Internet Appendix for:
"Issuer Quality and the Credit Cycle"*

Robin Greenwood and Samuel G. Hanson
Harvard University

June 2011

— Contents —

* Previously circulated as "Issuer Quality and Corporate Bond Returns."

A: Calculations for Section II

Assume that all random variables are independent. For simplicity, we assume that we forecast the excess returns on bonds with $\beta_\theta = 1$, so that $s_t = \pi_t + \delta_t$ and $E_t[rx_{t+1}] = \delta_t$. It trivially follows that the magnitude of regression coefficients will be larger for high default-risk firms than for low default-risk firms since $E_t[rx_{\theta,t+1}] = \beta_\theta \delta_t$ and $\beta_L < \beta_H$.

A.1 Time-series forecasting regressions of excess corporate bond returns

The coefficient from a univariate forecasting regression of rx_{t+1} on quality ($d_H - d_L$) is

$$b_{d_H - d_L} = \frac{-\frac{\beta_H - \beta_L}{\gamma}\sigma_\delta^2}{2\sigma_\varepsilon^2 + \left(\frac{\beta_H - \beta_L}{\gamma}\right)\sigma_\delta^2} < 0. \tag{A1}$$

The coefficient from a univariate forecasting regression of returns on total issuance ($d_H + d_L$) is:

$$b_{d_H + d_L} = \frac{-\frac{\beta_H + \beta_L}{\gamma}\sigma_\delta^2}{4\sigma_\xi^2 + 2\sigma_\varepsilon^2 + \left(\frac{\beta_H + \beta_L}{\gamma}\right)\sigma_\delta^2} < 0. \tag{A2}$$

We next consider a multivariate forecasting regression of rx_{t+1} on $d_H - d_L$ and $d_H + d_L$. We have

$$\begin{bmatrix} b_{d_H - d_L} \\ b_{d_H + d_L} \end{bmatrix} = \frac{-(\sigma_\delta^2 / \gamma)}{2\sigma_\varepsilon^2\left(4\sigma_\xi^2 + 2\sigma_\varepsilon^2 + \left(\frac{\beta_H + \beta_L}{\gamma}\right)\sigma_\delta^2\right) + \left(\frac{\beta_H - \beta_L}{\gamma}\right)^2 \sigma_\delta^2 (4\sigma_\xi^2 + 2\sigma_\varepsilon^2)} \begin{bmatrix} (4\sigma_\xi^2 + 2\sigma_\varepsilon^2)(\beta_H - \beta_L) \\ 2\sigma_\varepsilon^2(\beta_H + \beta_L) \end{bmatrix}. \tag{A3}$$

As σ_ξ^2 grows large, or as $\sigma_\varepsilon^2 \to 0$, aggregate debt issuance becomes less informative and relative issuance (i.e., issuer quality) becomes more informative about variation in δ_t.

The coefficient in a univariate forecasting regression of rx_{t+1} on spreads s_t is given by

$$b_s = \frac{\sigma_\delta^2}{\sigma_\pi^2 + \sigma_\delta^2} > 0. \tag{A4}$$

Next consider a multivariate regression of rx_{t+1} of $d_H - d_L$ and spreads s_t. We have

$$\begin{bmatrix} b_{d_H - d_L} \\ b_s \end{bmatrix} = \frac{\sigma_\delta^2}{\det[\mathbf{V}]} \begin{bmatrix} -\sigma_\pi^2(\beta_H - \beta_L)/\gamma \\ 2\sigma_\varepsilon^2 \end{bmatrix}. \tag{A5}$$

Where $\det[\mathbf{V}] > 0$ is the determinant of the variance-covariance matrix of $d_H - d_L$ and spreads s_t. As σ_π^2 grows or as σ_ε^2 falls, credit spreads become less informative and quality becomes more informative about δ_t.

A.2 Deriving ISS$_t$

Suppose that firm-level debt issuance is given by $d_{i,t} = \xi_t + \varepsilon_{i,t} - (\beta_i / \gamma) \cdot \delta_t$. Under the simplifying assumption β_i and $\varepsilon_{i,t}$ are independent and normally distributed, we can show that

$$ISS_t = E_t[\beta_i \mid High\ d^*_{i,t}] - E_t[\beta_i \mid Low\ d^*_{i,t}] = -\frac{\phi(\Phi^{-1}(0.20))}{0.10} \frac{(\sigma^2_\beta \delta_t)/(\gamma \sigma_\varepsilon)}{\sqrt{1 + (\delta_t / \gamma)^2 (\sigma^2_\beta / \sigma^2_\varepsilon)}}. \quad \text{(A6)}$$

It is easy to check that ISS_t is a decreasing function of δ_t. Furthermore, as σ^2_ε grows large relative to σ^2_β, so that individual firm debt issuance decisions are a noisy signal of expected returns, ISS_t becomes approximately linear in δ_t. Under these conditions, ISS_t is proportional to $B^{xs}_t = -\delta_t / \gamma$, the coefficient from a cross-sectional regression of $d_{i,t}$ on β_i.

To derive (A6), first note that the coefficient from a cross-sectional regression of β_i on $d_{i,t}$ is $-[\sigma^2_\beta(\delta_t / \gamma)] / [\sigma^2_\varepsilon + \sigma^2_\beta(\delta_t / \gamma)^2]$. Therefore, since β_i and $\varepsilon_{i,t}$ are normally distributed we have:

$$E_t[\beta_i \mid High\ d^*_{i,t}] - E_t[\beta_i \mid Low\ d^*_{i,t}] = -\frac{\sigma^2_\beta(\delta_t / \gamma)}{\sigma^2_\varepsilon + \sigma^2_\beta(\delta_t / \gamma)^2} \left(E_t[d^*_{i,t} \mid High\ d^*_{i,t}] - E_t[d^*_{i,t} \mid Low\ d^*_{i,t}] \right). \quad \text{(A7)}$$

Finally, equation (A6) follows from noting that

$$E_t[d^*_{i,t} \mid High\ d^*_{i,t}] = E_t[\varepsilon_{i,t} - \beta_i(\delta_t / \gamma) \mid \varepsilon_{i,t} - \beta_i(\delta_t / \gamma) \geq d^{0.80}_t] = -(\delta_t / \gamma) + \sqrt{\sigma^2_\varepsilon + \sigma^2_\beta(\delta_t / \gamma)^2}\ \frac{\phi(\Phi^{-1}(0.80))}{1 - \Phi(\Phi^{-1}(0.80))}$$

$$E_t[d^*_{i,t} \mid Low\ d^*_{i,t}] = E_t[\varepsilon_{i,t} - \beta_i(\delta_t / \gamma) \mid \varepsilon_{i,t} - \beta_i(\delta_t / \gamma) \leq d^{0.80}_t] = -(\delta_t / \gamma) - \sqrt{\sigma^2_\varepsilon + \sigma^2_\beta(\delta_t / \gamma)^2}\ \frac{\phi(\Phi^{-1}(0.20))}{\Phi(\Phi^{-1}(0.20))},$$

where we have made using of the express ions for the means of left- and right- truncated normal random variables (see e.g., Greene (2003), p. 759).

B: Data Definitions

Where applicable, we provide the relevant Compustat data items from the Fundamentals Annual file.

B.1 Compustat measures of issuance

We follow Baker and Wurgler (2002) and define net equity issues as the change in book equity minus the change in balance sheet retained earnings divided by lagged assets. Net debt issues in equation (8) are defined as the residual change in assets (the change in book assets minus the change in book equity), divided by lagged assets. Issuance in calendar year *t* is based on firm fiscal year-ends that fall in year *t*.

Book equity is stockholder's equity, plus balance sheet deferred taxes (item *TXDB*) and investment tax credits (*ITCB*) each when available, minus preferred stock. For stockholder’s equity we use *SEQ*; if *SEQ* is missing we use the book value of common equity (*CEQ*) plus the book value of preferred stock (*PSTK*); finally, we use total assets (*AT*) minus total liabilities (*LT*) minus minority interest (*MIB*). For preferred stock we use redemption value (*PSTKRV*), liquidation value (*PSTKL*), and book value (*PSTK*) in that order.

We obtain similar results if we use a more narrow definition of debt issuance which excludes non-bond and non-loan liabilities such as trade credit. This measure is defined as the change in debt in current liabilities (*DLC*) plus long-term debt (*DLTT*) divided by lagged assets.

B.2 Characteristic definitions

Firm characteristics that use CRSP market data are measured as of December *t*. Data from financial statements are from firm fiscal year ends that fall in *t*. For instance, *EDF* is computed using market data through December *t* and, in the case of firms with December fiscal year-ends, debt issuance is the normalized change in debt over the prior 12 months. Since we measure *EDF* at year-end, it reflects the change in debt over the prior year and captures any incremental risk that creditors are assuming. This is appropriate: if a transaction significantly raises leverage, we no longer want to say that the firm is low risk. This corresponds to the agency practice of rating new debt issues *pro-forma* for the amount of debt that the firm is adding.

Expected Default Frequency (*EDF*): *EDF* is computed following the procedure in Bharath and Shumway (2008). For each firm-year, we calculate $EDF_{i,t} = \Phi[-(\ln[(E_{i,t} + F_{i,t})/F_{i,t}] + (\mu_{i,t} - 0.5\sigma^2_{V_{i,t}}))/\sigma_{V_{i,t}}]$, where $E_{i,t}$ is the market value of the firm’s equity as of December, $F_{i,t}$ is the face value of the firm’s debt computed as short-term debt (*DLC*) plus one-half of long-term debt (*DLTT*), $\mu_{i,t}$ is the firm’s asset drift, $\sigma_{V_{i,t}}$ is the asset volatility, and $\Phi(\cdot)$ is the standard normal CDF. Following Bharath and Shumway (2008), we estimate $\mu_{i,t}$ using the firm's cumulative stock return over the prior 12 months, and estimate asset volatility using $\sigma_{V_{it},Naive} = (E_{it} / (E_{it} + F_{it}))\sigma_{E_{it}} + (F_{it} / (E_{it} + F_{it}))(0.05 + 0.25 \cdot \sigma_{E_{it}})$ where $\sigma_{E_{i,t}}$ is the annualized volatility of monthly stock returns over the prior year. As such, this construction is not an exact implementation of the Merton (1974) model, but Bharath and Shumway show that it is slightly better at forecasting defaults than the more complicated version which requires solving a system of nonlinear equations.

Shumway Distress (*SHUM*): We use the bankruptcy hazard rate estimated by Shumway (2001), $SHUM = \exp(H) / (1 - \exp(H))$ where:

$H = -13.303 - 1.982 \cdot (NI / A) + 3.593 \cdot (L / A) - 0.467 \cdot RELSIZE - 1.809 \cdot (R - R_M) + 5.791 \cdot \sigma.$

(*NI/A*) is net income over period-end assets, (*L/A*) is total liabilities over assets, *RELSIZE* is the log of a firm's market equity divided by the total capitalization of all NYSE and AMEX stocks, $R - R_M$ is firm's cumulative return over the prior 12-months minus the cumulative return on the value-weighted NYSE/AMEX index, and σ is volatility of residuals from trailing 12-month market-model regression.

Interest Coverage: Annual *EBITDA* divided by annual interest expense (*XINT*).

Leverage: Book debt (*DLC* plus *DLTT*) divided by book assets (*AT*).

CAPM β and σ: β and σ are estimated from a trailing 24-month CAPM regression. We require that a firm has valid returns for at least 12 of the previous 24 months.

Size (*ME*): Size is market equity (*ME*) at the end of December in year *t*.

Age: Age is number of years since the first appearance of a firm (PERMCO) on CRSP.

Dividends (*Div*)**:** *Div* is a dummy variable equal to one for dividend payers (firms for which *DVPSXF*>0) and zero for non-payers.

C: Time-series Robustness Checks

C.1 Non-parametric HAC standard errors and small sample inference

Our results are robust to alternate choices for the Newey-West bandwidth parameter. In our baseline specifications we use a bandwidth of k years in the k-year return forecasting regressions. As shown by Andrews (1991), the bandwidth, m, must grow at a rate proportional to $T^{1/3}$ in order for HAC standard errors to be consistent. If the scores, $w_t = x_t e_t$, follow an AR(1) (i.e., $w_t = \phi w_{t-1} + \xi_t$), Newey and West (1994) have shown that the MSE-minimizing bandwidth choice for the Bartlett Kernel is $m^* = 1.8171 \times (\phi^2)^{1/3} \times T^{1/3}$. We estimate AR(1) coefficients of no greater than 0.2, suggesting an optimal bandwidth of approximately 2 for $T = 47$. However, we obtain highly significant t-statistics for larger choices of m. We also obtain similar standard errors in our k-year forecasting regressions if we use Hansen-Hodrick (1980) standard errors which are robust to serial correlation at up to k-1 lags.

