Inflation Uncertainty, Asset Valuations, and the Credit Spreads Puzzle*

Alexander David
Haskayne School of Business
University of Calgary
November 2006

Abstract

Investors’ learning of the state of future real fundamentals from current inflation leads to macroeconomic state dependence of asset valuations and solvency ratios of firms within given rating categories. Since credit spreads are convex functions of solvency ratios, average spreads are higher than spreads at average solvency ratios. Macroeconomic shocks carry risk premiums so that expected default losses are more sensitive to changes in the price of risk than are credit spreads. By incorporating state dependence and increasing the price of risk, the econometrician obtains high credit spreads while maintaining average default losses at historical levels — the credit spreads puzzle.

Key Words: learning, uncertainty, proxy-hypothesis, through-the-business-cycle rating, state-dependent solvency ratios, convexity, stochastic volatility

JEL Classification: G12, G13, G14, C3, C5

*I am grateful to an anonymous referee, Yacine Ait-Sahalia, the Editor at RFS, Kerry Back, Michael Brandt, Michael Brennan, Mark Carey, Darrell Duffie, Pierre Collin-Dufresne, Phil Dybvig, Bob Goldstein, Alfred Lehar, Vadim Linetsky, Francis Longstaff, Mark Lowenstein, Gordon Sick, and Pietro Veronesi for helpful comments, and to seminar participants at the Mathematical Finance Conference in honor of Robert J. Elliott, the Credit Risk and Asset Pricing Conference at the Wharton School of the University of Pennsylvania, the 16th Annual Conference on Financial Economics and Accounting at UNC Chapel Hill, HEC-Montreal Finance Department, the McGill-IFM Risk Management Conference, and the Western Finance Association Meetings at Keystone, Colorado. I gratefully acknowledge a research grant from the SSHRC. Address: 2500 University Drive NW, Calgary, Alberta T2N 1N4, Canada. Phone: (403) 220-6987. Fax: (403) 210-3327. E. Mail: adavid@ucalgary.ca.
1 Introduction

Intuition developed in pricing options led to the development of the structural form approach for the pricing of risky corporate debt. The approach has been useful in understanding the implications of a firm’s capital structure, and the value and volatility of its asset value process on the pricing of its risky corporate debt.\(^1\) In summary, the approach assumes that a firm defaults on all its liabilities when the value of its assets at maturity is below a threshold. The approach has led to a recent revolution in measuring and managing credit risk with the focus shifting from traditional accounting-based measures of credit risk to market-based measures that use the information on asset values and their volatilities.

However, recent empirical testing has cast doubts on the ability of the entire structural form family of models to simultaneously generate both default probabilities and credit spreads that match historical experience.\(^2\) Elton, Gruber, Agrawal, and Mann (2001) argue that measures of expected default losses based on empirical frequencies of default implicit in credit ratings and historical recovery rates are too low to be consistent with the high observed credit spreads. Using related calibration methodologies, Delianedis and Geske (2001) and Amato and Remolona (2003) reach a similar conclusion. Extending this logic, Huang and Huang (2003) (henceforth HH) calibrate various structural form models to match the average solvency ratios of firms of different rating categories, and in addition choose free parameters to match the historical default probabilities and recovery rates on defaulted bonds, and find that the model-generated spread accounts for less than a third of the total spread over Treasuries on investment-grade corporate bonds. By constraining the model to match historical default losses, the calibration methodology of these authors is seemingly robust to the specification of the asset value process or the default threshold since changing either assumption leads to roughly equivalent variation in expected default losses and credit spreads. This problem has been called the credit spreads puzzle.

Additionally, Chen, Collin-Dufresne, and Goldstein (2006) (henceforth CCG) have contributed to a deepening of the puzzle. These authors argue that the credit spreads and equity premium puzzles...
are related since they measure risk premiums of different liabilities on the same underlying asset value process. It is therefore natural to consider models in the literature that have been successful in generating a large equity premium and assess their ability in addressing the credit spreads puzzle. Intuitively, and as implied by standard asset pricing theory, it is fairly clear that for a given level of expected default losses, a model’s bond prices will be lower (and thus credit spreads higher) if there is a large positive covariance between the losses from default and investors’ pricing kernel (marginal utility of consumption). Similarly, the equity premium is higher in a model when the covariance between the pricing kernel and stock returns is strongly negative. However, quite surprisingly, by strictly applying the calibration methodology of HH, CCG find that two well-established models in the literature with high prices of risk (volatility of the marginal utility of consumption), the Campbell and Cochrane (1999) (henceforth CC) habit formation model, and the Bansal and Yaron (2004) (henceforth BY) long-term risks model, can only generate small credit spreads after controlling for the expected loss from default. In addition, the CC model counterfactually predicts that conditional default probabilities of firms decline in recessions. In this sense, the credit spreads puzzle is more puzzling than the equity premium puzzle. CCG point out that besides being unable to match the level of credit spreads, the standard structural form model also has difficulty in explaining the high observed volatility in time series of credit spreads, which these authors call the credit spreads volatility puzzle. We will discuss in Section 5.4 how CCG are able to improve the performance of the CC model in both respects by incorporating systematic variation in firms’ default boundaries.

To address these puzzling phenomena, we construct a generalization of Merton’s basic structural form model, with three distinctive features. The first key aspect of our model is its unobserved regime-switching structure for fundamentals. Drifts of inflation and earnings growth jump erratically, and investors learn about these drifts over time. Second, the asset value process, which is normally exogenously specified, is endogenously determined in a classic asset pricing framework. Asset valuation ratios (which are constant in most existing credit risk models), and asset volatilities vary over time as investors update their beliefs about the hidden states of fundamentals and reassess the prospects of future real growth of fundamentals. In periods when investors are less optimistic about strong future earnings growth asset valuations fall. When they are more uncertain about the current state of fundamentals, they revise their beliefs more rapidly, leading to greater volatility of asset values (the discounted value of future fundamentals). David (1997) shows that

\[ \text{Our model is related to but distinct from the seminal work on incomplete information and credit risk of Duffie and Lando (2001) in which the asset value process is observed imperfectly by agents due to accounting noise. In our model the asset value process is formulated by investors conditional on their imperfect information of fundamentals.} \]
such a Bayesian learning process can generate several stylized facts about volatility processes in the GARCH literature. Third, we model the joint shifting distribution of inflation and real fundamentals and find that states of strong real fundamental growth are more persistent when inflation is low so that inflation predicts real growth rates. This is the signaling role of inflation (also called the “proxy hypothesis” by Fama 1981). In macroeconomics, there is robust empirical support that credit quality of firms weakens in periods after increases in short nominal rates (for a survey of this literature see Bernanke and Gertler 1995). Consistent with this literature, as seen in the top panel of Figure 1, most periods of enhanced spreads in the past forty years have been preceded by bouts of rising short rates.4

Using the options logic, we provide a formula for the defaultable bond price in this model as a function of investors’ beliefs about the hidden states with a combination of Fourier Transform and projection methods. An attractive feature of our model is the availability of closed-form solutions for the term-structure of interest rates and asset values.5 This enables us to calibrate our model parameters and investors’ beliefs to observed data on both fundamental and financial variables, the latter consistent with a rich information set for agents.

If an empirical asset value series were available, we could adapt standard econometric methods such as GMM or SMM to fit the parameters of the regime switching model. However, we face the added complexity (confronting all implementations of structural form models) of not having available such an observable process. We resolve the problem by treating the asset value process as “missing data” and adapting the well-known recursive Expectations Minimization (EM) algorithm to the framework. It is a generalization of the method used by Moody’s KMV to iteratively back out a series of asset values and volatilities from observed equity data which uses the Black-Scholes-Merton (BSM) formula, and so is only consistent with constant volatility and interest rates. The equivalence of the KMV methodology with the EM algorithm and a maximum likelihood procedure to fit the BSM model has been shown by Duan, Gauthier, and Simonato (2004).

To assess the ability of our model to generate empirically plausible credit spreads we make the assumption that the rating agencies adopt the “through-the-business-cycle” approach to assigning

---

4A notable exception is the increase in spreads in 1998 following the Russian government default and the LTCM crisis. While various contagion channels have been suggested for the increase in credit spreads from Russia to spill over to the U.S. market, the default in Russia itself was preceded by a resurgence in inflation in 1997 and early 1998.

5In recent years, several authors have written regime-switching models of the term structure (see, e.g., Bansal and Zhou 2002, Ang and Bekaert 2002, Dai, Singleton, and Yang 2003). Our work is different in two respects: (i) we do not model regimes of interest rates, but instead fundamental regimes (both real and inflationary), and (ii) the regimes are unobserved, so that volatility of interest rates is endogenously determined in contrast to the above papers. David and Veronesi (2006) pursue a similar direction in discrete time, while Ang and Piazzesi (2003) construct a term-structure model with fundamentals in an affine setting.
firms to a given rating category. That is, each firm is rated not based on its current conditions but rather on its average probability of default through good and bad years of the business cycle. A brief description of this method of rating is in Carey and Treacy (1998):

Under this philosophy, the migration of borrowers’ rating up and down the scale as the overall economic cycle progresses will be muted: Ratings will change mainly for those firms that experience good or bad shocks that affect long-term condition or financial strategy and for those whose original downside scenario was too optimistic. (box on page 899)

In keeping with this tradition, the solvency ratio for our “representative firm” will fluctuate through the business cycle with investors’ expectations of fundamental growth rates; however, we will assign it to a given rating category based on its average solvency ratio. For example, HH point out that the average solvency ratio for Baa-rated firms is 2.5. The bottom panel of Figure 1 shows time series of our model-based solvency ratio and its empirical counterpart constructed from aggregated series of liabilities of firms and aggregate prices for equity and debt. Indeed, the ratio has shown considerable variation in the sample, falling as low as 1.4 and rising at its peak to almost 5. The ratio fluctuates so much because asset valuations affecting the numerator are far more variable than debt in the denominator.

For addressing the credit spreads puzzle we make two two key departures in our calibration methodology relative to the analysis in HH. First, consistent with the through-the-cycle rating approach described above, we do not assume that within a rating category firms’ credit risk is constant over time. This is implicitly the approach taken by these authors when they assume constant solvency ratios for such firms. Using a result in Bergman, Grundy, and Weiner (1996), we show that credit spreads are convex functions of firms’ solvency ratios, and hence average spreads are larger than the spread evaluated at the average solvency ratio. By taking into account this convexity effect from time-varying solvency ratios we are able to justify credit spreads that are 60 percent higher than those reported by HH, who report the spread at the average solvency ratio for alternative models. Our second departure from HH is to build in explicitly the effect of stochastic asset volatility, a feature that these authors abstract from for reasons of tractability. In our model asset volatility is inversely related to asset valuations and makes the impact of credit risk convexity even larger, providing another 50 percent increase over their benchmark. Overall, our analysis justifies a default-related spread of Baa bonds of 106 basis points, which is close to the average historical

\[ \text{The growth rate of aggregated liabilities has an annual volatility of only about 4 percent. The volatility of the implied asset value process from equity data is significantly higher at about 22 percent.} \]
spread of Baa bonds over Aaa bonds. We attribute the remaining portion of the spread of Baa bonds over Treasuries, which is the spread of Aaa bonds over Treasuries, to issues such as illiquidity and taxes of corporate bonds that are not modeled in this paper.

It is relevant to point out why the convexity effect described above has a non-equivalent impact on credit spreads and default probabilities. Indeed, explaining this differential impact is the key to our resolution of the credit spreads puzzle, which requires models to generate large spreads while holding expected default losses at historically observed levels. In a nutshell, our explanation is as follows: an econometrician who takes into account the state dependence in credit risk of firms will, by the convexity effect, obtain higher expected default losses and credit spreads for given levels of the prices of risk relative to the case where he ignores the state dependence. Increasing the prices of risk in our model leads to a simultaneous decline in asset volatility and an increase in the asset risk premium. Both effects lead to a decline in the expected default losses, which are computed under investors’ objective measure, but only the former lowers credit spreads. Therefore, the expected default losses are more sensitive than credit spreads are to changes in the prices of risk. By incorporating the state dependence and increasing the prices of risk the econometrician can maintain expected default losses at historical levels and still have spreads at a higher level relative to the case of no state dependence in credit risk.

As further support for our approach, we show that an econometrician who takes into account the state dependence of solvency ratios and asset volatilities is able to reconcile the high credit spreads and the high equity premium, while one who ignores the state dependence would require too low an equity premium to calibrate the model to match historical expected default losses. We also show that the state dependence is able to rationalize the volatility of credit spreads. Finally, we show that our model-implied credit spread explains nearly two-thirds of the variation in the historical Moody’s Baa-Aaa spread, and is a very significant predictor of business cycles.

The plan for the remainder of this paper is as follows: In Section 2 we present the basic structure of the model and provide closed-form solutions for asset values and the term structure. In Section 3, we formulate the structural form model for credit risk in our setup. In Section 4, we provide a calibration of investors’ beliefs of the underlying states that uses information in both fundamentals and financial variables. Section 5 provides an analysis of the credit spreads puzzle and its synthesis with the equity premium puzzle. Section 6 concludes. Proofs of all propositions (unless explicitly stated) are in Appendix 1. Appendix 2 contains the SMM methodology for estimating the asset value process and the parameters of the model, and Appendix 3 provides a detailed description of
the projection method used to solve the PDE for corporate bond prices. All three appendices will be made available to readers.

2 Structure of the Model

We consider one of the simplest defaultable corporate bond structural-form models. We assume that the firm continuously issues a single class of discount debt with maturity $T$, there is no chance of default prior to maturity of the debt, there are no bankruptcy costs, and finally, there are no taxes. These assumptions simplify the exposition, but are not responsible for main results. Under these assumptions, one can use results similar to Merton (1974) and use option pricing methods to price defaultable bonds.

**Assumption 1:** The price of the single homogeneous good in the economy, $Q_t$, evolves according to the log-normal process

$$\frac{dQ_t}{Q_t} = \beta_t dt + \sigma_Q dW_t,$$

where the process followed by $\beta_t$ is described below and $\sigma_Q = (\sigma_{Q,1}, \sigma_{Q,2}, 0)$ is a $1 \times 3$ constant vector known by investors.

**Assumption 2:** Real earnings, $E_t$, evolves according to the log-normal process

$$\frac{dE_t}{E_t} = \theta_t dt + \sigma_E dW_t,$$

where $W_t = (W_{1t}, W_{2t}, W_{3t})'$ is a three-dimensional vector of independent Weiner processes, the $1 \times 3$ constant vector $\sigma_E = (0, \sigma_{E,2}, 0)$ is assumed known and constant. The process for $\theta_t$ is described below.

We use these fundamentals to price all assets in our model. In order to do so we need to specify a stochastic discount factor to be used to discount real payoffs. We make the following assumption:

**Assumption 3:** There exists a real pricing kernel $M_t$ taking the form:

$$\frac{dM_t}{M_t} = -k_t dt - \sigma_M dW_t,$$

where $\sigma_M = (\sigma_{M,1}, \sigma_{M,2}, \sigma_{M,3})$ is a $1 \times 3$ constant vector of the market prices of risk, and $k_t = \alpha_0 + \alpha_\theta \theta_t + \alpha_\beta \beta_t$, is the real short rate of interest. $M$ is used to price real claims and determine the expected real returns of all securities. We restrict the real rate of interest to be a linear function of the two (hidden) state variables of the model.

The kernel is observed by investors but not by the econometrician. Using it, agents price a generic real payout flow of $\{d_t\}$ as
where $\mathcal{F}_t$ is investors’ filtration to be described in Assumption 5. As in several recent papers (see for example Berk, Green, and Naik 1999, Brennan and Xia 2002) we have specified an exogenous pricing kernel process to formulate equilibrium relationships among endogenous financial variables. The linear dependence of the real rate on the drifts of fundamentals has a theoretical basis in general equilibrium models: For example, in a Lucas (1978) economy where investors have power utility $U(C,t) = e^{-\delta t \sigma_{C}^{\gamma}}$, we would have $C_t = E_t$, $M_t = U'(E_t)$, and hence $k_t = \phi + \gamma \theta_t + \frac{1}{2} \gamma (1 - \gamma) \sigma_E \sigma_E'$ and $\sigma_M = \gamma \sigma_E$. In this case, the real rate is not affected by the inflation drift, that is, $\alpha_\beta = 0$. Our specification generalizes the real rate process to economies where expected inflation affects the real cost of borrowing. The specified real rate can be supported in an equilibrium if we allow a storage technology that can be bought or sold, is in perfectly elastic supply, and offers a safe instantaneous return of $k_t$. We provide further motivation for its specified functional form next.

