News-Driven Uncertainty Fluctuations

Dongho Song
Boston College

Jenny Tang*
Federal Reserve Bank of Boston

This Version: April 30, 2017

VERY PRELIMINARY

Abstract

We present a two-state Markov-switching growth model in which economic agents learn true parameters from economic data and noisy news about the future state. News has two effects in this environment. First, it increases the accuracy of forecasts and parameter estimates. However, in contrast to standard Gaussian setups, our discrete-state environment allows news to increase forecast uncertainty when it contradicts existing beliefs. We develop a novel filtering technique that uses both actual GDP growth and recession probability forecasts from the Survey of Professional Forecasters to solve this sequential learning problem. Using data starting in 1969, we obtain the full joint posterior distribution of the model primitives including the filtered distribution of news. We find that the filtered probability of news correlates strongly with measures of consumer sentiment. Furthermore, we find that after each of the recent recession, our estimates of forecast uncertainty tend to remain elevated due to bad news. We embed the estimated time series of posterior beliefs into an asset pricing model in which agents prefer an early resolution of uncertainty and find that news shocks can generate sizable fluctuations in both risk-free rates and equity risk premia, thus providing a better fit to the data. This effect is particularly strong in expansion states where including news shocks in the model increases risk premium volatility by 5.5 times.

Key words: Bayesian updating, parameter learning, sequential particle learning, news shocks, survey forecasts, uncertainty, recursive utility, risk premium.

*Correspondence: Department of Economics, Boston College, 140 Commonwealth Avenue, Maloney Hall 382, Chestnut Hill, MA 02467. Email: dongho.song@bc.edu. Research Department, Federal Reserve Bank of Boston, 600 Atlantic Avenue, T-9 Boston, MA 02210. Email: jenny.tang@bos.frb.org. The views expressed in this paper are those of the authors and do not necessarily represent the views of the Federal Reserve Bank of Boston or the Federal Reserve System.

1 Introduction

The macroeconomics literature has often appealed to news shocks to explain business cycles and stock price movements.[1] The literature has mainly focused on the ability of news shocks to drive variations in macroeconomic and financial variables through its effect on the expected future path of the economy. However, not much work has been done so far on investigating two other channels through which news can have an effect. One is the direct impact of news about the future of the economy on higher-order moments of agents' beliefs, in particular their subjective uncertainty. Secondly, news can also have an effect through its impact on agents' beliefs about parameters governing the economy when they are not assumed to be known.

We contribute to this literature by considering a two-state Markov-switching growth model in which all of the channels mentioned above are present. The agents in our model are uncertain about both model parameters and future states and update their beliefs rationally using Bayes' rule as new data arrive.

To the extent that news shocks contain information about future states, news generally improves state prediction and parameter learning accuracy. However, due to the discrete nature of the state space, news can increase ex-ante subjective forecast uncertainty in contrast to many existing studies where agents update beliefs about Gaussian processes. We show that this tends to occur in our setting when the news realization contradicts agents' beliefs about the future state that were formed prior to observing the news. In such cases, agents' posterior beliefs about future states become more uniform after the news is observed. This generates state-dependent effects of news on uncertainty. For reasonable parameter values, we show that bad news increases uncertainty in good states while it has the opposite effect in bad states.

We use our model to obtain filtered estimates of the joint distribution of historical regimes and news shocks from the data. In order to do so, we make an important technical innovation. We develop a novel filtering technique that allows us to use both real output growth rates and recession probability forecasts collected by the Survey of Professional Forecasters from 1969:Q1 to 2016:Q3 to solve the sequential learning problem and identify the news component. Specifically, we extend the particle learning algorithm developed by Carvalho, Johannes, Lopes, and Polson (2010). By assuming that news is included in the professionals' information set, we are able to obtain the full joint posterior distribution of the model primitives (unknowns) including the filtered distribution of news. As evidence of the external validity of our exercise, we show that the filtered probability of bad news correlates strongly negatively with aggregate measures of consumer sentiment.

[1] See Beaudry and Portier (2006), Jaimovich and Rebelo (2009), Barsky and Sims (2011), Schmitt-Grohe and Uribe (2012), and Malkhozov and Tamoni (2015).

Our main empirical findings are that subjective beliefs vary significantly over time and, in particular, there is a pattern of higher 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.

To quantify the role of news in the economy, we turn to an asset pricing application in which agents prefer an early resolution of uncertainty and must also learn state transition probabilities. Johannes, Lochstoer, and Mou (2016) show that revisions in beliefs constitute permanent shocks to investors' expectations and therefore become a quantitatively important source of subjective long-run risks.[2] Adding news to this model changes both the risks coming from parameter uncertainty, by aiding in learning about unknown parameters, as well as the risks coming from regime uncertainty through the mechanisms described above. We extend the solution method used in Collin-Dufresne, Johannes, and Lochstoer (2016) to accommodate an additional news shock and obtain a fully nonlinear solution of the model by iterating the policy functions for both the wealth-consumption and price-dividend ratios backwards from a distant endpoint that is approximated by an economy in which all parameters are known.

Within this environment, we find that news has little effect on the average risk-free rate and equity risk premium due to the two opposing effects of higher average uncertainty about the future state but faster parameter learning. However, we find that news can greatly increase the volatility in risk-free rate and equity risk premia, particularly in expansions, thus improving the model's ability to match the data in terms of second moments.

Literature Review. This paper is at the intersection of several literatures. First, it's related to papers that study news shocks theoretically and empirically in the context of 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)) as well as asset pricing (such as Beaudry and Portier (2006), Kurmann and Otrok (2013), and Malkhozov and Tamoni (2015)). Unlike most of this existing work which focuses on the effect of news on mean forecasts of future economy activity, we focus on the additional effects that news shocks have on uncertainty and parameter learning.

In terms of papers that estimate models witih news shocks, Milani and Rajrhandari (2012), Hirose and Kurozumi (2012), and Miyamoto and Nguyen (2015) are the most closely related to ours since these papers all incorporate forecast data into the estimation of news shocks. Nonetheless, there remain two key differences. First, we develop new econometric methods that allow us to include recession probability forecasts in the estimation while these papers focus on only mean forecasts. Furthermore, these studies estimate DSGE models with shocks

[2]See also Bansal and Yaron (2004) and Collin-Dufresne, Johannes, and Lochstoer (2016).

to news about particular structural shocks (e.g., TFP, demand, monetary policy, etc.). In contrast, we remain agnostic about the types of structural shocks that news may pertain to and therefore our measure of news summarizes all information that is imbedded in the recession probability forecasts about the future growth regime and that is not already captured by the current regime.

Another related literature is the one on uncertainty fluctuations. There is a large literature showing that uncertainty has negative effects on the economy through a variety of mechanisms. Some examples include Bloom (2009), Bachmann, Elstner, and Sims (2013), Fernández-Villaverde, Guerrón-Quintana, Rubio-Ramírez, and Uribe (2011), Basu and Bundick (2017), Gilchrist, Sim, and Zakrajek (2014), and Leduc and Liu (2016). The role of news in our model has the potential of introducing an additional dimension to the relationship between uncertainty and economic outcomes due to its state-dependent relationship with uncertainty. To the extent that uncertainty is bad for the economy, the negative effects of bad news can be amplified by the uncertainty channel when the economy is currently in an expansion state. On the other hand, when the economy is already in a bad state, it's good news that raises uncertainty about the future. This higher uncertainty works in opposition to the expansionary effect that good news has through raising the mean forecast of future growth. Therefore, combined with the results of this literature, our model suggests the possibility of a stronger reduced-form relationship between uncertainty and negative economic outcomes during expansions rather than recessions.

