Internet Appendix

for

Risk and the Cross-Section of Stock Returns

This Version: 20-Jul-2011

This Internet Appendix provides supplemental material to accompany the above-named paper. These pages provide additional insights into the economics behind our variable. There is a review of the Admati (1985) model and proofs associated with the derivation of our variable.

Section G of this Internet Appendix contains a numerical analysis in which we generate a multi-asset market. We show, cross-sectionally, that our variable is positively related to expected returns. The numerical analysis provides support for using our measure in a predictive linear regression.

Importantly, there are discussions of related variables that have been used to explain the cross-section of returns including *PIN*, *FSRV* and *Delay1*. Finally, we provide robustness tests related to the paper's main results from Table 2. These robustness checks can be found in Appendices L and M. There are additional specifications in Appendices N and O.

Table of contents

Appendix A – Overview of this paper's research approach

Theory Section

Step		Notes
a)	Start with an existing multi-asset, one-period, CARA-normal model.	*Positives*: The Admati (1985) model produces closed-form solutions for prices and the framework is well accepted in the literature. *Negatives*: Comes with all the caveats of a CARA-normal framework as well as those of one-period models (like the CAPM).
b)	Define a stock *i*'s return	***b-1)*** In the main text, we follow convention and define a stock's return as its price difference. Note that one can choose parameters such that $P_{0,i} = 1$. In this case, stock *i*'s price difference and its (conventionally defined) return are the same. ***b-2)*** In Appendix E.2, we define a stock's return as the ratio of its future payoff to today's price (and minus one.) This is the conventional way to define returns.
c)	Define R^2_i as the projection of stock *i*'s return from *t*=0 to *t*=1 on prices at *t*=0.	We carry out this step for returns defined both ways: Using the definition in b-1 above (see main text and Appendix D.) Using the definition in b-2 above (see Appendix E.2.)
d)	Solve for the relation between a stock's $E[r_i]$ and its R^2_i.	We solve for cases when the relation between a stock's $E[r_i]$ and R^2_i / (1-R^2_i) is unambiguously positive (see Appendix E). We do this for returns defined under both ***b-1*** and ***b-2*** above. For other cases, we conduct a numerical analysis (Appendix G). We also appeal to our empirical results.

Empirical Section

Step		Notes
e)	Estimate R^2_i using recently observed data.	We invoke the (typical) assumption that recently-observed data can be used to estimate predictor variables. ***e-1)*** Project CRSP returns onto start-of-period prices (see Equation 5 in the main text.) ***e-2)*** Project price differences onto start-of-period prices (see Table 2.d. in Appendix N.)
f)	Use our R^2_i as a predictor variable in cross-sectional regressions. We have CRSP monthly returns on the LHS.	This step consists of a typical Fama-MacBeth type regression. We test with both the logit transformation of our R^2_i measure shown in Equation (6) as well as a "raw" R^2_i measure. Please see Appendix N. We report both unweighted coefficients as well as precision-weighted coefficients.

B Additional economic insights

B.1 Time-series regressions

To gain intuition about our variable, consider a single-asset, single-period REE model. In REE models, investors receive information about future dividends from two sources: *i*) Private signals; and *ii*) The market-clearing mechanism/Walrasian auctioneer. Readers who are familiar with REE models typically refer to the market-clearing mechanism or Walrasian auctioneer as the "publicly observable price signal". We choose to use the terminology "market-clearing mechanism" in an effort to avoid confusion with terms used in non-REE papers. In particular, focusing too narrowly on the term "informativeness of stock prices" is problematic and should be avoided when comparing our paper with non-REE papers. Appendix I has a note explaining where the confusion arises.

When investors' private information signals are imprecise and noisy, today's price (relative to expected future dividends) tends to be low, and the risk premium is high. The amount of information gleaned from the market-clearing mechanism is high relative to the amount of private information. In other words, the market-clearing mechanism (today's price) conveys a lot of information to investors. In terms of a time-series regression, increased uncertainty leads to a stronger (negative) relation between the stock's return and its price. The stronger relation leads to a higher fit (our time-series R^2) and to higher values of our variable.

When investors receive precise information about future dividends, today's price tends to be high, the risk premium is low, and the market-clearing mechanism conveys relatively little information to investors. The example in this paragraph is focused on the precision of the private signals. REE models also include supply shocks. Thus, readers can think about the precision of signal-to-noise. In this second example, there is a weaker (but still negative) time-series relation between the stock's return and its price. As investors' private information becomes increasingly precise, information about future dividends becomes perfectly incorporated into the stock's price, the expected return converges to the riskfree rate, the information risk premium goes to zero, and there is no role for the market-clearing mechanism to convey additional information. In such cases, there is zero covariance between the stock's return and its price (our time-series $R^2 \to 0$) and leads to lower values of our variable.

B.2 Comparisons and contrasts with a frictionless CAPM

To understand the economics behind our variable, it is helpful to compare and contrast a REE model with the frictionless Sharpe-Lintner-Mossin capital asset pricing model (CAPM).

In the frictionless, one-period, two-date, CAPM framework, a stock's future dividend is random, while its price (today) is a deterministic function of model parameters. These parameters include the stock's expected dividend and the variance-covariance matrix of all dividends. Neither parameter is a random quantity. Therefore, in the frictionless CAPM framework, the covariance between a stock's return over the t=0 to t=1 interval and its price at t=0 is zero due to the fact that prices are deterministic—i.e., the time-series $R^2 = 0$ from a regression of a stock's return on its own price.

In a noisy REE framework, future dividends are random as in the frictionless CAPM. Importantly, a stock's price is a random variable because it is a function of aggregate information and noise—both of which are random variables. Information variables are, by definition, linked to future dividends. The link induces a non-zero covariance between a stock's price and its return. Thus, in all but degenerate REE cases, $R^2 \neq 0$ for a time-series regression of stock i's return on its price.

If we consider cross-sectional relations, and not a time series analysis as in the two paragraphs above, the covariance between stocks' expected returns and prices will be negative regardless of whether prices are deterministic (as in the frictionless CAPM) or random (as in noisy REE models). Berk (1995) discusses that financial economists *should expect* to find such cross-sectional relations. In the first sentence of this paragraph, prices are "normalized" so that we can compare firms with similar levels of dividends (though riskiness may vary).

B.3 A very rough analogy

To further understand our paper, we can make a loose comparison: The frictionless CAPM produces stocks' betas. Sorting stocks by estimates of their betas should allow an econometrician to predict cross-sectional dispersion in average returns. Our paper produces a variable that is theoretically correlated with expected returns. Sorting stocks by estimates of our variable can be used to predict cross-sectional dispersion in realized returns. Empirically, our variable does a much better job predicting return dispersion than the frictionless CAPM beta does.

B.4 Disentangling information and noise

In a noisy REE model, and as Eq.(1) shows, information and noise cannot be separated when thinking about expected returns. There are strands of literature that attempt to isolate one of the variables (e.g., information) by looking at an observable quantity such as the number of analysts following a stock.

The derivation of our predictor variable does not rely on separating information and noise parameters. The derivation in Section 2 produces a single observable variable (our R_i^2) that is correlated with $\mathbb{E}[r]$ while allowing information and noise parameters to remain intertwined. We see this result as a major strength of our paper.

B.5 A note on public signals

Adding public signals does not alter the conclusions reached in this paper. Consider a public signal received by all investors in an Admati-type framework—the model is outlined in Appendix C of this document. The public announcement reduces the variance of the signal $\tilde{Y}_a$ that a given investor receives. Thus, investor a's matrix, $\mathbf{S}_a$, is more precise than it would have been without the public signal. A more precise $\mathbf{S}_a$ parameter leads to a more precise $\mathbf{Q}$ parameter but does not alter our derivations and conclusions.

C Theory set-up

The Admati (1985) model assumes a continuum of economic agents each of whom invests his initial wealth in a riskless asset and n risky assets. Assets are traded at date 0 and agents consume at date 1. Risky asset i pays $\tilde{F}_i$ units of the single consumption good or "dividend" at date 1. The $n \times 1$ vector of dividends is $\tilde{F} = (\tilde{F}_1, \ldots, \tilde{F}_n)'$. The mean and variance of dividends are given by $\bar{F} = \mathbb{E}[\tilde{F}]$ and $\mathbf{V} = Var[\tilde{F}]$. Boldface variables indicate matrices. The riskfree rate is denoted r_f.

Each agent "a" maximizes his utility of final consumption. The utility function exhibits constant absolute risk aversion with risk tolerance ρ_a. The average risk tolerance in the economy is $\bar{\rho} = \int \rho_a da$. Each agent receives an independent signal about future dividends in the form $\tilde{Y}_a = \tilde{F} + \tilde{\varepsilon}_a$. The final term $(\tilde{\varepsilon}_a)$ is a mean-zero random variable with variance-covariance matrix $\mathbf{S}_a$. The weighted average of the signal precision matrices is $\mathbf{Q} = \int \rho_a \mathbf{S}_a^{-1} da$.

Finally, and as is common in rational expectations models, the supply per capita is given by the random variable $\tilde{Z}$. The mean and variance of the supply are given by $\bar{Z} = \mathbb{E}[\tilde{Z}]$ and $\mathbf{U} = Var[\tilde{Z}]$. The closed-form solution for prices at date 0 is given by Theorem 3.1 on page 637 of Admati (1985) and contains three constants. All are shown below.

$$\tilde{P} = A_0 + \mathbf{A}_1 \tilde{F} - \mathbf{A}_2 \tilde{Z} \tag{8}$$

$$\begin{aligned}
A_0 &= \frac{\bar{\rho}}{1+r_f} \left(\bar{\rho}\mathbf{V}^{-1} + \bar{\rho}\mathbf{Q}\mathbf{U}^{-1}\mathbf{Q} + \mathbf{Q}\right)^{-1} \left(\mathbf{V}^{-1}\bar{F} + \mathbf{Q}\mathbf{U}^{-1}\bar{Z}\right) \\
\mathbf{A}_1 &= \frac{1}{1+r_f} \left(\bar{\rho}\mathbf{V}^{-1} + \bar{\rho}\mathbf{Q}\mathbf{U}^{-1}\mathbf{Q} + \mathbf{Q}\right)^{-1} \left(\mathbf{Q} + \bar{\rho}\mathbf{Q}\mathbf{U}^{-1}\mathbf{Q}\right) \\
\mathbf{A}_2 &= \frac{1}{1+r_f} \left(\bar{\rho}\mathbf{V}^{-1} + \bar{\rho}\mathbf{Q}\mathbf{U}^{-1}\mathbf{Q} + \mathbf{Q}\right)^{-1} \left(\mathbf{I}_n + \bar{\rho}\mathbf{Q}\mathbf{U}^{-1}\right)
\end{aligned}$$

The price vector is a function of the two $n \times 1$ vectors of random variables. The first vector is the sum of investors' information signals and equals the vector of dividends/payoffs at date 1 (denoted $\tilde{F}$). The second vector is the per-capita supply of risky assets (denoted $\tilde{Z}$). The equality in Eq.(8) arises from assuming the market has a continuum of investors which implies investors' signals are unbiased on average.

In REE models, investors can also observe the market-clearing mechanisms as well as their own private signals. A given investor is thus able to (imperfectly) infer other investors' information. We measure the relative importance of the market-clearing mechanism (public price signals) versus private signals in determining an asset's equilibrium price. The market-clearing mechanism (publicly observed price signals) plays a relatively larger role in determining an asset's equilibrium price when the asset's expected return is high.

Following convention, returns are defined to be the change in prices between date 0 and date 1 and given by an $n \times 1$ vector. The return of stock i is the i^{th} element of this vector. Below, $\mathbf{I}_n$ is the $n \times n$ identity matrix. As mentioned near Eq.(1) of the main text, we have assumed the riskfree rate is zero throughout the derivations to reduce notation.

$$\begin{aligned} r &\equiv \tilde{F} - \tilde{P} \\ &= -A_0 + (\mathbf{I}_n - \mathbf{A}_1)\tilde{F} + \mathbf{A}_2\tilde{Z} \end{aligned} \tag{9}$$

The return vector in Eq.(9) is a function of the same two random variables ($\tilde{F}$ and $\tilde{Z}$) found in the price vector—see Eq.(8). The two random variables provide a link between stock i's return and its price—i.e., the covariance of stocks' returns and prices are non-zero in all but rarest cases. Corollary 3.5 on page 640 of Admati (1985) can be used to derive an expression for the $n \times 1$ vector of expected prices. Thus, we can write expected returns as a function of investors' precisions and supply uncertainty with $\bar{P} = \mathbb{E}[\tilde{P}]$. We have shown the expression below as Eq.1) in the main text.

$$\begin{aligned} \mathbb{E}[r] &\equiv \bar{F} - \bar{P} \\ &= \left(\bar{\rho}\mathbf{V}^{-1} + \mathbf{Q} + \bar{\rho}\mathbf{Q}\mathbf{U}^{-1}\mathbf{Q}\right)^{-1}\bar{Z} \end{aligned} \tag{1}$$

In Eq.(1), the term in parentheses is positive definite. On average, low levels of investor precisions ($\mathbf{Q}$) are associated with high expected returns (provided $\bar{Z} > 0$ which is true for stocks.) High levels of supply uncertainty ($\mathbf{U}$) are associated with high expected returns. There may exist some individual stocks for which these relations do not hold. However, such situations are anomalous and studied in Admati (1985).