It is well known that HAC estimators exhibit size distortions in finite samples. We address this concern in two ways. First, we compute p-values using the asymptotic theory from Kiefer and Vogelsang (2005) which has better finite sample properties than traditional asymptotic theory. Second, we compute bootstrapped p-values using a moving-blocks bootstrap.

The usual asymptotic theory for HAC inference is derived under the assumption that $m \to \infty$ and $m/T \to 0$. Kiefer and Vogelsang proceed under the assumption that $m = bT$ for some $b \in (0,1]$. That is, they assume that the bandwidth is a fixed fraction of the sample size. Letting $B(r)$ denote a standard Brownian motion and $\tilde{B}(r) = B(r) - rB(1)$, with the Bartlett kernel they show that $t_b \xrightarrow{d} B(1)/\sqrt{Q(b)}$ where $Q(b) = (2/b)\int_0^1 \tilde{B}(r)dr - (2/b)\int_0^{1-b} \tilde{B}(r + b)\tilde{B}(r)dr$ and $Q(b) \xrightarrow{p} 1$ as $b \to 0$, so that their "fixed b asymptotics" are equivalent to standard asymptotics in the limit. They simulate this distribution to obtain the relevant critical values.[1] In Table A.1, we use asterisks to denote coefficients that are significant at the 10%, 5%, and 1% level using their critical values.

Gonglaves and Vogelsang (2008) show that inference based on this "fixed b" approach is asymptotically equivalent to inference based on a moving-block bootstrap. However, for small sample sizes they argue that better approximations may be obtained via the block bootstrap with

[1] The critical value for a 2-sided test with 95% confidence is $cv(b) = 1.9600 + 2.9694 \cdot b + 0.4160 \cdot b^2 - 0.5324 \cdot b^3$.

a suitably chosen block length. Thus, we also use a block bootstrap to estimate the empirical distribution of our *t*-statistics. For the b^{th} iteration of the bootstrap we create a pseudo time series using a moving-blocks resampling technique as described below. We estimate our regression and compute a HAC standard errors using the pseudo time series, saving the resulting *t*-statistic. Finally, we compute bootstrapped *p*-values by comparing the actual *t*-statistic to the distribution of bootstrapped *t*-statistics.

To preserve the time-series dependence of the data, we create pseudo time series using the stationary block bootstrap of Politis and Romano (1994). Let $B_{t,k}^{S} = \{\mathbf{z}_t, \mathbf{z}_{t+1}, \ldots, \mathbf{z}_{t+k-1}\}$ be the block of length *k* starting from *t*. If $t+i>T$ for some $i \leq k-1$, we let $\mathbf{z}_{t+i} = \mathbf{z}_{t'}$ where $t' = \text{mod}\{t+i, T\}$. For instance, if *T* = 10 and *k* = 2, then $B_{10,2}^{S} = \{\mathbf{z}_{10}, \mathbf{z}_1\}$, so we "wrap the data around the circle". Letting $\{L_j\}$ be a sequence of *iid* draws from the geometric distribution with probability *q* and $\{I_j\}$ be a sequence of *iid* draws from the discrete uniform distribution on $\{1, 2, \ldots, T\}$, we create a pseudo time series by re-sampling blocks of *random* length as $\{B_{I_1,L_1}, B_{I_2,L_2}, \ldots\}$. This process is stopped once *T* observations have been selected.

These results are presented in Table A.1. We use 10,000 replications for each regression and a parameter of *q* = 1/8, so that the average block length is 8 years. Similar results obtain for other choices of *q*. While the *p*-values derived from the bootstrap-*t* procedure are larger (i.e., less significant) than those based on asymptotic theory, we find that our 1- and 2-year forecasting results using ISS^{EDF} are significant at the 1% level or better. Thus, *t*-statistics as large as those shown in Table A.1 are highly unlikely to obtain by chance.

C.2 Parametric HAC standard errors

Following the suggestion of Cochrane (2008) and Bates (2010), we also compute parametric HAC standard errors under the assumption that regression residuals follow an ARMA(*p*,*q*) process. As noted by these authors, if true 1-period *expected* returns follow an AR(1) process (i.e., $\mu_t = \tau \cdot \mu_{t-1} + \omega_t$) and realized returns are expected returns plus white noise (i.e., $r_{t+1} = \mu_t + \varepsilon_{t+1}$), then realized 1-period returns follow an ARMA(1,1) and realized *k*-period cumulative returns follow an ARMA(1,*k*). Thus, if we are interested in testing the null that $H_0^{tc} = \{\beta_k = 0, \mu_t \text{ time-varying}\}$, we must take the serial correlation of residuals into account.[2] In other words, once we abandon the null of zero return predictability, we should expect serial correlation even in a non-overlapping return forecasting regression under the null that a given predictor has no forecasting power. Furthermore, since the variance of unexpected returns is likely to be large relative to the time-variation in expected returns, traditional non-parametric estimators or parametric VARHAC estimators which assume the residuals follow an AR(*p*) may fail to capture this dependence.

Letting $w_t = x_t e_{t+1}$ denote the OLS scores, we fit an ARMA(*p*,*q*) model for the w_t via maximum likelihood, $w_t = \phi_1 w_{t-1} + \ldots + \phi_p w_{t-p} + \xi_t + \theta_1 \xi_{t-1} + \ldots + \theta_p \xi_{t-p}$. Our ARMA-HAC variance estimator is

$$\hat{V}_{ARMA(p,q)}^{HAC}(b) = \frac{\left(1+\hat{\theta}_1+\cdots+\hat{\theta}_q\right)^2\left(\sum_t \hat{\xi}_t^2\right)}{\left(1-\hat{\phi}_1-\cdots-\hat{\phi}_p\right)^2\left(\sum_t x_t^2\right)^2}. \tag{C1}$$

The more familiar VARHAC estimator which assumes that the scores follow an ARMA(*p*) process obtains as a special case in which one assumes $\hat{\theta}_k = 0$. To implement (A1) in our univariate specifications, we apply this procedure after first demeaning both the left- and right-hand side variables. In order to side-step the problem of multivariate ARMA (“VARMA”) estimation, we implement the correction in multivariate specifications by exploiting the Frisch-Waugh theorem (i.e., we regress returns and the predictor of interest on the controls and then apply the procedure to a univariate regression of orthogonalized returns on the orthogonalized predictor). As shown in Table A.1, this ARMA-HAC procedure yields *t*-statistics that are similar in magnitude to those based on Newey-West standard errors.

C.3 Stambaugh bias

We next consider the potential impact that the small-sample bias described in Stambaugh (1999), so called “Stambaugh bias”, may have on our results. Specifically, as noted by Baker,

[2] Under the classical null, $H_0^C = \{\beta_k = 0, \mu_t = \bar{\mu}\}$ the residuals are serially uncorrelated and this problem does not arise.

Wurgler, and Taliaferro (2006), one might worry that, since corporations tend to issue debt securities following high past excess returns, this might result in what Butler, Grullon, and Weston (2005) have called "aggregate-pseudo market timing". Specifically, it is not that low quality issuance negatively forecasts excess credit returns, it is simply that lower quality firms issue following high past returns and, due to the existence of Stambaugh bias, we might mistakenly conclude that it has negative forecasting power in small samples.

Formally, consider a univariate forecasting regression of the form

$$r_t = \alpha + \beta x_{t-1} + u_t \,, \tag{C2}$$

$$x_t = \theta + \rho x_{t-1} + v_t \,, \tag{C3}$$

where u_t and v_t are jointly normal. Stambaugh (1999) shows that

$$E[b-\beta] = \frac{\sigma_{u,v}}{\sigma_v^2} E[\hat{\rho}-\rho] = -\frac{\sigma_{u,v}}{\sigma_v^2}\frac{1+3\rho}{T} + O(T^{-2})\,. \tag{C4}$$

Therefore, if $\sigma_{u,v} > 0$, b will exhibit a downward bias in small samples. In our univariate forecasting regressions, we use the techniques in Amihud and Hurvich (2004) to compute bias-adjusted estimates of β and standard error for these bias-adjusted estimates. For multivariate regressions, we use the simulation procedure in Baker and Stein (2004) to compute bias-adjusted estimates and to compute *p*-values for our OLS estimates under the null that the coefficient is 0.[3]

As shown in Table A1, Stambaugh (1999) bias is not a significant concern. Neither ISS^{EDF} nor *HYS* is highly persistent, having first order auto-correlations of roughly 0.55, significantly lower the scaled price ratios familiar from the equity premium forecasting literature. Interestingly, Table A.1 shows that, while Stambaugh bias is negative at a 1-year forecasting horizon as one would expect, the bias shrinks and often changes sign in our longer horizon forecasting regressions. As explained by Bates (2010), this is because while innovations in our

[3] We perform two simulations for each regression, the first to generate a bias-adjusted estimate and the second to generate *p*-values under the null of no predictability. We first simulate the multivariate analogs of (C2) and (C3) recursively, using the OLS coefficient estimates and drawing with replacement from the empirical distribution of errors, *u* and **v**. We throw out the first 100 draws and draw *T* additional observations. We estimate (C2) on each simulated sample, giving us a set of coefficients $\mathbf{b}^*$. Our bias-adjusted estimate is $\mathbf{b}_{\text{adj}} = \mathbf{b} - (\overline{\mathbf{b}^*} - \mathbf{b})$- i.e., we adjust the OLS estimate by subtracting off the bootstrap bias estimate: the mean of $\mathbf{b}^*$ minus the OLS estimate. Next, we run separate simulation for each covariate, repeating the above steps but imposing the null that $\beta_k = 0$. This gives us a set of coefficients, b_k^{**}, which we use to compute the probability of observing a coefficient as large b_k when $\beta_k = 0$.

predictors are positively related to current realizations of unexpected returns, they are negatively related to innovations to future expected returns (i.e., with current "discount rate *news*"). When forecasting 1-year returns, only the positive correlation with unexpected returns comes into play, generating the expected downward bias. However, when forecasting longer-horizon returns the negative correlation with discount rate news generates an offsetting bias.

D: Additional Empirical Results

D.1Additional tests of the null that expected returns are non-negative

Figure A.1 shows the forecasting exercise discussed in Section V of the text. Each year we forecast *k*-period cumulative excess returns, compute the standard error of the fitted value, and count the number of years in which expected returns are negative with 95% confidence. Figure A.1 shows that ISS^{EDF} has forecast significantly negative 3-year cumulative excess returns in 14 years since 1962, and all but one of these years was actually followed by negative excess returns. ISS^{EDF} has also forecast significantly negative excess returns at a 1-year and 2-year horizon in 7 and 14 sample years, respectively.

We next report tests of the hypothesis that $a+b\cdot ISS_t^{EDF}=0$ at various sample quantiles of ISS^{EDF}. The tests are based on our estimates for 2-year cumulative excess high yield returns. We have t = –0.65 at the 50th percentile, t = –2.52 at the 75th percentile, t = –3.32 at the 90th percentile, and t = –4.46 at the sample maximum. In summary, our OLS estimates indicate the expected excess returns are significantly negative for values of ISS^{EDF} above the 70th percentile of sample values.

We next estimate nonlinear forecasting models which nest the null that expected excess returns are *always* non-negative, allowing us to further assess the statistical significance of the above findings. While a variety of theories predict that expected excess returns should be non-negative, they are silent on the exact function form so we experiment with a few different possibilities. For starters, we assume that $E[rx_{t+2}^{HY} \mid ISS_t^{EDF}]=\max\left\{a+b\cdot ISS_t^{EDF},c\right\}$. We can estimate this model via nonlinear least squares and test the hypothesis that c = 0. Since nonlinear least squares is a GMM estimator, this *t*-test (a Wald test) is asymptotically equivalent to a Lagrange multiplier test based on the restricted estimator that imposes $c = 0$, or a likelihood ratio test that compares the criterion functions evaluated at the constrained and unconstrained estimates (see Newey and McFadden 1994). When we estimate this specification, we obtain c = –6.8 and t = –3.71 using Newey-West standard errors. Alternately, we can assume that $E[rx_{t+2}^{HY} \mid ISS_t^{EDF}] = a + b \cdot ISS_t^{EDF} \times 1\{ISS_t^{EDF} < -a/b\} + c \cdot ISS_t^{EDF} \times 1\{ISS_t^{EDF} \geq -a/b\}$, so the regression function is piecewise linear with a kink at the point where $a + b \cdot ISS_t^{EDF} = 0$. Estimating this specification, we obtain b = –19.7 (t = –3.54) and c = –13.8 (t = –2.66). In summary, the data reject the hypothesis that expected excess returns are always non-negative.