The literature on the effect of expected inflation on the real rate of return is extensive, and it is beyond the scope of this paper to build in all the effects without significantly complicating our analysis. At the micro level there is a fairly large literature on various tax and accounting channels through which expected inflation affects the real return on capital (Feldstein 1980a, Feldstein 1980b, Auerbach 1979, Cohen and Hassett 1999). Among the several monetary channels that lead to the non-neutrality of money, of particular relevance are cash-in-advance models in which expected inflation is a tax on money balances and raises the real cost of transactions, thus affecting the real interest rate, capital accumulation and business cycle variation (see, e.g., Svensson 1985, Lucas and Stokey 1987). Cooley and Hansen (1989) calibrate such a model to the US economy and report a positive relationship between expected inflation and the marginal product of capital (the real rate).\footnote{The papers by Feldstein point out that the typical depreciation allowance is based on the original or ‘historic’ cost of the asset rather than its current value. When prices rise, this historic cost method of depreciation causes the real value of depreciation to fall and the real value of taxable profits to rise, increasing the real return on invested capital. This effect can be countered by the effect of expected inflation on the value of inventories depending on the method of accounting used. If firms use the first-in-first-out method of accounting, then the cost of maintaining inventory levels is understated in the presence of inflation, hence lowering the real return on capital. The opposite effect holds with the last-in-first-out accounting method. Auerbach (1979) and Cohen and Hassett (1999) extend this analysis by studying further distortions on the choice of maturity of real assets and investment.\footnote{The cash-in-advance constraint applies only to the consumption good while leisure and investment are modeled as credit goods. These authors use a constant returns to scale production function of the form $f(K,N) = K^\theta N^{1-\theta}$, and calibrate the response of a representative agent to increases in the expected rate of inflation. They report that both capital (K) and labor (N) choices decrease as an optimal response to an increase in the expected rate of inflation, which could lead to either an increase or decrease in the marginal product of capital. For the calibrated economy, the net result is positive (see Table 2 of Cooley and Hansen 1989).}
Given the multiplicity of channels that affect the relationship between expected inflation and the real short rate, we treat the sign and size of this relationship as an empirical question that can be estimated from the joint time series variation of fundamentals and asset prices. Our chosen functional form of the real rate can be seen as a reduced form for the above noted effects.\footnote{Using related but different empirical methodologies from ours, Geske and Roll (1983) and more recently Brennan and Xia (2002) both find a positive relationship. Geske and Roll (1983) (page 12) write: The recent experience from a period (1981-82) characterized by large deficits and a low rate of Federal Reserve monetization seems to be that real rates have increased dramatically. At least, journalists, investment bankers, and foreign government leaders seem to think so.}

For notational convenience we stack the fundamental processes (1) and (2). Let $Y_t = (Q_t, E_t)'$, so that
\[ \frac{dY_t}{Y_t} = \varrho_t dt + \Sigma_2 dW_t, \]
where $\frac{dY_t}{Y_t}$ is to be interpreted as “element-by-element” division, $\varrho_t = (\beta_t, \theta_t)'$, and $\Sigma_2 = (\sigma'_Q, \sigma'_E)'$

Similarly, we find it useful to add the kernel process to the stacked vector and write $X_t = (Q_t, E_t, M_t)'$, which has the drift vector $\nu_t = (\beta_t, \theta_t, -k_t)'$, and volatility matrix $\Sigma = (\sigma'_Q, \sigma'_E, -\sigma_M)'$

**Assumption 4:** $\nu_t$ follows an $N$-state, continuous-time finite state Markov chain with generator matrix $\Lambda$,\footnote{The transition matrix over states in a finite interval of time, $s$, is $\exp(\Lambda s)$ (see e.g. Karlin and Taylor 1982).} that is, over the infinitesimal time interval of length $dt$
\[ \lambda_{ij} dt = \text{prob}(\nu_{t+dt} = \nu_j | \nu_t = \nu_i), \quad \text{for} \quad i \neq j, \quad \lambda_{ii} = -\sum_{j \neq i} \lambda_{ij}. \]

**Assumption 5:** Investors do not observe the realizations of $\nu_t$ but know all the parameters of the model.

Since investors do not observe $\nu_t$, they need to infer it from the observations of past earnings, inflation, and the pricing kernel. This will generate a distribution on the possible states $\nu_1, \ldots, \nu_N$ that in turn generates changes in “uncertainty” as they learn about the current state. At time $t$ investors’ distribution over hidden states is summarized by the posterior probabilities
\[ \pi_{it} = \text{prob}(\nu_t = \nu_i | F_t) \]
where $F_t$ is the filtration generated by observing the entire path $(X_s)_{0 \leq s \leq t}$. Let $\pi_t = (\pi_{1t}, \ldots, \pi_{Nt})$ be the vector of beliefs.

**Lemma 1** *Given an initial condition $\pi_0 = \tilde{\pi}$ with $\sum_{i=1}^N \tilde{\pi}_i = 1$ and $0 \leq \tilde{\pi}_i \leq 1$ for all $i$, the probabilities $\pi_{it}$ satisfy the $N$-dimensional system of stochastic differential equations:
\[ d\pi_{it} = \mu_i(\pi_t) dt + \sigma_i(\pi_t) d\tilde{W}_t, \]
\[ \mu_i(\pi_t) = [\pi_i \lambda_j], \quad \sigma_i(\pi_t) = \pi_{it} \left[ \nu_i - \mathcal{V}(\pi_t) \right]' \sum^{-1}, \quad (7) \]

\[ \mathcal{V}(\pi_t) = \sum_{i=1}^{N} \pi_{it} \nu_i = E_t \left( \frac{dX_t}{X_t} \right), \quad \text{and} \]

\[ d\tilde{W}_t = \sum^{-1} \left( \frac{dX_t}{X_t} - E_t \left( \frac{dX_t}{X_t} \right | \mathcal{F}_t \right) = \sum^{-1} (\nu_t - \mathcal{V}(\pi_t))dt + dW_t. \quad (8) \]

Moreover, for every \( t > 0 \), \( \sum_{i=1}^{N} \pi_t = 1. \)

**Proof.** See Wonham (1964), Liptser and Shiryaev (1977), or David (1993).

The filtering theorem for jumps in the underlying drift was first derived (to the best of our knowledge) in Wonham (1964). David (1993) provides a proof using the limit of Bayes’ rule in discrete time. The first application of this theorem in financial economics, as well as several properties of the filtering process, are derived in David (1997).

We make the following summary comments about the updating process (6): (i) The elements of the generator matrix, \( \Lambda \), completely capture the transition probabilities between states. Absent any new information, beliefs tend to mean-revert to the unconditional stationary probabilities that are completely determined by \( \Lambda \). For example, if there are only two states with \( \lambda_{12} = 2 \) and \( \lambda_{21} = 1 \), then the belief that the economy is in state 1 mean reverts to \( 1/3 \). (ii) The diffusion term describes the change in beliefs due to new information. When they have high uncertainty about the current state vector, they react more to news, and the volatility of their beliefs increases. (iii) Conditional on a given level of uncertainty, the volatility of beliefs is proportional to the parameters determining the ratio of signal (difference in drifts across states) to noise (\( \Sigma \)). (iv) Agents update their beliefs about underlying states by observing not only the path of fundamentals, but also their pricing kernel. This is analogous to the Lucas economy mentioned above in which the agent would learn about the drift rate of the marginal utility of consumption.

It is useful to note that investor’s beliefs change with their inferred shocks, \( d\tilde{W} \), in equation (8) as opposed to the true shocks, \( dW \), that affect fundamentals. Similarly, they infer that the process \( \{X_t\} \) follows: \( dX_t/X_t = \nu(\pi_t)dt + \Sigma d\tilde{W}_t \). At each point in time, substituting the definition of \( d\tilde{W}_t \), we have \( dX_t/X_t = \nu(\pi_t)dt + \Sigma [\sum^{-1} (\nu_t - \mathcal{V}(\pi_t))dt + dW_t] = \nu_t dt + \Sigma dW, \) which is the same process as in Assumptions 1-3. In the filtering literature \( d\tilde{W}_t \) is an “innovations” process under the investors’ filtration and under the separation principle it can be used for dynamic optimization. See David (1997) for a discussion. As a special case we will write

\[ \frac{dY_t}{Y_t} = \bar{\phi}(\pi_t)dt + \Sigma_2 d\tilde{W}_t = \phi(\pi_t)dt + \Sigma_2 dW_t, \quad (9) \]
and the kernel under investors’ filtration as \(dM_t = -\bar{k}(\pi_t)dt - \sigma_M d\bar{W}_t\), where the real rate in the economy, \(\bar{k}(\pi_t)\), is its expected value conditional on investors’ filtration.

### 2.1 Asset Value Process and the Term-Structure of Interest Rates

For evaluating nominal claims and nominal risk premiums we will also use the nominal pricing kernel, \(N_t = M_t/Q_t\), which follows

\[
\frac{dN_t}{N_t} = -r_{it}dt - \sigma_N dW_t,
\]

where \(r_{it} = k_{it} + \beta_{it} - \sigma_N \sigma'_Q\) and \(\sigma_N = \sigma_M + \sigma_Q\). The nominal rate differs from the real rate by the sum of expected inflation and the inflation risk premium, which is the covariance between inflation and the nominal pricing kernel. With unobserved states, the projected nominal interest rate at time \(t\) is \(\sum_{i=1}^{N} r_i \pi_{it}\). The following proposition provides expressions for the value-to-earnings (henceforth V/E) ratio and the nominal bond price:

**Proposition 1**

(a) The V/E ratio at time \(t\) is

\[
\frac{V_t}{E_t}(\pi_t) = \sum_{j=1}^{N} C_j \pi_{jt} \equiv C \cdot \pi_t,
\]

where the vector \(C = (C_1, \ldots, C_N)\) satisfies \(C = A^{-1} \cdot 1_N\),

\[
A = \text{Diag}(k_1 - \theta_1 + \sigma_M \sigma'_E, k_2 - \theta_2 + \sigma_M \sigma'_E, \cdots, k_N - \theta_N + \sigma_M \sigma'_E) - \Lambda.
\]

(b) The price of a nominal zero-coupon bond with maturity \(\tau\) is

\[
B_t(\pi_t, t, T) = \sum_{i=1}^{N} \pi_{it} B_i(t, t + \tau),
\]

\[
B_i(t, t + \tau) = E \left( \frac{M_T}{M_t} \cdot \frac{Q_T}{Q_t} | \nu_t = \nu_i \right) = \left[ \sum_{i=1}^{N} \Omega_i e^{\omega_i \tau} \right] \cdot (\Omega^{-1} 1_N),
\]

where \(\Omega_i\) and \(\omega_i, i = 1, \cdots, N\), are the \(i\)th eigenvector and \(i\)th eigenvalue respectively of the matrix \(\hat{\Lambda} = \Lambda - \text{Diag}(r_1, r_2, \cdots, r_N)\).

Notice that an `averaging’ result holds for linear securities: prices conditional on investors’ filtration are the averages of prices that would hold in the \(N\) regimes if they were observed by investors. This ‘averaging’ property does not hold for non-linear securities prices such as defaultable bonds that are derived in the next section. The proof follows from minor modifications to the discrete-time proof in David and Veronesi (2006). Also see Veronesi (2000), Veronesi and Yared (1999), and David and Veronesi (2002) for derivations in continuous time with similar assumptions.
In (a) each constant $C_i$ represents investors’ expectation of future earnings growth conditional on the state being $\nu_i$ today, discounted using the process for the pricing kernel process $\{M_t\}$, and normalized to make it independent of the current payout and time. Hence, a high $C_i$ implies that investors assess a high value relative to current earnings in state $\nu_i$. Since they do not actually observe the state $\nu_i$, they weight each $C_i$ by its conditional probability $\pi_i$ thereby obtaining (11). Notice in particular that the form of the constant vector $C$ suggests that: (i) if $\alpha_\theta < 1$, a higher growth rate of earnings implies a higher V/E, (ii) if $\alpha_\beta > 0$, a higher inflation state implies a lower V/E, which is the real rate effect of inflation; (iii) in addition, the V/E ratio in a given state of growth depends on the future sustainability of the growth rate and is determined by the transition probabilities $\lambda_{ij}$ as shown in the solution to the $N$-equations in (a).

Similarly, the bond price is a weighted average of the nominal bond prices that would prevail in each state $\nu_i$. Since investors do not actually observe the current state, they price the bond as a weighted average. Both higher inflation and higher growth rate of earnings lead to lower long term bond prices when $\alpha_\beta > -1$ and $\alpha_\theta > 0$. It’s useful to notice that all asset prices follow continuous paths even though the drift rates for earnings and inflation jump between a discrete set of states. This results from the continuous updating process.

Using the characterization of inflation and earnings, processes under the observed filtration, we are now able to specify the processes for the real and nominal value of the firm’s asset value.

**Proposition 2**

(a) The nominal asset return process under the investor’s filtration is given by:

$$
\frac{dV_t^Q}{V_t^Q} = (\mu_V^Q(\pi_t) - \delta(\pi_t)) dt + \sigma_V^Q(\pi_t) d\tilde{W}_t,
$$

where $\delta(\pi_t) = \frac{1}{C_t, \pi_t}$, the expression for $\mu_V^Q(\pi_t)$ is in Appendix 1, its volatility is

$$
\sigma_V^Q(\pi_t) = \sigma_E + \sigma_Q + \sum_{i=1}^N C_i \pi_{it} (\mu_t - \bar{\pi}(\pi_t))'(\Sigma')^{-1},
$$

and the vector $C$ is determined in Proposition 1. The volatility of the real asset process satisfies:

$$
\sigma_V(\pi_t) = \sigma_V^Q(\pi_t) - \sigma_Q.
$$

(b) The real and nominal risk premiums satisfy $\mu_V(\pi_t) - k(\pi_t) = \sigma_V(\pi_t)\sigma_M'$, and

$$
\mu_V^Q(\pi_t) - r(\pi_t) = \sigma_V^Q(\pi_t)\sigma_N'
$$

respectively.

This proposition shows that asset volatilities have an exogenous component due to noise in the fundamental process and a learning-based endogenous component. The latter is larger when investors are more uncertain about the state of underlying fundamentals because in such times investors react more to incoming fundamental news. Nominal (real) risk premiums are determined
by investors as the negative of the covariances of the nominal (real) asset value with the nominal (real) pricing kernel.

The following proposition states conditions under which the volatility of the firm’s asset value with respect to the first shock (inflation) is mostly negative, while the volatility with respect to the second shock (earnings) is positive. We will say state vectors $x$ and $y$ are (weakly) positively related when $x_i \geq x_j$ iff $y_i \geq y_j$.

**Proposition 3** Suppose the inflation state vector, $\beta$, and the asset V/E state vector, $C$, are negatively related, and the earnings state vector, $\theta$, and $C$ are positively related, and all volatilities, $\sigma_{Q,1}$, $\sigma_{Q,2}$, and $\sigma_{E,2}$ are all positive. Then, $\sigma_{V,1} \leq 0$, $\sigma_{V,1}^Q \leq \sigma_{Q,1}$, $\sigma_{V,2} \geq 0$, and $\sigma_{V,2}^Q \geq 0$.

The proposition is intuitive: if the V/E ratio is lower in states when the inflation drift is higher, then if realized inflation is higher than expected, investors’ conditional probability of higher inflation increases and asset values decline. Such a negative correlation between realized real asset returns and realized inflation has been noted by several authors. In our empirical section we examine the estimated values of the fundamentals states and V/E ratios in the hidden states and verify that the stated conditions in the proposition are satisfied for the estimated parameters. In addition, we see that the exogenous component of inflation volatility, $\sigma_{Q,1}$, is ‘small’ — of the order of one percent at an annual rate — so that in most periods $\sigma_{V,1}^Q < 0$. The sign of $\sigma_{V,3}$ is in general indeterminate; however, we will find that for our estimated parameters this volatility is negative in each period in our sample.

The risk of regime switches in the model is systematic and affects the prices of defaultable bonds in the next section. As for all non-traded assets in equilibrium models, for risk-neutral valuation, the drifts of the belief processes must be reduced by their market prices of risk (see, e.g., Cox, Ingersoll, and Ross 1985, Hull 1993). As in these frameworks, the market price of risk of $\pi_i$ is the negative of its covariance with the pricing kernel of the economy. The analytical expression is given for these market prices in the following proposition.

**Proposition 4** The nominal market price of risk of the state variable $\pi_{it}$, for each $i \in \{1, \cdots, N\}$, investors’ conditional probabilities of the economy being in state $i$, is time-varying and is given by

$$r_t(\pi_t) = \pi_{it} \cdot (r_{it} - \bar{r}(\pi_t)).$$

(16)

It is evident that even though the market prices of risks associated with the two shocks are constant, the market prices of risks for the beliefs are time-varying, and increase in the fundamental
uncertainties of earnings and inflation. The risk adjustments in a sense distort the mean reversion of the belief processes and effectively change the generator elements in Assumption 4.  

3 Price of Defaultable Debt

The firm issues a single class of zero coupon debt. If the real face value of debt remains constant credit spreads in structural form models tend to zero as maturity increases (Collin-Dufresne and Goldstein 2001, Ericsson and Reneby 2002). We will instead assume that the firm issues nominal debt at a pace that ensures that there is no trend in the distance between the nominal asset value and the nominal face value of debt outstanding. Black and Cox (1976) motivate the same assumption as a safety covenant for bondholders. In support of this assumption, the empirical work in Eom, Helwege, and Huang (2004) suggests no additional role for maturity (besides very short maturity bonds) in fitting spreads beyond the other inputs in structural form models.

Assumption 6: The nominal face value of debt at time $t$ is given by

$$D_t^Q = D_0^Q \cdot \exp\left[ \int_{s=0}^{t} (r(\pi_s) - \delta(\pi_s)) ds \right],$$  \hspace{1cm} (17)

where $r(\pi_t)$ is the nominal rate of interest defined in Proposition 1(b), and $D_0^Q$ is the initial face value of debt of the firm, and $\delta(\pi_t)$ is the earnings-yield in Proposition 2. The net growth rate of debt $r(\pi_t) - \delta(\pi_t)$ can result from gross issuance at the riskless rate and retirement at the earnings yield, or any such combination.

Assumption 7: Similar to Merton (1974), we assume the firm defaults at maturity, $T$, if the market value of the firm’s asset in nominal dollars at maturity is below the face value of its debt in nominal dollars $D_T^Q$.

Assumption 8: The asset value of the firm is a traded security.

The assumption and Proposition 2 imply that the nominal asset value of the firm can be written as

$$dV_t^Q/V_t^Q = (r(\pi_t) - \delta(\pi_t)) dt + \sigma_t^Q dW_t^*, \text{ where } dW_t^* = d\tilde{W}_t + \sigma_N dt,$$

where $\sigma_N = \sigma_M + \sigma_Q$.