D Derivation of our measure

Our goal is to derive a measure that is correlated with expected returns and that can be estimated from recently observed data. Without loss of generality, the riskfree rate is assumed to be zero. The derivation takes five steps.

Step i: We start with the fit from a multi-variate time-series regression of stock i's return on the prices of all stocks.

$$\begin{aligned} R_i^2 &= 1 - \frac{Var[\tilde{F}_i - \tilde{P}_i|\tilde{P}]}{Var[\tilde{F}_i - \tilde{P}_i]} \\ &= 1 - \frac{i_i' \cdot Var[\tilde{F} - \tilde{P}|\tilde{P}] \cdot i_i}{i_i' \cdot Var[\tilde{F} - \tilde{P}] \cdot i_i} \end{aligned} \tag{10}$$

Step ii: Computing the variance and conditional variance terms in Eq.(10) is difficult, therefore we re-write the equation in matrix-vector form. Above, i_i denotes a $n \times 1$ vector of zeros with a one (1) in the i^{th} position while $\mathbf{I}_n$ is an $n \times n$ identity matrix. We define "**Matrix** $\mathbf{R}^2$" to be the matrix of regression fits:

$$\textbf{Matrix } \mathbf{R}^2 = \mathbf{I}_n - Var\left[\tilde{F} - \tilde{P}|\tilde{P}\right] \cdot Var^{-1}\left[\tilde{F} - \tilde{P}\right] \tag{11}$$

Step iii: We expand the conditional variance in Eq.(11) into four terms. There are three unique terms (denoted "a", "b", and "c") each of which can be simplified.

$$\begin{aligned} Var[\tilde{F} - \tilde{P}|\tilde{P}] &= Var\left[\tilde{F} - \tilde{P}\right] - Cov\left[\tilde{F} - \tilde{P}, \tilde{P}\right] \cdot Var^{-1}\left[\tilde{P}\right] \cdot Cov\left[\tilde{F} - \tilde{P}, \tilde{P}\right]' \\ &= a - b \cdot c \cdot b' \end{aligned}$$

Step iv: To simplify the "a", "b", and "c" terms shown in Step iii, we define $\mathbf{K} \equiv (\bar{\rho}\mathbf{V}^{-1} + \mathbf{Q} + \bar{\rho}\mathbf{Q}\mathbf{U}^{-1}\mathbf{Q})^{-1}$. The variable $\mathbf{K}$ is part of expected returns shown in Eq.(1). We also note that from Eq.(8) we have $\tilde{F} - \tilde{P} = -A_0 + (\mathbf{I}_n - \mathbf{A}_1)\tilde{F} + \mathbf{A}_2\tilde{Z}$.

Term "a": We simplify the expression $Var[\tilde{F} - \tilde{P}]$

$$\begin{aligned} Var[\tilde{F} - \tilde{P}] &= Var\left[\left(\mathbf{I}_n - \mathbf{A}_1\right)\tilde{F} + \mathbf{A}_2\tilde{Z}\right] \\ &= (\mathbf{I}_n - \mathbf{A}_1)\mathbf{V}(\mathbf{I}_n - \mathbf{A}_1)' + \mathbf{A}_2\mathbf{U}\mathbf{A}_2' \\ &= \mathbf{K}\bar{\rho}\mathbf{V}^{-1}\mathbf{V}(\mathbf{K}\bar{\rho}\mathbf{V}^{-1})' + \mathbf{A}_2\mathbf{U}\mathbf{A}_2' \\ &= \mathbf{K}\left(\mathbf{U} + \bar{\rho}\mathbf{Q} + \bar{\rho}\mathbf{K}^{-1}\right)\mathbf{K}' \end{aligned}$$

Term "b": We simplify the expression $Cov[\tilde{F} - \tilde{P}, \tilde{P}]$. Note that the matrices $\mathbf{K}$, $\mathbf{U}$, and $\mathbf{Q}$ are all positive definite. Therefore, the covariance matrix is negative definite. Economically, lower prices today imply higher returns (on average) matching intuition in Berk (1995).

$$\begin{aligned} Cov[\tilde{F} - \tilde{P}, \tilde{P}] &= Cov[(\mathbf{I}_n - \mathbf{A}_1)\tilde{F} + \mathbf{A}_2\tilde{Z}, \mathbf{A}_1\tilde{F} - \mathbf{A}_2\tilde{Z}] \\ &= (\mathbf{I}_n - \mathbf{A}_1)V\mathbf{A}_1' - \mathbf{A}_2U\mathbf{A}_2' \\ &= (\mathbf{I}_n - \mathbf{K}(\mathbf{Q} + \bar{\rho}\mathbf{Q}\mathbf{U}^{-1}\mathbf{Q}))\mathbf{V}(\mathbf{K}(\mathbf{Q} + \bar{\rho}\mathbf{Q}\mathbf{U}^{-1}\mathbf{Q}))' \\ &\quad -(\mathbf{K}(\mathbf{I}_n + \bar{\rho}\mathbf{Q}\mathbf{U}^{-1}))\mathbf{U}(\mathbf{K}(\mathbf{I}_n + \bar{\rho}\mathbf{Q}\mathbf{U}^{-1}))' \\ &= -\mathbf{K}\left(\mathbf{U} + \bar{\rho}\mathbf{Q}\right)\mathbf{K}' \end{aligned}$$

Term "c": We simplify the expression $Var^{-1}[\tilde{P}]$

$$\begin{aligned} Var[\tilde{P}] &= Var\left[\mathbf{A}_1\tilde{F} - \mathbf{A}_2\tilde{Z}\right] \\ &= \mathbf{A}_1\mathbf{V}\mathbf{A}_1' + \mathbf{A}_2\mathbf{U}\mathbf{A}_2' \\ &= \left(\mathbf{K}(\mathbf{U} + \bar{\rho}\mathbf{Q})\mathbf{U}^{-1}\mathbf{Q}\right)\mathbf{V}\left(\mathbf{K}(\mathbf{U} + \bar{\rho}\mathbf{Q})\mathbf{U}^{-1}\mathbf{Q}\right)' \\ &\quad +\mathbf{K}(\mathbf{U} + \bar{\rho}\mathbf{Q})\mathbf{U}^{-1}\mathbf{U}(\mathbf{K}(\mathbf{U} + \bar{\rho}\mathbf{Q})\mathbf{U}^{-1})' \\ &= \mathbf{K}(\mathbf{U} + \bar{\rho}\mathbf{Q}) \cdot (\mathbf{U}^{-1}\mathbf{Q}\mathbf{V}\mathbf{Q}\mathbf{U}^{-1} + \mathbf{U}^{-1}) \cdot (\mathbf{U} + \bar{\rho}\mathbf{Q})\mathbf{K}' \end{aligned}$$

Step v: We combine terms "a", "b", and "c", re-write **Matrix $\mathbf{R}^2$**, and solve for R_i^2:

$$\begin{aligned} \textbf{Matrix } \mathbf{R}^2 &= \mathbf{I}_n - Var\left[\tilde{F} - \tilde{P}|\tilde{P}\right] \cdot Var^{-1}\left[\tilde{F} - \tilde{P}\right] \\ &= \mathbf{K}\left(\mathbf{U}^{-1}\mathbf{Q}\mathbf{V}\mathbf{Q}\mathbf{U}^{-1} + \mathbf{U}^{-1}\right)^{-1}\left(\mathbf{U} + \bar{\rho}\mathbf{Q} + \bar{\rho}\mathbf{K}^{-1}\right)^{-1}\mathbf{K}^{-1} \end{aligned} \tag{12}$$

When matrices $\mathbf{Q}$, $\mathbf{U}$, and $\mathbf{V}$ are diagonal, we simplify Eq.(12) to get Eq.(3) in the main paper. R_i^2 is the i^{th} element on the main diagonal of $\mathbf{Matrix\ R}^2$:

$$\begin{aligned} R_i^2 &= i_i' \cdot \left(\mathbf{Matrix\ R}^2\right) \cdot i_i \\ &= i_i' \cdot \mathbf{K}\left(\mathbf{U}^{-1}\mathbf{QVQU}^{-1} + \mathbf{U}^{-1}\right)^{-1}\left(\mathbf{U} + \bar{\rho}\mathbf{Q} + \bar{\rho}\mathbf{K}^{-1}\right)^{-1}\mathbf{K}^{-1} \cdot i_i \\ &= K_i\left(U_i^{-1}Q_iV_iQ_iU_i^{-1} + U_i^{-1}\right)^{-1}\left(U_i + \bar{\rho}Q_i + \bar{\rho}K_i^{-1}\right)^{-1}K_i^{-1} \\ &= \left(\frac{Q_i^2V_i}{U_i^2} + \frac{1}{U_i}\right)^{-1}\left(\frac{U_i^2V_i}{U_iV_i} + \frac{\bar{\rho}Q_iU_iV_i}{U_iV_i} + \frac{\bar{\rho}^2U_i + \bar{\rho}^2Q_i^2V_i + \bar{\rho}Q_iU_iV_i}{U_iV_i}\right)^{-1} \\ &= \left(\frac{U_i^2}{Q_i^2V_i + U_i}\right)\left(\frac{U_iV_i}{V_iU_i^2 + 2\bar{\rho}Q_iU_iV_i + \bar{\rho}^2U_i + \bar{\rho}^2Q_i^2V_i}\right) \qquad (3) \end{aligned}$$

To more easily see the relation between Q_i and R_i^2, Eq.(3) can be factored. From the expression below, we can easily show that stocks with higher (lower) values of Q_i have lower (higher) values R_i^2. For a stock with a very large value of Q_i, the R_i^2 is close to zero. If Q_i is close to zero, the expression below is close to $(U_i^3V_i)/(V_iU_i^3 + \bar{\rho}^2U_i^2)$ which is less than one.

$$R_i^2 = \frac{U_i^3V_i}{Q_i^4(\bar{\rho}^2V_i^2) + Q_i^3(2\bar{\rho}U_iV_i^2) + Q_i^2(V_i^2U_i^2 + 2\bar{\rho}^2U_iV_i) + Q_i(2\bar{\rho}U_i^2V_i) + (V_iU_i^3 + \bar{\rho}^2U_i^2)}$$

To more easily see the relation between U_i and R_i^2, Eq.(3) can again be factored. Stocks with higher (lower) values of U_i, have higher (lower) R_i^2. For a stock with a very large value of U_i, the R_i^2 is close to one. For a stock with U_i close to zero, the R_i^2 is close to zero.

$$R_i^2 = \frac{U_i^3V_i}{U_i^3(V_i) + U_i^2(Q_i^2V_i^2 + \bar{\rho}^2 + 2\bar{\rho}Q_iV_i) + U_i(2\bar{\rho}Q_i^3V_i^2 + 2\bar{\rho}^2Q_i^2V_i) + (\bar{\rho}^2Q_i^4V_i^2)}$$

To show the relations between R_i^2 and expected returns, and when assets are uncorrelated, we start with Eq.(1) written for asset i:

$$\mathbb{E}\left[r_i\right] = \frac{\bar{Z}_i}{\bar{\rho}V_i^{-1} + Q_i + \bar{\rho}Q_i^2U_i^{-1}} \qquad (13)$$

We square both sides of the above equation, divide by $U_i^2V_i^2$, and simplify to get:

$$\frac{\left(\mathbb{E}\left[r_i\right]\right)^2}{U_i^2V_i^2} = \frac{\left(\bar{Z}_i\right)^2}{Q_i^4\bar{\rho}^2V_i^2 + Q_i^3 2\bar{\rho}U_iV_i^2 + Q_i^2U_i^2V_i^2 + Q_i 2\bar{\rho}U_i^2V_i + \bar{\rho}^2U_i^2 + Q_i^2 2\bar{\rho}^2U_iV_i}$$

Using Eq.(3), the denominator of the right-hand side of the above equation, can be factored to show the relation between $\mathbb{E}[r_i]$ and R_i^2 matching Eq.(4) in the main paper:

$$\frac{\left(\mathbb{E}[r_i]\right)^2}{U_i^2 V_i^2} = \frac{\left(\bar{Z}_i\right)^2}{U_i^3 V_i \left(\frac{1}{R_i^2} - 1\right)}$$

$$\mathbb{E}[r_i] = \left(\frac{V_i}{U_i} \cdot \frac{R_i^2}{1 - R_i^2}\right)^{1/2} \bar{Z}_i \qquad (4)$$

E Analysis of relations between $\mathbb{E}[r_i]$ and R_i^2

E.1 Returns defined as price differences

We consider a market with two stocks and show that $Cov\left[\mathbb{E}[r_i], \frac{R_i^2}{1-R_i^2}\right] > 0$ if there is sufficient (cross-sectional) dispersion along one of the three dimensions $\mathbf{Q}$, $\mathbf{U}$ or $\mathbf{V}$. More precisely, if Stock 2 has sufficiently less uncertainty than Stock 1, then we have $Cov\left[\mathbb{E}[r_i], \frac{R_i^2}{1-R_i^2}\right] > 0$. Examples of such conditions include Stock 2's information precision is sufficiently higher than that of Stock 1 ($Q_2 > Q_1$), or its noise is sufficiently lower ($U_2 < U_1$), or the variance of its payoff is sufficiently lower ($V_2 < V_1$).