D.2 Unpacking issuer quality

In this section we decompose the forecasting power of issuer quality. We first note that fluctuations in issuer quality can be due to either between- or within-firm variation in credit quality. The *between-firm* effect is obvious: during bad times, low quality firms may be unable to borrow or may find credit to be prohibitively expensive. The *within-firm* effect is more subtle: during booms individual firms may add enough leverage to diminish their own credit quality. For example, rapid debt-financed growth might significantly raise a firm's probability of default.

Since $EDF_{i,t}$ is impacted by leverage-increasing transactions during year t, ISS^{EDF} combines between- and within-firm effects. We can recalculate ISS^{EDF} using *pre-issuance EDF* (i.e., replacing $EDF_{i,t}$ with $EDF_{i,t-1}$ in equation (8)) to isolate the between-firm effect. The coefficient on $ISS^{Pre\text{-}EDF}$ in a univariate forecasting regression of 2-year high yield returns is b = -11.419 (t = -3.25) versus our baseline result of b = -15.254 (t = -5.29). Although the coefficients are similar, the R^2 drops from 26% to 15%. Thus, both between- and within-firm variation contribute to our findings. Going further, since $EDF_{i,t} = EDF_{i,t-1} + \Delta EDF_{i,t}$, we can decompose $ISS^{Post\text{-}EDF} = ISS^{Pre\text{-}EDF} + ISS^{\Delta EDF}$ and we find that both terms have independent forecasting power: the coefficients in a bivariate forecasting regression are $b_{Pre\text{-}EDF}$= -15.240 (t = -5.17) and $b_{\Delta EDF}$ = -15.361 (t = -3.34). This suggests that our results are partially driven by periods in which the creditworthiness of low quality borrowers is further eroded by rising debt burdens.

In a related exercise, we ask whether variation in issuer quality is driven by industry-level debt issuance waves or within-industry variation in issuer quality. To do so, we first introduce a regression-based approach to measure the quality of issuance. Each year we run a cross-sectional regression of debt issuance decile on *EDF* decile, $d_{i,t} = A_t + B_t \cdot EDF_{i,t} + v_{i,t}$. The slope coefficient B_t is high in years when high *EDF* firms are issuing relatively more debt than low *EDF* firms. We then use the time series of estimated coefficients to forecast future high yield returns. We would expect this procedure to yield nearly identical results to ISS^{EDF} and this is what we find: we obtain b = -0.994 (t = -5.26) in univariate and b = -0.886 (t = -4.58) in multivariate forecasting regressions for 2-year excess returns. This is hardly surprising since the resulting B_t series is 0.99 correlated with ISS^{EDF}.

It is simple to explore the impact of industry-level issuance waves using this methodology. Specifically, each year we estimate $d_{i,t} = A_{t,IND(i)} + B_t \cdot EDF_{i,t} + v_{i,t}$, including a full set of industry effects, so B_t is identified using only within-industry variation in debt issuance and *EDF*. In the second stage, we obtain we obtain b = -1.044 (t = -4.18) in univariate and b = -0.941 (t = -4.20) in multivariate forecasting regressions. Thus, the results remain quite strong if we restrict attention to within-industry variation, suggesting that our results are not primarily driven by industry-level debt issuance waves.

D.3Forecasting equity market and equity factor returns

While ISS^{EDF} is a reliable forecaster of excess credit returns, the Table A.2 shows that this variable has little ability to forecast stock market returns. However, we do find that ISS^{EDF} has some ability to negatively forecast the Fama and French (1993) *HML* and *SMB* factors. Nonetheless, as previously shown in Table 5, the coefficient and significance of ISS^{EDF} when forecasting high yield excess returns are largely unchanged even if we control for contemporaneous realizations of the Fama and French (1993) factors or the term premium. Table A.2 also shows that ISS^{EDF} is a reliable negative forecaster of the returns on distressed stocks (firms with high *EDF*s) relative to those on non-distressed stocks. Thus, while our results suggest that there is an important degree of segmentation between equity and credit markets, the stocks of distressed firms are, perhaps unsurprisingly, sensitive to credit market factors.[4]

D.4Subsample forecasting results for log(*HYS*)

Tables A.3 and A.4 present univariate and multivariate subsample forecasting results for log(*HYS*). Specifically, the results are shown separately for 1926-1943, 1944-1982, 1983-2007, 1944-2007, and the full 1926-2007 sample. The results are generally quite strong for the 1944-1982 and 1983-2007 subsamples as well as the combine 1944-2007 post-war period. The single exception is the 1926-1943 subsample which, as noted in the main text, is heavily influenced by the outlying 1933 observation. However, the results remain significant even when we splice all four series together and examine the full 1926-2008 sample.

D.5Quantity and quality of corporate bond offerings

Here we show that the findings that (i) issuer quality contains incremental information over and above the total quantity of issuance and (ii) that the quantity of low quality issuance is particularly useful for forecasting returns emerge using our corporate bond issuance data from 1944-2008. These results complement the findings in Table 4 where we used *EDF* to measure firm credit quality. Here we measure credit quality using Moody's credit ratings, adopting the traditional investment grade versus speculative grade classification.

Table A.5 shows these results using the growth in issuance. Specifically, we compute the growth in total issuance $b_t^{TOT} = \ln[(B_t^{TOT})/(\Sigma_{l=1}^{5} B_{t-l}^{TOTAL}/5)]$ as well as the growth in high yield and investment grade issuance, $b_t^{HY} = \ln[(B_t^{HY})/(\Sigma_{l=1}^{5} B_{t-l}^{HY}/5)]$ and $b_t^{IG} = \ln[(B_t^{IG})/(\Sigma_{l=1}^{5} B_{t-l}^{IG}/5)]$.[5] We run forecasting horseraces between (i) log(*HYS*) and b_t^{TOT} and (ii) b_t^{HY} and b_t^{IG}. These results are shown separately for 1926-1943, 1944-1982, 1983-2007, 1944-2007, and the full

[4] Another exercise suggesting equity and credit market segmentation compares the quality of high and low *equity* issuers using *EDF*. Each year we compute measures as in equation (8), but we now compare the credit quality of high and low net *equity* issuers. These measures of equity issuer credit quality are only modestly correlated with ISS^{EDF} and *HYS*. Equity issuer quality does not forecast excess credit returns.

[5] The level of nominal issuance is deflated using the CPI deflator so these represent real growth rates. Baker, Wurgler, and Taliaferro (2006) find that a similar variable based on equity issues forecasts market-wide stock returns.

1926-2007 sample. Panel A shows forecasting regressions without controls and Panel B adds our additional time-series controls. For instance, column (5) shows a forecasting horserace between log(*HYS*) and b_t^{TOT} for the 1944-1982 sample. We find that log(*HYS*) remains a strong forecaster of excess bond returns even controlling for the aggregate growth in total bond issuance. The regressions in columns (6) through (8) suggest that b_t^{HY} is a more reliable forecaster of excess corporate bond returns than b_t^{IG}. Columns (9) through (12) repeat this analysis for the 1983-2007 sample, the results for 1944-2007 are in columns (13) to (16), and the full sample results are in columns (17) to (20) The same broad patterns emerge the later 1983-2007 sample as well as the 1944-2007 and 1926-2007 samples.

Table A.6 presents a parallel analysis using measures of bond issuance scaled by *GDP*. For instance, column (5) shows that log(*HYS*) remains significant in the 1944-1982 sample even if we control for the aggregate level of bond issues $\ln(B_t^{TOT}/GDP_t)$. The results in columns (6), (7), and (8) suggest that $\ln(B_t^{HY}/GDP_t)$ is generally a more reliable forecaster of bond returns than $\ln(B_t^{IG}/GDP_t)$. Similar results obtain for 1944-1982, 1983-2007, and 1944-2007. However, when we include the early subsample which contains 1933, this generally strengthens the forecasting power of $\ln(B_t^{IG}/GDP_t)$. In summary, (i) the quality of bond issuance contains incremental information over and over aggregate issuance quantities and (ii) much of the forecasting power of aggregate issuance comes from issuance by low quality firms.

D.6The determinants of the high yield share

Table A.7 analyses the determinant of *HYS* as well as 1-year and 2-year changes in *HYS* over the 1944-2008 period. Thus, the analysis parallels Table 8 in the text which analyzed the determinants of ISS^{EDF}. Specifically, Table A.7 presents regressions of the form:

$$HYS_t = a + b \cdot y_{S,t}^G + c \cdot (y_{L,t}^G - y_{S,t}^G) + d \cdot rx_t^{HY} + e \cdot DEF_t^{HY} + u_t. \tag{D1}$$

We also run this regression in changes rather than levels:

$$\Delta_k HYS_t = a + b \cdot \Delta_k y_{S,t}^G + c \cdot \Delta_k (y_{L,t}^G - y_{S,t}^G) + d \cdot rx_{t-k\to t}^{HY} + e \cdot \Delta_k DEF_t^{HY} + \Delta_k u_t, \tag{D2}$$

where Δ_k denotes the *k*-year difference. We focus on the 1944-2008 subsample to minimize the effect of the outlying 1933 observation for *HYS*: including 1933 in the subsequent analysis meaningfully weakens the results.

Unfortunately, the levels regressions shown in columns (1) through (5) are less informative due to the structural break in *HYS* in the early 1980s. However, when we estimate these regressions in first or second differences the results for *HYS* generally parallel those from Table 8 for ISS^{EDF}. Specifically, Table A.6 indicates that *HYS* tends to rise when (i) the short-term interest rates or the term spread decline or (ii) when high yield defaults fall or the excess returns on low-grade bonds are high.

E: An Extrapolative Model of Credit Cycles

Here we show that a simple model with extrapolative beliefs can match several of the stylized facts documented in the paper, namely, that (1) excess credit returns are mean-reverting and expected returns can be negative, (2) issuer quality predicts returns even after controlling for credit spreads, and (3) past default rates lead to changes in issuer quality. Our objective is to provide a simple account of the credit cycle in which extrapolation plays a role.

The model is based on the idea that investors are extrapolative rather than forward looking, and form their expectations of future defaults by looking at recent default patterns. As such, the idea is inspired by the accounts of Hickman (1958) and Grant (1992, 2008) who emphasize the "perils of tranquility" in which investors come to believe that good times will persist indefinitely during booms. To formalize these intuitions, we borrow from Barberis, Shleifer, and Vishny (1998, hereafter "BSV") who model equity market investors who are capable of either under-reacting to or over-extrapolating patterns in firm earnings. We adapt BSV's modeling approach to explain aggregate credit market dynamics.

In the model, investors form their expectations of future defaults by extrapolating recent default patterns. Specifically, the economy evolves according to a simple Markov process, switching between good times in which few firms default and bad times in which a higher fraction of firms default. However, investors think that the economy either evolves according to a more or less persistent process. After a series of consecutive good states, investors begin to believe that the process governing aggregate defaults is more persistent than it truly is, causing them to under-estimate future default probabilities.[6] These expectations will be revised after a period of high corporate defaults, resulting in a sharp decline in bond prices. And if these bad times persist for long enough, investors will begin to over-estimate future default probabilities. As in BSV, the model generates short-term return continuation and longer-term return reversals in corporate bond returns.

We then introduce a set of issuing firms into the model, allowing us to link the quality of corporate debt issuance to future bond returns. The mechanism is as follows: low quality firms respond to narrow spreads by issuing more debt during booms, raising their leverage and default probabilities. Investors understand that leverage impacts default probabilities. However, investors' growing belief that good times are likely to persist leads them to underestimate the impact of rising leverage on long-run default probabilities. Following a string of low aggregate defaults, investors become willing to lend to more highly levered firms for a given spread. Because spreads mean different things at different times, both spreads and issuer quality are useful for forecasting returns in the model. Specifically, controlling for the level of spreads, a lower level of issuer quality is associated with greater over-optimism about future default rates and, hence, lower expected returns. We provide a numerical illustration of the model in which investor biases are modest, but where there is meaningful mispricing nonetheless.