Our paper is also very 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 have stochastic volatility. The recent work by Berger, Dew-Becker, and Giglio (2017) uses an estimated 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 in our model fluctuates with realizations of actual economic data and news shocks about future regimes.

On the methodological side, Bianchi (2016) and Bianchi and Melosi (2016) provide analytical characterizations of uncertainty in Markov-switching models under both perfect information and settings where agents 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 effects of allowing for news on asset pricing moments.

Our asset-pricing application is related to a number of works featuring asset-pricing models where 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 regimes. Furthermore, these papers do not use forecast data in their estimations while we do.

In Section 2, we first 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 Model

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. 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 definition of χ and σ_z^2 will be provided shortly.

We introduce news that provides information about 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.[3] We assume that n_t can predict the future state with probability χ. The accuracy of the prediction increases with χ.

The synchronization of two discrete Markov chains, one for the fundamental variable in (1) and the other for the news component in (2), results in the following four-state Markov chain process:

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

The associated companion form transition matrix Π_B is provided in (A-15). We use interchangeably κ_t and $\{S_t, n_t\}$ to indicate the current regime.

2.2 News Implications

We demonstrate how the prediction of the future state and the estimates of model parameters are affected by the accuracy of news. For ease of illustration, we proceed with the assumption that model parameters and the history of states are known at time t. The predictive distribution of the future state conditional on news can be expressed as

$$p(S_{t+1} = 1 | n_t, S^t, \Pi, \theta) \quad = \quad \begin{cases} \frac{\chi q_{11}}{(1-\chi)+(2\chi-1)q_{11}}, & \text{if} \quad S_t = 1, n_t = 1 \\ \frac{(1-\chi)q_{11}}{\chi-(2\chi-1)q_{11}}, & \text{if} \quad S_t = 1, n_t = 2 \\ \frac{\chi(1-q_{22})}{\chi-(2\chi-1)q_{22}}, & \text{if} \quad S_t = 2, n_t = 1 \\ \frac{(1-\chi)(1-q_{22})}{1-\chi+(2\chi-1)q_{22}}, & \text{if} \quad S_t = 2, n_t = 2 \end{cases} \tag{4}$$

and the posterior mean of the Markov-switching transition probability can be expressed as

$$E(q_{ii} | n_t, y^t, S^t, \theta) \quad = \quad \begin{cases} \frac{\chi(a_{i,t}+1)+(1-\chi)b_{i,t}}{\chi a_{i,t}+(1-\chi)b_{i,t}} \cdot \frac{a_{i,t}}{a_{i,t}+b_{i,t}+1}, & \text{if} \quad S_t = i, n_t = i \\ \frac{(1-\chi)(a_{i,t}+1)+\chi b_{i,t}}{(1-\chi)a_{i,t}+\chi b_{i,t}} \cdot \frac{a_{i,t}}{a_{i,t}+b_{i,t}+1}, & \text{if} \quad S_t = i, n_t = j \\ \frac{a_{i,t}}{a_{i,t}+b_{i,t}}, & \text{if} \quad S_t = j, n_t \in \{i, j\} \end{cases} \tag{5}$$

where $i, j \in \{1, 2\}$. The details of derivation are provided in Appendix A and Appendix B.[4]

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

[4] Expressions (4) and (5) corresponding to the case in which we condition on the entire history of n^t instead of just the current n_t are equivalent. This is because once we have conditioned on the entire history of S^t, the past history of news n^{t-1} does not contain any additional information.

We consider two boundary values of $\chi \in \{0.5, 1\}$. We start from $\chi = 0.5$ in which news contain no information about the future state (because it assigns equal probability to both states). We can see this by plugging in the value of $\chi = 0.5$ into (4) to obtain

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

Analogously, we can deduce from (5) that the news has no impact on posterior mean when $\chi = 0.5$ since

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

On the other hand, if news contains certain information about future state, that is, $\chi = 1$, then knowledge about the current state S_t no longer plays a role in predicting the future state. (4) can be expressed as

$$p(S_{t+1} = 1 | n_t, S^t, \Pi, \theta) = p(S_{t+1} = 1 | n_t, \theta) = \begin{cases} 1, & \text{if} \quad n_t = 1 \\ 0, & \text{if} \quad n_t = 2. \end{cases}$$

We next examine the implications of news for parameter learning. We show that (5) reduces to

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

which follows from the laws of motion

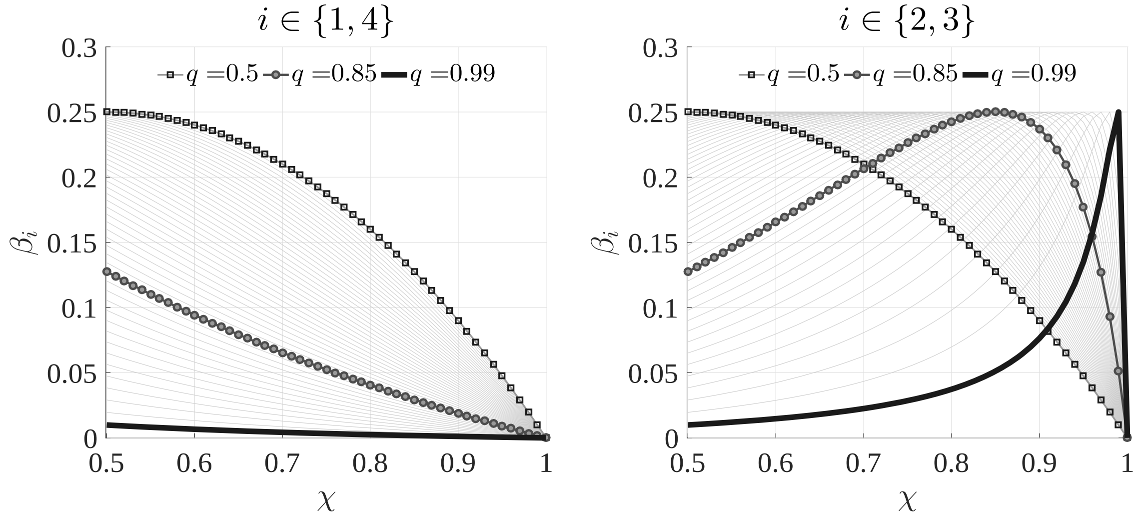
$$a_{i,t+1} \quad = \quad a_{i,t} + \mathbb{I}_{\{S_{t+1}=i\}} \mathbb{I}_{\{S_t = i\}}, \quad b_{i,t+1} = b_{i,t} + (1 - \mathbb{I}_{\{S_{t+1}=i\}}) \mathbb{I}_{\{S_t=i\}}.$$

The posterior mean is identical to the one that would have been obtained in time $t+1$ in the absence of news. For example, in the $\{S_t = i, n_t = j\}$ case, news with certainty ($\chi = 1$) implies that we will be switching into $S_{t+1} = j$ with probability 1. In this case, $a_{i,t+1} = a_{i,t}$ and $b_{i,t+1} = b_{i,t} + 1$. Similarly, it is straightforward to see that the identity holds for the remaining cases, $\{S_t = i, n_t = i\}$ and $\{S_t = j, n_t \in \{i, j\}\}$.