Note, our goal is to show positive relations exist between $\mathbb{E}[r_i]$ on the left-hand side and $\frac{R_i^2}{1-R_i^2}$ on the right-hand side of Eq.(4) from the main text.

$$\mathbb{E}[r_i] = \left(\frac{V_i}{U_i} \cdot \frac{R_i^2}{1-R_i^2}\right)^{1/2} \bar{Z}_i$$

We use $K_i = \bar{\rho}V_i^{-1} + Q_i + \bar{\rho}Q_i^2U_i^{-1} \quad \text{for } i \in \{1,2\}$. Eq.(13) of this Internet Appendix allows us to write the expression below and to the left. Eq.(4) of the main paper and Eq.(13) of this Internet Appendix allow us to write the expression below and to the right.

$$\mathbb{E}[r_i] = \tfrac{1}{K_i}\bar{Z}_i \qquad\qquad \tfrac{R_i^2}{1-R_i^2} = \tfrac{U_i/V_i}{K_i^2}$$

We deduce:

$$\begin{aligned}
Cov\left[\mathbb{E}[r_i], \frac{R_i^2}{1-R_i^2}\right] &= Cov\left[\frac{1}{K_i}\bar{Z}_i, \frac{U_i/V_i}{K_i^2}\right] \\
&= Cov\left[\frac{1}{K_i}, \frac{U_i/V_i}{K_i^2}\right] \\
&= \frac{1}{2}\left[\frac{1}{K_1} - \frac{1}{2}\left(\frac{1}{K_1}+\frac{1}{K_2}\right)\right]\cdot\left[\frac{U_1}{V_1}\frac{1}{K_i^2} - \frac{1}{2}\left(\frac{U_1}{V_1}\frac{1}{K_1^2}+\frac{U_2}{V_2}\frac{1}{K_2^2}\right)\right] \\
&\quad +\frac{1}{2}\left[\frac{1}{K_2} - \frac{1}{2}\left(\frac{1}{K_1}+\frac{1}{K_2}\right)\right]\cdot\left[\frac{U_2}{V_2}\frac{1}{K_2^2} - \frac{1}{2}\left(\frac{U_1}{V_1}\frac{1}{K_1^2}+\frac{U_2}{V_2}\frac{1}{K_2^2}\right)\right] \\
&= \frac{1}{4}\left(\frac{1}{K_1}-\frac{1}{K_2}\right)\cdot\left(\frac{U_1}{V_1}\frac{1}{K_i^2} - \frac{U_2}{V_2}\frac{1}{K_2^2}\right)
\end{aligned}$$

E.1.1 Condition on Q_2 such that $Cov\left[\mathbb{E}[r_i], \frac{R_i^2}{1-R_i^2}\right] > 0$

As $Q_2 \to \infty$, we have $K_2 \to \infty$ and consequently $Cov\left[\mathbb{E}[r_i], \frac{R_i^2}{1-R_i^2}\right] \to \frac{1}{4}\left(\frac{1}{K_1}\right) \cdot \left(\frac{U_1}{V_1}\frac{1}{K_1^2}\right)$. Since the right-hand side of the previous expression is positive, we conclude that the covariance is positive if Q_2 is sufficiently high.

More precisely, the covariance is positive if: (a) $\frac{1}{K_1} > \frac{1}{K_2}$ and (b) $\frac{U_1}{V_1}\frac{1}{K_1^2} > \frac{U_2}{V_2}\frac{1}{K_2^2}$.

Define l_1 to be a real number such that $Q_2 > l_1$ implies $\frac{1}{K_1} > \frac{1}{K_2}$. Such a number exists since $\frac{1}{K_2} \to 0$ when Q_2 becomes large. Define l_2 to be a real number such that $Q_2 > l_2$ implies $\frac{U_1}{V_1}\frac{1}{K_1^2} > \frac{U_2}{V_2}\frac{1}{K_2^2}$. Such a number exists since $\frac{1}{K_2} \to 0$ when Q_2 becomes large. We conclude, $Q_2 > max(l_1, l_2)$ implies $Cov\left[\mathbb{E}[r_i], \frac{R_i^2}{1-R_i^2}\right] > 0$.

E.1.2 Condition on U_2 such that $Cov\left[\mathbb{E}[r_i], \frac{R_i^2}{1-R_i^2}\right] > 0$

As $U_2 \to 0$, we have $K_2 \to \infty$ and consequently $Cov\left[\mathbb{E}[r_i], \frac{R_i^2}{1-R_i^2}\right] \to \frac{1}{4}\left(\frac{1}{K_1}\right) \cdot \left(\frac{U_1}{V_1}\frac{1}{K_1^2}\right)$. Since the right-hand side of the previous expression is positive, we conclude that the covariance is positive if U_2 is sufficiently low. Using similar steps as above, and defining two real numbers l_1 and l_2, we conclude, $U_2 < min(l_1, l_2)$ implies $Cov\left[\mathbb{E}[r_i], \frac{R_i^2}{1-R_i^2}\right] > 0$.

E.1.3 Condition on V_2 such that $Cov\left[\mathbb{E}[r_i], \frac{R_i^2}{1-R_i^2}\right] > 0$

As $V_2 \to 0$, we have $K_2 \to \infty$ and consequently $Cov\left[\mathbb{E}[r_i], \frac{R_i^2}{1-R_i^2}\right] \to \frac{1}{4}\left(\frac{1}{K_1}\right) \cdot \left(\frac{U_1}{V_1}\frac{1}{K_1^2}\right)$. Since the right-hand side of the previous expression is positive, we conclude that the covariance is positive if V_2 is sufficiently low. Using similar steps as above, and defining two real numbers l_1 and l_2, we conclude, $V_2 < min(l_1, l_2)$ implies $Cov\left[\mathbb{E}[r_i], \frac{R_i^2}{1-R_i^2}\right] > 0$.

We conclude by noting that similar proofs exist for $Cov\left[\mathbb{E}[r_i], R_i^2\right]$ and are available upon request.

E.2 Returns defined as price ratios

We are interested in computing $\mathbb{E}[r_i]$ and $\frac{R_i^2}{1-R_i^2}$ when returns are calculated using ratios as opposed to price differences. Problems arise when defining $r_i \equiv \frac{\tilde{F}_i}{\tilde{P}_i} - 1$ as we must deal with a ratio of normal variables. We follow, Hayya, Armstrong, and Gressis (1975, p. 1,339). The second-order Taylor expansion for r_i gives the approximations shown in Eq.(14) and Eq.(15) below:

$$\begin{aligned}
\mathbb{E}[r_i] &= \mathbb{E}\left[\frac{\tilde{F}_i}{\tilde{P}_i} - 1\right] \\
&= \left(\frac{\mathbb{E}[\tilde{F}_i]}{\mathbb{E}[\tilde{P}_i]} + Var[\tilde{P}_i]\frac{\mathbb{E}[\tilde{F}_i]}{\mathbb{E}[\tilde{P}_i]^3} - \frac{Cov[\tilde{F}_i, \tilde{P}_i]}{\left(\mathbb{E}[\tilde{P}_i]\right)^2}\right) - 1 \qquad (14)
\end{aligned}$$

And:

$$\begin{aligned}
Var[r_i] &= Var\left[\frac{\tilde{F}_i}{\tilde{P}_i} - 1\right] \\
&= Var[\tilde{P}_i]\frac{(\mathbb{E}[\tilde{F}_i])^2}{(\mathbb{E}[\tilde{P}_i])^4} + \frac{V_i}{(\mathbb{E}[\tilde{P}_i])^2} - 2Cov[\tilde{F}_i, \tilde{P}_i]\frac{\mathbb{E}[\tilde{F}_i]}{(\mathbb{E}[\tilde{P}_i])^3} \qquad (15)
\end{aligned}$$

We then consider a market with two stocks and establish conditions under which $Cov\left[\mathbb{E}[r_i], \frac{R_i^2}{1-R_i^2}\right]$ is positive. Stocks are subscripted with $i \in \{1, 2\}$. Without loss of generality and for simplification purposes, we assume that for all i, $\mathbb{E}[\tilde{F}_i] = \bar{F}_i = 1$, $\mathbb{E}[\tilde{Z}_i] = \bar{Z}_i = 1$, and $\bar{\rho} = 1$.

A key variable for the derivation is the matrix $\mathbf{K} = (\bar{\rho}\mathbf{V}^{-1} + \mathbf{Q} + \bar{\rho}\mathbf{Q}\mathbf{U}^{-1}\mathbf{Q})^{-1}$ (see Step *iv* in Appendix D). For a given stock i, we have $K_i = (\bar{\rho}V_i^{-1} + Q_i + \bar{\rho}Q_i^2U_i^{-1})^{-1}$. Eq.(8) in Appendix C allows us to write the following for stock i:

$$\tilde{P}_i = A_{0,i} + \mathbf{A}_{1,i}\tilde{F}_i - \mathbf{A}_{2,i}\tilde{Z}_i$$

$$\begin{aligned}
A_{0,i} &= \bar{\rho}K_i\left(V_i^{-1}\bar{F}_i + Q_iU_i^{-1}\bar{Z}_i\right) \\
A_{1,i} &= K_i\left(Q_i + \bar{\rho}Q_iU_i^{-1}Q_i\right) \\
A_{2,i} &= K_i\left(1 + \bar{\rho}Q_iU_i^{-1}\right)
\end{aligned}$$

We can thus write:

$$\begin{aligned}\mathbb{E}[\tilde{P}_i] &= A_{0,i} + A_{1,i}\bar{F}_i - A_{2,i}\bar{Z}_i \\ Var[\tilde{P}_i] &= A_{1,i}^2 V_i + A_{2,i}^2 U_i \\ Cov[\tilde{F}_i, \tilde{P}_i] &= V_i A_{1,i}\end{aligned} \tag{16}$$

E.2.1 Computation of $\mathbb{E}[r_i]$

We substitute the system of equations in (16) into Eq.(14) to get the expression below. Notice, the expression below is the counterpart from Eq.(13) shown in Appendix D:

$$\mathbb{E}[r_i] = \frac{\left[V_i(U_i+Q_i)(U_i+Q_iV_i)(U_i+Q_iV_iU_i+Q_i^2V_i)+(U_i+Q_iU_iV_i+Q_i^2V_i-U_iV_i)^2\right]\cdot U_iV_i}{(U_i+Q_iU_iV_i+Q_i^2V_i-U_iV_i)^3}$$

E.2.2 Computation of $\frac{R_i^2}{1-R_i^2}$

We know that:

$$\begin{aligned}R_i^2 &= 1 - \frac{Var[r_i|\tilde{P}_i]}{Var[r_i]} \\ Var[r_i|\tilde{P}_i] &= Var[r_i] - \frac{(Cov[r_i,\tilde{P}_i])^2}{Var[\tilde{P}_i]} \\ Cov[r_i,\tilde{P}_i] &= E[R_i\tilde{P}_i] - \mathbb{E}[r_i]\cdot\mathbb{E}[\tilde{P}_i] \\ &= \mathbb{E}[\tilde{F}_i - \tilde{P}_i] - \mathbb{E}[r_i]\cdot\mathbb{E}[\tilde{P}_i]\end{aligned}$$

Using Eq.(16) and Eq.(14), and after some manipulations, we find:

$$Cov[r_i,\tilde{P}_i] = \frac{Cov[\tilde{F}_i,\tilde{P}_i]}{\mathbb{E}[\tilde{P}_i]} - Var[\tilde{P}_i]\frac{\bar{F}_i}{(\mathbb{E}[\tilde{P}_i])^2}$$

Taking the equation directly above, Eq.(15), the equations in (16), and executing a rather lot of algebraic steps, we get:

$$\begin{aligned}\frac{R_i^2}{1-R_i^2} &= \frac{(U_i+Q_iV_i)^2U_iV_i}{(U_i+Q_iU_iV_i+Q_i^2V_i-U_iV_i)^2} \\ R_i^2 &= \frac{(U_i+Q_iV_i)^2U_iV_i}{(U_i+Q_iU_iV_i+Q_i^2V_i-U_iV_i)^2+(U_iQ_iV_i)^2U_iV_i}\end{aligned} \tag{17}$$

The above equation represents the counterpart to Eq.(3) in the main text.