E.1 Defaultable perpetuities and aggregate default dynamics

There is single class of risk neutral investors with discount rate r who purchase defaultable perpetuities. Perpetuities pay a coupon of c each period prior to default and recover $(1-\ell)$ upon default where $\ell \in (0,1]$ is the loss-given-default. As a simple benchmark, first suppose the default probability is constant over time and equal to π. In this case, the price of the perpetuity must be the same at any date prior to default and satisfies the present value relation

$$P = [(1-\pi)(c+P) + \pi(1-\ell)]/(1+r). \tag{E1}$$

Solving (E1) for P, we obtain

[6] As noted by, Caballero and Krishnamurthy (2008) and Gennaioli, Shleifer, and Vishny (2010), it is plausible to think that this tendency might be most pronounced for newer and less familiar credit market instruments. Thus, it is not surprising that many credit market booms have featured a different set of instruments than prior booms.

$$P=(1+r)^{-1}\sum_{j=0}^{\infty}\left((1+r)^{-1}(1-\pi)\right)^{j}[(1-\pi)c+\pi(1-\ell)]=(r+\pi)^{-1}[(1-\pi)c+\pi(1-\ell)], \tag{E2}$$

which is increasing in c and decreasing in π, ℓ, and r.[7]

We introduce aggregate default dynamics in a simple way. We suppose that there are two possible macro states S_t: a high-default state H, and low-default state L. If the economy is in the high-default state at time t, the default probability is $\pi_t=\pi_H$; otherwise, $\pi_t=\pi_L$ if $S_t=L$, where $\pi_L<\pi_H$. The economy switches between the H and L default states according to a Markov chain with transition matrix $\mathbf{T}_0$, where

$$\mathbf{T}_0=\begin{array}{c c} & \begin{array}{cc} S_{t+1}=H & S_{t+1}=L \end{array} \\ \begin{array}{c} S_t=H \\ S_t=L \end{array} & \begin{bmatrix} 1-\theta & \theta \\ \gamma & 1-\gamma \end{bmatrix} \end{array}. \tag{E3}$$

We assume that $\theta+\gamma<1$ and $\theta>\gamma$, implying that the economy is in the low default state most of the time (i.e., $\Pr(S_t=L)=\theta/(\gamma+\theta)>1/2$).

E.2 Rational prices

First consider how rational prices would evolve in this setting. (For simplicity, we carry out the analysis in terms of prices assuming a fixed coupon of c. Of course, prices can be mapped to spreads, s, using the convention that $s = c/P - r$.) Rational prices in the two states satisfy:

$$P_H=(1+r)^{-1}\left\{(1-\theta)[(1-\pi_H)(c+P_H)+\pi_H(1-\ell)]+\theta[(1-\pi_L)(c+P_L)+\pi_L(1-\ell)]\right\} \tag{E4}$$

$$P_L=(1+r)^{-1}\left\{\gamma[(1-\pi_H)(c+P_H)+\pi_H(1-\ell)]+(1-\gamma)[(1-\pi_L)(c+P_L)+\pi_L(1-\ell)]\right\}.$$

The solution to (E4) is given by

$$\mathbf{p_0}=\left[(1+r)\mathbf{I}_2-\mathbf{T}_0\mathrm{diag}(\mathbf{1}-\boldsymbol{\pi}_\mathbf{0})\right]^{-1}\mathbf{T}_0\left[(\mathbf{1}-\boldsymbol{\pi}_\mathbf{0})c+\boldsymbol{\pi}_\mathbf{0}\left(1-\ell\right)\right]. \tag{E5}$$

where $\mathbf{p_0}=\left[P_H \quad P_L\right]'$, $\boldsymbol{\pi}_\mathbf{0}=\left[\pi_H \quad \pi_L\right]'$, and $\mathbf{I}_2$ is the 2×2 identity matrix. P_H and P_L are the only two possible prices, and the price of non-defaulted claims only changes when the economy transitions between states. Since required returns are constant, differences between P_H and P_L arise solely from time-variation in conditional default probabilities and the expected timing of

[7] Technically, we assume $(1-\ell)<c+(c/r)$ (i.e., investors lose money when the firm defaults) to ensure that $\partial P/\partial\pi<0$.

defaults.[8] The set-up can be seen as a simplified version of models such as Duffie and Singleton (1999) which emphasize time-varying default arrival rates.[9]

E.3 Investor beliefs and equilibrium prices

While the true macro state is generated by (E3), investors incorrectly believe that the macro state is either generated by a less persistent regime (Regime 1) or a more persistent regime (Regime 2). In Regime 1, the perceived transition matrix for the macro state is $\mathbf{T}_1$; in Regime 2, is it $\mathbf{T}_2$. We assume

$$\mathbf{T}_1 = \begin{array}{c} \\ S_t = H \\ S_t = L \end{array} \begin{array}{c} \begin{array}{cc} S_{t+1}{=}H & S_{t+1}{=}L \end{array} \\ \begin{bmatrix} 1-\theta(1+\varepsilon) & \theta(1+\varepsilon) \\ \gamma(1+\varepsilon) & 1-\gamma(1+\varepsilon) \end{bmatrix} \end{array}, \quad \mathbf{T}_2 = \begin{array}{c} \\ S_t = H \\ S_t = L \end{array} \begin{array}{c} \begin{array}{cc} S_{t+1}{=}H & S_{t+1}{=}L \end{array} \\ \begin{bmatrix} 1-\theta(1-\varepsilon) & \theta(1-\varepsilon) \\ \gamma(1-\varepsilon) & 1-\gamma(1-\varepsilon) \end{bmatrix} \end{array}, \tag{E6}$$

where $\varepsilon \in [0,1)$, and $(1+\varepsilon)(\theta+\gamma)<1$. Higher values of ε indicate a greater scope for biased beliefs about transition dynamics. Specifically, in Regime 1, investors think the macro state is less persistent that it truly is; in Regime 2, investors think the state is more persistent than it really is. Finally, investors believe that the economy switches between Regime 1 and Regime 2 according to the following Markov process

$$\mathbf{\Lambda} = \begin{array}{c} \\ R_t = 1 \\ R_t = 2 \end{array} \begin{array}{c} \begin{array}{cc} R_{t+1}{=}1 & R_{t+1}{=}2 \end{array} \\ \begin{bmatrix} 1-\lambda_1 & \lambda_1 \\ \lambda_2 & 1-\lambda_2 \end{bmatrix} \end{array}, \tag{E7}$$

where $\lambda_1+\lambda_2<1$ and the regime is assumed to be independent of the macro state.

Investors observe the macro state S_t and form conjectures about the current regime R_t. Their estimate of the probability of being in Regime 1, denoted by $q_t \equiv \Pr[R_t = 1|\mathbf{S}^t]$, is based on the history of macro states observed up to and including t. While investors entertain two incorrect regimes that bracket reality, they update their belief about the current regime in a Bayesian fashion. Thus, the law of motion for q_t is

$$q_{t+1} = \frac{[(1\text{-}\lambda_1)q_t + \lambda_2(1\text{-}q_t)]\cdot \Pr[S_{t+1} \mid S_t, R_{t+1}=1]}{[(1\text{-}\lambda_1)q_t + \lambda_2(1\text{-}q_t)]\cdot \Pr[S_{t+1} \mid S_t, R_{t+1}=1] + [\lambda_1 q_t + (1\text{-}\lambda_2)(1\text{-}q_t)]\cdot \Pr[S_{t+1} \mid S_t, R_{t+1}=2]}. \tag{E8}$$

[8] We have $P_L > P_H$ so long as ℓ is not so small that investors are made better off by default. For instance, if $\ell = 1$, we have $P_L - P_H \propto c(\pi_H - \pi_L)[r(1-\theta-\gamma) + (1-\theta\pi_L - \gamma\pi_H)] > 0$.

[9] For instance, following Duffie and Singleton (1999), these models often decompose credit spreads as $s_t = \lambda_{Q,t}L_{Q,t} + l_t$ where $\lambda_{Q,t}$ is the risk-neutral default arrival rate at time t, $L_{Q,t}$ is the risk-neutral loss given default, and l_t is an illiquidity premium.

The estimated probability of being in the less persistent Regime 1, q, rises when S_{t+1} differs from S_t and falls when S_{t+1} is the same as S_t.

As discussed by BSV, this set-up can be seen as reflecting two psychological findings about biases in human inference. First, there is evidence of "conservatism" in which subjects underweight new evidence that conflicts with existing beliefs. Second, subjects often use a "representativeness" heuristic and act as if they believe in a "law of small numbers" leading to over-extrapolation at intermediate horizons. Gennaioli and Shleifer (2010) present a model in which agents have limited recall and represent hypotheses using a subset of representative scenarios. As explored in Gennaioli, Shleifer, and Vishny (2010), variation in the set of scenarios that come to mind can lead to mispricing as agents alternately under- and over-estimate the probability of rare bad states. However, one does not need to be swayed by the psychology literature to find our assumptions reasonable. An equally plausible institutional interpretation is that intermediaries use backwards-looking risk management systems such as Value-at-Risk when extending credit over the cycle, leading to under-reaction at short horizons and over-reaction at longer horizons.

As shown in Section *E.6* below, prices take the form

$$P(S_t, q_t) = (\mathbf{q}_t^{S_t})'\mathbf{p}_1, \tag{E9}$$

where $\mathbf{q}_t^H = \begin{bmatrix} q_t & 0 & 1-q_t & 0 \end{bmatrix}'$, $\mathbf{q}_t^L = \begin{bmatrix} 0 & q_t & 0 & 1-q_t \end{bmatrix}'$, $\mathbf{p}_1 = \begin{bmatrix} P_{H,1} & P_{L,1} & P_{H,2} & P_{L,2} \end{bmatrix}'$ is

$$\mathbf{p}_1 = \left[(1+r)\mathbf{I}_4 - \mathbf{T}\mathrm{diag}(\mathbf{1-\pi}_1)\right]^{-1}\mathbf{T}\left[(\mathbf{1-\pi}_1)c + \boldsymbol{\pi}_1\left(1-\ell\right)\right]. \tag{E10}$$

In (E10), $\boldsymbol{\pi}_1 = \begin{bmatrix} \pi_H & \pi_L & \pi_H & \pi_L \end{bmatrix}'$, $\mathbf{I}_4$ is the 4×4 identity matrix, and $\mathbf{T}$ is the 4×4 transition matrix for the combined state given by

$$\mathbf{T} = \begin{bmatrix} (1-\lambda_1)(1-\theta(1+\varepsilon)) & (1-\lambda_1)\theta(1+\varepsilon) & \lambda_1(1-\theta(1-\varepsilon)) & \lambda_1\theta(1-\varepsilon) \\ (1-\lambda_1)\gamma(1+\varepsilon) & (1-\lambda_1)(1-\gamma(1+\varepsilon)) & \lambda_1\gamma(1-\varepsilon) & \lambda_1(1-\gamma(1-\varepsilon)) \\ \lambda_2(1-\theta(1+\varepsilon)) & \lambda_2\theta(1+\varepsilon) & (1-\lambda_2)(1-\theta(1-\varepsilon)) & (1-\lambda_2)\theta(1-\varepsilon) \\ \lambda_2\gamma(1+\varepsilon) & \lambda_2(1-\gamma(1+\varepsilon)) & (1-\lambda_2)\gamma(1-\varepsilon) & (1-\lambda_2)(1-\gamma(1-\varepsilon)) \end{bmatrix}. \tag{E11}$$

For instance, $\Pr(S_{t+1} = H, R_{t+1} = 2 \mid S_t = H, R_t = 1) = \lambda_1(1-\theta(1-\varepsilon))$.

For example, in the high default state, the price is

$$P(H,q_t) = q_t P_{H,1} + (1-q_t)P_{H,2}, \tag{E12}$$

a weighted average of high-default probability prices in Regimes 1 and 2. Under conditions given in Section *E.6*, $P_{H,2} < P_{H,1} < P_{L,1} < P_{L,2}$ so we can interpret q_t as a conditional measure of investor sentiment. Specifically, in the low default state, a *low* value of q_t means that investors are overly optimistic that good times will last. Conversely, in the high default state, a *high* value of q_t means that investors are overly optimistic that good times will return.

While investors' required returns are constant and given by r, expected returns as perceived by an unbiased outside observer are not constant. Prices under-react to an initial transition from the low-default to the high-default state, and vice versa, but over-react to sustained spells in the low or high state.[10] For instance, a sustained spell in the low default state causes investors to underestimate the long-run probability of default and over-value risky debt. Conversely, sustained spells in the high state lead to over-estimation of long-run default probabilities and under-valuation of debt. Thus, excess bond returns exhibit short-term continuation and longer-term reversals.

E.4 Adding a corporate sector

To understand why issuer quality may be informative, we could now just invoke the reduced form model developed in Section 2.2 of the paper. Specifically, the debt financing costs faced by high default risk firms would be more exposed to movements in conditional investor sentiment, q_t, so issuer quality would be useful for forecasting bond returns. However, to simplify the analysis we just assume that issuer quality follows a simple law of motion. This analysis also shows that issuer quality might have significant forecasting power even if corporate managers are not particularly "smart" and issue according to a simple rule-of-thumb.

Each period t, we assume a new cohort of bonds is issued with quality D_t. D is an unconditional scaling of default probabilities: doubling D doubles the probability of default in both the high and low states (i.e. for an issuer with quality D, default probabilities in the high and low state are $D \cdot \pi_H$ and $D \cdot \pi_L$). Thus, it is convenient to think of D as reflecting the leverage of

[10] To deliver under-reaction we need to assume $\lambda_2 > \lambda_1$, so the perceived probability of being in Regime 1 exceeds ½.

the average issuing firm in each cohort.

