{j,T}$ and then iterate back by repeating the same two steps:

1. If $S_t = j$, we again iterate τ_j back one step as follows:

$$\begin{aligned}
&\begin{bmatrix} PD(\kappa_t = 2j-1, \lambda_{j,t}, \tau_{j,t}) \\ PD(\kappa_t = 2j, \lambda_{j,t}, \tau_{j,t}) \end{bmatrix} \\
= \;& \beta^{\alpha} \Pi^{T}_{B,t,ii,[2j-1:2j]} \times \Big(e^{(\rho-\gamma)\tilde{\mu}' + \frac{1}{2}(\rho-\gamma)^2 (\tilde{\sigma}')^{\circ 2} + (1-\rho)\bar{\mu}(q_{ii}, \hat{q}_{jj})} \\
& \odot \left((\mathbf{PC}' + 1) \oslash \mathbf{PC} \right)^{\circ(\alpha-1)} \odot \left(\begin{bmatrix} PD(\kappa_t = 1, \lambda_{j,t+1}, \tau_{j,t} + 1) \\ PD(\kappa_t = 2, \lambda_{j,t+1}, \tau_{j,t} + 1) \\ PD(\kappa_t = 3, \lambda_{j,t+1}, \tau_{j,t} + 1) \\ PD(\kappa_t = 4, \lambda_{j,t+1}, \tau_{j,t} + 1) \end{bmatrix} + 1 \right) \Big).
\end{aligned}$$

   In this expression, $\hat{q}_{jj}$ denotes $E\left[q_{jj} | I_t\right]$.

2. If $S_t = i$, we use the result from step 1 to solve for PD for the remaining values of κ_t and the same $\{\lambda_{j,t}, \tau_{j,t}\}$.

$$\begin{aligned}
\begin{bmatrix} PD(\kappa_t = 2i-1, \lambda_{j,t}, \tau_{j,t}) \\ PD(\kappa_t = 2i, \lambda_{j,t}, \tau_{j,t}) \end{bmatrix} \;=\; & \beta^{\alpha} \Pi^{T}_{B,t,ii,[2i-1:2i]} \\
& \times \Big(e^{(\rho-\gamma)\tilde{\mu}' + \frac{1}{2}(\rho-\gamma)^2 (\tilde{\sigma}')^{\circ 2} + (1-\rho)\bar{\mu}(q_{ii}, \hat{q}_{jj})} \\
& \odot \left((\mathbf{PC}' + 1) \oslash \mathbf{PC} \right)^{\circ(\alpha-1)} \odot (\mathbf{PD}' + 1) \Big),
\end{aligned}$$

   where $\mathbf{PD}'$ is formed by stacking

$$\begin{bmatrix} PD(\kappa_t = 2i-1, \lambda_{j,t}, \tau_{j,t}) \\ PD(\kappa_t = 2i, \lambda_{j,t}, \tau_{j,t}) \end{bmatrix}, \quad \begin{bmatrix} PC(\kappa_t = 2j-1, \lambda_{j,t}, \tau_{j,t}) \\ PC(\kappa_t = 2j, \lambda_{j,t}, \tau_{j,t}) \end{bmatrix}$$

   appropriately given the values of i and j.

F.5 Both $\{q_{11}, q_{22}\}$ unknown

Using the solutions for both cases of a single known transition probability, we can obtain the solution for the case in which both transition probabilities are unknown. The transition matrix is now the one given in (A-54). We solve for the equilibrium wealth-consumption and price-dividend ratios on a five-dimensional grid of values for

$\{\kappa, \lambda_1, \tau_1, \lambda_2, \tau_2\}$, where τ_1 and τ_2 are both truncated at large values $\tau_{1,T}$ and $\tau_{2,T}$, respectively.

F.5.1 Wealth-consumption ratio

We start with the following two approximations:

$$\begin{aligned} PC\left(\kappa_T, X_T, \theta\right) &= PC\left(\kappa_T, \lambda_{2,t}, \tau_{2,t}, q_{11}, \theta\right) \text{ for some large } \tau_{1,T} \text{ and all } \left(\lambda_{2,t}, \tau_{2,t}\right) \\ PC\left(\kappa_T, X_T, \theta\right) &= PC\left(\kappa_T, \lambda_{1,t}, \tau_{1,t}, q_{22}, \theta\right) \text{ for some large } \tau_{2,T} \text{ and all } \left(\lambda_{1,t}, \tau_{1,t}\right). \end{aligned}$$

From here, we iterate back in both the τ_1 and τ_2 dimensions alternatingly as follows:

1. If $S_t = 1$, $\lambda_{1,t+1}$ and $\tau_{1,t+1}$ will update as above for both $S_{t+1} = 1$ and $S_{t+1} = 2$. This allows us to iterate back one step for τ_1:

$$\begin{aligned} &\begin{bmatrix} PC(\kappa_t = 1, \lambda_{1,t}, \tau_{1,t}, \lambda_{2,t}, \tau_{2,t})^\alpha \\ PC(\kappa_t = 2, \lambda_{1,t}, \tau_{1,t}, \lambda_{2,t}, \tau_{2,t})^\alpha \end{bmatrix} \\ = \;& \beta^\alpha \Pi^T_{B,t,ii,[1:2]} \times \Bigg(e^{(1-\gamma)\tilde{\mu}' + \frac{1}{2}(1-\gamma)^2 (\tilde{\sigma}')^{\circ 2}} \\ & \odot \left(\begin{bmatrix} PC(\kappa_t = 1, \lambda_{1,t+1}, \tau_{1,t} + 1, \lambda_{2,t}, \tau_{2,t}) \\ PC(\kappa_t = 2, \lambda_{1,t+1}, \tau_{1,t} + 1, \lambda_{2,t}, \tau_{2,t}) \\ PC(\kappa_t = 3, \lambda_{1,t+1}, \tau_{1,t} + 1, \lambda_{2,t}, \tau_{2,t}) \\ PC(\kappa_t = 4, \lambda_{1,t+1}, \tau_{1,t} + 1, \lambda_{2,t}, \tau_{2,t}) \end{bmatrix} + 1 \right)^{\circ \alpha} \Bigg). \end{aligned}$$

2. If $S_t = 2$, we analogously iterate back one step for τ_2:

$$\begin{aligned} &\begin{bmatrix} PC(\kappa_t = 3, \lambda_{1,t}, \tau_{1,t}, \lambda_{2,t}, \tau_{2,t})^\alpha \\ PC(\kappa_t = 4, \lambda_{1,t}, \tau_{1,t}, \lambda_{2,t}, \tau_{2,t})^\alpha \end{bmatrix} \\ = \;& \beta^\alpha \Pi^T_{B,t,ii,[3:4]} \times \Bigg(e^{(1-\gamma)\tilde{\mu}' + \frac{1}{2}(1-\gamma)^2 (\tilde{\sigma}')^{\circ 2}} \\ & \odot \left(\begin{bmatrix} PC(\kappa_t = 1, \lambda_{1,t}, \tau_{1,t}, \lambda_{2,t+1}, \tau_{2,t} + 1) \\ PC(\kappa_t = 2, \lambda_{1,t}, \tau_{1,t}, \lambda_{2,t+1}, \tau_{2,t} + 1) \\ PC(\kappa_t = 3, \lambda_{1,t}, \tau_{1,t}, \lambda_{2,t+1}, \tau_{2,t} + 1) \\ PC(\kappa_t = 4, \lambda_{1,t}, \tau_{1,t}, \lambda_{2,t+1}, \tau_{2,t} + 1) \end{bmatrix} + 1 \right)^{\circ \alpha} \Bigg). \end{aligned}$$

F.5.2 Price-dividend ratio

The price-dividend ratio solution is obtained analogously again starting with the following two approximations:

$$\begin{aligned} PD\left(\kappa_T, X_T, \theta\right) &= PD\left(\kappa_T, \lambda_{2,t}, \tau_{2,t}, q_{11}, \theta\right) \text{ for some large } \tau_{1,T} \text{ and all } \left(\lambda_{2,t}, \tau_{2,t}\right) \\ PD\left(\kappa_T, X_T, \theta\right) &= PD\left(\kappa_T, \lambda_{1,t}, \tau_{1,t}, q_{22}, \theta\right) \text{ for some large } \tau_{2,T} \text{ and all } \left(\lambda_{1,t}, \tau_{1,t}\right). \end{aligned}$$

From here, we iterate back in both the τ_1 and τ_2 dimensions alternatingly as follows:

1. If $S_t = 1$, we iterate back one step for τ_1:

$$\begin{aligned} &\begin{bmatrix} PD(\kappa_t = 1, \lambda_{1,t}, \tau_{1,t}, \lambda_{2,t}, \tau_{2,t}) \\ PD(\kappa_t = 2, \lambda_{1,t}, \tau_{1,t}, \lambda_{2,t}, \tau_{2,t}) \end{bmatrix} \\ = \;& \beta^{\alpha} \Pi^{T}_{B,t,ii,[1:2]} \times \left(e^{(\rho-\gamma)\tilde{\mu}' + \frac{1}{2}(\rho-\gamma)^2 (\tilde{\sigma}')^{\circ 2} + (1-\rho)\bar{\mu}(q_{ii}, \hat{q}_{jj})} \odot \left((\mathbf{PC}' + 1) \oslash \mathbf{PC} \right)^{\circ(\alpha-1)} \right. \\ & \left. \odot \left(\begin{bmatrix} PD(\kappa_t = 1, \lambda_{1,t+1}, \tau_{1,t} + 1, \lambda_{2,t}, \tau_{2,t}) \\ PD(\kappa_t = 2, \lambda_{1,t+1}, \tau_{1,t} + 1, \lambda_{2,t}, \tau_{2,t}) \\ PD(\kappa_t = 3, \lambda_{1,t+1}, \tau_{1,t} + 1, \lambda_{2,t}, \tau_{2,t}) \\ PD(\kappa_t = 4, \lambda_{1,t+1}, \tau_{1,t} + 1, \lambda_{2,t}, \tau_{2,t}) \end{bmatrix} + 1 \right) \right). \end{aligned}$$

2. If $S_t = 2$, we iterate back one step for τ_2:

$$\begin{aligned} &\begin{bmatrix} PD(\kappa_t = 3, \lambda_{1,t}, \tau_{1,t}, \lambda_{2,t}, \tau_{2,t}) \\ PD(\kappa_t = 4, \lambda_{1,t}, \tau_{1,t}, \lambda_{2,t}, \tau_{2,t}) \end{bmatrix} \\ = \;& \beta^{\alpha} \Pi^{T}_{B,t,ii,[3:4]} \times \left(e^{(\rho-\gamma)\tilde{\mu}' + \frac{1}{2}(\rho-\gamma)^2 (\tilde{\sigma}')^{\circ 2} + (1-\rho)\bar{\mu}(q_{ii}, \hat{q}_{jj})} \odot \left((\mathbf{PC}' + 1) \oslash \mathbf{PC} \right)^{\circ(\alpha-1)} \right. \\ & \left. \odot \left(\begin{bmatrix} PD(\kappa_t = 1, \lambda_{1,t}, \tau_{1,t}, \lambda_{2,t+1}, \tau_{2,t} + 1) \\ PD(\kappa_t = 2, \lambda_{1,t}, \tau_{1,t}, \lambda_{2,t+1}, \tau_{2,t} + 1) \\ PD(\kappa_t = 3, \lambda_{1,t}, \tau_{1,t}, \lambda_{2,t+1}, \tau_{2,t} + 1) \\ PD(\kappa_t = 4, \lambda_{1,t}, \tau_{1,t}, \lambda_{2,t+1}, \tau_{2,t} + 1) \end{bmatrix} + 1 \right) \right). \end{aligned}$$

F.6 Asset pricing moments

Using our five-dimensional solutions for PC and PD, we can obtain analogous solution arrays for the risk-free rate and ex-ante equity risk premium on the same grid of $\{\kappa, \lambda_1, \tau_1, \lambda_2, \tau_2\}$ values using the following equations.

- Risk-free rate:
  The risk-free rate in the model is

$$\begin{aligned}\frac{1}{1+R_t^f} &= E_t\left[M_{t+1}\right] \\ &= \beta^\theta E_t\left[e^{-\gamma\tilde{\mu}_{\kappa_{t+1}}+\frac{1}{2}\gamma^2\tilde{\sigma}^2_{\kappa_{t+1}}}\left(\frac{PC_{t+1}+1}{PC_t}\right)^{\theta-1}\right].\end{aligned}$$

  Therefore, the log risk-free rate can be approximated by:

$$rf_t \approx \ln\left(1+R_t^f\right) = -\ln\left(\beta^\theta E_t\left[\exp\left\{-\gamma\tilde{\mu}_{\kappa_{t+1}}+\frac{1}{2}\gamma^2\tilde{\sigma}^2_{\kappa_{t+1}}\right\}\left(\frac{PC_{t+1}+1}{PC_t}\right)^{\theta-1}\right]\right).$$