E.2.3 Analyzing the link between $\mathbb{E}[r_i]$ and $\frac{R_i^2}{1-R_i^2}$

We compute the following covariance and show the conditions under which it is unambiguously positive

$$
\begin{aligned}
Cov\left[\mathbb{E}[r_i], \frac{R_i^2}{1-R_i^2}\right] &= \frac{1}{2}\left(\mathbb{E}[r_1] - \frac{\mathbb{E}[r_1]+\mathbb{E}[r_2]}{2}\right)\left(\frac{R_1^2}{1-R_1^2} - \frac{\frac{R_1^2}{1-R_1^2}+\frac{R_2^2}{1-R_2^2}}{2}\right) \\
&\quad +\frac{1}{2}\left(\mathbb{E}[r_2] - \frac{\mathbb{E}[r_1]+\mathbb{E}[r_2]}{2}\right)\left(\frac{R_2^2}{1-R_2^2} - \frac{\frac{R_1^2}{1-R_1^2}+\frac{R_2^2}{1-R_2^2}}{2}\right) \\
&= \frac{1}{4}\left(E[r_1]-E[r_2]\right)\left(\frac{R_1^2}{1-R_1^2} - \frac{R_2^2}{1-R_2^2}\right)
\end{aligned}
$$

Conditions on Q_2 for $Cov\left[\mathbb{E}[r_i], \frac{R_i^2}{1-R_i^2}\right]$ to be positive:

If $Q_2 \to \infty$, then $E[r_2] \to 0$ and $\frac{R_2^2}{1-R_2^2} \to 0$. Thus, $Cov[\mathbb{E}[r_i], \frac{R_i^2}{1-R_i^2}] \to \frac{1}{4}\left(E[r_1]\right)\left(\frac{R_1^2}{1-R_1^2}\right)$.

$\frac{R_1^2}{1-R_1^2} > 0$ thus if $E[r_1] > 0$, then $Cov\left[\mathbb{E}[r_i], \frac{R_i^2}{1-R_i^2}\right]$ is positive. The numerator of $E[r_1]$ is always positive. The denominator is positive under a number of cases.

The denominator of $\mathbb{E}[r_1]$ can be written as $[U_1 + Q_1U_1V_1 + Q_1^2V_1 - U_1V_1]^3 = [U_1(1+Q_1V_1 - V_1) + Q_1^2V_1]^3$. We have $Q_1 > 1 - V_1^{-1}$ as a sufficient condition for $E[r_1] > 0$ and thus, $Cov\left[\mathbb{E}[r_i], \frac{R_i^2}{1-R_i^2}\right] > 0$.

Conditions on U_2 for $Cov\left[\mathbb{E}[r_i], \frac{R_i^2}{1-R_i^2}\right]$ to be positive:

If $U_2 \to 0$, then $E[r_2] \to 0$ and $\frac{R_2^2}{1-R_2^2} \to 0$. Thus, $Cov[\mathbb{E}[r_i], \frac{R_i^2}{1-R_i^2}] \to \frac{1}{4}\left(E[r_1]\right)\left(\frac{R_1^2}{1-R_1^2}\right)$. We have $E[r_1]$ positive in several cases. One such condition is $Q_1 > 1 - V_1^{-1}$. This is a sufficient condition for $E[r_1] > 0$. A similar condition holds if $V_2 \to 0$.

We conclude by noting that similar proofs exist for $Cov\left[\mathbb{E}[r_i], R_i^2\right]$ and are available upon request.

F The case of correlated assets

Assume matrices **Q**, **U**, and/or **V** are not diagonal. We re-combine terms "a", "b", and "c" from Appendix D to give the expression below, where $\mathbf{K} \equiv (\bar{\rho}\mathbf{V}^{-1} + \mathbf{Q} + \bar{\rho}\mathbf{Q}\mathbf{U}^{-1}\mathbf{Q})^{-1}$ and i_i is a $n \times 1$ vector of zeros with a one (1) in the i^{th} position.

$$R_i^2 \;\; = \;\; 1 - \frac{i_i' \cdot \left(\mathbf{V}^{-1} + \mathbf{Q}\mathbf{U}^{-1}\mathbf{Q}\right)^{-1} \cdot i_i}{i_i' \cdot \left(\mathbf{K}\left(\mathbf{U} + \bar{\rho}\mathbf{Q}\right)\mathbf{K} + \bar{\rho}\mathbf{K}\right) \cdot i_i}$$

In broad terms, high values of **Q** are associated with low fit (with the understanding that we are speaking about elements in a matrix and therefore are being a bit imprecise.) High values of **U** are associated with high fit.

The relation between **Q**, **U**, and R_i^2 can be complicated by off-diagonal elements of the **Q**, **U**, or **V** matrices. Unfortunately, there is no closed-form solution for the relation between a stock's expected return and R_i^2 in the most general case. To better understand the relations, we turn to a numerical analysis. The set-up and results of the numerical analysis are shown in Appendix G.

G Numerical analysis

We numerically compare stocks with different precisions of investors' information and different levels of supply uncertainty. We consider a market with 25 stocks and the following parameter values. The average risk tolerance is $\bar{\rho} = 0.15$ and the mean of the supply vector is $\bar{Z} = 1$. As in the theory section, we set r_f=0 for notational simplicity. The average precisions of investors' private information signals are:

$$\mathbf{Q} = \begin{pmatrix} 5.0 & 0 & 0 & 0 & 0 \\ 0 & 6.0 & 0 & 0 & 0 \\ 0 & 0 & 7.0 & 0 & 0 \\ 0 & 0 & 0 & 8.0 & 0 \\ 0 & 0 & 0 & 0 & 9.0 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

The variance-covariance matrix of supply shocks is:

$$\mathbf{U} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 0.60 & 0 & 0 & 0 & 0 \\ 0 & 0.80 & 0 & 0 & 0 \\ 0 & 0 & 1.00 & 0 & 0 \\ 0 & 0 & 0 & 1.20 & 0 \\ 0 & 0 & 0 & 0 & 1.40 \end{pmatrix}$$

The simpler case: The variance of asset payoffs (the "**V**" matrix) is diagonal with the number 0.10 on the main diagonal. We provide three figures (G.1, G.2, G.3) that graphically illustrate the results from the numerical analysis of the simpler case. The figures can be found starting two pages from this point.

The first figure (G.1) shows cross-sectional relations between expected returns and unobserved model parameters (information precision and supply uncertainty). The second figure (G.2) shows cross-sectional relations between *Proxy E[r]* measure and the same (unobserved) model parameters. The third figure (G.3) shows the relations between expected returns and our *Proxy E[r]* measure. It is the third figure that supports using our *Proxy E[r]* measure in linear cross-sectional regressions.

The more complicated case: The variance of asset dividends (the "**V**" matrix) is based on empirical data. We first calculate the returns for 25 industry portfolios constructed at the 2-digit SIC level. We use the returns ("r") to estimate the correlation matrix (shown below). We then multiply each element of the correlation matrix shown below by 0.10 to produce the **V** matrix.

$$Corr[r] = \begin{pmatrix} 1.0000 & 0.4280 & 0.2634 & \cdots & 0.4808 \\ 0.4280 & 1.0000 & 0.2497 & \cdots & 0.5401 \\ 0.2634 & 0.2497 & 1.0000 & \cdots & 0.5211 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0.4808 & 0.5401 & 0.5211 & \cdots & 1.0000 \end{pmatrix}$$

The chart below provides a partial overview of the "stocks" studied in the more complicated case. The precision of investors' information is shown in Column 2. The level of supply uncertainty for stock i only is shown in Column 3. The fit from a time-series regression of stock i's return on the prices of all 25 stocks is shown in Column 4. Our *Proxy E[r]* measure is defined as the logistic transformation of the fit (R^2) and shown in Column 5. Stock i's expected return is shown in Column 6.

Stock #	**Q**	**U**	R_i^2	Proxy $E[r_i]$	$\mathbb{E}[r_i]$
1	5	0.60	4.3%	-3.1	8.2%
2	5	0.80	7.7%	-2.5	9.2%
3	5	1.00	11.7%	-2.0	10.1%
4	5	1.20	15.5%	-1.7	10.7%
5	5	1.40	19.4%	-1.4	11.0%
6	6	0.60	3.6%	-3.3	6.2%
7	6	0.80	5.6%	-2.8	7.5%
8	6	1.00	8.9%	-2.3	8.3%
9	6	1.20	11.9%	-2.0	8.8%
10	6	1.40	15.1%	-1.7	8.8%
...	...	...	...	...	...
25	9	1.40	8.9%	-2.3	6.5%

We again provide three figures (G.4, G.5, G.6) that graphically illustrate the results from the numerical analysis. The final figure (G.6) shows the relations between expected returns and our *Proxy E[r]* measure. Again, it is this figure that supports using our *Proxy E[r]* measure in linear cross-sectional regressions. Cross-sectional studies focus on general patterns across stocks. Figure G.6 also shows that, for two randomly chosen stocks, some relations may not hold.

Figure G.1
Expected Returns
(Simpler Case)

This figure shows stocks' expected returns as a function of two model parameters: i) the precision of investors' information about stocks' future dividends denoted "**Q**"; and ii) supply uncertainty denoted "**U**".

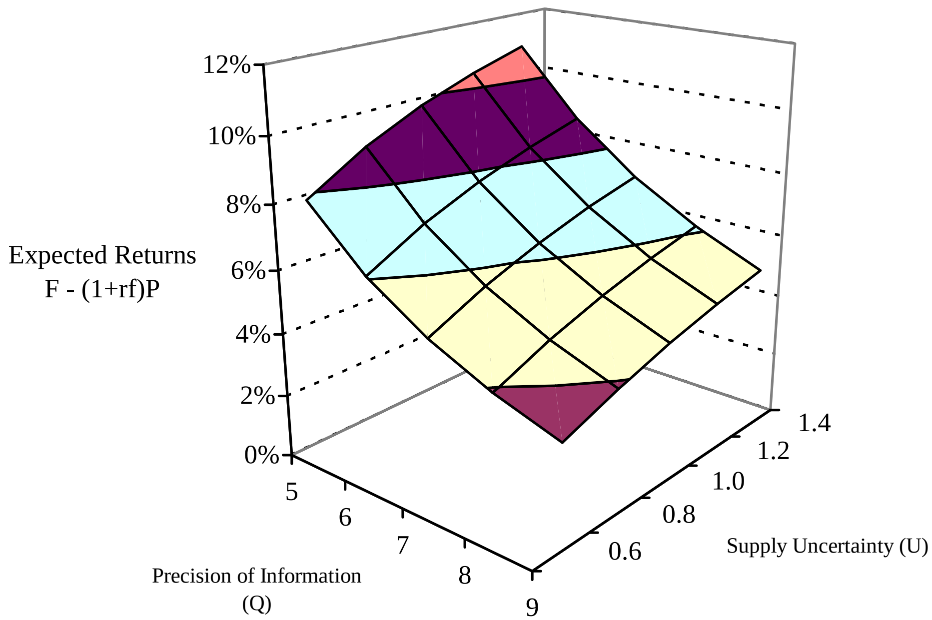


Figure G.2
Our Proxy E[r] Variable
(Simpler Case)

This figure shows our *Proxy E[r]* variable as a function of two model parameters: i) the precision of investors' information about stocks' future dividends denoted "**Q**"; and ii) supply uncertainty denoted "**U**".

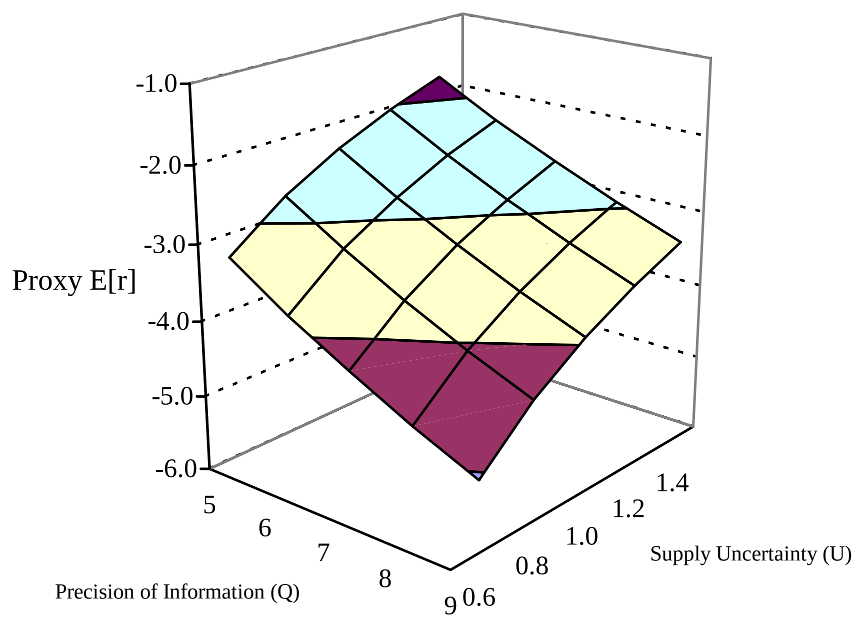


Figure G.3
Expected Returns and our *Proxy E[r]* Measure
(Simpler Case)

This figure depicts the relation between expected returns and our *Proxy E[r]* measure. We consider a market with 25 stocks. *"Proxy E[r]"* is the logistic transformation of our fit measure (R^2). A "best fit" or regression line is included in the scatter plot and shows the relation is approximately linear. Thus, our cross-sectional regression in Equation (7) in the main text is well specified in this regard.