We assume issuer quality evolves according to

$$D_t = \max\left\{D_{\min}, \min\left\{D_{\max}, D_{t-1} + b \cdot (P_{t-1} - \bar{P})\right\}\right\}, \tag{E13}$$

where $b > 0$, $0 < D_{\min} < D_{\max}$, and $\bar{P}$ is the long-run average price. Equation (E13) captures the idea that low quality firms borrow more when spreads are tight, raising their leverage and future default probabilities. Because (E13) can be seen as a description of leverage for a *representative* low quality firm, it best captures the "within-firm" quality dynamics discussed in Appendix D. However, we can interpret (E13) as description the quality of the *average issuing* firm, in which case (E13) would reflect both within- and between-firm effects. Although tighter spreads are assumed to lead to lower quality issuance, nothing in (E13) requires firm manager to be particularly "smart." For instance, equation (E13) could arise through a simple cost of capital channel in which tighter spreads induce low quality firms to borrow and invest more.

The price of each new cohort of bonds is given by

$$P(S_t, q_t, D_t) = (\mathbf{q}_t^{S_t})' \left[(1+r)\mathbf{I}_4 - \mathbf{T}\text{diag}(\mathbf{1} - D_t \boldsymbol{\pi}_1)\right]^{-1} \mathbf{T}\left[(\mathbf{1} - D_t \boldsymbol{\pi}_1)c + D_t \boldsymbol{\pi}_1 \left(1 - \ell\right)\right]. \tag{E14}$$

Investors understand that higher leverage leads to higher default probabilities and factor this into their valuations. But spreads can either rise or fall during a spell in the low default state: spreads fall if investors' growing belief in a low-default paradigm outweighs the rise in firm leverage.

The realized return on a large portfolio of bonds with quality D_t from t to $t+1$ is given by

$$r_{t+1} = [(1 - D_t \pi_{S_{t+1}})(c + P(S_{t+1}, q_{t+1}, D_t)) + D_t \pi_{S_{t+1}}(1 - \ell)] / P(S_t, q_t, D_t) - 1. \tag{E15}$$

We are interested in how expected returns, $E[r_{t+1} \mid S_t, P_t, D_t]$, vary according to the issuer quality, D_t. Because D_t varies over time, knowledge of P_t and S_t does not fully reveal conditional investor sentiment, q_t, the state variable that drives expected returns. Practically speaking, this means that issuer quality contains additional information about future returns that is not contained in credit spreads. For instance, if the economy is in the low default state, prices can either be high because leverage is low and future default probabilities are low, or because investors are overly optimistic about future defaults (i.e., q_t is low). More formally, it can be shown that D_t is negatively related to future returns even after controlling for the level or prices or spreads, so that

$\partial E[r_{t+1} \mid S_t, P_t, D_t] / \partial D_t < 0$ and $\partial E[r_{t+1} \mid S_t, P_t, D_t] / \partial P_t < 0.$ Intuitively, credit spreads under-react to the deterioration in issuer quality (i.e. increase in leverage) during credit booms, so issuer quality itself becomes useful for forecasting returns.

E.5 Numerical example

We simulate the model using the issuer quality rule in (E13) and generate returns using (E15). Consistent with the historical behavior of high yield default rates, we assume $\pi_L = 1\%$, $\pi_H = 10\%$, $\gamma = 5\%$, and $\theta = 20\%$, so 80% of the time is spent in the low-default state. We assume r = 0%, c = 1.25%, ℓ = 50%, and b = 0.2: leverage is increasing in past prices and remains bounded between $D_{\min}$ = 0.5 and $D_{\max}$ = 2. Finally, we assume ε = 75%, λ_1 = 0.5%, and λ_2 = 1%. Investors believe that regime changes are rare; if they believe that regime changes are frequent, q_t does not fluctuate much over time. While the differences between Regime 1 and 2 are non-trivial, investor biases are still fairly modest in the simulation. For instance, the perceived probability of transitioning from the L to H state ranges from 1.25% to 8.75%, so biases never exceeds 3.75%. Under these assumptions, $P_{H,1}$ = 93.52%, $P_{L,1}$ = 98.96%, $P_{H,2}$ = 80.50%, and $P_{L,2}$ = 113.83% when D = 1. Table A.8 lists expected returns for different combinations of lagged prices and lagged issuer quality separately in both the low-default and high-default macro states. The table shows that expected one-period returns are monotonically decreasing in both lagged prices and issuer quality. Although investors in the example under-estimate low probability events in good times, the mistakes they make are not unreasonable. Modest biases in assessing low probability events can generate meaningful mispricing which varies over time and can even generate negative conditional expected returns.

Table A.8 also shows that the variation in conditional expected returns is largest in the high default state. Expected returns are lowest when the economy first enters the high default state at the end of a long debt boom (i.e., after a spell of low default realizations). At these times, leverage is elevated due to the long boom and prices can still be high. Investors are often overly-optimistic that the low default state will return, leading them to under-react to the initial bad news.[11] These expectations will often be disappointed, resulting in large negative returns. Conversely, the highest expected returns occur following a long spell of high default realizations, which is marked by declining issuer quality and depressed prices.

E.6 Model proofs

Suppose the investor knows the regime R_t and let $P_{S,R}$ denote the price in state (S,R). For instance, $P_{H,1}$, the price in the high default state in the less persistent regime 1, satisfies

$$\begin{aligned} P_{H,1} &= (1+r)^{-1}\left\{(1-\lambda_1)(1-\theta(1+\varepsilon))[(1-\pi_H)(c+P_{H,1})+\pi_H(1-\ell)]+(1-\lambda_1)\theta(1+\varepsilon)[(1-\pi_L)(c+P_{L,1})+\pi_L(1-\ell)]\right\} \\ &+(1+r)^{-1}\left\{\lambda_1(1-\theta(1-\varepsilon))[(1-\pi_H)(c+P_{H,2})+\pi_H(1-\ell)]+\lambda_1\theta(1-\varepsilon)[(1-\pi_L)(c+P_{L,2})+\pi_L(1-\ell)]\right\} \end{aligned}$$

Letting $\mathbf{p}_1 = \begin{bmatrix} P_{H,1} & P_{L,1} & P_{H,2} & P_{L,2}\end{bmatrix}'$ and $\boldsymbol{\pi}_1 = [\pi_H \quad \pi_L \quad \pi_H \quad \pi_L]'$, the system can be written as

$$\mathbf{p}_1 = (1+r)^{-1}\mathbf{T}[\mathrm{diag}(\mathbf{1}-\boldsymbol{\pi}_1)(\mathrm{c}\mathbf{1}+\mathbf{p}_1)+\boldsymbol{\pi}_1(1-\ell)], \tag{E16}$$

where $\mathbf{T}$ is given by (E11). Solving we have

$$\mathbf{p}_1 = \left[(1+r)\mathbf{I}_4 - \mathbf{T}\mathrm{diag}(\mathbf{1}-\boldsymbol{\pi}_1)\right]^{-1}\mathbf{T}\left[(\mathbf{1}-\boldsymbol{\pi}_1)c+\boldsymbol{\pi}_1(1-\ell)\right], \tag{E17}$$

where $\mathbf{I}_4$ is a 4×4 identity matrix. It follows that

$$P(S_t, q_t) = (\mathbf{q}_t^{S_t})'\mathbf{p}_1, \tag{E18}$$

where $\mathbf{q}_t^H = [q_t \quad 0 \quad 1-q_t \quad 0]'$ and $\mathbf{q}_t^L = [0 \quad q_t \quad 0 \quad 1-q_t]'$ – e.g., if $S_t = H$, the price is $P(H,q_t) = q_tP_{H,1} + (1\text{-}q_t)P_{H,2}$, a weighted average of high-default probability prices in regimes 1 and 2.

Letting $\tilde{\tau} > t$ denote the random default time and noting that $E[(S_{t+j}, R_{t+j}) \mid \mathbf{S}^t] = (\mathbf{q}_t^{S_t})'\mathbf{T}^j$, an alternate derivation of prices follows from

$$\begin{aligned} P(S_t,q_t) &= E[\Sigma_{j=1}^{\infty}(1+r)^{-j}(c\cdot 1\{t+j<\tilde{\tau}\}+(1-\ell)\cdot 1\{t+j=\tilde{\tau}\} \mid S_t,q_t] \\ &= (1+r)^{-1}(\mathbf{q}_t^{S_t})'[\Sigma_{j=0}^{\infty}[(1+r)^{-1}\mathbf{T}\mathrm{diag}(\mathbf{1}-\boldsymbol{\pi}_1)]^j]\mathbf{T}\left[(\mathbf{1}-\boldsymbol{\pi}_1)c+\boldsymbol{\pi}_1(1-\ell)\right] \\ &= (\mathbf{q}_t^{S_t})'\left[(1+r)\mathbf{I}-\mathbf{T}\mathrm{diag}(\mathbf{1}-\boldsymbol{\pi}_1)\right]^{-1}\mathbf{T}\left[(\mathbf{1}-\boldsymbol{\pi}_1)c+\boldsymbol{\pi}_1(1-\ell)\right]. \end{aligned} \tag{E19}$$

So long as (i) $\pi_L < \pi_H$ (ii) $\lambda_1 + \lambda_2 < 1$, (iii) $\varepsilon \in [0,1)$, (iv) and $(1+\varepsilon)(\theta+\gamma) < 1$, it is tedious but straightforward to show that $P_{H,2} < P_{H,1} < P_{L,1} < P_{L,2}$, implying that $\partial P(H,q_t)/\partial q_t > 0$ and $\partial P(L,q_t)/\partial q_t < 0.$ Thus, q_t is a measure of conditional investor sentiment, noting that in the L

[11] At the end of a long spell of low default realizations q_t will be low as investors come to believe that the macro state is quite persistent. As a result, q_t will jump following a unexpected transition to the H state and, believing that they are now in Regime 1, investors will overestimate the probability of a return to good times – i.e. they overestimate the likelihood of a "soft landing."

state a low value of q_t is associated with positive sentiment while in the H state a high value of q_t is associated with positive sentiment.

When we allow for time-varying leverage, prices take the form

$$P(S_t, q_t) = (\mathbf{q}_t^{S_t})'\mathbf{p_1}(D_t) = (\mathbf{q}_t^{S_t})'\left[(1+r)\mathbf{I} - \mathbf{T}\text{diag}(\mathbf{1} - D_t\boldsymbol{\pi}_1)\right]^{-1}\mathbf{T}\left[(\mathbf{1} - D_t\boldsymbol{\pi}_1)c + D_t\boldsymbol{\pi}_1(1-\ell)\right]. \quad \text{(E20)}$$

Using the rules for vector and matrix differentiation (see e.g. Lax 1997), we have

$$\begin{aligned}\frac{\partial \mathbf{p_1}(D_t)}{\partial D_t} = &-\left[(1+r)\mathbf{I} - \mathbf{T}\text{diag}(\mathbf{1} - D_t\boldsymbol{\pi}_1)\right]^{-1}\mathbf{T}\text{diag}(\boldsymbol{\pi}_1)\left[(1+r)\mathbf{I} - \mathbf{T}\text{diag}(\mathbf{1} - D_t\boldsymbol{\pi}_1)\right]^{-1}\mathbf{T}\left[(\mathbf{1} - D_t\boldsymbol{\pi}_1)c + D_t\boldsymbol{\pi}_1(1-\ell)\right] \\ &-\left[(1+r)\mathbf{I} - \mathbf{T}\text{diag}(\mathbf{1} - \boldsymbol{\pi}_1)\right]^{-1}\mathbf{T}\boldsymbol{\pi}_1\left(c - (1-\ell)\right) < \mathbf{0},\end{aligned}$$

so long as ℓ is not too small (e.g. when $\ell = 1$ the inequality is immediate since all elements of $\left[(1+r)\mathbf{I} - \mathbf{T}\text{diag}(\mathbf{1} - \boldsymbol{\pi}_1)\right]^{-1}$ are non-negative). Put simply, prices are decreasing in leverage (or default probabilities) holding fixed the coupon and loss-given-default.