The point is made best in the 2-state case. In this case, $\pi_1$ mean reverts to $\pi_1^* = \lambda_{21}/(\lambda_{12} + \lambda_{21})$ under the objective measure. However, under the risk-neutral measure, $\pi_1$ mean reverts to the positive root of $(\lambda_{12} + \lambda_{21})(\pi_1^* - \pi) - (r_1 - r_2) \pi (1 - \pi_1)$, which is smaller than $\pi_1^*$ when $r_1 > r_2$. Therefore, less probability is assigned to higher interest rate states. Such an effective change in the generator element under the risk-neutral measure is also evident in the analysis of Pan (2002) who prices systematic jump risk in derivative securities, and the ratings-based credit risk model of Jarrow, Lando, and Turnbull (1997). An advantage of our approach is that the functional form of these risk adjustments is endogenously derived.
are nominal market prices of risk. Define the solvency ratio of the firm as \( Z_t = V^Q_t / D^Q_t \), and \( z_t = \log Z_t \). Under the nominal risk-neutral measure the solvency ratio follows:

\[
\frac{dZ_t}{Z_T} = \sigma^Q_V(\pi_t) \, d\tilde{W}_t. \tag{18}
\]

By standard no-arbitrage pricing (Duffie 1996, Chapter 5.G) every contingent claim \( f(t, z_t, \pi_t) \) must satisfy the fundamental partial differential equation (PDE)

\[
rf(\pi_t) = \frac{\partial f}{\partial t} - \frac{1}{2} \frac{\partial^2 f}{\partial z^2} \sigma^2_V(\pi_t) + \sum_{j=1}^N \frac{\partial f}{\partial \pi_j} (\mu_j(\pi_t) - \rho_j(\pi_t)) + \sum_{j=1}^N \frac{\partial^2 f}{\partial \pi_j^2} \sigma^2_j(\pi_t) + \frac{1}{2} \sum_{k=1}^N \sum_{j=1}^N \frac{\partial^2 f}{\partial \pi_k \partial \pi_j} \sigma_{kj}(\pi_t), \tag{19}
\]

where \( \mu_i(\pi) \) and \( \sigma_j(\pi) \) are given in eqs. (7), \( \sigma^Q_V(\pi_t) \) is in eq. (15), and \( \rho_i(\pi_t) \) is given in (16). In addition, natural boundary conditions at 0 and 1 for each belief process and terminal conditions determined by the final payoff of the contingent claim are imposed.

At maturity, the payoff for each dollar of face value of debt outstanding is \( \text{Min}(Z_T, 1) \). That is, in default, the recovery on each dollar of defaultable debt equals the solvency ratio. We extend the Fourier Transform methodology that has so far been used in the affine framework (see, e.g., Heston 1993) to solve for defaultable bond prices in the following proposition.

**Proposition 5** The defaultable bond price for a firm with nominal asset value \( V^Q_t \), nominal debt outstanding \( D^Q_t \), and hence \( Z_t = V^Q_t / D^Q_t \), and when investors’ beliefs are \( \pi_t \) satisfies

\[
P(Z_t, \pi_t, t, T) = B(\pi_t) \cdot \Pi_2(Z_t, \pi_t, t, T) + G(\pi_t, z_t) \cdot (1 - \Pi_1(Z_t, \pi_t, t, T)), \tag{20}
\]

where \( \Pi_1(\cdot, \cdot) \) and \( \Pi_2(\cdot, \cdot) \) are the time-t prices of Arrow-Debreu securities that pay a dollar if the firm is solvent at \( T \) (\( Z_T > 1 \)) under different measures. These two probabilities can be found by fast Fourier Inversion of the characteristic function of the discounted log solvency ratio under the risk-neutral measure

\[
f(z_t, \omega_1, \pi_t, T - t) = E[ e^{-\int_t^T r(s) \, ds} e^{j \omega_1 z_T} | z_t, \pi_t].
\]

\( B(\pi_t) = f(z_t, 0, \pi_t, T - t) \) and \( G(\pi_t, z_t) = f(z_t, 1/i, \pi_t, T - t) \) are obtained by simply evaluating \( f(\cdot, \cdot) \) at two points, and are the bond price, and forward solvency ratio, respectively, under the risk-neutral measure.

**Proof:** In Appendix 2.
The characterization of the risky-bond price is analogous to that in Merton (1974), with the extension to the case of stochastically growing debt of the firm. The call option value (equity of the firm) for the case where \( \pi_t \) is 1-dimensional was solved in David and Veronesi (2002). In this paper, we extend the solution to the general case where \( \pi_t \) is \( N \)-dimensional, and in addition, interest rates and the default boundary are both stochastic. An \( M = 220 \) term polynomial approximation of \( f(\cdot, \cdot) \) is provided in eq. (48) in the proof.

For convenience, we also write the \( T \)-period spread on the defaultable bond as

\[
s(Z_t, \pi_t, t, T) = \frac{1}{T} \log \left[ \Pi_2(\cdot, \cdot) + \frac{G(\pi_t, z_t)}{B(\pi_t)} \cdot (1 - \Pi_1(\cdot, \cdot)) \right].
\] (21)

We see in the following corollary that the credit spread is a decreasing and convex function of the solvency ratio, \( Z_t \), of the firm. As we will see in Section 5, the convexity property will partially resolve the credit spreads puzzle.

**Corollary 1** (i) The defaultable bond price \( P(Z_t, \pi_t, t, T) \) is an increasing and concave function of the solvency ratio, \( Z_t \). (ii) The credit spread \( s(Z_t, \pi_t, t, T) \) is a decreasing and convex function of the solvency ratio.

We end this section with some comments on the various channels through which unexpected inflation (as measured by inflation volatility) and expected inflation (as measured by the filtered beliefs) will affect the firm’s probability of default under the risk-neutral measure in our model. As seen from Proposition 2, the volatility of firms’ nominal asset value equals the sum of its real volatility and inflation volatility, while by Assumption 6, debt growth is locally smooth. Its direct implication is that credit risk for firms increases with inflation volatility as the defaultable bond price is the value of the safe bond price less a put option on the nominal asset value of the firm written by bondholders and purchased by equity holders. An increase in inflation volatility increases the value of this put option.

By Corollary 1(ii), credit spreads are decreasing in \( Z_t(\pi_t) = \frac{\sum_{i=1}^{N} C_i \pi_{it} E_t Q_t}{D_t^{\pi}} \). Therefore, the comments on the V/E ratios in the \( N \) states, \( C \), below Proposition 1 apply. Their implications are that *ex-ante*, the following effects hold:

1. For \( \alpha_{\beta} > 0 \), an increase in expected inflation increases the real rate of interest. This has an effect on lowering \( Z_t \) and increasing spreads.

2. If the transition probabilities to lower earnings states are higher in higher earnings states then an increase in expected inflation lowers \( Z_t \) and raises spreads.
3. In addition, if the transitions between earnings states are more likely in higher inflation states, higher expected inflation raises investors’ forecast of the endogenous component of future asset volatility, and since spreads are convex in $Z_t$, causes them to increase. David and Veronesi (2006) examine the volatility forecasting power of inflation-based uncertainty measures.

4 Calibration

In this section we provide a brief description of the calibration methodology and the parameter estimates of our model.

4.1 Calibration Methodology

We calibrate the model by using information in both fundamental and financial variables to obtain time series of investors’ beliefs over fundamental states as well as the underlying parameters. Suppose the V/E series is observable: then using it, the data on fundamentals, and the Treasury yields, we estimate the parameters using a SMM method as described below.

Let $\Psi$ denote the set of structural parameters in the fundamental processes of earnings growth and inflation in (1), (2), and the pricing kernel in (3). Let the likelihood function for the fundamentals data observed at discrete points of time (quarterly) be $L$. The simulation-based procedure used to compute $L$ also generates series of filtered probabilities that investors have of each of the underlying states. To extract information about investors’ beliefs from fundamentals as well as equity prices and Treasury bond yields, we add moments of asset prices and use the pricing formulae for the V/E ratio and Treasury bond prices in Proposition 1 to generate model-determined moments.

Let $\{e(t)\}$ denote the errors of the pricing variables, and let $\epsilon(t) = (e(t)', \frac{\partial \Sigma}{\partial \psi}(t))'$ where the second term is the score of the likelihood function of fundamentals with respect to $\Psi$. We now form the SMM objective:

$$c = \left( \frac{1}{T} \sum_{t=1}^{T} \epsilon_t \right)' \cdot \Omega^{-1} \cdot \left( \sum_{t=1}^{T} \frac{1}{T} \cdot \epsilon_t \right). \quad (22)$$

The details of the procedure are in Appendix 2.

Since the asset value series is in fact not observed, our estimation problem falls under the class of “missing data” problems, for which the Expectations Minimization (EM) procedure is applicable. The procedure is recursive. In brief, we use estimates of bond prices, the parameter estimates, and beliefs at the nth stage to estimate the asset value at time $t$ as $V_t^{Q(n)} = P_t^E + P_t^{(n)} \cdot D_t^Q$, where $P_t^E$ is the market value of the S&P 500 at each date, $D_t^Q$ is formulated using Assumption 6 with $D_0^Q$ set so
that the initial solvency ratio of the firm is $k$, when the initial V/E ratio of the firm is its unconditional value of $C \cdot \pi^*$, and $k$ is the long-run solvency ratio of a given rating category. The estimated stage $n$ V/E ratio at date $t$ is simply $V_t^{Q(n)}/E_t$, and the solvency ratio is $Z_t^{(n)} = V_t^{Q(n)}/D_t^Q$. Now the SMM estimation procedure is used to minimize (22) conditional on the $n$th stage bond prices, and to find the optimal parameter values of stage $n + 1$. New estimates of bond prices are obtained using Proposition 5. The steps are repeated until convergence. Further details are provided in Appendix 2.

The first formal description of the general EM methodology is in Dempster, Laird, and Rubin (1977). More recently, Duan, Gauthier, and Simonato (2004) provide an equivalence between the EM procedure used in industry by Moody’s KMV (see Crosbie and Bohn 2002) and and the maximum likelihood (ML) procedure to estimate structural credit risk models in Duan (1994). Ericsson and Reneby (2005) extend the ML approach to allow for stochastic interest rates and more general default boundaries. Our method generalizes the EM procedure in addition to account for both a stochastic term structure and volatility. We note that upon convergence, our empirical procedure for fitting the asset value process is internally consistent in that bond prices at each date are computed using a formula that incorporates both features.

The simultaneous joint estimation of the unobserved asset value and asset volatility has posed a conundrum to both academics and practitioners, especially when asset volatility varies stochastically over time. The most popular approach has been the two equation and two unknown problem: (a) equity is a call option on the asset value, $E = \text{Call Option}(V, \sigma_V)$, and (b) equity and asset volatilities are related by $\sigma_E = \sigma_V E_V V/E$ (see, e.g., Jones, Mason, and Rosenfeld 1984). Both relationships hold only when the volatilities are constant. In their implementation, however, researchers back out time-varying volatilities. More recently some authors have used historically inferred asset values from the call option inversion using (a) (Vassalou and Xing 2004), or have used historical volatilities of equities and transformed to asset volatilities using (b) (Eom, Helwege, and Huang 2004), to obtain current estimates of asset volatility. In addition to being internally inconsistent, these methods are necessarily backward looking. In contrast, our method uses forward-looking information in asset prices to back out investors’ current beliefs, formulates stochastic asset volatility as in eq. (15) in Proposition 2, and uses the bond pricing formula in Proposition 5 that is consistent with this stochastic volatility. Finally, our calibration method is similarly consistent with the valuation effects of a stochastic term structure of interest rates.
4.2 Estimation Results for the Regime Switching Model

In this subsection, we briefly describe the results of the estimation of the regime switching model for the fundamental variables — earnings and inflation. We start with the description of the data series used. Equity values are estimated using the S&P 500 operating earnings-per-share, and the price-to-operating earning ratio series from 1960-2001 are from Standard and Poor’s. The time series of nominal earnings is deflated using the Consumer Price Index series, which is also used to compute the time series of inflation levels. The time series of constant maturity zero coupon yields are from the Federal Reserve Board.

We estimate a model with three regimes for inflation, $\beta_1 < \beta_2 < \beta_3$, and and two regimes for earnings growth, $\theta_1 < \theta_2$, which lead us to six composite states overall. Some further comments on the selection of number of regimes in each series are in Appendix 2. In our estimation, we found that the unconstrained estimates of the transition matrix led to several zero elements, leading to a more parsimonious four state model $\{(\beta_1, \theta_2), (\beta_2, \theta_1), (\beta_3, \theta_2), (\beta_3, \theta_1)\}$. That is, we found that the $(\beta_1, \theta_1)$ and $(\beta_3, \theta_2)$ states had close to zero probability of occurring in the sample. Overall, the four state and six state model lead to almost the same value for the SMM objective function. Gray (1996) and Bansal and Zhou (2002) use a similar criterion for the choice among alternative regime specifications.

The top panel of Table 1 reports the means and volatilities of fundamentals, as well as the pricing kernel parameters. The middle panel reports the transition probability matrix as well as the asset prices in the four states. We estimate that inflation averages 1.3, 4.3 and 0.3 percent in the three states, while earnings growth averages negative 9 percent and 5.1 percent in the low and high growth states respectively. A few comments are in order: First, while all the states have statistically significant estimates, the standard errors for the inflation states are much smaller than for the earnings growth states. The volatility parameters and the top panels of Figure 3 show that inflation is much less noisy than the earnings series, leading to better estimates for the former series. Second, the high growth rate of earnings is far more persistent in the low inflation state: from the (LI-HG) state, there is a 98 percent chance of returning to this state, and a 2 percent chance of transitioning to the (MI-HG) state in a year, therefore there is almost a zero chance of growth slowing in a quarter in which there is also low inflation. From the (MI-HG) state on the other hand there is an 8.2 percent chance of a transition to the (MI-LG) state in the following quarter. This is the signaling role of inflation – it provides an early warning of unsustainable high growth of fundamentals. The overall SMM objective function value, which has a chi-squared distribution
with six degrees of freedom, is 9.23, implying a p-value larger than 16 percent, so we fail to reject our model.

We next turn to the kernel parameter estimates, the fourth set of parameters in Table 1. We notice immediately that \( \alpha_\theta \) and \( \alpha_\beta \) are both significantly positive, which means that the real rate is higher in states of faster growth of real fundamentals as well as states of higher inflation. *Ceteris paribus* we will have therefore lower V/E ratios and higher Treasury yields in such periods. These observations are confirmed in the bottom panel, which reports the implicit parameters for the V/E and bond yields across the states using the pricing formulas for stocks and bonds in eqs. (11) and (13) respectively.

As we will see in the next section, the market prices of risk in our model play an important role in determining the relation between the credit spreads and equity premium puzzles. Notice that \( \sigma_{M,1} \) and \( \sigma_{M,2} \) are both positive. By Proposition 3 we know that the volatility of the asset value with respect to the inflation shock is negative, but with respect to earning shock is positive (it is straightforward to verify from Table 1 that the inflation and V/E state vectors are negatively related as is needed in this proposition), which implies that the risk premiums associated with inflation shocks is *negative*, and with earnings shocks is *positive*. The negative covariance between expected inflation and expected returns for stocks has been noted by several authors (see, e.g., Fama 1981, Stulz 1986). While our model is not consumption based, our interpretation is that positive shocks to inflation occur during periods of weak earnings growth and therefore investors are willing to accept a negative risk premium to bear them. Positive earnings shocks occur in periods of strong real fundamentals and investors therefore require a positive risk premium to bear these. Finally, the price of risk of the kernel shock and its volatility are both negative, leading to a positive (but small) risk premium.

Besides the kernel parameters, the transition probabilities also have a large effect on V/E\( \text{s}. \) Consider the two states of medium inflation growth. If there were no regime switches, the V/E computed using Proposition 1 would be around 6.6 and 16.7 in the low and high growth rate states. However, with the estimated switching frequencies, the V/E\( \text{s} \) are 11.8 and 13.3, that is, the differences in V/E are to a great extent eliminated. This is of course due to the instability of high growth in the medium inflation state. The implied parameters also show why the slope of the term structure is another proxy for this signal. There is an extensive literature documenting the forecasting power of the slope of the term structure (see, e.g., Estrella and Mishkin 1998). In our calibrated model, the slope is positive in the LI-HG, and MI-LG states, but is negative in the MI-HG, and MI-HG states, in each of which the 3-month rate is in double digits. From the transition probability matrix we see
that in each of these latter states there is a large probability of low growth in the following quarter. Therefore in the model the slope is a proxy for investors’ assessed risk of the economy shifting to a low growth state.

The time series of the filtered probabilities are shown in Figure 2. The 1960s and 1990s were decades when the low inflation and high earnings state prevailed. There were several switches within the medium inflation states in the 1970s, with two bouts of high inflation and low growth in the mid-1970s and early 1980s. The mid to late 1980s were characterized by high probabilities of medium inflation and high growth, before the low inflation boom years of the 1990s. Among more recent dates, by late 1999 the state of medium inflation and high growth had become a likely possibility, that was followed by a spike of the medium inflation and low growth probability in the Spring of 2000, causing a severe drop in the model implied V/E. These filtered beliefs and the above implied parameters are used to generate time series of model-implied time series of fundamentals and financial variables in Figure 3. We report the explanatory power of our model for these variables in the footnote to this figure. As suggested earlier, we are much more successful in measuring the state of inflation ($R^2$ of 69 percent) than real earnings growth ($R^2$ of 16 percent). Among financial variables, the fit for the V/E ratio is about 50 percent, although notably, the model failed to fit the high valuations of the late 1990s. The fits for historical yield series have $R^2$s of between 44 and 51 percent, with the better fits for the short maturities. Despite the changes in regimes, the model fails to match the high yields during the early 1980s when the Federal Reserve experimented with monetary policy, although the fits are quite accurate outside of this period. The $\beta$ coefficients of each variable are significant at the one percent level and are all close to one (not reported).

5 The Credit Risk Convexity Effect and the Credit Spreads

Puzzle

For constructing model-based spreads of a given rating category, we take the approach that the rating agencies adopt the “through-the-business-cycle” approach, which was described in the introduction, to assigning firms to a given rating category based on their average solvency ratio through good and bad times. The time path of the solvency ratio is constructed as described in Section 4.1. The spread at time $t$ is generated using the calibrated $\pi_t$ series as $s(Z_t, \pi_t, t, T)$ in eq. (21).

As seen in Figure 4, default probabilities under the risk-neutral measure change quite significantly with the state of fundamentals. We plot the 10-year default probabilities for various levels
of the long-run solvency ratio, \( k \), conditional on having perfect knowledge of fundamentals being in each of the four possible composite states. In the plot, we set \( Z_t = C_i / (C \cdot \pi^*) \). \(^{12}\) Default probabilities for each given level of the solvency ratio shift dramatically with the state of inflation, and only slightly with the state of earning growth. Take the case where \( k = 2.5 \): in the (LI-HG) state the default probability is about 3.5 percent, in the (MI-LG) and (MI-HG) states it is around 26 and 19 percent respectively, and in the (HI-LG) state it is about 37 percent. Using the unconditional probabilities of the four states, \( \pi^+ \) as shown Table 1, we obtain an unconditional average of the default rate of a firm with a long-run solvency of 2.5 to be about 13.5 percent. The average default probability under the objective measure is lower at 4.6 percent due to the effect of the positive asset risk premium of about 7.3 percent that increases the drift of the solvency ratio.