The above examples show that news can lead to improvement in prediction and parameter learning accuracy. We now examine its implication on forecast uncertainty. In this model, one measure of ex-ante forecast uncertainty is the conditional variance of one-period-ahead output growth defined as

$$\begin{aligned} Var(y_{t+1} | y^t, S^t, n^t, \Pi, \theta) &= \int (y_{t+1} - y_{t+1|t})^2 p(y_{t+1} | y^t, S^t, n^t, \Pi, \theta) dy_{t+1} \qquad (6) \\ &= \underbrace{(q_{1i}^B + q_{2i}^B)\sigma_1^2 + (q_{3i}^B + q_{4i}^B)\sigma_2^2}_{\text{due to second moment}} + \underbrace{(q_{1i}^B + q_{2i}^B)(q_{3i}^B + q_{4i}^B)(\mu_1 - \mu_2)^2}_{\text{due to first moment}} \end{aligned}$$

Figure 1: Forecast Uncertainty: β_i



Notes: The expression for β_i is provided in (8). Squared-lines indicate when $q = 0.5$, circled-lines are when $q = 0.85$, and black-solid lines are when $q = 0.99$.

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 $\sum_{j=1}^{4} q_{ji}^B = 1$. We refer to (A-15) for relationship between q_{ji} and q_{ji}^B. Note that (6) is comprised of two components: The first component captures the uncertainty with respect to innovation variances and the second component arises from uncertainty about mean values. To understand this, assume for simplicity that the innovation variance is identical across states (i.e., that there is no stochastic volatility), that is, $\sigma^2 = \sigma_1^2 = \sigma_2^2$. Then, (6) simplifies to

$$Var(y_{t+1}|y^t, S^t, n^t, \Pi, \theta) = \sigma^2 + (q_{1i}^B + q_{2i}^B)(q_{3i}^B + q_{4i}^B)(\mu_1 - \mu_2)^2. \tag{7}$$

If both states are absorbing states, i.e. $(q_{1i}^B + q_{2i}^B)(q_{3i}^B + q_{4i}^B) = 0$, or the distribution of future output growth is identical in both states, i.e. $(\mu_1 - \mu_2)^2 = 0$, then forecast uncertainty reduces to $Var(y_{t+1}|y^t, S^t, n^t, \Pi, \theta) = \sigma^2$. Thus, it is clear that uncertainty with respect to different mean values $\mu_1 \neq \mu_2$ gives rise to forecast uncertainty above and beyond the usual variance channel. Suppose now that persistence of each regime is identical $q = q_{11} = q_{22}$. We can further simplify (7) to

$$Var(y_{t+1}|y^t, S^t, n^t, \Pi, \theta) = \sigma^2 + \beta_i(\mu_1 - \mu_2)^2, \tag{8}$$

$$\beta_i = \begin{cases} \frac{\chi(1-\chi)q(1-q)}{(1-\chi+(2\chi-1)q)^2}, & \text{if} \quad i \in \{1,4\} \\ \frac{\chi(1-\chi)q(1-q)}{(\chi-(2\chi-1)q)^2}, & \text{if} \quad i \in \{2,3\}. \end{cases}$$

We refer to (3) for the definition of states, $\kappa_t \in \{1, 2, 3, 4\}$.

Figure 1 displays β_i as a function of $\chi \in [0.5, 1)$ for different values of $q \in [0.5, 1)$. We are only considering moderately persistent q values and excluding the end points ($q = 1, \chi = 1$) which would remove all state uncertainty. Note that by considering 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.

It is interesting to observe that if news suggests that the current regime will persist next period, that is $i \in \{1, 4\}$, then forecast uncertainty monotonically decreases as χ increases. On the contrary, if news contradicts the implication of the current state alone and suggests switching into different state, that is $i \in \{2, 3\}$, then forecast uncertainty increases for $\forall q \leq \chi$ and decreases $\forall q > \chi$. The key takeaway is that in contrast to standard Gaussian setups, our discrete-state environment allows news to increase forecast uncertainty when it contradicts existing beliefs.

3 State and Parameter Learning

3.1 Finding the Empirical Proxy for the News Component

Our model features news that arrives in discrete form, but there is not a direct empirical proxy that has this feature. Instead, we look for a forward-looking variable that contains information about agents' beliefs regarding future states. Imagine that there exists forecasters who know the true model parameters and current states, and in addition to this, receive noisy news about the future state. Let $I_t = \{n^t, S^t, y^t, \Pi, \theta\}$ be the information set for these forecasters. Here, we are making the assumption that these forecasters are endowed with full structural knowledge of the economy, as in standard rational expectations model. Suppose that we can observe their probability forecast[5]

$$z_t = p(S_{t+1} = 1|I_t). \tag{9}$$

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 (9). Professional forecasters have an information set that is closest to I_t, but in practice, it is possible that their information set may not be perfect. Their forecasts could be biased or inefficient. We

[5]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 function of $p(S_{t+1} = i|I_t)$. Instead, we rely on a direct measure of $p(S_{t+1} = i|I_t)$.

allow for this by modeling the anxious index as

$$z_{\text{SPF},t} = \Phi\left(b_z + \Phi^{-1}\Big(p(S_{t+1} = 1|I_t)\Big) + u_t\right), \quad u_t \sim N(0, \sigma_z^2) \tag{10}$$

where $\Phi(\cdot)$ is a cdf of the standard normal distribution.[6] 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 (9).

3.2 Information

In Section 2, we used an assumption that model parameters and the full history of states are known at time t to illustrate the implication of news on parameter learning, state prediction, and forecast uncertainty. Here, we relax the assumption in our inference problem by making model parameters and states unknown in the econometrician's sequential learning problem.[7] Consider an econometrician that observes past and current output growth and recession probability forecasts. She updates her beliefs about states and parameters using Bayes's rule as she obtains new data.

3.3 Solving the 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 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 to 2016:Q3, we obtain the full joint posterior distribution of the model primitives including the filtered distribution of discrete states at each point in time. Specifically, we rely on particle methods to directly sample from the particle approximation to

$$p(\theta, \Pi, \kappa^t | x^t) = \underbrace{p(\theta, \Pi | \kappa^t, x^t)}_{\text{(i) parameter learning}} \times \underbrace{p(\kappa^t | x^t)}_{\text{(ii) state filtering}} . \tag{11}$$

To sample from (i) and (ii) jointly, we extend the particle learning algorithm developed by Carvalho, Johannes, Lopes, and Polson (2010), which is a generalization of the mixture Kalman filter of Chen and Liu (2000). The details are provided in Appendix C and Appendix E.

As explained in Johannes, Lochstoer, and Mou (2016), the joint learning of states and parameters is a high-dimensional problem which incurs confounding effects arising from multiple

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

[7]We will distinguish agent's learning problem from econometrician's problem when we introduce an asset pricing model.