Now suppose $S_t = L$, so that $P_t = q_t P_{L,1}(D_t) + (1-q_t)P_{L,2}(D_t)$. Consider the experiment of varying P_t while holding fixed D_t and vice versa. Since $P_{L,1}(D_t) - P_{L,2}(D_t) < 0$, it follows that

$$\frac{\partial q_t}{\partial P_t} = \left(P_{L,1}(D_t) - P_{L,2}(D_t)\right)^{-1} < 0 \text{ and } \frac{\partial q_t}{\partial D_t} = -\frac{q_t(\partial P_{L,1}(D_t)/\partial D_t) + (1-q_t)(\partial P_{L,2}(D_t)/\partial D_t)}{P_{L,1}(D_t) - P_{L,2}(D_t)} < 0.$$

Since low values of q_t are associated with more favorable sentiment when $S_t = L$, it follows that expected returns are decreasing in both P_t and D_t. Next suppose that $S_t = H$, so $P_t = q_t P_{H,1}(D_t) + (1-q_t)P_{H,2}(D_t)$. Since $P_{H,1}(D_t) - P_{H,2}(D_t) > 0$, we have

$$\frac{\partial q_t}{\partial P_t} = \left(P_{H,1}(D_t) - P_{H,2}(D_t)\right)^{-1} > 0 \text{ and } \frac{\partial q_t}{\partial D_t} = -\frac{q_t(\partial P_{H,1}(D_t)/\partial D_t) + (1-q_t)(\partial P_{H,2}(D_t)/\partial D_t)}{P_{H,1}(D_t) - P_{H,2}(D_t)} > 0.$$

Since high values of q_t are associated with more favorable sentiment when $S_t = H$, it follows that expected returns are decreasing in both P_t and D_t. Thus, we have shown that expected returns are decreasing in both P_t and D_t in both default states.

References

Amihud, Yakov, and Clifford M. Hurvich, 2004, Predictive Regressions: A Reduced-bias Estimation Method, *Journal of Financial and Quantitative Analysis* 39, 813-841.

Andrews, Donald W K, 1991. Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation, *Econometrica* 59, 817-858.

Baker, Malcolm and Jeremy Stein, 2004, Market Liquidity as a Sentiment Indicator, *Journal of Financial* Economics 7, 271-299.

Baker, Malcolm, Ryan Taliaferro, and Jeffrey Wurgler, 2006, Predicting Returns with Managerial Decision Variables: Is there a small-sample bias?, *Journal of Finance* 61, 1711-1730.

Baker, Malcolm, and Jeffrey Wurgler, 2006, Investor Sentiment and the Cross-Section of Stock Returns, *Journal of Finance* 61, 1645-1680.

Bates, Brandon, 2010, A Predictability Predicament, Harvard University Working Paper.

Butler, Alexander, Gustavo Grullon and James Weston, 2005, Can Managers Forecast Aggregate Market Returns?, *Journal of Finance* 60, 963-986.

Caballero, Ricardo J. and Arvind Krishnamurthy, 2008, Collective Risk Management in a Flight to Quality Episode, *Journal of Finance* 63, 2195-2230.

Cochrane, John, 2008, Comments on “Bond Supply and Excess Bond Returns by Robin Greenwood and Dimitri Vayanos.

Gennaioli, Nicola and Andrei Shleifer, 2010, What comes to mind? *Quarterly Journal of Economics*, forthcoming.

Gennaioli, Nicola, Andrei Shleifer, and Robert Vishny, 2010, Neglected Risks, Financial Innovation, and Financial Fragility, Harvard University Working Paper.

Goncalves, Silvia and Timothy J. Vogelsang, 2008, Block Bootstrap HAC Robust Tests: The sophistication of the naive bootstrap, Working Paper.

Greene, William H., 2003, Econometric Analysis (5th edition), Prentice Hall, Upper Saddle River, NJ

Hansen, Lars P., and Robert J. Hodrick, 1980, Forward Exchange-Rates As Optimal Predictors of Future Spot Rates - An Econometric-Analysis, *Journal of Political Economy* 88: 829-853.

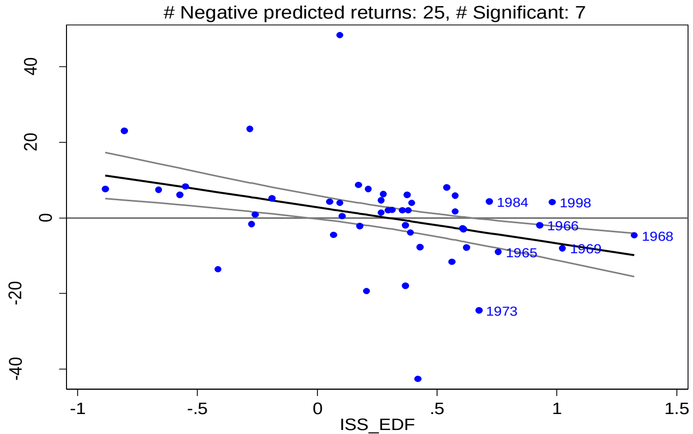
Lax, Peter D., 1997, Linear Algebra, John Wiley & Sons, New York.

Newey, Whitney K. and Kenneth D. West, 1994, Lag Selection in Covariance Matrix Estimation, *Review of Economic Studies* 61, 631-653.

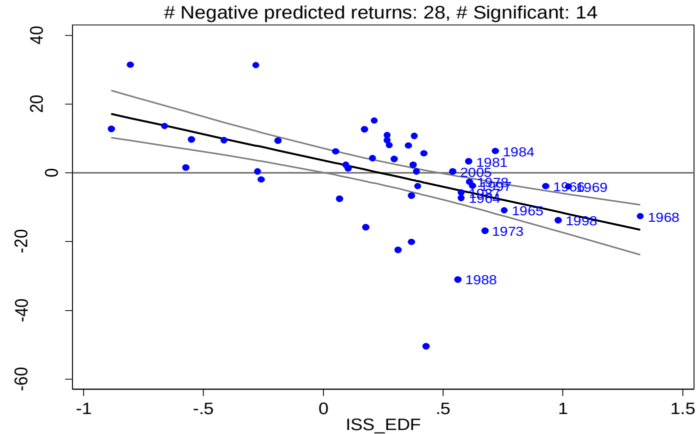
Newey, Whitney K., and Daniel McFadden, 1994, Large Sample Estimation and Hypothesis Testing, Handbook of Econometrics Vol. 4, McFadden and Engle, (eds), Elsevier, North Holland, 2111-2245.

Politis, Dimitris N. and Joseph P. Romano, 1994, The Stationary Bootstrap, *Journal of the American Statistical Association* 89, 1303-1313.

Panel A: 1-year expected excess returns



Panel B: 2-year expected excess returns



Panel C: 3-year expected excess returns

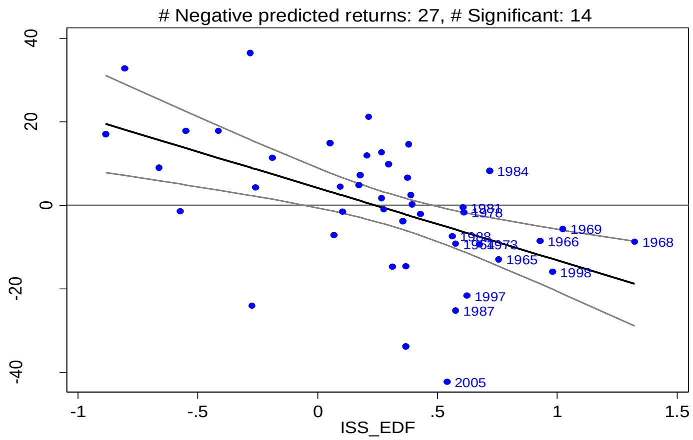


Figure A.1:Univariate Forecasts of High Yield Excess Returns. These figures plot ISS^{EDF} (horizontal axis) against cumulative 1-, 2- and 3-year future high yield excess returns (vertical axis). The darker solid line is the univariate forecast corresponding to regressions in Table 2.2 shown with 95% confidence bands. In each panel, the caption summarizes the number of years with negative predicted excess returns, and the number of years where the prediction is significantly negative at the 95% level. Years in which the prediction is significantly negative are labeled. Standard errors are based on Newey-West (1987) standard errors.

Table A.1
Time-Series Robustness Checks

Time-series forecasting regressions of log excess returns on speculative-grade bonds on debt issuer quality ISS^{EDF}, including controls for the short-rate, the term spread, the credit spread, and lagged excess returns

$$rx^{HY}_{t+k} = a_k + b \cdot ISS^{EDF}_t + c \cdot (y^G_{L,t} - y^G_{S,t}) + d \cdot y^G_{S,t} + e \cdot (y^{BBB}_{L,t} - y^G_{L,t}) + f \cdot rx^{HY}_t + u_{t+k}.$$

t-statistics based on Newey-West (1987) are shown in brackets. *, **, *** denotes significance at the 10%, 5%, and 1% level, respectively, based on the fixed-*b* asymptotics developed by Kiefer and Vogelsang (2005). We next report bootstrapped *p*-values using the stationary moving-blocks bootstrap of Politis and Romano (1994) using a average block length of 8 and 10,000 bootstrap replications. We also report *t*-statistics based on parametric standard errors which assume that the scores for *k*-period returns follow an either ARMA(1,*k*) and ARMA(2,*k*) process, respectively. We then study the impact of Stambaugh (1999) bias on our baseline results. The corrections that we consider require us to drop the final observation for 2008, so we first report the OLS coefficient omitting the final sample year. We next report a bootstrap bias-adjusted estimate and a bootstrap *p*-value using the approach of Baker and Stein (2004) and Baker, Taliaferro, and Wurgler (2006). For the univariate specifications, we also report the bias-adjusted estimates and associated standard error using the methods in Amihud and Hurvich (2004). Last, we decompose the bias as $BIAS(b_{OLS}) = \phi \times BIAS(\rho)$ following Amhiud and Hurvich (2004).

	1-yr returns: rx^{HY}_{t+1}				2-yr returns: rx^{HY}_{t+2}			
	-9.534***	-7.636***	-8.617***	-6.282**	-15.254***	-11.022***	-18.052***	-13.890***
[t] Newey-West	[-3.97]	[-3.45]	[-2.97]	[-2.40]	[-5.29]	[-3.45]	[-4.60]	[-4.54]
Bootstrapped p-value	0.0005	0.000	0.001	0.003	0.001	0.017	0.006	0.001
[t] ARMA(1,*k*)	[-2.95]	[-4.02]	[-2.29]	[-3.05]	[-5.1]	[-3.92]	[-4.25]	[-5.73]
[t] ARMA(2,*k*)	[-5.08]	[-3.76]	[-3.66]	[-2.76]	[-4.93]	[-3.69]	[-4.63]	[-6.34]
Controls	None	Rates	Credit	All	None	Rates	Credit	All
Stambaugh Bias								
[b_{OLS}] (dropping 2008)	-8.842	-6.868	-9.584	-8.031	-15.392	-10.939	-18.172	-14.198
[b] bootstrap bias-adjusted	-8.783	-6.608	-8.808	-6.780	-15.815	-11.317	-17.651	-13.307
Bootstrap *p*-value	0.006	0.065	0.017	0.063	0.000	0.019	0.001	0.005
[b] AH (2004) bias-adjusted	-8.746				-15.816			
[t] AH (2004) bias-adjusted	[-2.70]				[-4.02]			
$BIAS(\mathbf{b_{OLS}})$ AH (2004)	-0.096				0.424			
$BIAS\ (\rho_{\mathbf{OLS}})$	-0.061				-0.062			
ϕ	1.581				-6.782			

Table A.2
Forecasting Equity Market and Equity Factor Returns

Annual time-series forecasting regressions of log excess equity returns on issuer quality ISS^{EDF}

$$rx^{E}_{t+k} = a + b \cdot ISS^{EDF}_{t} + c \cdot (y^{G}_{L,t} - y^{G}_{S,t}) + d \cdot y^{G}_{S,t} + e \cdot (y^{BBB}_{L,t} - y^{G}_{L,t}) + f \cdot rx^{E}_{t} + u_{t+k},$$

where rx^E is cumulative excess return on an equity portfolio over the next *k*-years. Control variables include the term spread, short-rate, credit spread, and lagged values of the dependent variable. We report only the coefficient on ISS^{EDF} and its associated t-statistic. Excess returns are alternately *MKTRF*, *SMB*, and *HML*, obtained from Ken French's web-site, or the return on the size-balanced value-weighted long-short portfolio based on *EDF*. *t*-statistics for *k*-period forecasting regressions are based on Newey-West (1987) standard errors, allowing for serial correlation up to *k*-lags.