For our calibration exercises in the following section we generate a model-implied Baa-Aaa spread series using the firm’s average solvency ratio of \( k = 2.5 \) in each state. The choice is based on the average leverage of BBB (equivalent to Moody’s Baa rating) firms as reported in Standard and Poor’s 1999 publication Corporate Ratings Criteria. For bonds with ten-year maturities we generate a credit spread of 31 b.p. in the (LI-HG) state, which balloons to over 257 b.p. in the stagflation state. In the ‘normal’ recession (MI-LG) state, it averages 205 b.p., while in the state of medium inflation but still solid growth, it is 157 b.p. Using the stationary distribution over states we obtain an unconditional average of the credit spread of 106 b.p. By comparison, the Baa-Aaa spread in the sample from 1960-2001 was about 97 b.p., while the Baa less 10-Year Treasury spread is larger at 172 b.p.

A substantial part of the spread is attributed to a credit risk convexity effect, which will arise in any other model with a time-varying solvency ratio. To illustrate the credit risk convexity effect in a simple and easily verifiable way, we adapt the BSM model to our structural assumptions of debt growth at the rate \( r - \delta \). We incorporate a proportional bankruptcy cost parameter, \( \alpha \), to enable the model to fit historical recovery rates as well; building this feature into our model is straightforward.\(^{13}\) For example, fixing the asset value-to-debt ratio to 2.5, choosing the asset-volatility

---

\(^{12}\) Implicit in this calculation is that the firm’s value at time \( t \) is \( V_t^Q = C_t E_t Q_t \). Its debt level at \( t \) to be in a rating category with a long-run solvency ratio of \( k \) is determined by \( \frac{(C \cdot \pi^*) E_t Q_t}{D_t} = k \). Now substituting for \( Z_t = V_t^Q / D_t^Q \) gives the stated level of the conditional solvency ratio. Alternatively, this is the level of the solvency ratio that the firm would have at \( t \) starting with \( Z_0 = k \) and the processes for the asset value and debt as specified.

\(^{13}\) The bond price for this case is given by

\[
\begin{align*}
P_t &= e^{-r(T-t)} \cdot \left[ N(d_{2t}) + (1 - \alpha) Z_t (1 - N(d_{1t})) \right], \\
d_{1t} &= \frac{\log Z_t + \sigma^2/2 \cdot (T-t)}{\sigma \cdot \sqrt{T-t}}.
\end{align*}
\]
parameter to equal 21.9 percent, which is the average volatility endogenously generated as in eq. (15) using the calibrated belief process, we obtain a credit spread for a bond with a 10-year maturity of 46 b.p. for the case \( \alpha = 0 \). Next, if we use our calibrated stationary probabilities and asset value-to-earnings ratios in the four states in the bottom panel of Table 1, then using the formula \( Z_t = C_t/(C \cdot \pi^*) \) 2.5 we obtain solvency ratios of 3.34, 1.60, 1.80, and 1.24. Calculating the spread using these initial ratios, and then averaging using the stationary probabilities, we obtain an average spread of 73 basis points, which is 59 percent higher than the spread using the average solvency ratio. Finally, if we use a market price of risk of 0.36 (this is the Sharpe ratio obtained for the asset value process in our model), we obtain an average default probability in the four states under the objective measure of 3.6 percent, which is 0.9 percent lower than the average 10-year default probability of BBB bonds as reported by Standard and Poor’s (and used in the calibration by HH).

The intuition behind the result is straightforward: the credit spread is convex in the solvency ratio. The top panel of Figure 5 illustrates the point. When solvency ratios vary over time, the average spread is higher than the spread at the average solvency ratio. HH in all their calibrations always report the latter. As such, the amount of the spread they justify by credit risk is too low. In our model, debt grows at a steady pace while asset valuations vary more with changing investors’ beliefs of the state of the economy which cause the solvency ratio to fluctuate. The “through-the-ratings” cycle approach implies that rating agencies do not change companies’ ratings based on such fluctuations. In calibrating structural form models for bonds within a given rating category, therefore, the econometrician too must account for changes in the solvency ratio over time.

The convexity effect of changing asset valuations and solvency ratios is inversely proportional to asset volatility because for higher levels of volatility, spreads are less sensitive to the initial solvency ratio. We note that in their calibrations HH use higher levels of asset volatilities, of about 26 percent for BBB-rated firms. In this case, the increased spread due to the convexity effect is lower, at 42

\[
d_{2t} = d_{1t} - \sigma \sqrt{T - t},
\]

and the credit spread is \( s_t = -1/T \log(P_t) - r \). Since \( \mu - r = \sigma_m \sigma \), where \( \mu \) is the drift of the asset value process under the objective measure, and \( \sigma_m \) is the market price of risk of the single shock with positive volatility, the default probability under the objective measure is \( 1 - N(d^2_{OMr}) \), where \( d^2_{OM} = d^2_{OMt} + \sigma_m \sqrt{T - t} \), for \( j = 1, 2 \). The recovery rate conditional on default and the expected loss per dollar of debt under the objective measure are given by

\[
R^O M_t = (1 - \alpha) Z_t e^{-\alpha} \sigma (T - t) (1 - N(d^2_{OM})) / (1 - N(d^2_{OMt})), \quad \text{and} \quad L^O M_t = (1 - N(d^2_{OM})) \cdot (1 - R^O M_t).
\]

For computing risk premiums, the volatilities of bonds and equities are

\[
\sigma^2_{P,t} = \frac{\partial^2 P_t}{\partial Z_t^2} \frac{Z_t}{P_t} \sigma \quad \text{and} \quad \sigma^2_{E,t} = \frac{1 - \partial^2 P_t}{1 - \partial Z_t^2} \sigma, \quad \text{where} \quad \frac{\partial P_t}{\partial Z_t} = e^{-r(T - t)} \left( (1 - \alpha) (1 - N(d_{1t})) + \frac{\alpha}{Z_t \sigma \sqrt{T}} \right).
\]
percent. In the right panel of Figure 5, we calculate the ratio of the average spread to the spread at the average solvency ratio — which we call the **convexity premium** for the case where the solvency ratios and their probabilities are given as in our calibrated model above. As another example, a firm with asset volatility of 16 percent, and the same asset valuation ratios as above, would have an average spread 2.7 times higher than the spread at the average solvency ratio.

In computing model-implied spreads HH not only matched the historical default probability under the objective measure, but also recovery rates. By having non-zero bankruptcy costs, we are able to match both estimates. Computing again the average spread with a time-varying solvency ratio as calibrated above we find that with a market price of risk of \( \sigma_m = 0.322 \), volatility of 21.7 percent and bankruptcy costs of \( \alpha = 0.35 \), we match the historical 10-year default probability of 0.045, the recovery rate of 0.51 cents per dollar, and obtain a credit spread of 131 b.p., more than twice as high as HH found in most models. Eom, Helwege, and Huang (2004) use estimates of \( \alpha \) in this range and point to recent empirical evidence supporting such values. Finally, incorporating bankruptcy costs (assumed zero in our model) will further increase the spread by another 50-70 b.p.

It is important to point out that the convexity effect approximated from the BSM model with changing solvency ratios was able to boost the credit spread to 73 b.p., while our average model spread is higher, at 106 b.p. A large proportion of the remainder is due to the fact that asset volatility is stochastic and negatively covaries with asset valuations (the former is higher and the latter is lower in higher inflation states). This further boosts spreads in low valuation states and lowers them in higher valuation states, enhancing the convexity effect. To quantify this effect we first find that the average volatility forecasted at the 10-year horizon, which is relevant for pricing, in the four states is 0.178, 0.318, 0.232, 0.299 respectively.\(^{14}\) Using the BSM model with the solvency ratios specified earlier, and these volatility forecasts, and averaging using the stationary probabilities provides a spread of 96 b.p. That is, this effect explains an incremental 23 b.p. This effect also increases the average default probability under the risk-neutral measure to 4.6 percent, at about the observed level for BBB bonds. It is also important to note that the effect of stochastic volatility on its own is small, raising spreads to only 54 b.p. when used with a constant solvency ratio of 2.5. Also notice from the bottom panel of Figure 1 that estimated solvency ratios in our model vary in a slightly wider range than in the approximation here that uses only the four mean states so that the convexity effect is larger in our model by the remaining 10 b.p.

\(^{14}\) Let \( \tilde{v}(t, T, \pi_t) = E[\int_t^T |\sigma^2_{\pi_t}(\pi_s)| ds] \) be the forecast of volatility to the maturity of the bond. Then \( \tilde{v}(t, T, \pi_t) \) follows the PDE: \( 0 = -\bar{v}_t + |\sigma^2_{\pi_t}(\pi)| + \bar{v}_s (\mu(\pi) - \rho(\pi)) + \frac{1}{2} \sigma(\pi)\sigma(\pi)' \). We solve this PDE with projection methods similar to the solution of the fundamental PDE in eq. (19).
It is also of interest to separate the impact of the real rate and signaling effects of inflation on average spreads. To achieve this, we fit the model by constraining $\alpha_\beta = 0$, and in this case find that the fitted $V/E$ ratios display smaller variation than the general case. Solvency ratios in the model are also less volatile thus lowering the convexity effect. By repeating the analysis above, we find that the average spread for Baa bonds is 16 b.p. lower (90 b.p rather than 106 b.p.). This implies that most of the impact of changes in expected inflation on average spreads is through the signaling role of inflation as opposed to its real rate effect.

5.1 Non-Equivalent Variation in Expected Default Losses and Credit Spreads

In the analysis so far we have shown that the convexity effect that arises from the state-dependence of solvency ratios and asset volatilities in our model is consistent with expected losses at historical levels and credit spreads at empirically observed levels. In this section we analyze the three following issues that shed further light on how this convexity effect helps to reconcile the credits spreads and equity premium puzzles. First, what inputs is the econometrician free to choose in matching low default losses and high spreads? Second, what is the role of state-dependence of the inputs of the structural form model? Finally, are models that lead to a high equity premium also consistent with high credit spreads?

To facilitate the discussion, following CCG, we find it fruitful to write the bond price in our model as

\[
P_t = E_t[\frac{N_T}{N_t}(1 - 1_{Z_T < 1}(1 - Z_T))] \\
= E_t[\frac{N_T}{N_t}]E_t[(1 - 1_{Z_T < 1}(1 - Z_T))] + \text{Cov}[\frac{N_T}{N_t}, 1 - 1_{Z_T < 1}(1 - Z_T)].
\]

The first term is the value of discounted expected losses, which are pinned down from data on historical defaults in model calibration. The second term shows that the bond price will be lower (and hence spreads higher) in models where the covariance of the pricing kernel and the solvency ratio is smaller at the 10-year horizon. Therefore, the discussion of raising spreads while controlling for default losses can be focused around this covariance term.

We first discuss the degrees of freedom the econometrician has at his disposal in matching both historical default losses and credit spreads. Consider the case where he holds the firm’s solvency ratio and asset volatility fixed for a given rating category. By changing the market price of risk, he can control the model’s expected default losses by altering the drift of the asset value process under
the objective measure. However, having matched its historical average, he has no further room to alter the covariance term. We immediately notice this difference between the credit spreads and equity premium puzzles: in the latter, the market price or risk is unconstrained, but in the former it is used to pin down the model’s expected losses from default under the objective measure. In formulating the credit spreads puzzle, HH have correctly constrained the average solvency ratio and asset volatility by historical data. If, for example, the constraint on a fixed level of average asset volatility is removed, obtaining higher credit spreads is no longer as challenging: by increasing the volatility parameter, credit spreads in the model increase, and by simultaneously increasing the price of risk, expected default losses are maintained at historical levels. High spreads can similarly be obtained by lowering the average solvency ratio and increasing the price of risk. In our analysis, the only variable we vary freely is the price of risk, while solvency ratios and asset volatilities are endogenous functions of the price of risk. With this single degree of freedom, we study the ability of the model to match historical averages of solvency ratios, asset volatilities, historical default losses, and credit spreads.

Our strategy of varying the price of risk to match moments of financial and real variables is analogous to changing investors’ preference parameters in general equilibrium complete market models. In such models, \( \sigma_m = \gamma \sigma_x \), where \( \gamma \) is the coefficient of relative risk aversion (CRRA), and \( \sigma_x \) is the volatility of aggregate consumption. Starting with the classic paper of Mehra and Prescott (1985), various papers have changed the market prices of risk by keeping \( \sigma_x \) fixed, but varying the CRRA. By raising the CRRA, the econometrician can match the observed equity premium. While we have not chosen to conduct our analysis within the complete markets-general equilibrium framework, we can back out the level of CRRA to be consistent with our calibrated market price of risk.\(^{15}\)

In attempting to calibrate to higher spreads, the econometrician can choose to ignore the state-dependence of solvency ratios by evaluating spreads at their historical average or choose to estimate their average values in \( N \) different states. To isolate the impact of the convexity effect that arises from the state dependence we will continue to use the BSM pricing formulas for spreads and expected losses in Footnote 13 with the assumption that there are \( N \) observable states with distinct

\(^{15}\)In our analysis, the vector \( \sigma_M \) is calibrated to match the valuation ratios of equities and bonds and leads to an average Sharpe ratio on equities of 0.32. In a single shock general equilibrium framework with aggregate consumption volatility of around two percent, we would require investors to have a CRRA of 16 to obtain a Sharpe ratio of this magnitude. One other way to obtain a high Sharpe ratio is to assume different preferences such as external habit-formation that conditionally raises investors’ CRRA for a low level of \( \gamma \) as in Campbell and Cochrane (1999). With these preferences, the average CRRA that justifies the Sharpe ratio on equities is above 45. David (2006) shows that another way to obtain a high Sharpe ratio is to model speculation among agents that leads to high per capita consumption volatility with a CRRA of only 0.5.
solvent ratios are formulated endogenously as functions of the prices of risk as shown in Propositions 1 and 2. In the remainder of this section we simply refer to these two functional relationships as “endogeneity.” As in our model, we now use a three-dimensional vector of market prices of risk, $\sigma_M$. However, due to our chosen orthogonalization in Assumption 2, the V/E ratio in Proposition 1 only depends on $\sigma_{M,2}$, the price or risk of the earnings shock, and we will only consider its variations. We also assume in this approximation that the variations in $\sigma_{M,2}$ have no influence on the time series of investors’ beliefs. For the large values of $\sigma_{M,2}$ considered, the signal-to-noise ratio in investors’ kernel is low and we find this to be a reasonable assumption. Average asset volatility across the four states is approximated as $0.45 \cdot ||\sigma_V^{(\pi^*)}||$ where $\pi^*$ is the vector of stationary probabilities in Table 1 and $\sigma_V^{(\pi)}$ is in (15). The factor 0.45 is used because $\sigma_V^{(\pi)}$ is a concave function of beliefs, and hence average volatility (from our fitted model) is lower than the volatility at the stationary probabilities. Forecasted volatility in the four states is then approximated by scaling this average volatility using the optimal forecasts for our calibrated model described in Footnote 14. The scaling factors for volatility in the four states are 0.81, 1.45, 1.06, and 1.37, respectively. The drift of the solvency ratio under the objective measure is $\sigma_V^{(\pi)}(\sigma_M + \sigma_Q)^{\prime}$. Given these relationships, the econometrician attempts to find a $\sigma_{M,2}$ to match average historical default losses.

There are three effects of an increase in $\sigma_{M,2}$ on $s$, $\bar{s}$, $L$, and $\bar{L}$.

1. Asset volatility with respect to the earnings shock (which contributes to a positive risk premium) is nearly insensitive to $\sigma_{M,2}$, while asset volatilities with respect to the inflation and kernel shocks, are both negative, and decline in absolute value as $\sigma_{M,2}$ increases (Figure 6, left panel). Proposition 3 provides intuition for the signs of these volatilities. The cumulative effect is a decline in total asset volatility (Figure 6, center panel) that lowers $s$, $\bar{s}$, $L$, and $\bar{L}$.

2. A mean preserving decline in spread of the solvency ratios in the $S$ states (Figure 6, right panel) that lowers $\bar{L}$ and $\bar{s}$ but leaves $L$ and $s$ unchanged. Therefore, the convexity effect on default losses and credit spreads both decline when the price of risk is raised.
3. The drift of the nominal asset value under the objective measure, \( Q V(M + Q) \) increases. This happens because the negative risk premium from the inflation shock declines more rapidly than the positive risk premium from the earnings shock as \( M_2 \) increases (see Section 4.2 for the signs of the prices of risk). This lowers \( LOM \) and \( LOM \), but leaves \( s \) and \( \bar{s} \) unchanged. In addition, the real risk premium on the asset value, \( \sigma V \), equities, \( \sigma E \), and defaultable bonds, \( \sigma B \), all increase (Figure 6, middle panel).