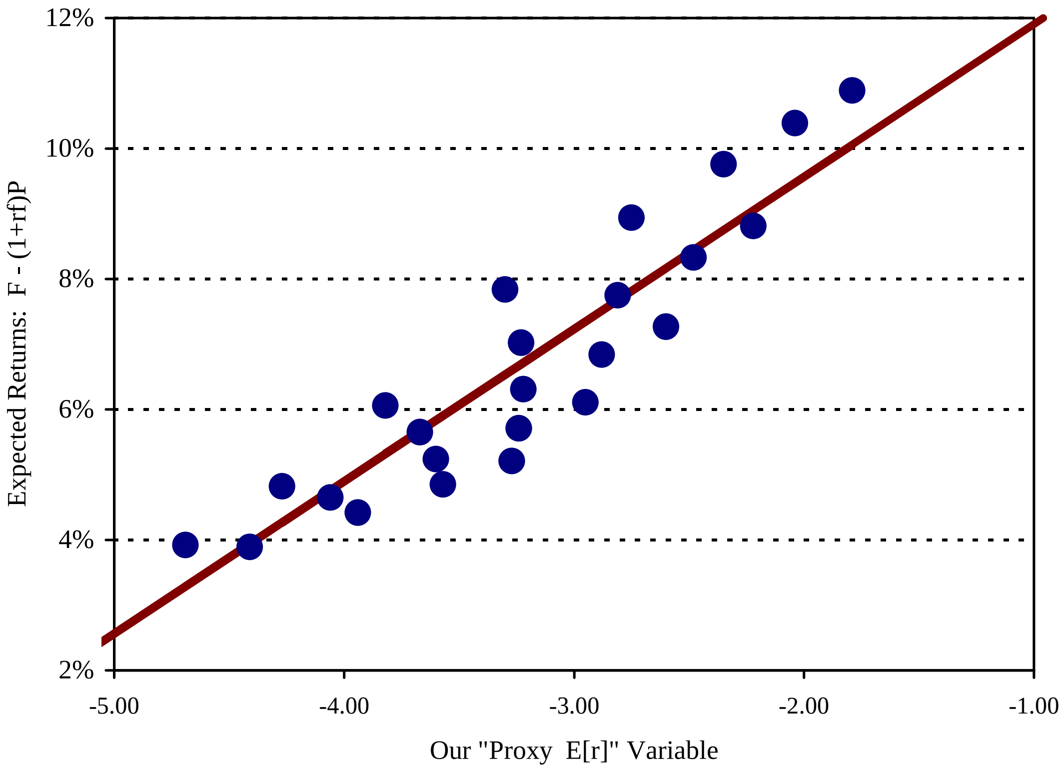


Figure G.4
Expected Returns
(More Complicated Case)

This figure shows stocks' expected returns as a function of two model parameters: i) the precision of investors' information about stocks' future dividends denoted "**Q**"; and ii) supply uncertainty denoted "**U**".

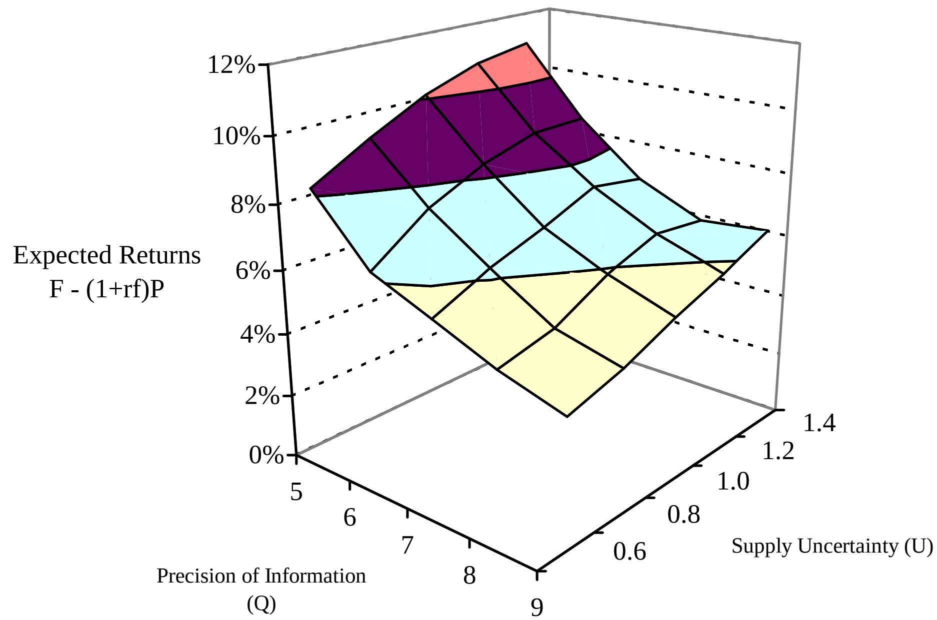


Figure G.5
Our Proxy E[r] Variable
(More Complicated Case)

This figure shows our *Proxy E[r]* variable as a function of two model parameters: i) the precision of investors' information about stocks' future dividends denoted "**Q**"; and ii) supply uncertainty denoted "**U**".

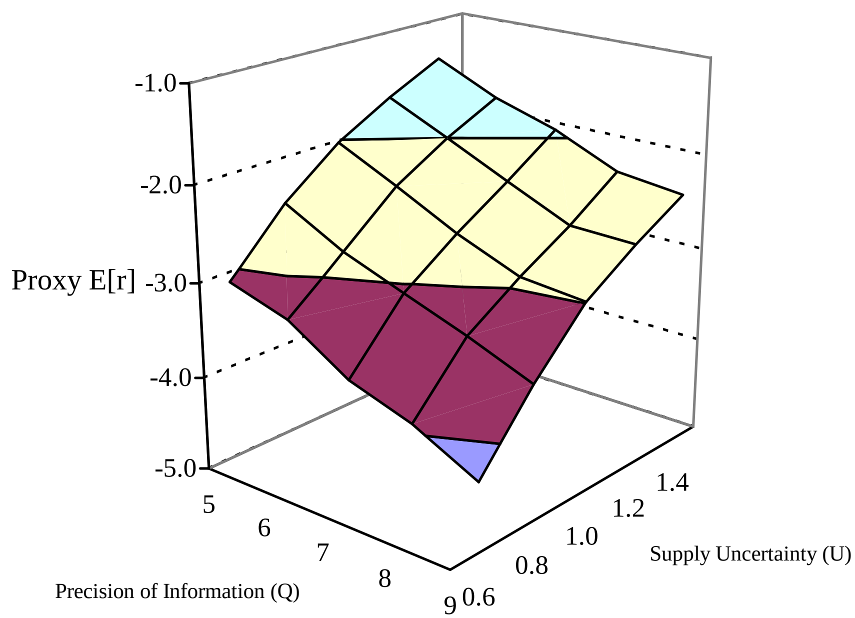
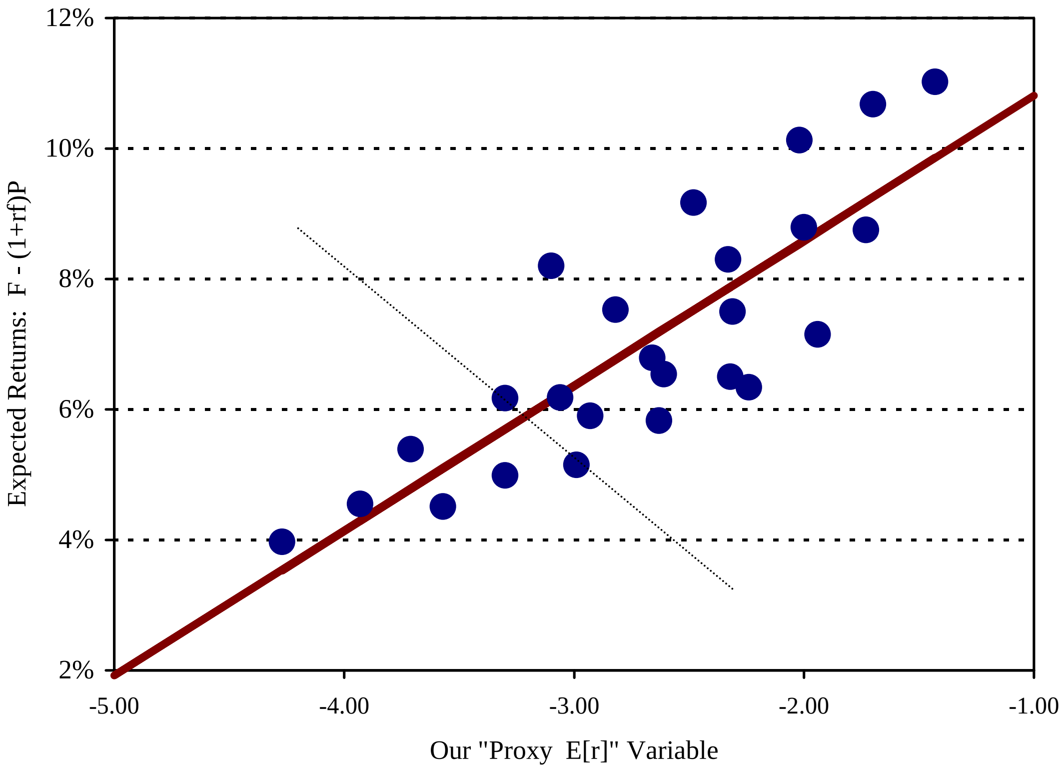


Figure G.6
Expected Returns and our *Proxy E[r]* Measure
(More Complicated Case)

This figure depicts the relation between expected returns and our *Proxy E[r]* measure. We consider a market with 25 stocks. *"Proxy E[r]"* is the logistic transformation of our fit measure (R^2). A "best fit" or regression line is included in the scatter plot and shows the relation is approximately linear. Thus, our cross-sectional regression in Equation (7) in the main text is well specified in this regard.



H Comparisons and contrasts with *PIN*

H.1 Theoretical comparisons and contrasts

The effect of information on expected returns is studied in Easley and O'Hara (2004). The authors present a multi-asset model that focuses on the role of public and private signals in determining a firm's cost of capital—i.e., expected returns. Private signals in their model are received only by a group of informed investors as in Grossman and Stiglitz (1980). Like the Admati (1985) model, cross-sectional differences in expected returns are complicated functions of unobservable parameters such as the precision of investors' signals.

Our paper is also motivated by a multi-asset equilibrium model. The main difference between the two papers is that our empirical variable follows directly from model equations. A second difference is that our measure makes use of many stocks in the market. Although Easley and O'Hara (2004) present a multi-asset model, *PIN* is ultimately estimated using data of one stock at a time.

H.2 Empirical comparisons and contrasts

Easley, Hvidkjaer, and O'Hara (2002) propose the empirical *PIN* variable. Microstructure data is used to estimate the arrival rates of informed trades, uninformed buy orders, and uninformed sell orders. The estimation is done on a stock-by-stock basis. Stocks with high *PIN* measures have high proportions of informed trades. These stocks have higher returns than stocks with low *PIN* measures.

A difference between our variable and *PIN* is that our variable is constructed using only standard return data and industry classification codes (i.e., no microstructure data). Therefore, it is possible to construct our variable for stocks in many markets around the world. In addition, our variable for stock i uses data from stock i <u>and</u> data from many other stocks. We believe our multi-variate approach explains why both our variable and *PIN* predict future returns when included in the same predictive regression.

H.3 Additional notes

The finding that both our *Proxy E[r]* and the *PIN* measures are significant predictors of cross-sectional return differences is of particular interest. The two measures appear to pick up different effects. Our proxy variable is motivated by a multi-asset model and the prices of many stocks are used in its construction. The *PIN* measure relies on analyzing trades of one stock at a time.

We believe the two measures are capturing complementary aspects of information. Our *Proxy E[r]* measure is based on multi-stock regressions while *PIN* is based only on the trades in stock i.

I Comparisons and contrasts with *FSRV*

I.1 Theoretical comparisons and contrasts

Our paper presents equations from an existing multi-asset equilibrium model. We derive our empirical variable directly from these equations. Durnev, Morck, and Yeung (2004) propose the Firm-Specific Return Variation (*FSRV*) measure but do not offer a specific model.

To be fair, the *FSRV* is grounded in arguments put forth by Roll (1988). In the original article, Roll assumes asset returns are explained by "systematic economic influences" and by "public firm specific news events." He observes low R^2 values from projections of returns on contemporaneous systematic and industry factors. Specific news about a firm decreases its R^2. Roll concludes that the portion of return variance unexplained by systematic risk factors can be partly due to trades by informed investors. He also leaves open the possibility that this effect is caused by an "occasional frenzy unrelated to concrete information"— i.e., noise.

Durnev, Morck, and Yeung (2004) assume that higher firm-specific return movement indicates that informed investors are actively trading. This implies that more information is conveyed by equilibrium prices. Consequently, prices are said to be more informative. Other related papers include Morck, Yeung, and Yu (2000) and Durnev, Morck, Yeung, and Zarowin (2003).