Table 1: Simulation: RMSE

	Simulation 1 $\chi = 0.8$			Simulation 2 $\chi = 0.5$		
	5%	50%	95%	5%	50%	95%
	Parameter Learning					
μ_1	0.005	0.061	0.131	0.009	0.059	0.142
σ_1^2	0.003	0.046	0.118	0.002	0.048	0.114
μ_2	0.010	0.065	0.134	0.006	0.060	0.131
σ_2^2	0.002	0.038	0.110	0.003	0.037	0.108
q_{11}	0.002	0.016	0.040	0.003	0.018	0.049
q_{22}	0.002	0.015	0.039	0.002	0.017	0.046
	Output Growth Forecast					
y_t	0.627	0.693	0.750	0.656	0.727	0.783

Notes: We have two sets of simulated data labeled by "Simulation 1" and "Simulation 2." In both data set, output growth data are assumed to be identical. The survey forecasts, however, are generated based on informative news $\chi = 0.8$ and uninformative news $\chi = 0.5$. We keep the number of estimated parameters to be identical in both cases and fix χ and σ_z^2 to their true values. 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%, 50%, 95% percentiles of RMSE distributions based on 100 Monte Carlo simulations.

sources of uncertainty. Before we move 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

We fix parameters at their prior median values in Table 2 and simulate output growth (y_t) and state (S_t) from (1). We simulate two different time series paths of news (n_t): one in which news is informative $\chi = 0.8$ and the other in which it is not, that is $\chi = 0.5$. Survey forecasts ($z_{\text{SPF},t}$) are simulated based on (10). The length of simulated data is set to match the estimation sample. 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. It is important to note that all simulation results are generated with unbiased priors.

We run the particle learning algorithm for two sets of simulated data labeled by "Simulation 1" and "Simulation 2." We keep the estimated parameters identical in both cases. We fix χ

and σ_z^2 to their true values to impose the same degree of uncertainty in both cases.

In the top panel of Table 1, we provide 5%, 50%, 95% percentiles of RMSE distributions based on 100 Monte Carlo simulations. RMSE 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 the good performance of the particle learning algorithm. Second, we find roughly 10% improvement in parameter learning accuracy for q_{11} and q_{22}. While the differences in absolute magnitudes are small, we find that RMSEs are uniformly smaller with informative news. The results for the rest of parameters are virtually identical across two sets of simulations. This is to be expected since news in this model is informative about the realization of the future state and not growth itself. Thus, it only provides information about the state transition and not about the distribution of growth conditional on a state.

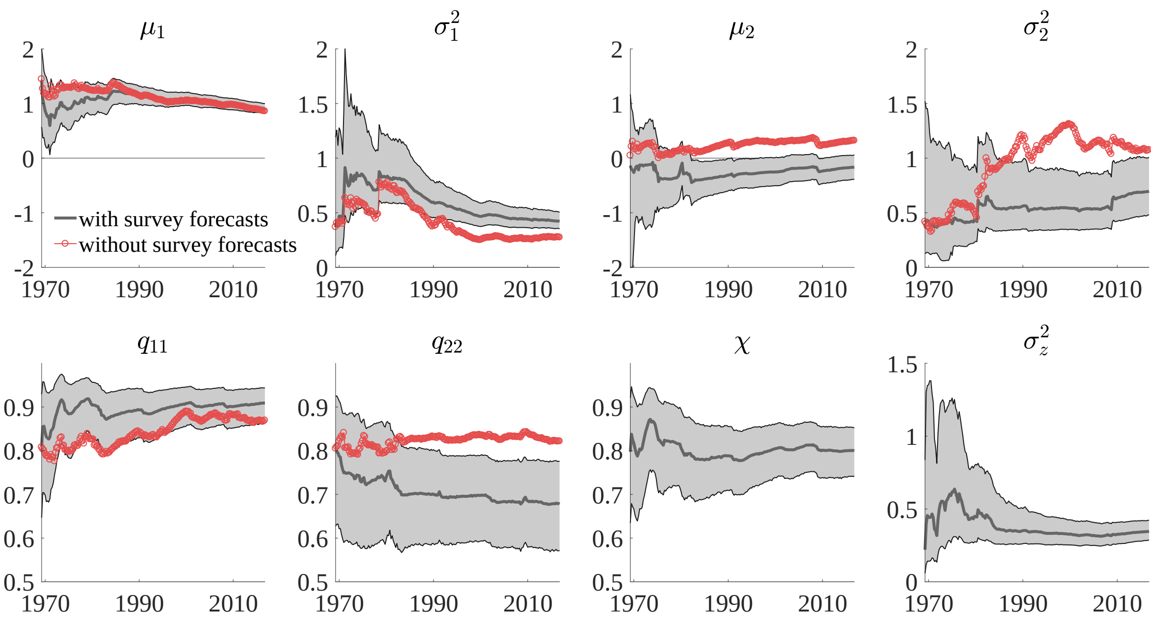
In the bottom panel of Table 1, we compute the output growth forecast RMSEs. On average, we find 5% improvement in the ex-post accuracy of output growth forecasts if we allow purely forward-looking news to be embedded in probability forecasts. In sum, the simulation results are consistent with the implications of news 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. Figure 2 provides the evolution of parameter learning. The credible intervals at time 0 correspond to the 90% prior intervals. As more information from observed data is incorporated into the posterior distributions over time, the 90% credible intervals shrink. Those at time T are posterior credible intervals one would obtain from the entire times series data. Table 2 reports 5%, 50%, 95% percentiles of posterior distributions in Figure 2. The benchmark estimation results are provided in panel (2) of Table 2. First, the expansion regime ($S_t = 1$) is identified with positive mean and the recession regime ($S_t = 2$) with negative mean. Note that we only impose $\mu_1 > \mu_2$ to deal with the label switching problem in the estimation. We also find that posterior median estimate of variance is larger in the recession regime than in the expansion 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. It is also reflected in much lower estimates of q_{22} than q_{11}. Third, the posterior median estimate of χ is 0.8 which implies that news extracted from surveys are quite informative about the future regime.