Using these inputs we obtain two major results: First, strikingly, credit spreads are decreasing in the price of risk of the earnings shock, while all premiums are increasing. Therefore, the credit spreads puzzle cannot be resolved simply by raising the market price of risk. Second, the expected (under the objective measure) default losses are more sensitive to changes in \( M_2 \) than credit spreads — that is, there is non-equivalent variation in the two measures. This ensures that the convexity effect remains positive even with the endogeneity described. Intuition for the result can be obtained from Figure 7. As seen \( LOM, LOM, s, \) and \( \bar{s} \) are all decreasing functions of \( M_2 \). However, since the third effect only affects the functions computed under the objective measure, these decline faster than the two spread functions. Consider an increase in \( M_2 \) from \( \sigma M_2 \) to \( \bar{\sigma} M_2 \). This increase in the price of risk is just sufficient to ensure that \( \bar{L}(\bar{\sigma} M_2) = L(\sigma M_2) = L H \). However, \( \bar{s}(\bar{\sigma} M_2) > s(\sigma M_2) \), since the decline in the spread is less rapid. The implication is that if the econometrician accounts for the state dependence of the solvency ratio, then he will calibrate to higher spreads, even though he raises the market price of risk sufficiently to lower expected default losses to their historical average level. The amount \( \bar{s}(\bar{\sigma} M_2) - s(\sigma M_2) \) is the convexity effect with the endogeneity. It is smaller than the convexity effect \( \bar{s}(\bar{\sigma} M_2) - \bar{s}(\bar{\sigma} M_2) \), which would result if the econometrician ignores the endogeneity. For example, using the calibrated parameters for Baa-rated bonds for our model, the former effect accounts for 29 b.p, while the latter for 46 b.p. To put these into perspective, we recall that the historical Baa-Aaa spread in our sample is 96 b.p.\(^{16}\)

Returning to the relation between the credit spreads and equity premium puzzles we note that \( \sigma M_1 \) and \( \sigma M_3 \) have been held fixed in our analysis in this subsection, and using the implied \( \sigma M_2 \) from this section and equity volatilities we obtain equity premia of 7.3 and 5.1 percent and Sharpe ratios of 0.32 and 0.22 for the cases with and without state dependence. Empirical estimates of Sharpe ratio for equities range between 0.3 and 0.5 (see, e.g., Cochrane 2001). Therefore, allowing

\(^{16}\)While we did not reestimate our model for alternative rating categories, we used the parameter estimates for the Baa case above and calibrated the solvency ratios to match their average values in HH. We then approximate the convexity effects with exogenous and with endogenous solvency ratios and volatilities at 32 and 16 basis points for the A rating, and 21 and 10 for the Aa rating, respectively. The historical spreads over Aaa bonds are 60 and 28 b.p. as reported in HH.
for state dependence of solvency ratios and asset volatilities has an added advantage of providing a reconciliation of the high Sharpe ratios on equities and high credit spreads.

5.2 The Time Series of Credit Spreads and the Credit Spreads Volatility Puzzle

Returning to our model results, one naturally questions if movements in the aggregate solvency ratio and asset volatility lead to spread movements that are consistent with time series of observed credit spreads. Using the calibrated solvency ratio series, \( \{Z_t\} \) and beliefs of the states, \( \{\pi_t\} \), we use eq. (21) to generate the time-series of model spreads, \( \text{CSM}(t) \). The historical Baa-Aaa spread obtained from Moody’s and the fitted series are in Figure 8 and are used in the regressions below:

\[
\text{CS}(t) = 0.0306 + 0.887 \text{CSM}(t - 1) \tag{24}
\]

\[
(0.535) \quad (17.537) \quad R^2 = 0.642, \tag{25}
\]

where t-statistics are in parentheses. We make the following summary comments: (i) The model spreads explain about 64 percent of the variation in the historical spread. While some readers have questioned the role of inflation in recent years, the fits in the first and second halves of the samples explain 54 and 72 percent of the variation respectively (regressions not shown), (ii) The insignificant intercept term follows from the fact that in our model there is no credit spreads puzzle, (iii) We have used the one-quarter lag of the model spread. Highly significant regressions are also obtained with the contemporaneous regression; however, the fit one-quarter ahead is better, suggesting that investors do not fully adjust to all incoming news on inflation instantaneously. (iv) Several authors in the past have presented the regressions in changes of spreads, due to the possible non-stationarity of historical spreads. We find that the evidence for non-stationarity of the spread is not strong.\(^{17}\) While we present only level regressions here we have also obtained very significant results for returns on corporate bond indexes, which are more tightly linked to the theory than changes in spreads (these are available upon request). In addition, there has been a recent resurgence in understanding the level of spreads (see Eom, Helwege, and Huang 2004, Huang and Huang 2003).

The credit risk convexity effect outlined above is related to the volatility of credit spreads, since more volatile spreads enhance this convexity. We find that the standard deviation of the historical Baa-Aaa spread is about 45 b.p., and the model fitted spread in Figure 8 has a standard deviation

\(^{17}\)The estimate for the autoregressive parameter using the Augmented Dickey-Fuller procedure (Dickey and Fuller 1979) with no time trend is 0.914, with a t-statistic of -2.667. The 5 and 10 percent critical values of the t-statistic are -2.876 and -2.568. The estimate for the autoregressive parameter using the Perron procedure (Perron 1989) is 0.909, with a t-statistic of -2.810. The 10 percent critical value is -3.130.
of 43 basis points. Therefore, the changing solvency ratio and asset volatility in our model are able to address the credit spreads volatility puzzle raised in CCG, who point out that for several credit risk models in which the solvency ratio is held constant in the calibration exercise, the predicted volatility of spreads is too low (around 10 b.p.). In addition, using our bond pricing formula and calibrated beliefs but always using the solvency ratio of 2.5 leads to a standard deviation of only 7 b.p.

Since both the level and volatility of credit spreads are related to the volatility of the solvency ratio, it is important to ascertain if empirical solvency ratios have been as volatile as required by our model to generate a convexity effect of the required magnitude. Our estimated solvency ratio series is shown in the bottom panel of Figure 1. We compare its variation to that of a similarly constructed series in CCG for the inverse of the solvency ratio, the leverage ratio. CCG report that the variation of the annual book leverage ratio series from 1975 to 1998 is 9 percent. Restricting to this subsample we find the volatility of our implied leverage ratio is similar at 9.2 percent. For the full sample the volatility is 10.1 percent. The autocorrelation coefficient of the leverage ratio is 0.86, so that as for the credit spreads series, it does not have a unit root. When decomposing the convexity effect in our model above we noted that statistics of leverage ratios are not sufficient to determine the moments of credit spreads. In addition the variation of forecasted asset volatility and its covariation with the leverage ratio is also needed. When the covariation is positive, as in our model, the convexity effect is larger. To shed further light on the roles of time variation of the leverage ratio and asset volatility, we examine the empirical distributions of credit spreads and the relevant state variables below.

The top panels of Figure 9 show that the empirical distributions of the historical credit spread and our model fitted series are both positively skewed with skewness coefficients of about 1.3. The positive skewness arises from the widening of spreads in medium and high inflation states. The bottom left panel shows that the leverage ratio displays a smaller level of skewness (skewness coefficient of only 0.11). The right panel shows that the distribution of the inverse distance-to-default (forecasted asset volatility divided by the solvency ratio) in our model has a level of positive skewness comparable to that of credit spreads. The distance-to-default is a widely used statistic in industry to measure the credit risk of firms (see, e.g., Crosbie and Bohn 2002) as it incorporates the effects of declining solvency ratios as well as increasing volatility during bad times. This double impact plays a role in boosting our model average spread to empirical levels as reported earlier.
5.3 Credit Spreads and Macroeconomic Recessions

As seen in Figure 8 credit spreads increase during economic recessions (see also Duffie and Singleton 2003). This might suggest that a two state observable regime switching model for credit spreads such as in Hackbarth, Miao, and Morellec (2005) might be adequate to capture most of the variation in credit spreads.\footnote{Their model provides some key insights into financing and default policies of firms through the business cycle. However, it is well known that market participants are unsure in real time of the state of the economy. Therefore, it would be hard for firms to adjust leverage and default policies in real time as advocated by HMM. The current paper explicitly builds in investors’ uncertainty of the state of the economy. Moreover, higher uncertainty itself can affect the level of spreads due to higher volatility in these times.} However, such a model provides limited success in fitting historical spreads: the regression of the historical Baa-Aaa spread on the NBER recession indicator yields an $R^2$ of only 9.6 percent, which might suggest that macroeconomic factors overall play a small role in explaining the time variation in spreads. In contrast, once we explicitly build the forecasting effect of inflation into investors’ information sets, the fit improves to 64% as noted in the previous subsection.

An additional point of interest is to test whether credit spreads can be used to forecast recessions or the reverse, whether recessions forecast spreads. We find that the historical spread does have limited forecasting power, but its effect is insignificant once we include the model-based spread. A standard Granger Causality test of the null hypothesis of our model spread not causing NBER recessions with four lags yields an F-statistic of over 11 with a p-value of the order $10^{-8}$, which means the null of no causality is strongly rejected. The p-value of the reverse test is 0.69. In this sense, our model spread leads recessions.

We attribute the forecasting power of the model-based spread to the signaling role of inflation, since the slope of the term structure, which is another proxy for this role, provides similar results. High growth rates states are less stable in higher inflation states causing asset valuations to fall and asset volatility to rise in periods of rising inflation expectations. Using a similar model, David and Veronesi (2006) find that measures of inflation uncertainty forecast stock and Treasury bond volatilities up to two years ahead. The latter are known to be strongly contemporaneously related to recessions (Schwert 1989).

5.4 Related Work on the Credit Spreads Puzzle

As seen above our key argument in addressing the credit spreads puzzle is to explicitly incorporate the cyclical variation of credit risk into the calibration exercise. This feature was ignored by authors
such as HH and Delianedis and Geske (2001). We now look at the work of other authors in addressing the credit spreads puzzle and determine the role of this cyclical variation in these alternative explanations.

We start with CCG’s analysis of the CC habit-formation model. In their calibration exercise CCG assume that the solvency ratio of a typical Baa-rated firm is constant at each period in their sample. Therefore, they do not account for the impact of cyclical variation in asset valuations on solvency ratios. However, they build in alternative channels that cause time variation in the credit risk of firms. In particular, in the CC model, the market price of risk is countercyclical due to variations in the surplus consumption ratio. In addition, the countercyclicality of the volatility of the surplus consumption ratio causes asset volatility to be countercyclical. Both channels lower the covariance term between the pricing kernel and solvency ratios in eq. (23); that is, they make it more likely that default losses occur during periods of high marginal utility and hence increase credit spreads. However, CCG find that while the CC model spread is higher and more volatile than in a simple benchmark model that holds the price of risk and asset volatility constant, it fails to match both moments of historical data. In addition, the habit-formation model implies that risk premiums are countercyclical, and hence, counterfactually, that the probability of default under the objective measure is lowest in recessions. CCG are able to substantially improve the performance of the CC model in both respects by postulating a default boundary that increases more rapidly than debt in recessions for all firms. The incorporation of this countercyclical time-varying boundary in effect causes larger systematic time variation in credit risk of firms within the rating category and leads to more defaults in bad times.

On a similar note, CCG find that due to a relatively low duration of corporate bonds relative to equities, neither the low frequency time variation in the growth rate to fundamentals nor the stochastic consumption growth volatility in the long-term risks model of BY can themselves lead to higher credit spreads. However, the BY model in addition has a countercyclical price of risk, which when combined with the other features leads to some improvement. Therefore, the BY model shares with our model and the CC model the common feature that the cyclical variation of the asset risk premium plays a role in increasing credit spreads, particularly in bad times.

Without questioning the validity of the various channels that are required to provide large credit spreads in the models analyzed in CCG, we summarize how our analysis is different.

1. The CC model requires larger variation in the default boundary than the leverage ratio. While volatility and risk premiums are higher in recessions in our model as well, solvency ratios are endogenous and fall in these periods, thus avoiding the procyclicality of default probabilities.
2. The economic mechanism that gives rise to countercyclical asset volatility in our model is quite different from that in CC. Volatility in CC is countercyclical and is driven by volatility of the surplus-consumption ratio that is specified exogenously. CC explain that the model is “reverse-engineered” to obtain a countercyclical stock volatility and risk premium.\textsuperscript{19} In our work the countercyclical asset volatility arises endogenously from investors’ learning about the state of fundamentals.

3. Both the BY and CC model require a time varying price of risk. Although we have used constant market prices of risk, we generate countercyclical asset risk premiums from the variation in asset volatility.

Cremers, Driessen, and Maenhout (2006) generalize the class of models that are tested in HH by allowing for both firm-specific and systematic jumps in asset values of firms. The novelty of their calibration exercise is to infer the size of jump risk premia from the prices of S&P 100 options. The large systematic jump mean under the risk neutral measure as opposed to the objective measure is able to reconcile the high spreads and the low expected default losses, while incorporating idiosyncratic jumps can control the clustering of defaults of individual firms. While the paper does not conduct a time series analysis of credit spreads, it admits to some difficulty in matching the stylized fact that the jump risk premium implicit in options prices increases sharply after the stock market crash of 1987 without a corresponding increase in credit spreads.

In other work, Tang and Yan (2006) are able to justify large default-related credit spreads in a flow-based structural form model. Defining default as shortfall of cash flow places different constraints on the empirical exercise from value-based structural form models. Longstaff, Mithal, and Neis (2006) calibrate a reduced form model to spreads in the credit default swaps market and support the view that default risk accounts for the predominant component of spreads.

6 Conclusion

We study the credit spreads puzzle: the inability of structural form credit risk models to reconcile low historical averages of losses from default and their high credit spreads for bonds within specified rating categories. Since both sets of averages are conditioned on ratings, this paper suggests that the rating methodology followed by the rating agencies should affect the way an econometrician evaluates the performance of structural form models in their ability to match historical data. An implication of the “through-the-cycle” rating procedure used by rating agencies is that credit risk

\textsuperscript{19}In fact, if the volatility of the surplus consumption ratio is held constant then Li (2006) shows that the asset volatility in the CC model is counterfactually procyclical.
for firms within a given rating category will fluctuate in response to macroeconomic shocks. Large firm-specific shocks, on the other hand, will cause the agencies to move the firm into an alternative rating category so that such shocks do not affect the time series averages for the specified rating category. Looking back at historical data with this perspective, the econometrician must recognize that the credit quality of all firms within a rating category weakens at certain stages of the business cycle and in these periods their credit spreads increase. Similarly during periods preceding strong periods of growth, credit spreads decrease. Since credit spreads are convex in the solvency ratio, this state dependence of credit risk leads to higher average spreads than the spreads if the solvency ratio and asset volatility were held constant at their historical averages. In earlier work authors such as HH ignore this cyclical variation of credit risk while some others, such as CCG, only account for it partially.

In addition to causing state dependence, macroeconomic shocks carry risk premiums so that they lead to non-equivalent variation in averages of expected default losses, which are computed using objective probabilities, and credit spreads, which are computed under risk-neutral probabilities. Firm-specific shocks cause equivalent variation in credit spreads and expected default losses and therefore cannot increase credit spreads once we control for expected default losses. Our modeling choice in this paper of having only macroeconomic shocks is predicated on our goal of characterizing the state dependence in credit risk of the representative firm of a given rating category.

We estimate the size of the effects of the state-dependence on historical averages using an unobservable regime-switching framework where investors’ changing expectations of the state of future real fundamental growth causes endogenous state-dependence of solvency ratios of firms and their asset volatilities. High real growth is unstable in higher inflation states causing asset valuations and solvency ratios to fall and asset volatility to rise rapidly. The average spread produced by the model is about 60 percent larger than the spread computed at the average solvency ratio by several authors — the credit risk convexity effect. The negative relation between asset valuations and volatility enhances this convexity, causing spreads to increase a further 50 percent. An econometrician who ignores the state-dependence of the solvency ratios requires too low a risk premium to keep expected default losses at their historical levels, making it more difficult to match the high equity premium and high credit spreads simultaneously despite varying the market prices of risk.

A natural extension of this paper is to take the implications for spreads to individual firm level data. It holds promise since recent work by Duffie, Saita, and Wang (2006) has revealed that interest rates are highly significant in predicting defaults even in the presence of several key explanatory variables such as leverage ratios.
References


35


37


Table 1: 4-State Model Calibration

<table>
<thead>
<tr>
<th>States</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\beta_3$</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0133</td>
<td>0.043</td>
<td>0.093</td>
<td>-0.090</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.005)</td>
<td>(0.014)</td>
<td>(0.020)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Volatilities</th>
<th>$\sigma_{Q,1}$</th>
<th>$\sigma_{Q,2}$</th>
<th>$\sigma_{E,2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.017</td>
<td>0.002</td>
<td>0.112</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.000)</td>
<td>(0.009)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Generator Elements</th>
<th>$\lambda_{13}$</th>
<th>$\lambda_{21}$</th>
<th>$\lambda_{23}$</th>
<th>$\lambda_{24}$</th>
<th>$\lambda_{32}$</th>
<th>$\lambda_{42}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.039</td>
<td>0.111</td>
<td>0.114</td>
<td>0.285</td>
<td>0.110</td>
<td>0.300</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.007)</td>
<td>(0.012)</td>
<td>(0.018)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pricing Kernel</th>
<th>$\alpha_0$</th>
<th>$\alpha_\beta$</th>
<th>$\alpha_\theta$</th>
<th>$\sigma_{M,1}$</th>
<th>$\sigma_{M,2}$</th>
<th>$\sigma_{M,3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.206</td>
<td>0.826</td>
<td>0.351</td>
<td>0.076</td>
<td>0.536</td>
<td>-0.137</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.023)</td>
<td>(0.046)</td>
<td>(0.040)</td>
<td>(0.100)</td>
<td>(0.021)</td>
</tr>
</tbody>
</table>

GMM Error Value ($\chi^2(6)$): 9.23  P-Value: 0.161

<table>
<thead>
<tr>
<th>Implied Quarterly Transition Probability Matrix</th>
<th>Implied Annual Transition Probability Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>(LI-HG)</td>
<td>(MI-LG)</td>
</tr>
<tr>
<td>LI-HG</td>
<td>0.995</td>
</tr>
<tr>
<td>MI-LG</td>
<td>0.026</td>
</tr>
<tr>
<td>MI-HG</td>
<td>0.000</td>
</tr>
<tr>
<td>HI-LG</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Implied Stationary Probabilities, V/E Ratio, 3-Month Treasury Yield
5-Year Treasury Yield, Slope of Term-Structure, and Credit Spreads

<table>
<thead>
<tr>
<th>State</th>
<th>$\pi^*$</th>
<th>$C_i$</th>
<th>$i_{25}$</th>
<th>$i_5$</th>
<th>$i_5 - i_{25}$</th>
<th>$s(Z_t = 2.5)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LI-HG</td>
<td>0.493</td>
<td>25.449</td>
<td>0.031</td>
<td>0.034</td>
<td>0.030</td>
<td>0.003</td>
</tr>
<tr>
<td>MI-LG</td>
<td>0.101</td>
<td>11.827</td>
<td>0.046</td>
<td>0.070</td>
<td>0.024</td>
<td>0.020</td>
</tr>
<tr>
<td>MI-HG</td>
<td>0.337</td>
<td>13.333</td>
<td>0.090</td>
<td>0.085</td>
<td>-0.005</td>
<td>0.015</td>
</tr>
<tr>
<td>HI-LG</td>
<td>0.067</td>
<td>8.964</td>
<td>0.141</td>
<td>0.105</td>
<td>-0.036</td>
<td>0.025</td>
</tr>
</tbody>
</table>