- Risk premium on equity:
  The expected return on equity is

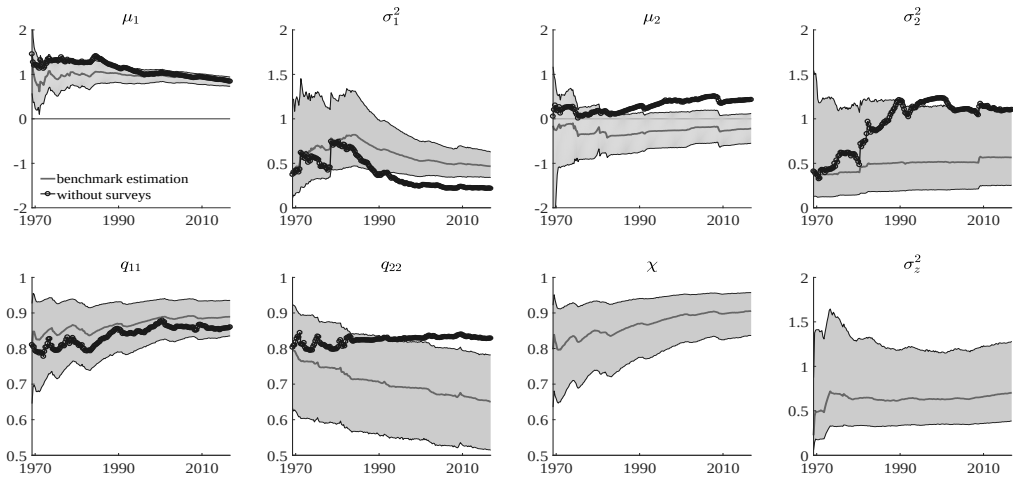
$$\begin{aligned}E_t\left[1+R_{e,t+1}\right] &= E_t\left[\frac{PD_{t+1}+1}{PD_t}\frac{D_{t+1}}{D_t}\right] \\ &= E_t\left[e^{(1-\rho)\bar{\mu}_{t+1}+\rho\tilde{\mu}_{\kappa_{t+1}}+\frac{1}{2}\rho^2\tilde{\sigma}^2_{\kappa_{t+1}}+\frac{1}{2}\sigma_d^2}\frac{PD_{t+1}+1}{PD_t}\right],\end{aligned}$$

  where $\bar{\mu}$ varies over time as beliefs about $\{q_{11}, q_{22}\}$ evolve. Therefore, the log ex-ante expected excess return on equity (adjusted by the Jensen's inequality term) is

$$\begin{aligned}&\ln\left(\frac{E_t\left[1+R_{e,t+1}\right]}{1+R_t^f}\right)-\frac{1}{2}\sigma_d^2 \\ &= \ln\left(E_t\left[\exp\left\{(1-\rho)\,\bar{\mu}_{t+1}+\rho\tilde{\mu}_{\kappa_{t+1}}+\frac{1}{2}\rho^2\tilde{\sigma}^2_{\kappa_{t+1}}\right\}\frac{PD_{t+1}+1}{PD_t}\right]\right)-rf_t.\end{aligned}$$

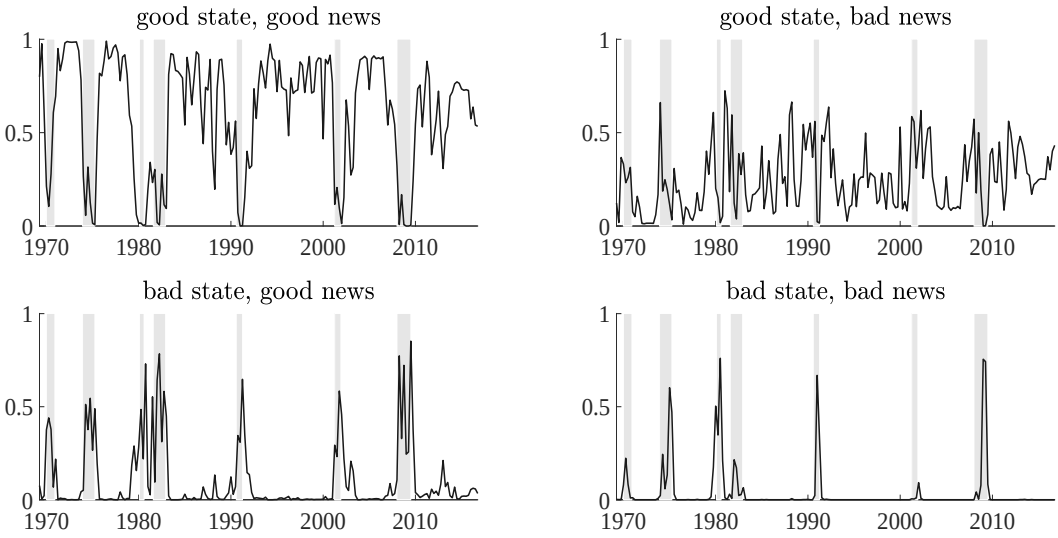
G Supplementary Figures

Figure A-1: Filtered estimates: Parameters



Notes: Gray solid lines are posterior median values that are overlaid with the 90 percent credible interval (gray shaded areas). Black circled lines are posterior median values obtained without survey forecasts. To deal with label switching problem, we impose that $\mu_1 > \mu_2$.

Figure A-2: Probability of each regime



Notes: Black solid lines are posterior median values.