Figure 2: Filtered Estimates: Parameters



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

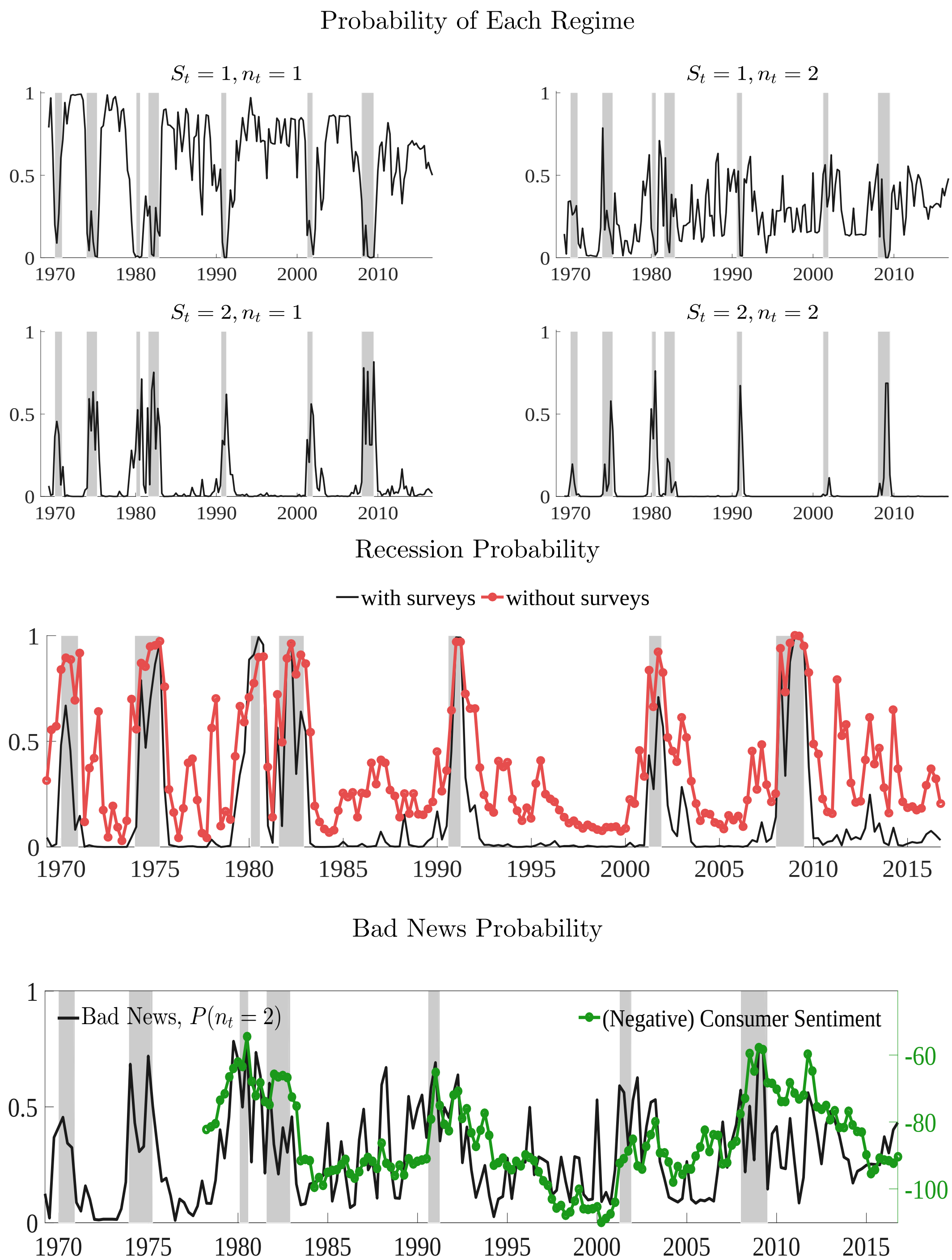
Table 2: Prior and Posterior Distributions: Parameters

	Prior (1)			Posterior (2) With Surveys Benchmark			Posterior (3) With Surveys $\chi = 0.5$			Posterior (4) Without Surveys		
	5%	50%	95%	5%	50%	95%	5%	50%	95%	5%	50%	95%
μ_1	-1.00	1.00	3.00	0.82	0.89	1.00	0.83	0.89	1.01	0.76	0.83	0.99
σ_1^2	0.12	0.40	1.50	0.35	0.42	0.51	0.33	0.39	0.47	0.20	0.29	0.36
μ_2	-2.00	0.00	2.00	-0.39	-0.12	0.06	-0.38	-0.12	0.06	0.05	0.39	0.61
σ_2^2	0.12	0.40	1.50	0.48	0.70	1.00	0.51	0.67	0.91	0.77	1.02	1.59
q_{11}	0.64	0.80	0.93	0.86	0.91	0.94	0.85	0.90	0.94	0.80	0.87	0.93
q_{22}	0.64	0.80	0.93	0.57	0.68	0.78	0.57	0.68	0.77	0.74	0.82	0.90
χ	0.64	0.80	0.93	0.74	0.80	0.85	-	0.50	-	-	-	-
σ_z^2	0.06	0.24	0.96	0.29	0.35	0.42	0.24	0.29	0.35	-	-	-

Notes: This table reports 5%, 50%, 95% percentiles of posterior distributions in Figure 2. The details of prior choices are provided in Table A-1.

To understand how news would influence the inference of model parameters, we repeat the estimation by fixing $\chi = 0.5$. The corresponding estimation results are provided in panel (3)

Figure 3: Filtered Estimates: State



Notes: In the top panel, the gray solid lines are posterior median values. In the second panel, we compare the two filtered recessions probabilities obtained with (black solid-line) and without surveys (red circled-line). In the bottom panel, the correlation of the model-implied $P(n_t = 2)$ (black solid-line) and the University of Michigan Consumer Sentiment Index (green circled-line) is 0.50.

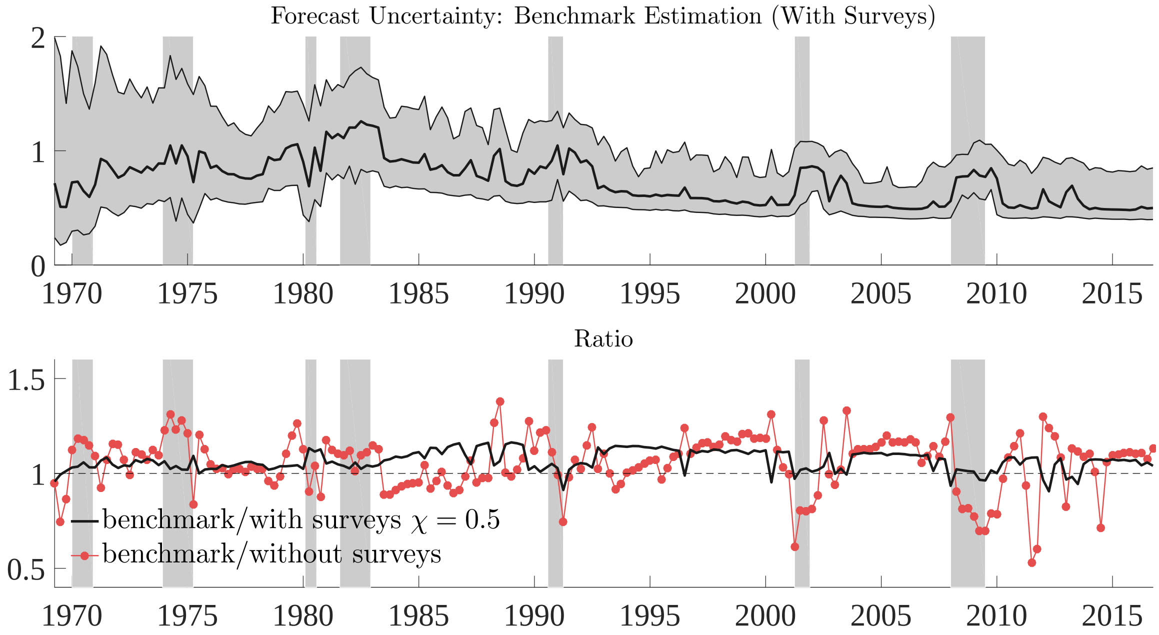
of Table 2. We find that posterior estimates are nearly identical to those reported in panel (2). This evidence suggests that the first-order improvement on the parameter estimates is achieved through the assumption that forecasters are endowed with (imperfect) structural knowledge of the economy rather than the assumption that these forecasters are receiving news about future state as confirmed in our simulation exercise.

Finally, we remove survey forecasts from the estimation to examine its impact on 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 the 90% credible intervals are larger for 4 out of 6 of the parameters governing the real GDP growth process. Furthermore, posterior median estimates associated with the recession state 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 larger. Perhaps surprisingly, q_{22} is estimated to be much more persistent without news. The distinction can be seen clearly in Figure 2 in which the sequential parameter posterior distributions are plotted.