The top panel reports SMM estimates of the following model for CPI, $Q_t$, real earnings, $E_t$, and the real pricing kernel, $M_t$:

\[
\frac{dQ_t}{Q_t} = \beta_t \ dt + \sigma_Q \ dW_t, \\
\frac{dE_t}{E_t} = \theta_t \ dt + \sigma_E \ dW_t, \\
\frac{dM_t}{M_t} = -k_t \ dt - \sigma_M \ dW_t,
\]

where $\sigma_Q = (\sigma_{Q,1}, \sigma_{Q,2}, 0)$, $\sigma_E = (0, \sigma_{E,2}, 0)$, $\sigma_M = (\sigma_{M,1}, \sigma_{M,2}, \sigma_{M,3})$, $k_t = \alpha_0 + \alpha_\theta \ \theta_t + \alpha_\beta \ \beta_t$, and the vector $\nu_t = (\beta_t, \theta_t, -k_t)'$, follows a four-state regime switching model with the generator matrix $\Lambda$ whose non-zero diagonal elements are shown as $\lambda_{i,j}$. The pricing kernel is observed by investors but not by the econometrician. Assets are priced using the formulas in Proposition 1. Estimates are obtained from data on the fundamentals as well six financial variables using the EM-SMM methodology described in Appendix 2. Standard errors are in parentheses. The middle panels report the quarterly and annual implied transition probability matrix between the four states. Rows may not sum to one due to rounding. The bottom panel reports implied asset prices for the calibrated. The V/E ratio and bond yields are computed as shown in Proposition 1. Spreads for defaultable bonds of maturity 10 years are computed using (21), using the conditional solvency ratio in state $i$ as $Z_i = (C_i)/(C \cdot \pi^*)$ $k$ (see Footnote 39 for an explanation).
The top panel shows time series of Moody’s Baa - Aaa yield spread and the 3-month Treasury Bill rate. The latter has been rescaled. The bottom panel shows time series of the solvency ratio, $Z_t$, for a firm with an average solvency ratio of 2.5. The solvency ratio defined as the value of a “representative” Baa firm’s assets divided by the face value of its debt outstanding, $Z_t = (P_t^E + D_t P_t)/D_t$. The value of equity is approximated as $P_t^E$, the value of the S&P 500, the price of debt is $P_t = e^{-y_{t,Baa}10}$, where $y_{t,Baa}$ is Moody’s Baa average bond yield, and $D_t$ is an approximation of the liabilities of Baa-rated firms. The growth of $D_t$ each quarter is the average growth rate of liabilities of all non-financial firms in Compustat from 1975:Q2 – 2001 (prior data is not reliable). The debt level at the first date is adjusted to match the solvency ratio implied from our analysis. A series in which debt growth and bond prices are from our calibrated model, but with market equity valuations for the sample from 1960-2001 is also shown, which is backed out from equity and Treasury market valuations using a generalization of Moody’s KMV methodology described in Section 4.1. Shaded areas denote NBER dated recessions.
We display the time series of filtered probabilities of investors of the four possible underlying states obtained using the SMM methodology in Appendix 2. The states are numbered as $(\beta_1, \theta_2)$, $(\beta_2, \theta_1)$, $(\beta_2, \theta_2)$, and $(\beta_3, \theta_1)$; the parameter estimates are in Table 1.
We display the time series of financial and fundamental variables series used in the SMM estimation of the model, and their fitted values. The calibrated values of the parameters are shown in Table 1. The filtered beliefs series of investors used to generate the fitted values are shown in Figure 2. As a measure of the fit of our model to the data we provide the $R^2$ of the following regressions $y(t) = \alpha + \beta x(t) + \epsilon(t)$, where $y$ is the variable under consideration and $x$ is its expected value based on model: Inflation — 68.9%, Earnings — 17.5%, V/E Ratio — 49.8%, 3-Month Treasury — 50.4%, 1-Year Treasury 51.0%, 3-Year Treasury 47.6%, 5-Year Treasury — 45.4%, 10-Year Treasury 43.6%.
This figure shows the risk-neutral probabilities of default at the 10-year horizon for different values of the firm’s long-run solvency ratio, $k$, and conditional on the investors knowing with certainty each of the four states of fundamentals: State 1, low inflation - high growth, State 2, medium inflation - low growth, State 3, medium inflation - high growth, and State 4, high inflation - low growth. The default probabilities are computed as in $\Pi_2(Z_t, \pi_t, t, T)$ in Proposition 5, where the solvency ratio in state $i$ is $Z_i = (C_i)/(C \cdot \pi^*) k$ (see Footnote 12 for an explanation).

The left panel shows credit spreads for bonds with 10 years to maturity assuming an asset volatility parameter of 21.9 percent and bankruptcy costs of zero in the Black-Scholes-Merton model as shown in Footnote 13. The right panel shows the convexity premium defined as the ratio of the average spread in the four states to the spread at the average solvency ratio of 2.5. The average spread is calculated using the solvency ratios in the four states defined as $Z_i = C_i/(C \cdot \pi^*)$ 2.5 (see Footnote 12) where the V/E ratios, $C$, and stationary probabilities, $\pi^*$ are in in the bottom panel of Table 1. Using the formula we obtain solvency ratios of 3.34, 1.60, 1.80, and 1.24 in the four states respectively.
Asset volatilities are plotted as functions of $\sigma_{M,2}$ as in (15). We plot their averages across the four states as approximated as $0.45 \cdot \sigma_V^O(\pi^*)$ where $\sigma_V(\pi_t)$ is in (15) in the left panel. The middle panel plots $||\sigma_V(\pi^*)||$ as well as the risk premiums on the asset value, equities, and bonds as shown in Footnote 13. The right panel shows that the solvency ratios in the four states are $Z_i = C_i/(C \cdot \pi^*)$ $2.5$, $i = 1, \ldots, 4$ (see Footnote 12) where the V/E ratios, $C$, are functions of $\sigma_{M,2}$ as shown in Proposition 1(a). All parameter estimates (besides $\sigma_{M,2}$) and the stationary probabilities, $\pi^*$ are in in Table 1.

Credit spreads and expected default losses under the objective measure are computed for bonds with 10 years to maturity and a long run solvency ratios of 2.5 using the Black-Scholes-Merton model as shown in Footnote 13. The average spread is calculated using the solvency ratios in the four states, asset risk premiums, and average volatilities as described in Figure 6. Forecasted volatility in the four states is approximated by scaling this average volatility using the optimal forecasts for our calibrated model described in Footnote 14. The scaling factors for volatility in the four states are 0.81, 1.45, 1.06, and 1.37, respectively. Average historical expected default losses per dollar of debt is $L_H$. $s$ and $L^{OM}$ are computed at the average solvency ratio, while $\bar{s} = \sum_i \pi_i^* s(Z_i)$ and $\bar{L}^{OM} = \sum_i \pi_i^* L^{OM}(Z_i)$. The drift of the nominal asset value under the objective measure is $\sigma_V^O(\sigma_M + \sigma_Q)^\prime$. An econometrician who accounts for the endogeneity of solvency ratios and asset volatilities would find a convexity effect of $\bar{s}(\bar{\sigma}_{M,2}^*) - s(\sigma_{M,2}^*)$, while one who ignores the endogeneity would find a convexity effect of $\bar{s}(\bar{\sigma}_{M,2}^*) - s(\sigma_{M,2}^*)$.
The plot shows time series of Moody’s Baa - Aaa yield spread and its model based counterpart. Model spreads are calculated using the “through-the-business-cycle” approach: The spread at time $t$ is generated using the calibrated $\pi_t$ series as $s(Z_t, \pi_t, t, T)$ in eq. (21), where the calibrated solvency ratio series, $\{Z_t\}$ and belief series, $\{\pi_t\}$ are shown in Figures 1 and 2 respectively. The model series explains 65 percent of the variation in the data series for the full sample. It explains 54 and 72 percent of the variation in the first and second halves of the sample respectively.

The top panels show the empirical distributions of Moody’s Baa-Aaa yield spread and its model based counterpart that is generated as described in the note to Figure 8. The bottom left panel shows the empirical distribution of the implied leverage ratio (the inverse of the solvency ratio) that is displayed in the bottom panel of Figure 1. The bottom right panel shows the distribution of the Inverse Distance-to-Default, which equals the forecasted asset volatility divided by the solvency ratio. Forecasted asset volatility for our calibrated model is calculated as described in Footnote 14.
Appendix 1:

Proof of Proposition 2. Let \( i_Q = (1, 0, 0) \), \( i_E = (0, 1, 0) \), and \( i_M = (0, 0, 1) \). These are useful since we can write \( \sigma_Q = i_Q \Sigma \), for example. Notice that the following covariances hold:

\[
d E d Q' &= E Q \sigma_Q \Sigma Q', \\
d \pi_i d \pi_j' &= \pi_i \pi_j(\nu_i - \bar{\nu})' (\Sigma \Sigma')^{-1} (\nu_j - \bar{\nu}) dt, \\
d E d \pi_i' &= E_i E \Sigma \Sigma^{-1} (\nu_j - \bar{\nu}) \pi_i dt = (\theta_i - \bar{\theta}) \pi_i E, \\
d Q d \pi_i' &= Q_i Q \Sigma \Sigma^{-1} (\nu_j - \bar{\nu}) \pi_i dt = (\beta_i - \bar{\beta}) \pi_i Q.
\]

(a) Recall that the real asset value \( V_t = (\sum_{i=1}^N C_{i} \pi_i) E_t \). Hence, from Ito’s lemma:

\[
d V = \sum_{i=1}^N C_i \pi_i d E + E \sum_{i=1}^N C_i d \pi_i + \sum_{i=1}^N C_i d E d \pi_i' + \sum_{i=1}^N C_i d \pi_i' \\
= V \left( \bar{\theta} dt + \sigma_E d \tilde{W} + \frac{\sum_{i=1}^N C_i [\pi A]_i}{\sum_{i=1}^N C_i \pi_i} dt + \frac{\sum_{i=1}^N C_i \pi_i (\nu_i - \bar{\nu})' (\Sigma \Sigma')^{-1} d \tilde{W}}{\sum_{i=1}^N C_i \pi_i} + \frac{\sum_{i=1}^N C_i \pi_i (\theta_i - \bar{\theta})}{\sum_{i=1}^N C_i \pi_i} dt \right) \\
= V \left( \bar{\theta} + \frac{\pi \Lambda C}{\sum_{i=1}^N C_i \pi_i} + \frac{\sum_{i=1}^N C_i \pi_i (\theta_i - \bar{\theta})}{\sum_{i=1}^N C_i \pi_i} \right) dt + V \left( \sigma_E + \frac{\sum_{i=1}^N C_i \pi_i (\nu_i - \bar{\nu})' (\Sigma \Sigma')^{-1}}{\sum_{i=1}^N C_i \pi_i} \right) d \tilde{W}. \tag{26}
\]

Using (12) we can write, \( \pi \Lambda C = -1 + \pi \text{diag}(k - \theta) C + \pi C \sigma_E \sigma_M \). Therefore, the drift equals

\[
\mu_V - \delta(\pi_t) = \bar{\theta} + \frac{\sum_{i=1}^N C_i \pi_i (k - \theta_i)}{\sum_{i=1}^N C_i \pi_i} + \frac{\sigma_M \sigma_E'}{\sum_{i=1}^N C_i \pi_i} + \frac{1}{\sum_{i=1}^N C_i \pi_i}. \tag{27}
\]

Turning now to the nominal value process, recall that the nominal asset value \( V_t' Q = (\sum_{i=1}^N C_{i} \pi_i) Q_t E_t \). Hence, from Ito’s lemma and the characterization of the fundamental processes under the observed filtration (9), \( d V' Q = \)

\[
= \sum_{i=1}^N C_i \pi_i Q d E + \sum_{i=1}^N C_i \pi_i Ed Q + E \sum_{i=1}^N C_i d \pi_i + \sum_{i=1}^N C_i d \pi_i d Q' + E \sum_{i=1}^N C_i d \pi_i d Q' + Q \sum_{i=1}^N C_i d \pi_i d E' \\
= V Q \left[ \bar{\theta} dt + \sigma_E d \tilde{W} + \beta dt + \sigma_Q d W + \frac{\sum_{i=1}^N C_i [\pi A]_i}{\sum_{i=1}^N C_i \pi_i} dt \right] \\
+ \frac{\sum_{i=1}^N C_i \pi_i (\nu_i - \bar{\nu})' (\Sigma \Sigma')^{-1} d \tilde{W}}{\sum_{i=1}^N C_i \pi_i} + \sigma_E \sigma_Q' + \frac{\sum_{i=1}^N C_i \pi_i (\theta_i - \bar{\theta})}{\sum_{i=1}^N C_i \pi_i} dt + \frac{\sum_{i=1}^N C_i \pi_i (\beta_i - \bar{\beta})}{\sum_{i=1}^N C_i \pi_i} dt \\
= V Q \left( \bar{\theta} + \beta + \sigma_E \sigma_Q' + \frac{\pi \Lambda C}{\sum_{i=1}^N C_i \pi_i} + \frac{\sum_{i=1}^N C_i \pi_i (\theta_i - \bar{\theta})}{\sum_{i=1}^N C_i \pi_i} + \frac{\sum_{i=1}^N C_i \pi_i (\beta_i - \bar{\beta})}{\sum_{i=1}^N C_i \pi_i} \right) dt \\
+ V Q \left( \sigma_E + \sigma_Q + \frac{\sum_{i=1}^N C_i \pi_i (\nu_i - \bar{\nu})' (\Sigma \Sigma')^{-1}}{\sum_{i=1}^N C_i \pi_i} \right) d \tilde{W}.
\]
Proof of Proposition 4.

Covar. volatilities are positive (note that the nominal pricing kernel of the economy. Therefore, using eqs. (7) and (10) we obtain
\[ \mu_V^Q(\pi_t) - \delta(\pi_t) = \tilde{\beta}(\pi_t) + \mu_V(\pi_t) - \delta(\pi_t) + \sigma_E \sigma'_Q + \frac{\sum_{i=1}^{N} C_i \pi_i (\beta_i - \tilde{\beta}(\pi_t))}{\sum_{i=1}^{N} C_i \pi_i}, \]
(28)
and clearly \( \sigma_V^Q(\pi_t) = \sigma_V(\pi_t) + \sigma_Q \), which completes the proof. \( \blacksquare \)

(b) By eq. (27) we have
\[ \mu_V(\pi_t) - k(\pi_t) = \sum_{i=1}^{N} C_i \pi_i (k_i - \bar{k}(\pi_t)) + \sigma_M \sigma'_E. \]
From (26) we have
\[ \sigma_V(\pi_t) \sigma'_M = \sigma_E \sigma'_M + \frac{\sum_{i=1}^{N} C_i \pi_i (\nu_i - \bar{\nu}(\pi_t))' (\Sigma')^{-1} \Sigma'_{M} \sigma'_M + \sum_{i=1}^{N} C_i \pi_i (k_i - \bar{k}(\pi_t))}{\sum_{i=1}^{N} C_i \pi_i}, \]
which proves that the relationship for the real risk premium holds. Using it and (28) we have
\[ \mu_V^Q(\pi_t) - r(\pi_t) = \mu_V(\pi_t) - \bar{k}(\pi_t) + \bar{k}(\pi_t) - r(\pi_t) + \sigma_E \sigma'_Q + \frac{\sum_{i=1}^{N} C_i \pi_i (\beta_i - \tilde{\beta}(\pi_t))}{\sum_{i=1}^{N} C_i \pi_i}, \]
\[ = \sigma_V \sigma'_M + \sigma_N \sigma'_Q + \sigma_V \sigma'_Q = (\sigma_Q + \sigma_V)(\sigma_M + \sigma_Q)' = \sigma_V^Q \sigma'_N, \]
which establishes the relationship for the nominal risk premium. Note that in the second equality we used \( \bar{r}(\pi_t) = \bar{k}(\pi_t) + \tilde{\beta}(\pi_t) - \sigma_N \sigma'_Q \), and similarly define \( \bar{\theta} \).

Proof of Proposition 3. As in Veronesi (2000), we define the value-weighted probabilities \( \pi^\circ_i = (\pi_i C_i) / (\sum_{j=1}^{N} \pi_j C_j) \). Notice that \( 0 \leq \pi^\circ_i \leq 1 \) and \( \sum_{j=1}^{N} \pi^\circ_j = 1 \). Further, let \( \tilde{\beta}^\circ = \sum_{j=1}^{N} \beta_j \pi^\circ_j \), and similarly define \( \bar{\theta}^\circ \).

Then, the asset volatilities in (15) can be written as
\[ \sigma_{V,1}^Q = \sigma_{Q,1} + \frac{\bar{\beta}^\circ - \bar{\beta}}{\sigma_{Q,1}} - \frac{(\bar{\theta}^\circ - \bar{\theta})}{\sigma_{Q,1} \sigma_{E,2}} \sigma_{Q,2}, \]
\[ \sigma_{V,2}^Q = (\sigma_{Q,2} + \sigma_{E,2}) + \frac{\tilde{\beta}^\circ - \bar{\theta}}{\sigma_{E,2}} \sigma_{Q,1}, \]
\[ \sigma_{V,3}^Q = -\frac{(\bar{\beta}^\circ - \bar{\beta}) \sigma_{M,1} \sigma_{E,2} + (\bar{k} - \bar{\theta}) \sigma_{Q,1} \sigma_{E,2} + (\bar{\theta}^\circ - \bar{\theta}) (\sigma_{M,2} \sigma_{Q,1} - \sigma_{M,1} \sigma_{Q,2})}{\sigma_{M,3} \sigma_{Q,1} \sigma_{E,2}}, \]
Now given the stated conditions on the drifts and the vector \( C, \beta^\circ \leq \tilde{\beta}, \) and \( \bar{\theta}^\circ \geq \bar{\theta} \), and if all the volatilities are positive (note that \( \sigma_{E,1} = 0 \)) then \( \sigma_{V,1} \leq \sigma_{Q,1}, \) and \( \sigma_{V,2} \geq 0 \), and by \( \sigma_V^Q = \sigma_V + \sigma_Q \) we get the signs on the real volatilities as claimed. \( \blacksquare \)

Proof of Proposition 4.