I.2 Empirical comparisons and contrasts

The *FSRV* measure is calculated by first projecting stock i's daily returns over the past twelve months on the returns of the market portfolio and the returns of stock i's three-digit industry portfolio (excluding stock i). When we estimate *FSRV*, we require a stock to have a minimum of 60 days of data.

$$r_{i,k} = \alpha + \beta_m r_{m,k} + \beta_s r_{sic3,k} + \varepsilon_{i,k}$$

$$FSRV \equiv ln\left(\tfrac{1-R^2}{R^2}\right)$$

Note that *FSRV* is estimated with contemporaneous returns as right-hand side variables while our *Proxy E[r]* measure is estimated with lagged prices (normalized) as right-hand side variables.

I.3 Alternative interpretation

There are alternative interpretations of firm-specific return variation. For example, Dasgupta, Gan, and Gao (2010) question whether the R^2 from a market model regression measures the extent to which firm-specific information is reflected in stock prices. The authors provide both a model and empirical evidence to show that the relations between R^2 and the informativeness of stock prices is more complicated than has previously been assumed.

I.4 Additional notes

Confusion between *FSRV* and our variable comes from variants of the words "prices" and "information". For example, Durnev, Morck, and Yeung (2004) use the words "informativeness of prices" to indicate that a stock's price has already incorporated priced information and informed investors are actively trading on firm-specific information. Thus, future movements in the stock's price is likely to come from firm-specific releases of information or from informed traders acting on firm-specific information. Such situations lead to low R^2 measures.

In our paper, investors look to the market-clearing mechanism to glean information that others' have about future dividends. When investors' private information is imprecise/noisy, they must rely heavily on the market-clearing mechanism. In a sense, the market-clearing mechanism (price) is highly informative (relative to private information). Such situations lead to high R^2 measures in our regressions. In conclusion, there is nothing inconsistent in how the two papers interpret their separate R^2 measures.

J Comparisons and contrasts with *Delay1*

J.1 Theoretical comparisons and contrasts

Our paper presents equations from an existing multi-asset equilibrium model. We derive our empirical variable directly from these equations. Hou and Moskowitz (2005) propose the *Delay1* measure but do not offer a specific equilibrium model. Their measure is motivated by strong and convincing economic arguments.

Hou and Moskowitz (2005) propose that frictions cause information to be incorporated into prices with delay. Stocks with greater frictions have higher expected returns. The authors estimate the degree of delay by regressing stock i's returns on contemporaneous and lagged market returns.

J.2 Empirical comparisons and contrasts

The *Delay1* measure of Hou and Moskowitz (2005) is calculated by first estimating two regressions using weekly data. The measure is defined using the ratio of the R^2 measures from the following two regressions. Note these regressions use contemporaneous and lagged market returns, while our measure is based on lagged normalized prices as right-hand side variables.

$$\text{(A)} \quad r_{i,w} = \alpha + \beta_m r_{m,w} + \varepsilon_{i,w}$$

$$\text{(B)} \quad r_{i,w} = \alpha + \beta_m r_{m,w} + \sum_{n=1}^{4} \delta^{(-n)} r_{m,w-n} + \varepsilon_{i,w}$$

$$\textit{Delay1} \equiv 1 - \frac{R^2(A)}{R^2(B)}$$

K Double sort results with *Proxy E[r]* and *PIN*

We use a double sort procedure to again test whether our *Proxy E[r]* measure and *PIN* can explain economically and statistically differences in returns. For each month t, we first sort stocks into quintiles based on their *PIN* measures. We next sort stocks into quintiles based on their *Proxy E[r]* measure. For each of the resulting 25 bins, we report the average return of the portfolio of stocks over month t+1.

Results from the double sort procedure are shown on the next page. We see that our *Proxy E[r]* measure is a significant predictor of returns when stocks are in the 4^{th} or $5^{th}(Hi)$ *PIN* quintile. To see this effect, consider stocks in the 4^{th} *PIN* quintile. When our *Proxy E[r]* measure is "*Lo*", stocks have an average return of 0.0045 the following month. When our *Proxy E[r]* measure is "*Hi*", stocks have an average return of 0.0119 the following month. The difference between the "*Hi*" and "*Lo*" is 0.0074 per month. This value is statistically significant with a 3.09 t-statistic.

If we form a portfolio that buys stocks when *PIN* and *Proxy E[r]* are both "*Hi*" and sells stocks when both are "*Lo*", the difference in returns is 0.0088 per month on average. The relevant portfolios are highlighted in the table. The four factor alpha of this portfolio is 0.0110 with a 3.80 t-statistic using R_m-R_f, *HML*, *SMB*, and *MOM* as factors (on a monthly basis and not reported in the table).

We conclude that *PIN* and *Proxy E[r]* help explain cross-sectional differences in returns. Using both variables together identifies stocks with return differences on the order of 11% per annum.

Appendix K
Economic Significance of *Proxy E[r]* and *PIN*

This table presents double sort results. Each month, we first sort stocks into quintiles by the probability of information-based trading (*PIN*) and then by our *Proxy E[r]* measure. For each of the 25 bins, we record the portfolio return for stocks in the bin. The table shows the average return for each of the 25 bins. Also shown are differences of returns (both across a given row and down a given column).

		Proxy E[r]					*Proxy E[r]* Effect	
		Lo	2	3	4	Hi	Hi – Lo	*(T-stat)*
	Lo	0.0065	0.0075	0.0069	0.0067	0.0058	-0.0007	(-0.44)
	2	0.0067	0.0059	0.0051	0.0056	0.0059	-0.0009	(-0.69)
PIN	3	0.0052	0.0050	0.0051	0.0070	0.0057	0.0005	(0.30)
	4	0.0045	0.0062	0.0071	0.0074	0.0119	0.0074	3.09
	Hi	0.0112	0.0099	0.0087	0.0093	0.0153	0.0041	1.91
	PIN Effect Hi – Lo	0.0047	0.0024	0.0019	0.0026	0.0095		
	(T-stat)	(1.70)	(0.90)	(0.68)	(0.93)	(2.50)		

L Litzenberger and Ramaswamy methodology

We recreate tables in the main text but report Litzenberger and Ramaswamy (1979) precision-weighted means. A variable's monthly precision is calculated based on its OLS standard error from each month's cross-sectional regression. The variable's precision equals one divided by the squared standard error. Monthly coefficient estimates are the same in the following tables as in the main text. The difference is in how we calculate weights (in the time series dimension).

Appendix L – Table 2
Return Regressions Using Individual Stocks + LR Methodology

This table presents time-series average coefficients from cross-sectional regressions of monthly excess stock returns on lagged stock characteristics. We use Litzenberger and Ramaswamy precision-weighted means where a variable's monthly precision is calculated based on its OLS standard error from each month's cross-sectional regression. Data start July 1965 and end December 2005 for a total of 486 months. There are 13,993 ordinary common stocks. *"Proxy E[r]"* is the logistic transformation of the fit (R^2) from a regression of returns on prices and defined in the text. *FSRV* is a measure of firm-specific return variation. *Delay(1)* is a measure of a stock's delayed price reaction. *PIN* is a stock's probability of information-based trading. All coefficients have been multiplied by 100. Estimated regression constants are not shown. T-statistics are shown in parentheses.

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Proxy E[r]	0.22	0.21	0.24	0.21	0.24	0.25
(T-stat)	*(4.61)*	*(4.69)*	*(5.57)*	*(4.83)*	*(4.40)*	*(4.72)*
Beta	-0.22	-0.23	-0.22	-0.23	-0.22	-0.23
(T-stat)	*(-3.22)*	*(-3.36)*	*(-3.56)*	*(-3.45)*	*(-2.09)*	*(-2.39)*
ln (*MktCap*)		0.08	0.11	0.08	0.23	0.20
(T-stat)		*(1.75)*	*(2.25)*	*(1.70)*	*(3.93)*	*(3.03)*
ln (*Book-to-Mkt*)		0.26	0.28	0.26	0.15	0.15
(T-stat)		*(6.45)*	*(6.55)*	*(6.37)*	*(1.88)*	*(1.85)*
FSRV			0.03			-0.07
(T-stat)			*(0.85)*			*(-1.63)*
Delay(1)				0.04		-0.15
(T-stat)				*(0.37)*		*(-1.02)*
PIN					4.33	4.49
(T-stat)					*(6.34)*	*(6.81)*
Adj R^2 (%)	0.90	3.24	3.69	3.42	3.05	3.43
# of Months	486	486	486	486	228	228

Appendix L – Table 3
Economic Significance of Predictor Variables + LR Methodology

This table presents estimates of economic significance. We calculate a predictor variable's economic significance as the difference in returns for stocks one standard deviation above the mean and stocks one standard deviation below the mean (see the 2×σ terms). *"Proxy E[r]"* is the logistic transformation of the fit (R^2) from a regression of returns on prices and defined in the text. *PIN* is a stock's probability of information-based trading.

	(1) Coefficient Estimate from Appendix L Table 2, Reg 5 (#)	**(1')** Unweighted Coefficient Estimate (γ)	**(2)** Average Cross-Sectional Stdev of the Variable (Monthly) (σ)	**(3)** Rough Cross-Sectional Estimate $2 \times \sigma \times \gamma$ (#)	**(4)** Time-Series Average of $2 \times \sigma_t \times \gamma_t$ (#)	**(5)** Annualized Economic Significance (#)
Proxy E[r]	0.24	0.22	0.716	0.32%	0.31%	3.78%
Beta	-0.22	-0.07	0.644	-0.09%	-0.07%	-0.80%
ln (*MktCap*)	0.23	0.15	2.053	0.62%	0.58%	7.20%
ln (*Book-to-Mkt*)	0.15	0.10	0.999	0.20%	0.20%	2.39%
PIN	4.33	2.63	0.080	0.42%	0.37%	4.49%

Appendix L – Table 4
Additional Return Regressions Using Individual Stocks + LR Methodology

This table presents time-series average coefficients from cross-sectional regressions of monthly excess stock returns on lagged stock characteristics. We use Litzenberger and Ramaswamy precision-weighted means where a variable's monthly precision is calculated based on its OLS standard error from each month's cross-sectional regression. Data start July 1965 and end December 2005 for a total of 486 months. There are 13,993 ordinary common stocks, an average of 3,167 stocks per month, and 1,539,436 stock-month observations. *"Proxy E[r]"* is the logistic transformation of the fit (R^2) from a regression of returns on prices and defined in the text. Additional control variables include lagged stock returns from *t-3:t-2*, from *t-6:t-4*, and from *t-12:t-7*. Also included are the standard deviation of a stock's returns, the natural log of turnover, Amihud's illiquidity measure, and the natural log of the reciprocal of price. *PIN* is a stock's probability of information-based trading. All coefficients have been multiplied by 100. Estimated regression constants are not shown. T-statistics are shown in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Proxy E[r]	0.21	0.21	0.22	0.21	0.10	0.10	0.21	0.18
(T-stat)	*(4.90)*	*(5.27)*	*(5.67)*	*(4.73)*	*(3.29)*	*(2.92)*	*(4.86)*	*(3.66)*
Beta	-0.24	-0.24	-0.22	-0.16	-0.07	-0.17	-0.19	-0.10
(T-stat)	*(-3.59)*	*(-3.75)*	*(-3.57)*	*(-2.69)*	*(-1.26)*	*(-2.55)*	*(-3.03)*	*(-1.34)*
ln (*MktCap*)	0.08	0.07	0.06	-0.06	0.07	0.51	-0.01	0.13
(T-stat)	*(1.77)*	*(1.66)*	*(1.52)*	*(-1.94)*	*(1.77)*	*(6.06)*	*(-0.21)*	*(1.47)*
ln (*B-to-M*)	0.25	0.24	0.25	0.22	0.23	0.26	0.25	0.20
(T-stat)	*(6.54)*	*(6.65)*	*(7.41)*	*(6.23)*	*(5.80)*	*(5.46)*	*(6.18)*	*(3.08)*
Ret *t-3* to *t-2*	0.52	0.59	0.57					0.92
(T-stat)	*(2.28)*	*(2.66)*	*(2.58)*					*(3.33)*
Ret *t-6* to *t-4*		1.04	1.03					1.37
(T-stat)		*(6.06)*	*(6.10)*					*(5.80)*
Ret *t-12* to *t-7*			0.63					0.98
(T-stat)			*(5.82)*					*(6.19)*
Std(*Ret*)				-13.26				-11.92
(T-stat)				*(-3.77)*				*(-2.53)*
ln(*Turnover*)					-0.45			-0.34
(T-stat)					*(-5.66)*			*(-2.69)*
Illiquid						0.33		-0.06
(T-stat)						*(7.04)*		*(-0.80)*
ln(*1/P*)							-0.29	0.30
(T-stat)							*(-2.96)*	*(3.18)*
PIN								2.70
(T-stat)								*(4.71)*
Adj R^2	3.76	4.27	4.74	4.72	4.51	3.95	4.34	6.19
# of Months	486	486	486	486	486	486	486	228

Appendix L – Table 5
Return Regressions Using Portfolios of Stocks + LR Methodology

This table presents time-series average coefficients from cross-sectional regressions of monthly excess portfolio returns on lagged characteristics. We use Litzenberger and Ramaswamy precision-weighted means where a variable's monthly precision is calculated based on its OLS standard error from each month's cross-sectional regression. Data start July 1965 and end December 2005 for a total of 486 months. "*Proxy E[r]*" is the logistic transformation of the fit (R^2) from a regression of returns on prices and defined in the text. All coefficients have been multiplied by 100. Estimated regression constants are not shown. T-statistics are shown in parentheses.