Regime Probabilities. A different way to see the differences in the estimations with and without survey data is to look at the filtered regime probabilities. The first panel of Figure 3 provides the probability of each regime in the benchmark estimation. It is interesting to observe that the economy spends most of the time in the $\{S_t = 1, n_t = 1\}$ regime. Interestingly, the probability of $\{S_t = 1, n_t = 2\}$ is non-negligible. This is the regime in which the economy is currently in the expansion state, but news suggests that the economy will enter recession in the following period. On the other hand, the economy assigns very small probabilities to the regimes $\{S_t = 2, n_t = 1\}$ and $\{S_t = 2, n_t = 2\}$. The second panel of Figure 3 combines the probabilities of $\{S_t = 2, n_t = 1\}$ and $\{S_t = 2, n_t = 2\}$ 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 (without surveys), 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 third panel of Figure 3 combines the estimated probabilities of $\{S_t = 1, n_t = 2\}$ and $\{S_t = 2, n_t = 2\}$ for our baseline case. It is interesting to observe that the correlation with the University of Michigan Consumer Sentiment Index and our filtered probability of bad news realizations is around 0.5.

Forecast Uncertainty. Figure 4 plots the evolution of forecast error variance over time. The bottom panel of Figure 4 displays the ratio of the forecast error variances computed from the benchmark estimation (with surveys) and two alternative estimations: with surveys (fixing $\chi = 0.5$) and without surveys. On average, the ratio is greater than one implying that news increases forecast uncertainty. There is a pattern of higher uncertainty driven by bad

Figure 4: Forecast Uncertainty



Notes: In the top panel, we plot posterior median forecast error variance with the 90% credible interval (shaded areas). In the bottom panel, we display the ratio of the median forecast error variances computed from the benchmark estimation over those obtained from the alternative estimation strategies.

news during expansions in our baseline case relative to the case where we assume that news is irrelevant by fixing $\chi = 0.5$.

4 Asset Pricing Implications

We have seen from Figure 2 that subjective beliefs vary significantly over time. Johannes, Lochstoer, and Mou (2016) show that revisions in beliefs constitute permanent shocks to investors' expectations and therefore become a source of subjective long-run risks (see Bansal and Yaron (2004) and Collin-Dufresne, Johannes, and Lochstoer (2016)). By embedding these beliefs in an asset pricing model in which agents prefer an early resolution of uncertainty, Johannes, Lochstoer, and Mou (2016) are able to replicate several asset pricing moments.[8] One key feature of Johannes, Lochstoer, and Mou (2016) is the generation of strongly countercyclical return volatility and a high equity premium in recessions. This is easy to explain because the infrequent nature of recessions implies greater parameter updating when the economy vis-

[8] 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.

its that state. Thus, the long-run shocks to parameter beliefs will be largest during recessions and this contributes to the high volatility of returns and equity premium in recessions.

The model-implied forecast uncertainty in the top panel of Figure 4 confirms the mechanics of Johannes, Lochstoer, and Mou (2016). That is, uncertainty is strongly countercyclical and largest in recessions. What is interesting is 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 within states through the addition of news shocks. This is most apparent in expansions, particularly in the period surrounding 1990 and the Great Recession (see the bottom panel of Figure 4). In light of this property of our model, we expect news-driven uncertainty fluctuations can be important for generating more variation in asset pricing variables within states.

To this end, we quantify the role of news in accounting for variations in returns and equity premia by solving an asset pricing problem that features sequential learning of unknown $\{q_{ii}\}_{i=1,2}$ where agents observe both the current true states as well as news about the one-period-ahead state. To alleviate the computational burden, we assume that agents are fully aware of the rest of the model parameters including the accuracy of news.[9] One reason to allow agents to learn $\{q_{ii}\}_{i=1,2}$ rather than the other parameters is that Collin-Dufresne, Johannes, and Lochstoer (2016) show that uncertainty about transition probabilities $\{q_{ii}\}_{i=1,2}$ has the largest asset pricing impact. Furthermore, as discussed in Section 3, since the news we're considering is about the future state and not about future growth rates, it only provides information about $\{q_{ii}\}_{i=1,2}$ and not the other parameters governing the consumption or dividend processes in the model. Therefore, news in this model should interact most strongly with learning about $\{q_{ii}\}_{i=1,2}$ and not about other parameters.

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

$$V_t = \max_{C_t} \left[(1-\beta) C_t^{\frac{1-\gamma}{\alpha}} + \beta (E_t V_{t+1}^{1-\gamma})^{\frac{1}{\theta}} \right]^{\frac{\alpha}{1-\gamma}}, \tag{12}$$

subject to budget constraint $W_{t+1} = W_t R_{c,t+1} - C_{t+1}$, where W_t is the wealth of the agent, $R_{c,t+1}$ is the return on wealth, β is the discount rate, γ is risk aversion, ψ is the intertemporal elasticity of substitution (IES), and $\alpha = \frac{1-\gamma}{1-1/\psi}$.

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

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

[9] The curse of dimensionality is a big issue in the full-fledged parameter learning model.

where X_t denotes our summary statistics

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

$E(q_{ii}|n^t, y^t, S^t, \theta)$ for $i \in \{1, 2\}$ are functions of $\{\kappa_t, X_t\}$ where $f(\cdot)$ is from Bayes's rule. The details are provided in Appendix A. Following Collin-Dufresne, Johannes, and Lochstoer (2016), we specify the exogenous dividend growth process below

$$\Delta d_{t+1} = \bar{\mu} + \rho(\Delta c_{t+1} - \bar{\mu}) + \sigma_d \epsilon_{d,t+1}, \quad \epsilon_{d,t+1} \sim N(0, 1), \tag{13}$$

and 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 agents beliefs about this quantity also evolved over time.

In this economy, agents' beliefs about $\{q_{11}, q_{22}\}$ converge to the truth and their uncertainty about these parameters vanish as time progresses. Therefore, the model is solved by iterating the policy functions for both the wealth-consumption and price-dividend ratios backwards from a distant endpoint that is approximated by an economy in which all parameters are known. Details on the numerical solution algorithm are provided in Appendix F.

Asset Pricing Moments. We simulate asset prices based on

$$\beta = 0.994, \quad \gamma = 10, \quad \psi = 3, \quad \rho = 2.5, \quad \sigma_d = 5.49$$

and posterior estimates of $\{q_{11}^{(i)}, q_{22}^{(i)}\}_{i=1}^{N}$ in Figure 2. All the simulated results are based on the filtered state estimates underlying Figure 3.[10]

In Table 3, we present moments from both the data as well as the 5th, 50th, and 95th percentiles of moments based on our model simulations. The table contains three cases. The first is our benchmark case while the second is a case in which we fix $\chi = 0.5$, thus making news shocks irrelevant, but keep all other parameters the same. That is, the parameters in both the "Benchmark" and $\chi = 0.5$ cases here come from the "Benchmark" posterior median estimates in Table 2. Comparing these two cases illustrates the effect of news shocks in this model conditional on our other benchmark parameters. The third case is one in which there are both 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

[10]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 which depend only on current observations, but they cannot be used to produce moments of ex-post returns which are a function of both current and lagged observations. For these moments, we use simulations created using the posterior median estimates of all parameters from the "Benchmark" case in Table 2.