The nominal market price of risk for any state variable is given by its covariance with the nominal pricing kernel of the economy. Therefore, using eqs. (7) and (10) we obtain \( \rho_i(\pi_t) = \text{Cov}(\nu_i t, \frac{dN}{N}) = \pi_i t (\nu_i t - \bar{\nu} t) \cdot (\Sigma')^{-1} \cdot \sigma'_N \). Using the fact that \( \sigma_N = (i_M + i_Q) \Sigma, \) where \( i_M = (0, 0, 1)' \), \( i_Q = (1, 0, 0)' \), and \( r_i = k_i + \beta_i - \sigma_N \sigma'_N \), provides the statement in the proposition. \( \blacksquare \)

Proof of Corollary 1.
(i) We appeal to Theorems 1 and 2 in Bergman, Grundy, and Weiner (1996). The solvency ratio \[ dZ_t = (\mu^Q_t(\pi_t) - r(\pi_t))dt + \sigma^Q_t d\tilde{W}_t \] and beliefs in (6) have the same structure as eqs. (2a) and (2b) in their paper. Notably different though, the beliefs are \( N \)-dimensional, while their general state variable \( y \) was of single dimension. However, the authors explain their results are valid for the case when \( y \) is a vector. The crux of the proof is that, subject to certain regularity conditions verified below, for given realizations of the innovations \( \{\tilde{W}_t\} \), and hence \( \{\pi_t\} \), for \( s \in (t, T] \), the risk-neutralized solvency ratio process in (18) satisfies the ‘no-crossing’ property, that is starting at a higher level for \( Z_t \) would yield a higher solvency ratio at maturity.

The payoff at maturity of the defaultable bond is an increasing and concave function of \( Z_T \). The bond price will be an increasing and concave function of \( Z_t \) if (a) \( \mu_i(\pi_t) \) and \( \sigma_i(\pi_t) \), the drift and volatility for the conditional probability of state \( i \) for \( i = 1 \cdots N \) do not depend on the level of \( Z_t \), and (b) the covariance between percentage changes in \( Z_t \) and each conditional probability process, \( \sigma^Q_i \sigma_i(\pi_t)' \), does not depend on the level of \( Z_t \). These sufficiency conditions are easily verified for each updating process in (6), and the volatility of \( Z_t \) in (15).

(ii) To establish the convexity of the spread let \( g(x, y) = \log(f(x, y)^{-1/T}) \), where \( f(x, y) \) is a positive function of scalar \( x \) and a vector \( y \) of dimension \( N \), with \( \frac{\partial^2 f(x)}{\partial x^2} < 0 \). Then, \[ \frac{\partial^2 g(x)}{\partial x^2} = \frac{1}{f^2} \left( \frac{\partial^2 f(x)}{\partial x^2} \right)^2 - f \frac{\partial^2 f(x)}{\partial x^2} > 0. \] It follows from (i) that the credit spread is convex is \( Z_t \).

\[ \textbf{Appendix 2: EM-SMM Estimation of the Regime Switching Model} \]

In this Appendix, we provide the details of the SMM estimation that uses information in both fundamentals and financial variables to estimate the parameter values and time-series of investors beliefs about the hidden states. Our procedure accounts for the fact the underlying process for fundamentals follows a continuous time model, but is observed at discrete points of time (quarterly).\(^{20}\) It also allows for the fact that the econometrician observes data only on fundamentals while agents in addition observe their pricing kernel as well and update their beliefs about the state of fundamental growth.

Since fundamentals are stationary in growth rates, we start by defining logs of variables: \( y_{2t} = \log(Y_{2t}) \), \( m_t = \log(M_t) \). Using (9) and (3) we can write
\[ dy_t = (\varphi(\pi_t) - \frac{1}{2}(\sigma_Q \sigma_Q' - \sigma_E \sigma_E')dt + \Sigma_2 d\tilde{W}_t, \]
\[ dm_t = (-\bar{\kappa}(\pi_t) - \frac{1}{2}\sigma_M \sigma_M')dt - \sigma_M d\tilde{W}_t. \]

It is immediate that investors beliefs \( \pi_t \) completely capture the state of the system \( (y_t, m_t) \) for forecasting future growth rates. The specification of the system is completed with the belief dynamics in (6).

The econometrician has data series \( \{y_{t_1}, y_{t_2}, \cdots, y_{t_K}\} \). Let \( \Psi \) be the set of parameters of the model. We start by specifying the likelihood function over data on fundamentals observed discretely using the procedure in the SML methodology of Brandt and Santa-Clara (2002). See also Duffie and Singleton (1993). Adapting their notation, let
\[ \mathcal{L}(\Psi) \equiv p(y_{t_1}, \cdots, y_{t_K}; \Psi) = p(\pi_{t_0}; \Psi) \prod_{k=1}^K p(y_{t_{k+1}} - y_{t_k}, t_{k+1}|y_{t_k}, t_k; \Psi), \]

\(^{20}\)Pioneering work in this area is in Ait-Sahalia (2002).
where $p(y_{t,k+1} - y_{t,k}, k+1 | \pi_{t,k}, t; \Psi)$ is the marginal density of fundamentals at time $t_{k+1}$ conditional on investors’ beliefs at time $t_k$. Since $\{\pi_{t_k}\}$, for $k = 1, \ldots, K$ is not observed by the econometrician, we maximize

$$E[\mathcal{L}(\Psi)] = \int \cdots \int \mathcal{L}(\Psi) f(\pi_{t_1}, \pi_{t_2}, \ldots, \pi_{t_K}) d\pi_{t_1}, d\pi_{t_2}, \ldots, d\pi_{t_K},$$

where the expectation is over all continuous sample paths for the fundamentals, $\tilde{y}_t$, such that $\tilde{y}_{t_k} = y_{t_k}, k = 1, \ldots, K$. In general along each path, the sequence of beliefs $\{\pi_{t_k}\}$ will be different.

As a first step, we need to calculate $p(y_{t,k+1} - y_{t,k}, t_{k+1} | \pi_{t,k}, t; \Psi)$. Following Brandt and Santa-Clara (2002), we simulate paths of the state variables over smaller discrete units of time using the Euler discretization scheme (see also Kloeden and Platen 1992):

$$\tilde{y}_{t+h} - \tilde{y}_t = \left( \varphi(\pi_t) - \frac{1}{2}(\sigma_Q \sigma_Q', \sigma_E \sigma_E') \right) h + \Sigma_2 \sqrt{h}\tilde{\epsilon}_t;$$

$$m_{t+h} - m_t = (-\tilde{\kappa}(\pi_t) - \frac{1}{2}\sigma_M \sigma_M') h - \sigma_M \sqrt{h}\tilde{\epsilon}_t;$$

$$\pi_{t+h} - \pi_t = \mu(\pi_t) h + \sigma(\pi_t) \sqrt{h}\tilde{\epsilon}_t$$

where $\tilde{\epsilon}$ is a $3 \times 1$ vector of standard normal variables, and $h = 1/M$ is the discretization interval. The Euler scheme implies that the density of fundamental growth over $h$ is bivariate (since $\sigma_{Q,3} = \sigma_{E,3} = 0$) normal.

We approximate $p(\cdot | \cdot)$ with the density $p_M(\cdot | \cdot)$, which obtains when the state variables are discretized over $M$ sub-intervals. Since the drift and volatility coefficients of the state variables in (6), (29) and (30) are infinitely differentiable, and $\Sigma \Sigma'$ is positive definite, Lemma 1 in Brandt and Santa-Clara (2002) implies that $p_M(\cdot | \cdot) \rightarrow p(\cdot | \cdot)$ as $M \rightarrow \infty$.

The Chapman-Kolmogorov equation implies that the density over the interval $(t_k, t_{k+1})$ with $M$ sub-intervals satisfies

$$p_M(y_{t_{k+1}} - y_{t_k}, t_{k+1} | \pi_{t_k}, t; \Psi) =$$

$$\int \int \int \phi(y_{t_{k+1}} - y; \varphi(\pi) h, \Sigma_2 \Sigma_2' h; \Psi) \times p_M(y - y_{t_k}, \pi, m, t_k + (M - 1)h | \pi_{t_k}, t_k) d\pi dy dm,$$

where $\phi(y; \text{mean, variance})$, denotes a bivariate normal density. Now $p_M(\cdot | \cdot)$ can be approximated by simulating $L$ paths of the state variables in the interval $(t_k, t_k + (M - 1)h)$ and computing the average

$$\hat{p}_M (y_{t_{k+1}} - y_{t_k}, t_{k+1} | \pi_{t_k}, t; \Psi) = \frac{1}{L} \sum_{l=1}^L \phi(y_{t_{k+1}} - y^{(l)}; \varphi(\pi^{(l)}) h, \Sigma_2 \Sigma_2' h; \Psi).$$

The Strong Law of Large Numbers (SLLN) implies that $\hat{p}_M \rightarrow p_M$ as $L \rightarrow \infty$.

To compute the expectation in (31), we simulate several paths of the system (32) – (34) ‘through’ the full time series of fundamentals. Each path is started with an initial belief, $\pi_{t_0} = \pi^*$, the stationary beliefs implied by the generator matrix $A$. In each time interval $(t_k, t_{k+1})$ we simulate (M-1) successive values of $\tilde{y}_t(s)$ using the discrete scheme in (32), and set $\tilde{y}_{t_k} = y_{t_k}$. The results in the paper use $M = 90$ for quarterly data, so that shocks are approximated at roughly a daily frequency. The pricing kernel and beliefs along the entire path of the $s^{th}$ simulation are obtained by
We approximate the expected likelihood as

\[ \hat{L}^{(s)}(\Psi) = \frac{1}{S} \sum_{s=1}^{S} \prod_{k=0}^{K-1} \hat{p}_M(y_{t_{k+1}}^{(s)} - y_{t_{k}}^{(s)}, t_{k+1} | \pi_{t_k}^{(s)}, t_k; \Psi), \]

where \( \hat{p}_M(\cdot) \) is the density approximated in (36). The SLLN implies that \( \hat{L}^{(s)}(\Psi) \to E[\mathcal{L}(\Psi)] \) as \( S \to \infty \). We often report \( \hat{\pi}_{tk} = 1/S \sum_{s=1}^{S} \pi_{tk}^{(s)} \) which is the econometrician’s expectation of investors’ belief at \( t_k \).

To extract investors’ beliefs from price data as well as fundamentals we extend the SML method to a simulated SMM method that in addition uses information contained in the time series of price-earning ratios and a set of Treasury bond yields observed at the same discrete set of dates. To extract model parameters and investors’ uncertainty from asset prices, we use the pricing formulas for the asset value and bond prices in Proposition 1 and proceed as follows: First, for given parameters \( \Psi \) of the regime shift model, we use the time series probabilities \( \{ \hat{\pi}_{t_1}, \cdots, \hat{\pi}_{t_K} \} \) filtered as shown above. From these, we can compute the time series of model-implied asset value-to-earning ratios and bond yields using the formulas derived in Proposition 1

\[ \frac{\widehat{V}/E_{tk}}{C \cdot \hat{\pi}_{tk}}, \quad \hat{\tau}_{tk}(\tau) = -\frac{1}{\tau} \log \left( B(\tau) \cdot \hat{\pi}_{tk} \right). \]

We note that the constants \( C \)'s and the functions \( B(\tau) \) depend both on the parameters of the fundamental processes, \( \Psi \), and, the kernel parameters, \( \Psi \). Hence, let the pricing errors be denoted

\[ e_{tk} = \left( \frac{\widehat{V}/E_{tk}}{C \cdot \hat{\pi}_{tk}} - \frac{\widehat{V}/E_{tk}}{C \cdot \hat{\pi}_{tk}} \right), \]

where, \( \tau_1 = 3 \) months, \( \tau_2 = 1 \) year, \( \tau_3 = 3 \) years, \( \tau_4 = 5 \) years, and, \( \tau_5 = 10 \) years. Also note that since the pricing formulas are linear in beliefs, \( 1/S \sum_{s=1}^{S} C \cdot \pi_{tk}^{(s)} = C \cdot \hat{\pi}_{tk} \) (and similarly for the bond yields), and no information is lost by simply evaluating the errors at the econometrician’s conditional mean of beliefs.

To estimate \( \Psi \) from data on fundamentals as well as financial variables, we form the overidentified SMM objective function as in (22). The moments used are the errors from financial variables and the scores of the log likelihood function from fundamentals. Since the number of scores in \( \frac{\partial \log \mathcal{L}(\hat{\pi})}{\partial \hat{\pi}(t_k)} \) equals the number of parameters driving the fundamental processes in \( \Psi \), and the number of pricing errors is six, the statistic \( e \) in (22) has a chi-squared distribution with six degrees of freedom. Since we find the processes of financial errors, \( \{ e_{tk} \} \), to be serially correlated, while the scores are not, we diagonally partition the matrix \( \Omega \) into two parts: \( \Omega_1 \), and \( \Omega_2. \) \( \Omega_1 \) is estimated using the Newey-West correction (see Hamilton 1994, Eq. 14.1.19).

We finally describe the steps of the recursive EM algorithm to estimate the model.

1. Let \( \{ P_{tk}^{(n)} \} \) be the estimated series of bond prices at each date from the estimation of iteration \( n \), starting with \( P_{tk}^{(0)} = 1 \) for all \( k \). Let \( \hat{\Psi}^{(n)} \) be the estimated parameters at the \( n \)th stage, \( \{ \hat{\pi}_{tk}^{(n)} \} \) be the series of filtered beliefs.

2. Approximate \( V_{tk}^{Q(n)} = P_{tk}^{E} + P_{tk}^{Q(n)} \cdot D_{tk}^{Q} \) at each date, where \( P_{tk}^{E} \) is the market value of the S&P 500 at each date. \( D_{tk}^{Q} \) is the face value of nominal debt at each period and is given by its

\[ \text{As has been noted by several authors, price earnings ratios and Treasury Yields follow highly persistent processes (see also Figure 3). We find the pricing errors remain correlated with lagged values of more that 24 quarters. For example, the autocorrelation of the 3-month Treasury Bill with its 24-quarter lagged value is 0.33. We therefore, use the parameter J in the Newey-West correction of 24.} \]
discrete approximation of Assumption 6 by $D_{tk}^Q = \exp\left(\sum_{k=1}^{K} (r_{tk} - \delta_t^{(n)}) \cdot 1/4\right) D_{tk}^Q$, where $r_{tk}$ is the annualized 3-month Treasury yield, and $\delta_t^{(n)} = E_{tk}/V_{tk}^{Q(n)}$ The empirical asset value-to-earnings ratio at date $t$ is $V_{tk}^{Q(n)}/E_{tk}$ and the solvency ratio is $Z_{tk}^{(n)} = V_{tk}^{Q(n)}/D_{tk}^Q$.

3. The expectation step of the algorithm is trivial since by Corollary 1(i) for a given $\{\pi_{tk}\}$ series, $P_{tk}^{(n)}$ is a monotonic function of $Z_{tk}^{(n)} = V_{tk}^{Q(n)}/D_{tk}^Q$. Therefore, the objective function to minimize is simply the SMM objective in (22) given by $c(\psi|\psi(n))$, where the dependence of the objective on the previous parameter estimates arises due to the assumed bond price series $\{P_{tk}^{(n)}\}$.

4. Using the SMM procedure above, estimate the parameters at the $n+1$st stage $\psi^{(n+1)}$.

5. Using Proposition 5, formulate the bond price at each date as

$$P_{tk}^{(n+1)} = P(Z_{tk}^{(n)}, \pi_{tk}^{(n)}, t_k, t_k + T; \psi^{(n+1)})$$

6. Repeat steps 1 to 5 until the convergence of $\psi^{(n)}$.

In step 4, $c(\psi^{(n+1)}|\psi(n)) \leq c(\psi(n)|\psi(n))$, which implies that the sequence $c(\psi(n)|\psi(n))$ is bounded and monotonically decreasing and hence converges.

### Selection of Number of Regimes

We test for the number of drift states for the two fundamental series in a univariate regime-switching framework. To do this, we use the likelihood ratio tests that adjust for the presence of nuisance parameters unidentified under the null hypothesis (for e.g., under the 1-state null hypothesis, the transition probabilities of a 2-state model are not identified). The likelihood function is estimated from each series using the SMM procedure above, without use of financial variables. Computing exact critical values for the alternative of two states over the null of a single state variable would be numerically very intensive. However, we found that the likelihoods are very similar to the case where the data are observed discretely and use the critical values computed by Garcia (1998). For earnings growth, the log likelihood ratio (LLR) for the two state over the single state specification attains a value 13.85, which has a $P$-value of less than one percent. The LLR for a 3-state over a 2-state specification is very small of the order of $10^{-3}$, suggesting insignificant gains in modeling a third state for the earnings drift. For inflation, the LLR for the two state specification over the single state specification is 16.54, which once again has a $P$-value of less than one percent. The LLR for a three state over a two state is 13.21, however, we do not have analogous critical values available in this case. The LLR for a four state over a three state specification has a small LLR of under $10^{-2}$.

### Appendix 3: Solving the N-State PDE by Projection Methods

We use projection methods described in Judd (1999) to solve the partial differential equation (PDE) that solves the density of asset returns under the risk-neutral measure. The nominal asset return process under the risk-neutral measure is $dV_t/V_t = r_t dt + \sigma_{\pi_t} \, dW^*$, where $\sigma_{\pi_t}$ is given in Proposition 2 (b). By usual no-arbitrage arguments, it can be shown that the risk-neutral asset return density function over primitive states under the risk-neutral measure can be characterized by the usual Fokker-Planck backward equation (c.f., for e.g., Karlin and Taylor 1982). The density function $f(z, \pi_t, t)$ satisfies the PDE in equation (19). The initial condition is $f(z_t, \pi_t, 0) = \delta(z - z_t)$, where $\delta(\cdot)$ stands for the Dirac function, and the last argument of $f(\cdot, \cdot)$ denotes time to maturity.
To simplify the notation, we define the vector function, \( \mu(\pi) = (\mu_1(\pi), \mu_2(\pi), \ldots, \mu_N(\pi))^T \), and, analogously for \( \sigma(\pi), \sigma_V(\pi), \) and \( \rho(\pi) \). We also omit time subscripts.