	Univariate			Including time-series controls		
	rx^E_{t+1}	rx^E_{t+2}	rx^E_{t+3}	rx^E_{t+1}	rx^E_{t+2}	rx^E_{t+3}
rx = MKTRF						
B	-4.570	-4.968	-7.814	-2.099	-5.263	-12.703
[*t*]	[-0.98]	[-0.95]	[-1.31]	[-0.36]	[-0.90]	[-2.33]
R^2	0.02	0.01	0.02	0.04	0.19	0.33
rx = SMB						
b	-6.187	-8.788	-9.729	-6.593	-12.621	-21.579
[*t*]	[-1.62]	[-1.20]	[-1.10]	[-1.35]	[-1.57]	[-2.45]
R^2	0.06	0.05	0.03	0.15	0.13	0.18
rx = HML						
b	-6.803	-10.635	-12.085	-3.816	-2.689	-5.332
[*t*]	[-2.06]	[-3.10]	[-3.58]	[-0.77]	[-0.68]	[-1.06]
R^2	0.07	0.10	0.10	0.13	0.25	0.31
rx = EDF (high-low)						
b	-6.169	-10.420	-14.484	-5.853	-11.438	-21.495
[*t*]	[-1.80]	[-2.00]	[-2.37]	[-1.56]	[-2.33]	[-3.44]
R^2	0.05	0.08	0.12	0.12	0.18	0.25

Table A.3
Univariate Subsample Results for log(*HYS*)

Univariate time-series forecasting regressions of log excess returns on log(*HYS*)

$$rx_{t+k} = a + b \cdot \log(HYS_t) + u_{t+k}.$$

The high yield share (*HYS*) is the fraction of non-financial corporate bond issuance with a high yield rating from Moody's. In Panel A, the dependent variable is the cumulative 1-, 2-, or 3-year excess return on high yield bonds. In Panel B, the dependent variable is the cumulative 1-, 2-, or 3-year excess return on BBB-rated corporate bonds. In Panel C, the dependent variable is the cumulative 1-, 2-, or 3-year excess return on AAA-rated corporate bonds. *t*-statistics for *k*-period forecasting regressions are based on Newey-West (1987) standard errors allowing for serial correlation up to *k*-lags.

	1926-1943			1944-1982			1983-2008			1944-2008			1926-2008		
	1-yr	2-yr	3-yr	1-yr	2-yr	3-yr	1-yr	2-yr	3-yr	1-yr	2-yr	3-yr	1-yr	2-yr	3-yr
	Panel A: High Yield Excess Returns (rx^{HY})														
b	5.523	0.862	-5.128	-2.940	-5.103	-6.323	-11.483	-14.264	-17.798	-2.029	-3.371	-4.100	-1.517	-2.917	-3.884
[t]	[0.86]	[0.06]	[-0.24]	[-5.65]	[-5.14]	[-3.48]	[-2.77]	[-4.23]	[-5.76]	[-2.52]	[-2.84]	[-2.74]	[-1.77]	[-1.98]	[-1.93]
R^2	0.04	0.00	0.01	0.19	0.34	0.38	0.15	0.21	0.28	0.05	0.11	0.13	0.02	0.04	0.05
	Panel B: BBB Excess Returns (rx^{BBB})														
b	3.352	1.676	-2.162	-2.440	-4.001	-4.503	-3.139	-3.344	-5.118	-1.138	-1.887	-2.099	-0.874	-1.656	-2.100
[t]	[1.18]	[0.25]	[-0.20]	[-5.81]	[-5.49]	[-4.59]	[-1.32]	[-2.46]	[-2.09]	[-2.41]	[-2.51]	[-2.37]	[-1.71]	[-1.86]	[-1.85]
R^2	0.05	0.00	0.00	0.31	0.43	0.44	0.05	0.06	0.13	0.06	0.12	0.12	0.02	0.04	0.05
	Panel C: AAA Excess Returns (rx^{AAA})														
b	0.453	0.136	0.133	-1.706	-2.763	-2.685	-0.030	-1.015	-1.920	-0.406	-0.600	-0.242	-0.310	-0.473	-0.108
[t]	[0.82]	[0.15]	[0.12]	[-6.32]	[-4.93]	[-3.05]	[-0.07]	[-1.07]	[-1.14]	[-1.40]	[-1.14]	[-0.36]	[-1.17]	[-1.03]	[-0.19]
R^2	0.02	0.00	0.00	0.36	0.36	0.21	0.00	0.02	0.05	0.03	0.02	0.00	0.02	0.01	0.00

Table A.4
Multivariate Subsample Results for log(HYS)

Time-series forecasting regressions of log excess returns on speculative-grade bonds on log(*HYS*), controlling for the term spread, short-rate, credit spread, and lagged excess returns $rx_{t+k}^{HY} = a + b \cdot \log(HYS_t) + c \cdot \left(y_{L,t}^{G} - y_{S,t}^{G}\right) + d \cdot y_{S,t}^{G} + e \cdot \left(y_{L,t}^{BBB} - y_{L,t}^{G}\right) + f \cdot rx_t^{HY} + u_{t+k}$.

	1-yr returns			2-yr returns			3-yr returns		
				Panel A: 1926-1943					
log(*HYS*)	12.272	3.365	8.823	10.092	-6.574	4.713	9.257	-16.305	4.319
	[1.88]	[0.45]	[1.06]	[1.08]	[-0.53]	[0.39]	[0.77]	[-0.88]	[0.33]
$y_{L,t}^{G} - y_{S,t}^{G}$	-6.369		-6.223	-4.871		-9.664	-9.448		-25.799
	[-1.81]		[-1.85]	[-0.72]		[-2.30]	[-0.79]		[-4.51]
$y_{S,t}^{G}$	-8.091		-6.346	-14.651		-13.481	-21.098		-23.719
	[-2.29]		[-2.17]	[-2.55]		[-2.78]	[-3.48]		[-4.25]
$y_{L,t}^{BBB} - y_{L,t}^{G}$		2.805	4.033		12.810	12.815		20.846	27.299
		[0.89]	[2.42]		[1.92]	[4.33]		[1.95]	[7.95]
rx_t^{HY}		0.596	0.489		1.127	0.818		1.225	0.882
		[1.77]	[2.08]		[1.51]	[1.77]		[1.16]	[1.66]
R^2	0.28	0.27	0.39	0.31	0.22	0.41	0.36	0.23	0.59
				Panel B: 1944-1982					
log(*HYS*)	-2.258	-2.807	-1.917	-4.774	-5.227	-4.587	-6.762	-6.725	-6.626
	[-3.00]	[-4.82]	[-2.20]	[-4.42]	[-4.57]	[-3.62]	[-3.35]	[-3.37]	[-2.91]
$y_{L,t}^{G} - y_{S,t}^{G}$	2.525		-1.461	1.256		1.572	-2.846		-0.703
	[1.12]		[-0.49]	[0.46]		[0.31]	[-0.67]		[-0.11]
$y_{S,t}^{G}$	-0.015		-1.155	0.005		-0.237	-0.368		-0.251
	[-0.04]		[-1.40]	[0.01]		[-0.21]	[-0.41]		[-0.17]
$y_{L,t}^{BBB} - y_{L,t}^{G}$		1.664	5.920		-0.103	0.595		-2.557	-1.595
		[1.44]	[1.60]		[-0.05]	[0.13]		[-0.62]	[-0.27]
rx_t^{HY}		0.041	0.053		-0.198	-0.241		-0.394	-0.382
		[0.35]	[0.39]		[-1.35]	[-1.33]		[-2.50]	[-1.80]
R^2	0.23	0.21	0.29	0.34	0.36	0.38	0.39	0.44	0.44
				Panel C: 1983-2008					
log(*HYS*)	-11.180	-6.632	-5.326	-11.889	-12.432	-9.525	-15.000	-17.516	-14.451
	[-2.89]	[-1.58]	[-1.20]	[-3.21]	[-3.12]	[-2.32]	[-5.05]	[-4.02]	[-5.19]
$y_{L,t}^{G} - y_{S,t}^{G}$	1.803		4.203	8.188		11.651	8.269		11.254
	[0.39]		[0.81]	[2.10]		[2.55]	[2.34]		[3.52]
$y_{S,t}^{G}$	-1.751		-0.225	-1.052		-1.173	-0.105		-0.430
	[-1.19]		[-0.15]	[-1.14]		[-1.11]	[-0.08]		[-0.34]
$y_{L,t}^{BBB} - y_{L,t}^{G}$		11.227	8.335		5.693	-4.489		-0.161	-9.068
		[2.42]	[1.56]		[0.53]	[-0.57]		[-0.01]	[-0.85]
rx_t^{HY}		0.138	-0.059		-0.025	-0.620		-0.077	-0.585
		[0.56]	[-0.27]		[-0.08]	[-2.20]		[-0.22]	[-2.79]
R^2	0.24	0.33	0.36	0.40	0.22	0.48	0.39	0.28	0.45
				Panel D: 1944-2008					
log(*HYS*)	-1.787	-2.210	-1.720	-3.260	-3.395	-3.094	-4.143	-4.068	-4.130
	[-2.23]	[-2.10]	[-1.75]	[-2.91]	[-2.51]	[-2.50]	[-2.62]	[-2.46]	[-2.46]
$y_{L,t}^{G} - y_{S,t}^{G}$	4.295		2.590	7.884		9.519	7.441		10.857
	[2.21]		[0.92]	[4.01]		[3.47]	[2.89]		[3.31]
$y_{S,t}^{G}$	-0.174		-0.810	0.356		0.227	0.520		0.893
	[-0.36]		[-1.33]	[0.78]		[0.31]	[0.90]		[0.84]
$y_{L,t}^{BBB} - y_{L,t}^{G}$		5.918	5.923		3.155	-0.674		1.399	-4.693
		[-2.36]	[2.26]		[1.48]	[-0.22]		[0.42]	[-0.91]
rx_t^{HY}		-0.030	-0.105		-0.122	-0.443		-0.215	-0.553
		[-0.24]	[-0.80]		[-0.76]	[-2.42]		[-1.21]	[-3.05]
R^2	0.13	0.20	0.30	0.27	0.15	0.35	0.24	0.16	0.32
				Panel E: 1926-2008					
log(*HYS*)	-1.495	-1.925	-1.717	-3.154	-3.438	-3.250	-4.067	-4.543	-4.254
	[-1.88]	[-2.32]	[-2.14]	[-2.51]	[-2.57]	[-2.63]	[-2.21]	[-2.23]	[-2.20]
$y_{L,t}^{G} - y_{S,t}^{G}$	2.201		0.275	6.269		5.577	5.555		3.076
	[1.20]		[0.12]	[2.47]		[1.46]	[1.31]		[0.51]
$y_{S,t}^{G}$	-0.375		-0.514	-0.054		-0.057	-0.240		-0.515
	[-0.78]		[0.94]	[-0.07]		[-0.06]	[-0.20]		[-0.38]
$y_{L,t}^{BBB} - y_{L,t}^{G}$		3.492	2.936		4.552	1.024		5.904	3.658
		[1.53]	[1.13]		[1.60]	[0.28]		[1.78]	[0.76]
rx_t^{HY}		0.172	0.131		0.307	0.130		0.201	0.067
		[1.06]	[0.80]		[1.27]	[0.49]		[0.73]	[0.21]
R^2	0.07	0.07	0.09	0.14	0.09	0.14	0.11	0.10	0.12

Table A.5
Quantity and Quality and Future Returns to Credit Using the Growth of Bond Issuance

Annual time-series regressions of the form $rx_{t+2}^{HY} = a + b \cdot X_t + c \cdot \left(y_{L,t}^{G} - y_{S,t}^{G}\right) + d \cdot y_{S,t}^{G} + e \cdot \left(y_{L,t}^{BBB} - y_{L,t}^{G}\right) + f \cdot rx_t^{HY} + u_{t+2}$, where rx^{HY} is the cumulative 2-year excess return on high yield bonds, *HYS* is the high yield share, $b_t^{TOTAL} = \ln[(B_t^{TOTAL})/(\Sigma_{l=1}^{5} B_{t-l}^{TOTAL}/5)]$ denotes real growth is total issuance: the log ratio of total corporate bond issuance in year *t* and average issuance over the prior 5 years (nominal issuance is deflated using the CPI deflator so this represents a real growth rate), $b_t^{IG} = \ln[(B_t^{IG})/(\Sigma_{l=1}^{5} B_{t-l}^{IG}/5)]$ and $b_t^{HY} = \ln[(B_t^{HY})/(\Sigma_{l=1}^{5} B_{t-l}^{HY}/5)]$ are the analogous constructions for investment grade and high yield issuance. To facilitate comparisons of the coefficients, b_t^{IG} and b_t^{HY} are standardized to have a standard deviation of 1 in each subsample. Panel A shows regressions without controls. Panel B shows regressions with controls. Control variables include the term spread, short-rate, credit spread, lagged excess high yield returns. *t*-statistics are based on Newey-West (1987) standard errors allowing for serial correlation up to 2-lags.