Table 3: 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					
$S_t = 1$	1.04	2.16	2.32	2.45	2.25	2.32	2.36	3.08	3.10	3.10
$S_t = 2$	0.44	-0.37	0.01	0.33	0.51	0.65	0.76	2.57	2.59	2.60
					Average Ex-Ante Equity Premium					
$S_t = 1$	7.18	2.67	2.95	3.25	2.72	2.80	2.91	1.02	1.04	1.07
$S_t = 2$	9.37	7.94	9.10	10.48	8.68	9.38	10.19	1.67	1.79	1.92
					Ex-post Sharpe Ratio					
Full Sample		0.08	0.31	0.54	0.08	0.31	0.54	-0.13	0.11	0.35
					Standard Deviation of Risk-Free Rate					
$S_t = 1$	1.81	0.91	1.02	1.19	0.23	0.29	0.42	0.03	0.04	0.05
$S_t = 2$	2.36	0.74	1.00	1.36	0.23	0.31	0.41	0.04	0.05	0.06
					Standard Deviation of Ex-Ante Equity Premium					
$S_t = 1$	2.00	1.82	2.05	2.31	0.28	0.37	0.50	0.06	0.08	0.11
$S_t = 2$	2.59	2.78	3.64	4.89	1.19	1.71	2.41	0.22	0.30	0.43

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.

Surveys" posterior median estimates in 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 to the benchmark illustrates the effect of also changing the econometric procedure used to infer parameters and regimes.

Focusing first on the average risk-free rate, ex-ante equity premium, and the ex-post Sharpe ratios, we see that adding news has little effect. This can be understood as a result of the fact that while news can increasing uncertainty regarding future states in our model, it tends to speed up learning about $\{q_{ii}\}_{i=1,2}$. This produces two opposing effects of news on average uncertainty so we don't see much difference in average risk-free rates or risk-premia. The largest difference, in terms of these first three moments, actually comes from including survey data in the estimation. We see from the third set of moments that, using the same set of asset pricing parameters but estimating the data-generating processes of exogenous quantities 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. The Sharpe ratio is also quite a bit lower (and even sometimes negative) in this third case. Given these asset pricing parameters, it's clear that the data-generating process estimated 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 would be 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 3.5 and 3.2 times relative to the case where news is irrelevant in the good and bad states, respectively, bringing the model much closer to matching the data in this dimension. In terms of the equity premia, we see a similar amplification of volatility, particularly in the good state where the within-state standard deviation grows by over 5.5 times relative to the $\chi = 0.5$ case and comes quite close to matching the corresponding data moment.

5 Conclusion

We study the effects of news about the future state of the economy, focusing on its impact on uncertainty and on parameter learning. We present in a discrete-state model where news can actually increase uncertainty about future growth when it contradicts agents' existing beliefs. At the same time, news helps agents to learn the true values of parameters faster.

Using a novel filtering technique, we obtain filtered estimates of both the historical regimes and news shock realizations using data on actual GDP growth and recession probability forecasts from the Survey of Professional Forecasters. As evidence of external validity, our estimated probabilities of bad news realizations correlate strongly negatively with measures of consumer sentiment.

One empirical finding is that bad news tends to persist during the recovery phases following recessions. This information contradicts the signal coming from the fact that good states tend to be quite persistent in the data and produces a period of elevated uncertainty that can last several quarters longer than the recession itself. Compared to an estimation that does not allow a role for news shocks, our baseline estimates feature higher uncertainty on average and particularly during expansion regimes.

We then assess the implications of these news shocks in an asset pricing model where agents must learn state transition probabilities and also have a preference for early resolution of uncertainty. We obtain asset pricing moments from this model using our estimates of parameters and filtered states. We find that news has little effect on the average risk-free rate and equity risk premium due to the two opposing effects of higher uncertainty about the future state but faster parameter learning. However, we find that news shocks can greatly increase the volatility in risk-free rates and equity risk premia, particularly in expansions, thus improving the model's ability to match to the data in terms of second moments.

References

Bachmann, R., S. Elstner, and E. Sims (2013): "Uncertainty and Economic Activity: Evidence from Business Survey Data ," *American Economic Journal: Macroeconomics*, 5(2), 217–249.

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.

Basu, S., and B. Bundick (2017): "Uncertainty Shocks in a Model of Effective Demand," *Econometrica*, Forthcoming.

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.

Bloom, N. (2009): "The Impact of Uncertainty Shocks," *Econometrica*, 77, 623–685.

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

Fernández-Villaverde, J., P. Guerrón-Quintana, J. Rubio-Ramírez, and M. Uribe (2011): "Risk Matters: The Real Effects of Volatility Shocks," *American Economic Review*, 101(6), 2530–2561.

Gilchrist, S., J. W. Sim, and E. Zakrajek (2014): "Uncertainty, Financial Frictions, and Investment Dynamics," Working Paper 20038, National Bureau of Economic Research, DOI: 10.3386/w20038.

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.

Kozeniauskas, N., A. Orlik, and L. Veldkamp (2016): "The Common Origin of Uncertainty Shocks," Working Paper 22384, National Bureau of Economic Research, DOI: 10.3386/w22384.

Kurmann, A., and C. Otrok (2013): "News Shocks and the Slope of the Term Structure of Interest Rates," *American Economic Review*, 103(6), 2612–32.

Leduc, S., and Z. Liu (2016): "Uncertainty Shocks are Aggregate Demand Shocks," *Journal of Monetary Economics*, 82, 20–35.

Malkhozov, A., and A. Tamoni (2015): "News Shocks and Asset Prices," Manuscript.

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, Social Science Research Network, Rochester, NY.

Schmitt-Grohe, S., and M. Uribe (2012): "Whats 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.

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 updates 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 which 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 data 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),

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

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

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

The posterior means are

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

and

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

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

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

and

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

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

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

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

$$
\begin{aligned}
p(S_{t+1}|n_t,S^t,\Pi,\theta) &= \frac{p(n_t|S_{t+1},S^t,\Pi,\theta)p(S_{t+1}|S^t,\Pi,\theta)}{p(n_t|S_t,\Pi,\theta)} \qquad \text{(A-14)} \\
&= \begin{cases}
\frac{\chi q_{11}}{(1-\chi)+(2\chi-1)q_{11}}, & \text{if } n_t=1, S_t=1, S_{t+1}=1 \\
\frac{(1-\chi)(1-q_{11})}{(1-\chi)+(2\chi-1)q_{11}}, & \text{if } n_t=1, S_t=1, S_{t+1}=2 \\
\frac{(1-\chi)q_{11}}{\chi-(2\chi-1)q_{11}}, & \text{if } n_t=2, S_t=1, S_{t+1}=1 \\
\frac{\chi(1-q_{11})}{\chi-(2\chi-1)q_{11}}, & \text{if } n_t=2, S_t=1, S_{t+1}=2,
\end{cases} \\
&= \begin{cases}
\frac{\chi(1-q_{22})}{\chi-(2\chi-1)q_{22}}, & \text{if } n_t=1, S_t=2, S_{t+1}=1 \\
\frac{(1-\chi)q_{22}}{\chi-(2\chi-1)q_{22}}, & \text{if } n_t=1, S_t=2, S_{t+1}=2 \\
\frac{(1-\chi)(1-q_{22})}{1-\chi+(2\chi-1)q_{22}}, & \text{if } n_t=2, S_t=2, S_{t+1}=1 \\
\frac{\chi q_{22}}{1-\chi+(2\chi-1)q_{22}}, & \text{if } n_t=2, S_t=2, S_{t+1}=2.
\end{cases}
\end{aligned}
$$

Putting these two back into (A-13) gives a 4-state transition matrix

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

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

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 get 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 one when $\kappa_t = i$ in and all other elements equal to zero (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,

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

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.[11]