Define the functions
\[
A_1(\pi) = (\mu(\pi) - \rho(\pi_i))(C \cdot \pi)^2, \\
A_2(\pi) = \sigma_V(\pi)(C \cdot \pi)^2, \\
A_3(\pi) = \sigma_V(\pi)(C \cdot \pi)^2, \quad \text{and} \\
A_4(\pi) = \sigma(\pi)(C \cdot \pi)^2.
\]

\( A_1(\pi) \) and \( A_3(\pi) \) are \( N \times 1 \) vector functions of \( \pi \), \( A_2(\pi) \) is a scalar function, and \( A_4(\pi) \) is an \( N \times N \) matrix. Notice that \( A_i(\pi), i = 1, \ldots, 4 \) are ordinary polynomials of finite length in \( \pi \). Now further define
\[
h_1(\pi) = \frac{1}{2} A_4(\pi) \\
h_2(\pi) = A_1(\pi) - i\omega_1 A_3(\pi), \\
h_3(\pi) = [-r(\pi) - \omega_2] \cdot (C \cdot \pi)^2 + \frac{i\omega_1}{2} A_2(\pi) - \frac{\omega_1^2}{2} A_2(\pi)
\]

Now taking a Fourier transform of the PDE (19) with respect to \( z \) and using \( \omega_1 \) as a frequency variable, and following it with a Laplace Transform with respect to the time to maturity, \( \tau \), with frequency variable \( \omega_2 \) implies that
\[
-e^{i\omega_1 z t}(C \cdot \pi)^2 = 1' \left( \hat{f}_{\pi,\pi} \odot h_1(\pi) \right) 1 + \hat{f}_{\pi} h_2(\pi) + \hat{f} h_3(\pi), \tag{38}
\]
where the partial derivatives are written w.r.t. the vector \( \pi \), for e.g., \( f_{\pi} = \{f_{\pi_1}, \ldots, f_{\pi_N}\} \), \( \odot \) stands for element-by-element multiplication, and \( 1 \) stands for the unit vector of length \( N \). For given \( \omega_1, \omega_2 \), (38) is a linear PDE of the parabolic type in \( \pi \) with no known explicit analytical solution. \( h_1(\pi) \) and \( h_3(\pi) \) are ordinary polynomials of order \( 6 \) in \( \pi_i \), and \( h_2 \) is of order \( 5 \). The frequency variables \( \omega_1 \) and \( \omega_2 \) can be thought of as parameters of the PDE. Anticipating a future Laplace Inversion with respect to \( \omega_2 \) of the solution to (38), we first solve the slightly simpler PDE
\[
-1(C \cdot \pi)^2 = 1' \left( \hat{f}_{\pi,\pi} \odot h_1(\pi) \right) 1 + \hat{f}_{\pi} h_2(\pi) + \hat{f} h_3(\pi). \tag{39}
\]

Due to homogeneity, the Inverse Laplace Transform of the solution of (38) with respect to \( \omega_2 \), will coincide with \( e^{i\omega_1 z \tau} \) the Inverse Laplace Transform of the solution of (39) with respect to \( \omega_2 \) (the logic is simply that \( e^{i\omega_1 z \tau} \) will pass out of the integral for the Laplace Inversion with respect to \( \omega_2 \)). The difficulty in formulating an analytical solution to (39) stems mainly due to the fact that the system cannot be uncoupled into \( N \) separate ODEs in \( \pi_i, i = 1, \ldots, N \).

We proceed by formulating an ‘approximate’ solution to (39) using projection methods (Judd 1999, Chapter 11).

**Step 1.** Choice of individual basis functions. We choose the Legendre polynomials in each of the \( N \) dimensions: The Legendre polynomials on \([-1, 1]\) for the basis in the \( n \)-th dimension, are given by
\[
q_{n,m}(\pi_n) = (-1)^m \frac{d^m}{dx^m}[(1 - \pi_n^2)^m],
\]
for \( m = 1, 2, \ldots \), which satisfy the recursive scheme
\[
q_{n,m+1}(x) = \frac{2m + 1}{m + 1} x q_{n,m}(x) - \frac{m}{m + 1} q_{n,m-1}(x) \tag{40}
\]
and to restrict them to the domain $[0, 1]$, use $p_{n,m}(\pi_n) = \frac{g_{n,m}(2\pi_n-1)}{\|g_{n,m}(\pi_n)\|}$, where $\| \cdot \|$ denotes the $L^2$ norm on the space $[0, 1]$. The family $\{p_{n,m}(\pi_n)\}_{m=1,2,\ldots}$ are orthonormal polynomials over $[0, 1]$.

**Step 2.** Choose a basis of ‘complete’ polynomials over the space $[0, 1]^N$. The basis of degree $M$ for $N$ dimensions is given by

$$P_M^N = \{p_{i_1i_1}(\pi_1), \ldots, p_{N,i_N}(\pi_N)\} \sum_{n=1}^{N} i_n \leq M, 0 \leq i_1, \ldots, i_N\}$$

As explained in (Judd 1999, pp. 239), the set of complete polynomials grows polynomially in $N$, as opposed to the tensor product basis which would use every possible product of the degree-$M$ individual basis functions and hence would grow at the rate of $M^N$. The complete polynomials asymptotically, as $M$ becomes large, provide as good an approximation as the tensor product, but with far fewer elements.\(^{22}\) Let $M^c$ be the length of the complete polynomial basis. Finally, write the complete basis as $\{\psi_m(\pi_1, \ldots, \pi_N)\}_{m=1,\ldots,M^c}$, where $\psi_m(\pi_1, \ldots, \pi_N) = p_{i_1}p_{i_2} \cdots p_{i_N}$, for $\sum_{n=1}^{N} i_n \leq M, 0 \leq i_1, \ldots, i_N$. Extend the $L^2$ norm over the $N$-dimensional space as the $N$-fold integral $\int_0^1 \cdots \int_0^1 \psi_m(\pi_1, \ldots, \pi_N) d\pi_1 \cdots d\pi_N$, and accordingly the inner product $<\psi_i, \psi_j> = \int_0^1 \cdots \int_0^1 \psi_i(\pi_1, \ldots, \pi_N) \psi_j(\pi_1, \ldots, \pi_N) d\pi_1 \cdots d\pi_N$. It can be verified that the basis of complete polynomials in orthonormal on $[0, 1]^N$.

**Step 3** Let $D(y)$ be the differential operator associated with the PDE (38), i.e.

$$D(y) = (C \cdot \pi)^2 + 1'(f_x \pi \circ h_1(\pi)) 1 + f_x h_2(\pi)' + \hat{f} h_3(\pi). \quad (41)$$

Then any solution to the PDE, $\hat{y}$, will be written as $D(\hat{y}) = 0$. Write the candidate solution as $\hat{y}(\pi_1, \ldots, \pi_N) = \sum_{m=1}^{M^c} a_m(\omega_1, \omega_2) \psi_m(\pi_1, \ldots, \pi_N)$.

**Step 4** The approximation is made by choosing coefficient functions $a_m(\omega_1, \omega_2)$ to solve the minimization problem:

$$\min_{a_1(\omega_1, \omega_2) \cdots a_{M^c}(\omega_1, \omega_2)} \int_0^1 \cdots \int_0^1 D(\hat{y})^2 d\pi_1 \cdots d\pi_N. \quad (42)$$

The minimization is nothing more than a standard least-squares problem, and yields the $M^c$ first order conditions (see the Galerkin Method in Boyd 1989)

$$<D(\hat{y}(\pi_1, \ldots, \pi_N), \psi_m(\pi_1, \ldots, \pi_N)> = 0,$$

for $m = 1, \ldots, M^c$. Because each $\psi_m(\pi_1, \ldots, \pi_N)$ is simply the product of linear combinations of simple polynomials (see Eq. (40), each of the f.o.c’s leads to a linear equation in the coefficient functions $a_m(\omega_1, \omega_2)$, $m = 1, \ldots, M^c$. Moreover, due to the linearity, these f.o.c’s can be completely characterized analytically.

David (1997) (Property 1) showed that each updating process $\{\pi_i\}, i = 1 \cdots, N$, has ‘entrance’ boundaries at 0 and 1 — neither boundary can be reached from the interior of the state space, but it is possible to consider processes that begin there. In such cases, it is impossible to impose arbitrary values for the candidate function at the boundaries. The $M^b$ entrance boundary conditions are:

\(^{22}\)For e.g., for $M = 4$, the number of terms required for the complete polynomial approximation is $M^c = 1 + N + N(N + 1)/2 + N(N - 1)(N - 2)/6$. 

8
For the four-state problems that we solve in this paper, there are 16 boundaries, \( B \), of the hypersimplex \( \pi_1 \geq 0, \sum_{i=1}^{N} \pi_i \leq 1 \).

Overall, for each value of the frequency variables, \( \omega_1 \) and \( \omega_2 \), we have an overidentified system of \( M^c + M^b \) linear equations in the \( M^c \) unknowns \( \hat{a}_m(\omega_1, \omega_2) \), \( m = 1, \ldots, M^c \), and constant terms. Denote the \( (M^c + M^b) \times 1 \) vector of constants from each equation as \( c(\omega_1, \omega_2) \), and the \( (M^c + M^b) \times M^c \) coefficient matrix as \( A(\omega_1, \omega_2) \). Analogous to regression coefficients, the best-fitting set of coefficients satisfies:

\[
\hat{a}(\omega_1, \omega_2) = \left( A^T(\omega_1, \omega_2) \cdot A(\omega_1, \omega_2) \right)^{-1} \cdot A^T(\omega_1, \omega_2) \cdot c(\omega_1, \omega_2).
\]

**Step 5** At this point, the solution to the PDE can be written as the Laplace Transform of a Fourier transform. We provide an approximate but very fast inversion procedure of the Laplace Transform of the coefficient function \( \hat{a}_m(\omega_1, t) \) defined implicitly in

\[
a_m(\omega_1, \omega_2) = \int_0^\infty e^{-\omega_2 t} \hat{a}_m(\omega_1, t) dt.
\]

Any other valid method of inverting will suffice. In practice, the inversion is performed on the Bromwich contour: If \( a(\omega_1, \omega_2) \) is analytic in the region \( \text{Re}(\omega_1) > c_0 \), then,

\[
\hat{a}_m(\omega_1, t) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} e^{\omega_2 t} a(\omega_1, \omega_2) d\omega_2,
\]

for any \( c > c_0 \). A fast approximation of (44) is obtained using the Euler method of Abate and Whitt (1995). The method has been used successfully in option pricing problems by Fu, Madan, and Wang (1997) and Linetsky (1999). Evaluating the integral by a trapezoidal rule, Abate and Whitt show that the integral in (44) can be written as the nearly alternating series

\[
\hat{a}_m(\omega_1, t) = \frac{e^{A/2}}{2t} \text{Re} \left[ a_m \left( \omega_1, \frac{A}{2t} \right) \right] + \frac{e^{A/2}}{t} \sum_{j=1}^{\infty} (-1)^j \text{Re} \left[ a_m \left( \omega_1, \frac{A + 2j\pi i}{2t} \right) \right],
\]

where \( c = A/(2t) \). The series is approximated by the weighted average of partial sums

\[
\hat{a}_1^{\bar{w}}(\omega_1, t) = 2^{-\bar{w}} \sum_{k=0}^{\bar{w}} C^\bar{w}_k s_{\bar{w}+k}(\omega_1, t),
\]

where

\[
s_{\bar{w}+k}(\omega_1, t) = \frac{e^{A/2}}{2t} \text{Re} \left[ a_m \left( \omega_1, \frac{A}{2t} \right) \right] + \frac{e^{A/2}}{t} \sum_{j=1}^{\bar{w}+k} (-1)^j \text{Re} \left[ a_m \left( \omega_1, \frac{A + 2j\pi i}{2t} \right) \right],
\]

and \( C^\bar{w}_k = \frac{\bar{w}!}{\bar{w}!(\bar{w}-k)!} \). The choice of parameters \( A = 2ct, \bar{w} \) and \( \bar{r} \) is dictated by the desired accuracy. 

\( A \) controls for the discretization error, which is bounded from above by \( \frac{\exp(-A)}{1-\exp(-A)} \). For example, \( A = 18.4 \) will lead to an upper bound of the order of \( 10^{-8} \). Abate and Whitt suggest starting with \( \bar{w} = 11 \), and \( \bar{r} = 15 \), and to adjust \( \bar{r} \) as needed. We have found these values to provide accurate

---

23 For the four-state problems that we solve in this paper, there are 16 boundaries: \( \pi_i = 0 \), and \( \pi_i = 1 \), and \( \pi_i = 0, \pi_j = 0 \), for \( j \neq i \) for \( i = 1, \ldots, 4 \).
values when comparing to Monte-Carlo simulations. Linetsky (1999) in his examples finds \( \bar{r} = 5 \) to give very accurate values for double-barrier options.

Using the coefficients \( a(\omega_1, T - t) \) in (44), we can write the Fourier transform of the asset returns simply as

\[
f(\omega_1, \pi, T - t; z_t) = e^{\omega_1 z_t} \cdot \sum_{m=1}^{M_c} a_m(\omega_1, t) \psi_m(\pi_1, \cdots, \pi_N). \tag{48}
\]

Finally, to perform the Fourier inversion with respect to the log asset value as well as the computation of defaultable bond prices we define the following functions as in Scott (1997):

\[
\begin{align*}
g_1(z_t, \omega_1, \pi, T - t) &= f(z_t, 1/i + \omega_1, \pi_t, T - t)/f(z_t, 1/i, \pi_t, T - t), \quad \text{and,} \\
g_2(z_t, \omega_1, \pi_t, T - t) &= f(z_t, w_1, \pi_t, T - t)/f(z_t, 0, \pi_t, T - t). \tag{49, 50}
\end{align*}
\]

Then, by the Fourier Inversion formula (c.f., for example, Kendall and Stuart 1977) the probability distribution functions can be written as

\[
\Pi_j(Z_t, \pi_t, t, T) = \frac{1}{2} + \frac{1}{\Pi} \int_0^\infty \text{Re} \left[ \frac{g_j(z_t, \omega_1, \pi_t, T - t)}{i \omega_1} \right] d\omega_1. \tag{51}
\]

\[\blacksquare\]

**Lemma 2** The partial delta of defaultable bond prices in Proposition 5 with respect to the log solvency is given by

\[
P_z(Z_t, \pi_t, t, T) = \frac{1}{Z} \cdot [B(\cdot, \cdot) \cdot \Pi_2(\cdot, \cdot) + G(\cdot, \cdot) \cdot (1 - \Pi_1(\cdot, \cdot) - \Pi_{1z}(\cdot, \cdot))]. \tag{52}
\]

The partial delta of the defaultable bond price with respect to beliefs is given by

\[
P_\pi(Z_t, \pi_t, t, T) = B(\cdot, \cdot) \cdot \Pi_2(\cdot, \cdot) + B(\cdot, \cdot) \cdot \Pi_{2\pi} + G(\cdot, \cdot) \cdot (1 - \Pi_1(\cdot, \cdot)) - G(\cdot, \cdot) \cdot \Pi_{1\pi}. \tag{53}
\]

The partial derivatives with respect to log solvency value, \( \Pi_{iz}(\cdot, \cdot) \), and, beliefs, \( \Pi_{i\pi}(\cdot, \cdot), i = 1, 2 \), are provided in the proof.

**Proof:** We use the \( M_c \) term polynomial approximation of the Fourier Transform \( f(\omega_1, \pi, T - t; z_t) \) in (48) and the characterization in Proposition 5. The partial derivatives of \( B(\cdot, \cdot) \) and \( G(\cdot, \cdot) \) are computed directly from those of \( f(\omega_1, \pi, T - t; z_t) \) at the two points \( \omega_1 = 0 \) and \( \omega_1 = 1/i \) respectively. Similarly, the functions \( g_1 \) and \( g_2 \) in (49) and (50) can be written as series whose partial derivatives are evident. Now by passing partial derivatives under the integral of the Fourier inversion, one obtains the desired functions. For example,

\[
\Pi_{j\pi}(z_t, K, \pi_t, T - t) = \frac{1}{\Pi} \int_0^\infty \text{Re} \left[ \frac{e^{-i \omega_1 \log K} g_{j\pi}(\omega_1, \pi_t, T - t; z_t)}{i \omega_1} \right] d\omega_1.
\]

We can similarly find \( \Pi_{jz}(z_t, K, \pi_t, T - t) \). Finally, using the product and chain rules of differentiation, and the partial derivatives above provide the stated expressions. \( \blacksquare \)

For evaluating expected returns and volatilities of bond returns we provide the following result.

**Corollary 2** The volatilities of \( dP(\cdot, \cdot)/P(\cdot, \cdot) \) with respect to the two shocks in the economy are given by

\[
\sigma_P(Z_t, \pi_t, t, T) = \frac{P_z(\cdot, \cdot) \cdot \sigma_Q^P(\pi_t) + P_\pi(\cdot, \cdot) \cdot \sigma(\pi_t)}{P(\cdot, \cdot)}, \tag{54}
\]

10
where $\sigma(\pi_t)$ are the volatilities of beliefs with respect to the two shocks as in (7), and, $\sigma^Q_V(\pi_t)$ are the volatilities of the nominal asset value as shown in (15). The partial derivatives of the bond price with respect to the solvency ratio and the beliefs are provided in the proof. The expected nominal return on bonds is

$$\mu_p(Z_t, \pi_t, t, T) = r(\pi_t) + \sigma_p(\cdot, \cdot) \cdot \sigma_N,$$

(55)

where $\sigma_N$ are the nominal market prices of risks defined in (10).

**Proof.** Using the partial derivatives in Lemma 2, the volatilities follows from a straightforward application of Ito’s Lemma, and the expected returns follow from the equilibrium condition that excess returns equal negative of the covariance of asset returns with the pricing kernel. □