	3-Digit SIC Indus. Portfolios		Portfolios Sorted on *Proxy E[r]* and *Beta*		Portfolios Sorted on *Proxy E[r]* and *MktCap*	
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)
Proxy E[r]	0.19	0.19	0.17	0.08	0.24	0.24
(T-stat)	*(2.41)*	*(2.55)*	*(2.87)*	*(1.10)*	*(3.81)*	*(4.05)*
Beta	-0.11	-0.19	-0.08	-0.09	0.04	-0.20
(T-stat)	*(-1.03)*	*(-1.73)*	*(-0.95)*	*(-0.93)*	*(0.16)*	*(-1.10)*
ln (*MktCap*)		0.11		-0.03		0.12
(T-stat)		*(2.39)*		*(-0.56)*		*(2.86)*
ln (*Book-to-Mkt*)		0.19		0.17		0.10
(T-stat)		*(3.73)*		*(3.25)*		*(2.54)*
Adj R^2 (%)	1.65	5.29	10.29	17.56	7.84	23.26
# of Portfolios	450	450	100	100	100	100
# of Months	486	486	486	486	486	486

Appendix L – Table 6
Robustness Checks and Size Results + LR Methodology

This table presents time-series average coefficients from cross-sectional regressions of monthly excess stock returns on lagged stock characteristics. We use Litzenberger and Ramaswamy precision-weighted means where a variable's monthly precision is calculated based on its OLS standard error from each month's cross-sectional regression. Data start July 1965 and end December 2005 for a total of 486 months. There are 13,993 ordinary common stocks, an average of 3,167 stocks per month, and 1,539,436 stock-month observations. "*Proxy E[r]*" is the logistic transformation of the fit (R^2) from a regression of returns on prices and defined in the text. "*Beta*" is an estimate of stock *i*'s beta. All coefficients have been multiplied by 100. Estimated regression constants are not shown. T-statistics are shown in parentheses.

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
	Month Has Min of 10 Trading Days	Data From 1965 – 1985	Data From 1986 – 2005	Deciles 1-3 NYSE Breakpoints	Deciles 4-7 NYSE Breakpoints	Deciles 8-10 NYSE Breakpoints	Stocks From NYSE & Amex	Stocks From Nasdaq
Proxy E[r]	0.22	0.10	0.34	0.31	0.04	-0.05	0.13	0.32
(T-stat)	*(5.10)*	*(1.86)*	*(4.69)*	*(5.80)*	*(0.88)*	*(-1.40)*	*(3.46)*	*(3.95)*
Beta	-0.27	-0.29	-0.20	-0.20	-0.10	-0.13	-0.19	-0.22
(T-stat)	*(-3.83)*	*(-2.86)*	*(-2.16)*	*(-3.23)*	*(-0.98)*	*(-1.04)*	*(-2.44)*	*(-2.91)*
ln (*MktCap*)	0.05	0.04	0.12	-	-	-	0.07	0.04
(T-stat)	*(1.18)*	*(0.57)*	*(1.94)*	-	-	-	*(1.67)*	*(0.70)*
ln (*B-to-M*)	0.30	0.19	0.39	0.27	0.23	0.12	0.16	0.38
(T-stat)	*(6.26)*	*(5.27)*	*(4.69)*	*(6.84)*	*(4.43)*	*(2.37)*	*(4.00)*	*(6.59)*
Avg R^2 (%)	3.56	4.22	2.24	1.16	2.78	4.40	3.56	1.97
# of Months	486	246	240	486	486	486	486	486

M Litzenberger and Ramaswamy + heteroskedastic consistent standard errors

We recreate tables in the paper but report Litzenberger and Ramaswamy precision-weighted means. A variable's monthly precision is calculated based on its heteroskedastic consistent, White (1980), standard error from each month's cross-sectional regression. The variable's precision equals one divided by the squared standard error. Monthly coefficient estimates are the same in the following tables as in the main text. The difference is in how we calculate weights (in the time series dimension).

Appendix M – Table 2
Return Regressions Using Individual Stocks + LR Methodology + Heteroskedastic Consistent Standard Errors

This table presents time-series average coefficients from cross-sectional regressions of monthly excess stock returns on lagged stock characteristics. We use Litzenberger and Ramaswamy precision-weighted means where a variable's monthly precision is calculated based on its heteroskedastic consistent (White) standard error from each month's cross-sectional regression. Data start July 1965 and end December 2005 for a total of 486 months. There are 13,993 ordinary common stocks. *"Proxy E[r]"* is the logistic transformation of the fit (R^2) from a regression of returns on prices and defined in the text. *FSRV* is a measure of firm-specific return variation. *Delay(1)* is a measure of a stock's delayed price reaction. *PIN* is a stock's probability of information-based trading. All coefficients have been multiplied by 100. Estimated regression constants are not shown. T-statistics are shown in parentheses.

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Proxy E[r]	0.19	0.21	0.23	0.21	0.22	0.24
(T-stat)	*(3.92)*	*(4.79)*	*(5.61)*	*(4.98)*	*(4.39)*	*(4.62)*
Beta	-0.24	-0.27	-0.25	-0.26	-0.29	-0.30
(T-stat)	*(-3.50)*	*(-3.82)*	*(-4.08)*	*(-3.95)*	*(-2.70)*	*(-3.04)*
ln (*MktCap*)		0.12	0.16	0.12	0.30	0.26
(T-stat)		*(2.81)*	*(3.20)*	*(2.71)*	*(5.11)*	*(4.06)*
ln (*Book-to-Mkt*)		0.24	0.26	0.23	0.17	0.17
(T-stat)		*(6.49)*	*(6.56)*	*(6.46)*	*(2.19)*	*(2.21)*
FSRV			0.03			-0.08
(T-stat)			*(0.83)*			*(-1.64)*
Delay(1)				0.07		-0.11
(T-stat)				*(0.70)*		*(-0.80)*
PIN					4.12	4.49
(T-stat)					*(6.18)*	*(6.70)*
Adj R^2 (%)	0.90	3.24	3.69	3.42	3.05	3.43
# of Months	486	486	486	486	228	228

Appendix M – Table 3
Economic Significance of Predictor Variables + LR Methodology + Heteroskedastic Consistent Standard Errors

This table presents estimates of economic significance. We calculate a predictor variable's economic significance as the difference in returns for stocks one standard deviation above the mean and stocks one standard deviation below the mean (see the 2×σ terms). *"Proxy E[r]"* is the logistic transformation of the fit (R^2) from a regression of returns on prices and defined in the text. *PIN* is a stock's probability of information-based trading.

	(1) Coefficient Estimate from Appendix M Table 2, Reg 5 (#)	**(1')** Unweighted Coefficient Estimate (γ)	**(2)** Average Cross-Sectional Stdev of the Variable (Monthly) (σ)	**(3)** Rough Cross-Sectional Estimate $2 \times \sigma \times \gamma$ (#)	**(4)** Time-Series Average of $2 \times \sigma_t \times \gamma_t$ (#)	**(5)** Annualized Economic Significance (#)
Proxy E[r]	0.22	0.22	0.716	0.32%	0.31%	3.78%
Beta	-0.29	-0.07	0.644	-0.09%	-0.07%	-0.80%
ln (*MktCap*)	0.30	0.15	2.053	0.62%	0.58%	7.20%
ln (*Book-to-Mkt*)	0.17	0.10	0.999	0.20%	0.20%	2.39%
PIN	4.12	2.63	0.080	0.42%	0.37%	4.49%

Appendix M – Table 4

Additional Return Regressions Using Individual Stocks + LR + Heteroskedastic Consistent Std Errors

This table presents time-series average coefficients from cross-sectional regressions of monthly excess stock returns on lagged stock characteristics. We use Litzenberger and Ramaswamy precision-weighted means where a variable's monthly precision is calculated based on its heteroskedastic consistent (White) standard error from each month's cross-sectional regression. Data start July 1965 and end December 2005 for a total of 486 months. There are 13,993 ordinary common stocks, an average of 3,167 stocks per month, and 1,539,436 stock-month observations. "*Proxy E[r]*" is the logistic transformation of the fit (R^2) from a regression of returns on prices and defined in the text. Additional control variables include lagged stock returns from *t-3:t-2*, from *t-6:t-4*, and from *t-12:t-7*. Also included are the standard deviation of a stock's returns, the natural log of turnover, Amihud's illiquidity measure, and the natural log of the reciprocal of price. *PIN* is a stock's probability of information-based trading. All coefficients have been multiplied by 100. Estimated regression constants are not shown. T-statistics are shown in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Proxy E[r]	0.20	0.21	0.22	0.21	0.10	0.10	0.20	0.20
(T-stat)	*(4.93)*	*(5.31)*	*(5.83)*	*(5.05)*	*(3.36)*	*(3.00)*	*(4.87)*	*(4.25)*
Beta	-0.27	-0.26	-0.24	-0.18	-0.09	-0.21	-0.21	-0.11
(T-stat)	*(-4.03)*	*(-4.06)*	*(-3.90)*	*(-2.94)*	*(-1.61)*	*(-3.00)*	*(-3.29)*	*(-1.45)*
ln (*MktCap*)	0.12	0.11	0.10	-0.07	0.11	0.54	-0.02	0.16
(T-stat)	*(2.80)*	*(2.63)*	*(2.55)*	*(-2.46)*	*(2.73)*	*(6.73)*	*(-0.53)*	*(1.83)*
ln (*B-to-M*)	0.23	0.22	0.23	0.19	0.20	0.24	0.23	0.22
(T-stat)	*(6.52)*	*(6.52)*	*(7.33)*	*(5.99)*	*(5.59)*	*(5.48)*	*(6.30)*	*(3.56)*
Ret *t-3* to *t-2*	0.34	0.40	0.39					0.90
(T-stat)	*(1.45)*	*(1.76)*	*(1.75)*					*(3.30)*
Ret *t-6* to *t-4*		0.87	0.86					1.20
(T-stat)		*(4.90)*	*(4.93)*					*(5.00)*
Ret *t-12* to *t-7*			0.55					0.90
(T-stat)			*(5.19)*					*(5.55)*
Std(*Ret*)				-27.90				-30.05
(T-stat)				*(-7.35)*				*(-7.10)*
ln(*Turnover*)					-0.47			-0.26
(T-stat)					*(-6.00)*			*(-2.09)*
Illiquid						0.33		-0.03
(T-stat)						*(7.29)*		*(-0.42)*
ln(*1/P*)							-0.42	0.35
(T-stat)							*(-4.43)*	*(4.19)*
PIN								2.23
(T-stat)								*(4.08)*
Adj R^2	3.76	4.27	4.74	4.72	4.51	3.95	4.34	6.19
# of Months	486	486	486	486	486	486	486	228

Appendix M – Table 5
Return Regressions Using Portfolios of Stocks + LR Methodology + Heteroskedastic Consistent Standard Errors

This table presents time-series average coefficients from cross-sectional regressions of monthly excess portfolio returns on lagged characteristics. We use Litzenberger and Ramaswamy precision-weighted means where a variable's monthly precision is calculated based on its heteroskedastic consistent (White) standard error from each month's cross-sectional regression. Data start July 1965 and end December 2005 for a total of 486 months. "*Proxy E[r]*" is the logistic transformation of the fit (R^2) from a regression of returns on prices and defined in the text. All coefficients have been multiplied by 100. Estimated regression constants are not shown. T-statistics are shown in parentheses.

	3-Digit SIC Indus. Portfolios		Portfolios Sorted on *Proxy E[r]* and *Beta*		Portfolios Sorted on *Proxy E[r]* and *MktCap*	
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)
Proxy E[r]	0.17	0.20	0.18	0.09	0.27	0.27
(T-stat)	*(2.22)*	*(2.97)*	*(3.25)*	*(1.36)*	*(4.63)*	*(4.87)*
Beta	-0.16	-0.30	-0.11	-0.13	0.10	-0.20
(T-stat)	*(-1.51)*	*(-2.87)*	*(-1.25)*	*(-1.41)*	*(0.45)*	*(-1.13)*
ln (*MktCap*)		0.15		0.01		0.11
(T-stat)		*(3.30)*		*(0.29)*		*(2.63)*
ln (*Book-to-Mkt*)		0.16		0.16		0.08
(T-stat)		*(3.65)*		*(3.50)*		*(2.48)*
Adj R^2 (%)	1.64	5.21	10.58	17.74	8.09	23.66
# of Portfolios	450	450	100	100	100	100
# of Months	486	486	486	486	486	486

Appendix M – Table 6
Robustness Checks and Size Results + LR Methodology + Heteroskedastic Consistent Standard Errors

This table presents time-series average coefficients from cross-sectional regressions of monthly excess stock returns on lagged stock characteristics. We use Litzenberger and Ramaswamy precision-weighted means where a variable's monthly precision is calculated based on its heteroskedastic consistent (White) standard error from each month's cross-sectional regression. Data start July 1965 and end December 2005 for a total of 486 months. There are 13,993 ordinary common stocks, an average of 3,167 stocks per month, and 1,539,436 stock-month observations. *"Proxy E[r]"* is the logistic transformation of the fit (R^2) from a regression of returns on prices and defined in the text. "*Beta*" is an estimate of stock *i*'s beta. All coefficients have been multiplied by 100. Estimated regression constants are not shown. T-statistics are shown in parentheses.