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

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

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 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

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

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 by element 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

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

and 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

$$\begin{aligned} p(\mu|\sigma^2, y^{t+1}, S^{t+1}, \Pi) &\propto p(y_{t+1}, S_{t+1}|\mu, \sigma^2, y^t, S^t, \Pi)p(\mu|\sigma^2, y^t, S^t, \Pi) \qquad \text{(A-30)} \\ &= p(y_{t+1}|S_{t+1}, \mu, \sigma^2, y^t, S^t, \Pi)p(S_{t+1}|\mu, \sigma^2, y^t, S^t, \Pi)p(\mu|\sigma^2, y^t, S^t, \Pi) \\ p(\sigma^2|y^{t+1}, S^{t+1}, \Pi) &\propto p(y_{t+1}, S_{t+1}|\sigma^2, y^t, S^t, \Pi)p(\sigma^2|y^t, S^t, \Pi) \qquad \text{(A-31)} \\ &= p(y_{t+1}|S_{t+1}, \sigma^2, y^t, S^t, \Pi)p(S_{t+1}|\sigma^2, y^t, S^t, \Pi)p(\sigma^2|y^t, S^t, \Pi) \end{aligned}$$

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

$$\begin{aligned}
p(\mu|\sigma^2, y^{t+1}, S^{t+1}) &\propto p(y_{t+1}|S_{t+1}, \mu, \sigma^2, y^t, S^t, \Pi)p(S_{t+1}|\mu, \sigma^2, y^t, S^t, \Pi)p(\mu|\sigma^2, y^t, S^t, \Pi) \qquad \text{(A-32)}\\
&\propto p(y_{t+1}|S_{t+1}, \mu, \sigma^2, y^t, S^t, \Pi)p(\mu|\sigma^2, y^t, S^t, \Pi)\\
&= \Big(\sum_{i=1}^{2} \mathbb{I}_{\{S_{t+1}=i\}} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\Big\{-\frac{1}{2\sigma_i^2}(y_{t+1}-\mu_i)^2\Big\}\Big)\\
&\times \frac{1}{\sqrt{2\pi\sigma_i^2 C_{i,t}}} \exp\Big\{-\frac{1}{2\sigma^2 C_{i,t}}(\mu_i - m_{i,t})^2\Big\} \times \frac{1}{\sqrt{2\pi\sigma_j^2 C_{j,t}}} \exp\Big\{-\frac{1}{2\sigma^2 C_{j,t}}(\mu_j - m_{j,t})^2\Big\}\\
&\propto \frac{1}{\sqrt{2\pi\big(1+\frac{1}{C_{i,t}}\big)^{-1}\sigma_i^2}} \exp\Big\{-\frac{1}{2\sigma_i^2}\Big(1+\frac{1}{C_{i,t}}\Big)\Big(\mu_i - \big(1+\frac{1}{C_{i,t}}\big)^{-1}(y_{t+1} + \frac{m_{i,t}}{C_{i,t}})\Big)^2\Big\}\\
&= N(m_{i,t+1}, C_{i,t+1}\sigma_i^2).
\end{aligned}$$

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

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

We re-express (A-32) as

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

from which we can deduce that

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

Transition Probabilities. At $t = 0$, the agent is given an initial (potentially truncated) Beta-distributed prior over each of these parameters and thereafter 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 $p(\theta, \Pi, \Lambda, F_{\theta,t}, F_{\Pi,t}, F_{\Lambda,t}, \kappa^t | x^t)$ than from $p(\theta, \Pi, \Lambda, \kappa^t | x^t)$.

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

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 $w_t^{(k)}$,

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

   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:**

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

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

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

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

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

4. **Draw Parameters:**

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

$$\begin{aligned} \sigma_i^{2,(k)} &\sim IG(\frac{v_{i,t+1}^{(k)}}{2}, \frac{K_{i,t+1}^{(k)}}{2}) \\ \mu_i^{(k)} &\sim N(m_{i,t+1}^{(k)}, \sigma_i^{2,(k)} C_{i,t+1}^{(k)}). \end{aligned} \tag{A-49}$$

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

$$\begin{aligned} q_{11}^{(k)} &\sim B(a_{1,t+1}^{(k)}, b_{1,t+1}^{(k)}) \\ q_{22}^{(k)} &\sim B(a_{2,t+1}^{(k)}, b_{2,t+1}^{(k)}). \end{aligned} \tag{A-50}$$

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

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

$$\sigma_z^{2,(k)} \sim IG(\frac{v_{z,t+1}^{(k)}}{2}, \frac{K_{z,t+1}^{(k)}}{2}). \tag{A-52}$$

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, multinomial distribution, respectively.

F Asset pricing solution

F.1 Sufficient statistics

In our asset pricing model, we assume that agents' information sets at time t are $I_t = \{\kappa^t, y^t, \theta\}$ and they rationally learn 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 which yield 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 $\{\kappa_t, \kappa_{t-1}\}$:

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

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

For the solution below, agents' 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:

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

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

where we use the short-hand 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 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 $\bar{\mu}(q_{11}, q_{22})$ 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 $X_{t+1} = F(S_{t+1}, S_t, X_t)$ 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 satisfies the following recursions:

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

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

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 backwards from the known parameters solution. Details on this procedure are provided below:

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

Despite no impact of news 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 3-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

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

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 multiplication and exponentiation. For fixed values of $\{\Pi, \theta\}$, this is simply a system of 4 nonlinear equations which 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:

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

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 4-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 since $\tau_{j,t}$ 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 $\tau_{j,T}$ since this relationship becomes exact as $\tau_{j,T} \to \infty$. We then iterate backwards 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 such that $S_t = j$ are a function of the values of PC with τ_j incremented by one. That is, we iterate τ_j backwards one step as follows:

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

where we use the fact that $\kappa_t = 2\left(S_t - 1\right) + n_t$ 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 LHS of the previous step, we can directly solve for PC for the remaining values of κ_t using this system of equations:

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

where $\mathbf{PC}'$ is formed by stacking $\begin{bmatrix} PC(\kappa_t = 2i-1, \lambda_{j,t}, \tau_{j,t}) \\ PC(\kappa_t = 2i, \lambda_{j,t}, \tau_{j,t}) \end{bmatrix}$ and $\begin{bmatrix} PC(\kappa_t = 2j-1, \lambda_{j,t}, \tau_{j,t}) \\ PC(\kappa_t = 2j, \lambda_{j,t}, \tau_{j,t}) \end{bmatrix}$ 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

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

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

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

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

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

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 $\{\lambda_{j,t}, \tau_{j,t}\}$.

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

where $\mathbf{PD}'$ is formed by stacking $\begin{bmatrix} PD(\kappa_t = 2i-1, \lambda_{j,t}, \tau_{j,t}) \\ PD(\kappa_t = 2i, \lambda_{j,t}, \tau_{j,t}) \end{bmatrix}$ and $\begin{bmatrix} PC(\kappa_t = 2j-1, \lambda_{j,t}, \tau_{j,t}) \\ PC(\kappa_t = 2j, \lambda_{j,t}, \tau_{j,t}) \end{bmatrix}$ appropriately given the values of i and j.

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

Using the solutions for both the cases of a single known transition probability, we can now obtain the solution for the case where 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 5-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:

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

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:

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

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

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

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:

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

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

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

F.6 Asset pricing moments

Using our 5-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

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

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

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

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

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

  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:

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