	1926-1943				1944-1982				1983-2007				1944-2007				1926-2007			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
Panel A: Univariate																				
log(*HYS*)	-12.41				-5.10				-13.76				-3.27				-3.07			
	[2.19]				[5.31]				[3.64]				[2.81]				[2.45]			
b^{TOT}	-34.14				-2.13				-4.02				-3.00				-15.91			
	[8.08]				[0.48]				[0.62]				[0.71]				[3.21]			
b^{IG}		-20.30		-16.16		-0.51		-0.61		-1.31		-0.96		-0.83		-0.74		-7.14		-6.28
		[5.68]		[5.19]		[0.26]		[0.34]		[0.54]		[0.33]		[0.56]		[0.49]		[3.33]		[3.25]
b^{HY}			-16.36	-8.42			-4.68	-4.69			-4.79	-4.72			-4.57	-4.55			-6.42	-5.42
			[2.93]	[1.87]			[3.55]	[3.45]			[1.65]	[1.58]			[3.09]	[3.10]			[3.49]	[3.84]
R^2	0.57	0.46	0.30	0.52	0.34	0.00	0.20	0.21	0.21	0.01	0.07	0.08	0.12	0.00	0.11	0.12	0.23	0.15	0.12	0.24
Panel B: Multivariate																				
log(*HYS*)	-6.55				-4.12				-6.86				-2.64				-3.25			
	[1.52]				[3.67]				[1.56]				[2.42]				[2.86]			
b^{TOT}	-28.15				-5.62				-11.09				-8.83				-15.93			
	[4.69]				[0.94]				[1.28]				[2.07]				[3.52]			
b^{IG}		-18.06		-15.22		-3.51		-2.52		-4.98		-4.50		-3.82		-3.19		-7.03		-6.41
		[3.42]		[2.54]		[1.51]		[1.12]		[1.71]		[1.50]		[2.49]		[1.98]		[3.52]		[3.47]
b^{HY}			-12.37	-5.60			-3.94	-3.17			-4.22	-3.57			-3.57	-2.90			-6.09	-5.33
			[2.19]	[1.85]			[2.20]	[2.24]			[2.03]	[1.43]			[2.70]	[2.27]			[3.41]	[3.52]
R^2	0.74	0.69	0.56	0.71	0.41	0.26	0.28	0.32	0.53	0.49	0.47	0.52	0.40	0.33	0.32	0.36	0.32	0.24	0.20	0.32

Table A.6
Quantity and Quality and Future Returns to Credit Using the Ratio of Issuance to GDP

Annual time-series regressions of the form

$$rx_{t+2}^{HY} = a + b\cdot X_t + c\cdot(y_{L,t}^{G} - y_{S,t}^{G}) + d\cdot y_{S,t}^{G} + e\cdot(y_{L,t}^{BBB} - y_{L,t}^{G}) + f\cdot rx_t^{HY} + u_{t+2},$$

where rx^{HY} is the cumulative 2-year excess return on high yield bonds, *HYS* is the high yield share, $\ln(B_t^{TOTAL}/GDP_t)$ is the log ratio of total corporate bond issuance in year *t* to GDP, $\ln(B_t^{IG}/GDP_t)$ and $\ln(B_t^{HY}/GDP_t)$ are the analogous constructions for investment grade and high yield corporate bond issuance. To facilitate comparisons of the coefficients, $\ln(B_t^{IG}/GDP_t)$ and $\ln(B_t^{HY}/GDP_t)$ are standardized to have a standard deviation of 1 in each subsample. Panel A shows regressions without controls. Panel B shows regressions with controls. Control variables include the term spread, short-rate, credit spread, lagged excess high yield returns. *t*-statistics are based on Newey-West (1987) standard errors allowing for serial correlation up to 2-lags.

	1926-1943				1944-1982				1983-2007				1944-2007				1926-2007			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
Panel A: Univariate																				
log(*HYS*)	-4.21				-5.73				-13.65				-3.54				-1.52			
	[-0.70]				[-5.72]				[-3.39]				[-3.54]				[-0.99]			
$\ln(B^{TOT}/GDP)$	-30.74				-5.91				-2.51				1.05				-9.09			
	[-5.27]				[-1.77]				[-0.46]				[0.33]				[-2.05]			
$\ln(B^{IG}/GDP)$		-21.76		-16.83		0.90		-0.26		-1.36		3.26		-0.14		1.90		-6.20		-4.74
		[-7.17]		[-4.78]		[0.56]		[-0.22]		[-0.43]		[0.97]		[-0.07]		[-1.18]		[-2.42]		[-1.72]
$\ln(B^{HY}/GDP)$			-18.22	-8.90			-6.78	-6.82			-7.02	-8.75			-4.12	-4.91			-5.44	-3.39
			[-4.62]	[-2.13]			[-6.35]	[-5.95]			[-3.44]	[-3.98]			[-2.17]	[-2.88]			[-2.75]	[-1.74]
R^2	0.58	0.53	0.37	0.59	0.38	0.01	0.42	0.42	0.21	0.01	0.16	0.19	0.11	0	0.09	0.11	0.13	0.12	0.09	0.14
Panel B: Multivariate																				
log(*HYS*)	-1.89				-5.24				-7.55				-1.90				-1.40			
	[-0.39]				[-4.47]				[-1.92]				[-1.67]				[-1.07]			
$\ln(B^{TOT}/GDP)$	-24.40				-14.26				-20.59				-5.97				-12.94			
	[-3.56]				[-3.63]				[-2.49]				[-1.63]				[-3.56]			
$\ln(B^{IG}/GDP)$		-17.52		-13.88		-2.74		-2.97		-8.55		-6.12		-4.81		-2.35		-8.23		-6.45
		[-3.38]		[-2.07]		[-1.33]		[-2.12]		[-2.09]		[-1.53]		[-2.06]		[-1.05]		[-3.88]		[-3.14]
$\ln(B^{HY}/GDP)$			-15.25	-6.37			-6.41	-6.47			-9.29	-7.52			-4.72	-3.81			-6.80	-4.27
			[-2.77]	[-1.71]			[-4.82]	[-5.17]			[-4.47]	[-4.79]			[-2.95]	[-2.34]			[-3.76]	[-2.66]
R^2	0.71	0.67	0.59	0.69	0.50	0.21	0.51	0.54	0.62	0.53	0.56	0.61	0.38	0.33	0.38	0.39	0.30	0.28	0.23	0.32

Table A.7
Determinants of the High Yield Share, 1944-2008

Time-series regressions of *HYS* on levels and past changes of interest rates:

$$HYS_t = a + b \cdot y^G_{S,t} + c \cdot (y^G_{L,t} - y^G_{S,t}) + d \cdot rx^{HY}_{t-1} + e \cdot DEF^{HY}_t + u_t \text{, or}$$

$$\Delta_k HYS_t = a + b \cdot \Delta_k y^G_{S,t} + c \cdot \Delta_k (y^G_{L,t} - y^G_{S,t}) + d \cdot rx^{HY}_{t-k-1 \to t-1} + e \cdot \Delta_k DEF^{HY}_t + \Delta_k u_t .$$

y_S^G denotes the short-term Treasury bill yield; y_L^G-y_S^G denotes the term spread, DEF^{HY} is the issuer-weighted high yield default rate from Moody's, r_L^{HY}-r_L^G is the excess high yield return, and Δ_k denotes the *k*-year difference. In columns (1) to (5) we regress the level of *HYS* on a number of covariates, columns (6) to (10) repeat this analysis in first differences, and columns (11) to (15) in second differences. In the last two columns in each block we add additional controls for lagged stock market returns and macroeconomic variables (the growth in industrial product, real consumption growth, and a recession indicator). Robust *t*-statistics are shown in brackets.

		HYS					$\Delta_1 HYS$					$\Delta_2 HYS$				
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Levels	$y^G_{S,t}$	1.305		0.582	0.656	0.617										
		[2.30]		[0.95]	[1.02]	[1.16]										
	$(y^G_{L,t} - y^G_{S,t})$	2.465		-0.703	-0.335	-0.159										
		[1.77]		[-0.32]	[-0.15]	[-0.08]										
	rx^{HY}_{t-1}		0.065	0.166	0.002	0.048										
			[0.50]	[0.90]	[0.01]	[0.31]										
	DEF_t		1.483	1.808	1.800	1.578										
			[2.44]	[1.57]	[1.56]	[1.81]										
1-year Changes	$\Delta_1 y^G_{S,t}$						-0.929		-0.142	-0.138	0.007					
							[-1.11]		[-0.23]	[-0.20]	[0.01]					
	$\Delta_1 (y^G_{L,t} - y^G_{S,t})$						-3.361		-0.442	-0.586	-0.176					
							[-2.07]		[-0.33]	[-0.43]	[-0.11]					
	rx^{HY}_{t-1}							0.336	0.255	0.380	0.262					
								[4.23]	[3.10]	[2.79]	[3.01]					
	$\Delta_1 DEF_t$							-1.527	-1.855	-1.741	-1.781					
								[-3.32]	[-2.90]	[-2.86]	[-2.65]					
2-year Changes	$\Delta_2 y^G_{S,t}$											-2.208		-0.715	-0.736	-0.191
												[-2.29]		[-1.28]	[-1.12]	[-0.23]
	$\Delta_2 (y^G_{L,t} - y^G_{S,t})$											-5.153		-0.753	-1.006	0.293
												[-2.95]		[-0.54]	[0.70]	[0.16]
	$rx^{HY}_{t-3 \to t-1}$												0.220	0.139	0.270	0.136
													[1.82]	[1.14]	[1.47]	[1.06]
	$\Delta_2 DEF_t$												-2.212	-2.951	-2.807	-2.927
													[-3.05]	[-3.68]	[-3.70]	[-3.21]
	Other controls	None	None	None	*MKT*	Macro	None	None	None	*MKT*	Macro	None	None	None	*MKT*	Macro
	R^2	0.06	0.09	0.11	0.12	0.33	0.09	0.40	0.41	0.43	0.42	0.15	0.41	0.47	0.49	0.53

Table A.8
Simulated Returns from an Extrapolative Model of Credit Cycles

Model simulation assuming c = 1.25%, r = 0%, ℓ = 50%, θ = 20%, γ = 5%, ε = 75%, λ_1 = 0.5%, λ_2 = 1%, b = 20%, $D_{\min}$ = 50%, and $D_{\max}$ = 200%. We simulate a history of 5,000,000 periods from the model and then compute the average return within each cell so long as there are at least 2,500 observations in the cell. In the simulation, the macro state evolves according to equation (EE) and investor beliefs about the persistence of the macro state (i.e. conditional sentiment) evolve according to equation (E8). Leverage is assumed to follow equation (E13) and we simulate returns according to equation (E15).

Panel A. Expected Returns (in %) in the High Default State $E[r_{t+1}|S_t=H,P_t,D_t]$

Lagged Leverage (D_t)	Lagged Price (P_t)																					
	78	**80**	**82**	**84**	**86**	**88**	**90**	**92**	**94**	**96**	**98**	**100**	**102**	**104**	**106**	**108**	**110**	**112**	**114**	**116**	**118**	**120**
0.70									4.82													
0.75							4.55	3.82	3.46													
0.80						4.35	3.20	1.57	0.55													
0.85					4.18	2.44	1.07	-0.06	-0.74	-1.24	-1.37											
0.90				3.79	1.97	0.60	-0.20	-0.95	-1.35	-1.44												
0.95			3.47	2.14	0.34	-0.37	-1.10	-1.43	-1.49													
1.00		2.99	1.83	0.32	-0.53	-1.10	-1.39	-1.51														
1.05		2.04	0.40	-0.44	-1.04	-1.42	-1.53															
1.10	1.89	0.74	-0.39	-1.03	-1.34																	
1.15	0.88	-0.25	-0.88																			
1.20	0.06	-0.65																				
1.25	-0.01																					

Panel B. Expected Returns (in %) in the Low Default State $E[r_{t+1}|S_t=L,P_t,D_t]$

Lagged Leverage (D_t)	Lagged Price (P_t)																					
	78	**80**	**82**	**84**	**86**	**88**	**90**	**92**	**94**	**96**	**98**	**100**	**102**	**104**	**106**	**108**	**110**	**112**	**114**	**116**	**118**	**120**
0.70																						
0.75																						
0.80																0.26	0.20	0.20	0.04	-0.13	-0.31	-0.40
0.85														0.34	0.30	0.27	0.14	0.05	-0.18	-0.38	-0.60	
0.90												0.36	0.34	0.31	0.25	0.07	-0.03	-0.26	-0.51			
0.95											0.38	0.34	0.30	0.21	0.04	-0.07	-0.43	-0.53				
1.00										0.37	0.34	0.28	0.17	0.06	-0.16	-0.49	-0.60					
1.05									0.42	0.34	0.27	0.17	-0.04	-0.13	-0.46	-0.69						
1.10									0.33	0.27	0.25		-0.26	-0.43	-0.59							
1.15								0.37	0.35				-0.47	-0.64								
1.20												-0.66	-0.62									
1.25												-0.71	-0.92									