	(1) Month Has Min of 10 Trading Days	**(2)** Data From 1965 – 1985	**(3)** Data From 1986 – 2005	**(4)** Deciles 1-3 NYSE Breakpoints	**(5)** Deciles 4-7 NYSE Breakpoints	**(6)** Deciles 8-10 NYSE Breakpoints	**(7)** Stocks From NYSE & Amex	**(8)** Stocks From Nasdaq
Proxy E[r]	0.22	0.08	0.37	0.29	0.03	-0.05	0.12	0.36
(T-stat)	*(5.10)*	*(1.59)*	*(5.39)*	*(5.55)*	*(0.82)*	*(-1.22)*	*(3.10)*	*(4.69)*
Beta	-0.32	-0.31	-0.24	-0.23	-0.12	-0.18	-0.25	-0.27
(T-stat)	*(-4.42)*	*(-2.80)*	*(-2.63)*	*(-3.63)*	*(-1.19)*	*(-1.48)*	*(-3.13)*	*(-3.54)*
ln (*MktCap*)	0.10	0.07	0.18	-	-	-	0.10	0.13
(T-stat)	*(2.29)*	*(1.14)*	*(3.07)*	-	-	-	*(2.50)*	*(2.13)*
ln (*B-to-M*)	0.27	0.17	0.40	0.23	0.22	0.10	0.16	0.36
(T-stat)	*(6.28)*	*(5.35)*	*(4.88)*	*(6.41)*	*(4.51)*	*(2.26)*	*(4.62)*	*(6.53)*
Avg R^2 (%)	3.56	4.22	2.24	1.16	2.78	4.40	3.56	1.97
# of Months	486	246	240	486	486	486	486	486

N Additional specifications

This table provides a number of robustness checks. All follow the form of Table 2 from the main text. The checks are:

Table 2.a	Including a stock's AR(1) coef and $\|AR(1)\|$ as explanatory variables
Table 2.b	Including a stock's β_{SMB} and β_{HML} as explanatory variables
Table 2.c	Using our raw R_i^2 measure as an explanatory variable
Table 2.d	Using *Proxy E[r]* based on price differences
Table 2.e	Using *Proxy E[r]* based on weekly data
Table 2.f	Applying the Shanken (1992) correction to the standard errors
Table 2.g	Including a stock's β_{SMB} and β_{HML} and reporting unweighted coefficients
Table 2.h	Using our raw R_i^2 and reporting unweighted coefficients

Appendix N – Table 2.a
Using a Stock's *AR(1)* Coefficient and |*AR(1)*| as RHS Variables

This table is similar to Table 2, Regression 2 in the main text except we include two new explanatory variables and report Litzenberger and Ramaswamy precision-weighted means of coefficients. Regression (1) uses a stock's *AR(1)* coefficient as an explanatory variable. Regression (2) uses the absolute value of the *AR(1)* coefficient as an explanatory variable. Regression (3) contains both. A stock's *AR(1)* coefficient is the first-order autocorrelation coefficient between stock returns. It is calculated for each stock *i* and each month *t* by projecting daily returns on the one-day lagged returns. Daily data come from months *t*-12 to *t*-1. We require a minimum of 60 daily returns to compute an *AR(1)* coefficient.

	(1)	**(2)**	**(3)**
Proxy E[r]	0.14	0.15	0.14
(T-stat)	*(3.71)*	*(3.98)*	*(3.77)*
Beta	-0.20	-0.21	-0.2
(T-stat)	*(-3.00)*	*(-3.09)*	*(-2.99)*
ln (*MktCap*)	0.10	0.09	0.10
(T-stat)	*(2.17)*	*(2.04)*	*(2.18)*
ln (*Book-to-Mkt*)	0.25	0.25	0.25
(T-stat)	*(6.43)*	*(6.43)*	*(6.40)*
AR(1)	-0.61		-0.28
(T-stat)	*(-3.04)*		*(-1.54)*
abs(AR(1))		0.80	0.28
(T-stat)		*(3.17)*	*(1.27)*
Adj R^2 (%)	3.48	3.41	3.56
# of Months	486	486	486

Appendix N – Table 2.b
Using a Stock's β_{SMB} and β_{HML} as RHS Variables

This table is similar to Table 2 in the main text except we use a stock's β_{SMB} in place of its log market capitalization and report Litzenberger and Ramaswamy precision-weighted means of coefficients. We also use a stock's β_{HML} in place of its log book-to-market ratio. Both β_{SMB} and β_{HML} are estimated for each stock *i* and each month *t* by regressing stock *i*'s monthly excess returns on the monthly excess returns of the CRSP Value-Weighted market portfolio and the monthly returns of the *SMB* and *HML* portfolios as constructed by Fama and French (1993). The regressions use data from the previous five years (months *t-1* to *t-60*). We require a minimum of 24 months of data for computing β_{SMB} and β_{HML}.

	(1)	(2)	(3)	(4)	(5)	(6)
Proxy E[r]	0.22	0.21	0.25	0.22	0.10	0.19
(T-stat)	*(4.61)*	*(5.14)*	*(7.15)*	*(5.86)*	*(1.86)*	*(3.67)*
Beta	-0.22	-0.37	-0.36	-0.36	-0.36	-0.43
(T-stat)	*(-3.22)*	*(-4.23)*	*(-4.36)*	*(-4.21)*	*(-2.68)*	*(-3.22)*
β_{SMB}		-0.24	-0.25	-0.23	-0.33	-0.29
(T-stat)		*(-3.20)*	*(-3.59)*	*(-3.20)*	*(-3.83)*	*(-3.37)*
β_{HML}		0.27	0.29	0.26	0.27	0.28
(T-stat)		*(4.21)*	*(4.30)*	*(4.10)*	*(2.78)*	*(2.89)*
FSRV			-0.01			-0.16
(T-stat)			*(-0.32)*			*(-3.45)*
Delay(1)				-0.08		-0.33
(T-stat)				*(-0.78)*		*(-2.39)*
PIN					1.83	3.18
(T-stat)					*(2.58)*	*(6.00)*
Adj R^2 (%)	0.90	3.58	4.07	3.81	3.02	3.61
# of Months	486	486	486	486	228	228

Appendix N – Table 2.c
Using Our *Raw* R_i^2 as a RHS Variable

This table is similar to Table 2 in the main text except we do not apply the logistic transformation to regression fits when creating our predictor variable and report Litzenberger and Ramaswamy precision-weighted means of coefficients. The untransformed variable is referred to "*Raw* R_i^2" because it is our R^2 measure from the time series regression of returns on lagged normalized prices.

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Raw R_i^2	3.34	3.17	3.44	3.27	4.67	3.33
(T-stat)	*(5.33)*	*(5.46)*	*(6.29)*	*(5.73)*	*(4.97)*	*(3.56)*
Beta	-0.21	-0.22	-0.21	-0.22	-0.22	-0.24
(T-stat)	*(-3.09)*	*(-3.24)*	*(-3.45)*	*(-3.33)*	*(-2.09)*	*(-2.46)*
ln (*MktCap*)		0.08	0.12	0.08	0.24	0.19
(T-stat)		*(1.85)*	*(2.30)*	*(1.79)*	*(4.09)*	*(3.04)*
ln (*Book-to-Mkt*)		0.26	0.28	0.25	0.15	0.14
(T-stat)		*(6.38)*	*(6.49)*	*(6.30)*	*(1.85)*	*(1.74)*
FSRV			0.03			-0.07
(T-stat)			*(0.84)*			*(-1.61)*
Delay(1)				0.03		-0.13
(T-stat)				*(0.32)*		*(-0.92)*
PIN					4.32	4.43
(T-stat)					*(6.32)*	*(6.69)*
Adj R^2 (%)	0.94	3.27	3.72	3.44	3.12	3.48
# of Months	486	486	486	486	228	228

Appendix N – Table 2.d
Using *Proxy E[r]* based on Price Differences as a RHS Variable

This table is similar to Table 2 in the main text except we estimate our proxy variable with price differences (as opposed to using returns) and report Litzenberger and Ramaswamy precision-weighted means of coefficients. Specifically, the left-hand side variable in the Stage A regression (Step 5, Equation 5 in the main text) has been replaced with the difference of (actual) prices as opposed to daily CRSP returns. Our R^2 measure comes from the time series regression of these differences on lagged normalized prices. The *Proxy E[r]* is defined as the logit transformation of our R^2 measure.

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Proxy E[r]	0.14	0.13	0.14	0.13	0.14	0.15
(T-stat)	*(2.75)*	*(3.15)*	*(3.67)*	*(3.17)*	*(2.73)*	*(2.90)*
Beta	-0.24	-0.24	-0.22	-0.23	-0.22	-0.24
(T-stat)	*(-3.43)*	*(-3.32)*	*(-3.58)*	*(-3.36)*	*(-2.09)*	*(-2.45)*
ln (*MktCap*)		0.07	0.10	0.07	0.22	0.19
(T-stat)		*(1.67)*	*(2.08)*	*(1.65)*	*(3.81)*	*(2.92)*
ln (*Book-to-Mkt*)		0.29	0.29	0.28	0.15	0.14
(T-stat)		*(6.78)*	*(6.71)*	*(6.75)*	*(1.82)*	*(1.78)*
FSRV			0.04			-0.07
(T-stat)			*(1.04)*			*(-1.56)*
Delay(1)				0.07		-0.13
(T-stat)				*(0.65)*		*(-0.91)*
PIN					4.30	4.44
(T-stat)					*(6.27)*	*(6.70)*
Adj R^2 (%)	0.96	3.30	3.67	3.49	3.04	3.40
# of Months	486	486	486	486	228	228

Appendix N – Table 2.e
Using *Proxy E[r]* based on Weekly Data as a RHS Variable

This table is similar to Table 2 in the main text except that our *Proxy E[r]* variable is based on a regression fit that uses weekly data. Our *Proxy E[r]* uses up to two years of weekly normalized prices. We require a minimum of 60 weeks to perform the regression. We report Litzenberger-Ramaswamy precision-weighted mean coefficients.

	(1)	(2)	(3)	(4)	(5)	(6)
Proxy E[r]	0.05	0.05	0.04	0.04	0.04	0.04
(T-stat)	*(3.40)*	*(2.88)*	*(2.64)*	*(2.86)*	*(1.69)*	*(1.78)*
Beta	-0.26	-0.25	-0.23	-0.25	-0.27	-0.28
(T-stat)	*(-3.55)*	*(-3.40)*	*(-3.61)*	*(-3.51)*	*(-2.54)*	*(-2.88)*
ln (MktCap)		0.06	0.11	0.07	0.21	0.18
(T-stat)		*(1.46)*	*(2.16)*	*(1.50)*	*(3.60)*	*(2.87)*
ln (Book-to-Mkt)		0.30	0.31	0.29	0.21	0.21
(T-stat)		*(6.70)*	*(6.84)*	*(6.65)*	*(2.70)*	*(2.71)*
FSRV			0.07			-0.05
(T-stat)			*(1.85)*			*(-1.08)*
Delay(1)				0.10		-0.11
(T-stat)				*(0.95)*		*(-0.79)*
PIN					4.34	4.41
(T-stat)					*(6.39)*	*(6.69)*
Adj R^2 (%)	0.71	3.26	3.66	3.45	2.89	3.23
# of Months	486	486	486	486	228	228

Appendix N – Table 2f
Return Regressions Using Individual Stocks and Shanken (1992) Standard Errors

This table is similar to Appendix L - Table 2 except t-statistics, shown in parentheses, are based on Skanken (1992) corrected time-series standard deviations of coefficient estimates.

	(1)	(2)	(3)	(4)	(5)	(6)
Proxy E[r]	0.22	0.21	0.24	0.21	0.24	0.25
(T-stat)	*(3.13)*	*(3.09)*	*(3.78)*	*(3.22)*	*(3.02)*	*(3.13)*
Beta	-0.22	-0.23	-0.22	-0.23	-0.22	-0.23
(T-stat)	*(-2.19)*	*(-2.22)*	*(-2.42)*	*(-2.3)*	*(-1.43)*	*(-1.59)*
ln (*MktCap*)		0.08	0.11	0.08	0.23	0.20
(T-stat)		*(1.15)*	*(1.53)*	*(1.14)*	*(2.69)*	*(2.01)*
ln (*Book-to-Mkt*)		0.26	0.28	0.26	0.15	0.15
(T-stat)		*(4.25)*	*(4.45)*	*(4.25)*	*(1.29)*	*(1.23)*
FSRV			0.03			-0.07
(T-stat)			*(0.58)*			*(-1.08)*
Delay(1)				0.04		-0.15
(T-stat)				*(0.25)*		*(-0.68)*
PIN					4.33	4.49
(T-stat)					*(4.34)*	*(4.52)*
Adj R^2 (%)	0.90	3.24	3.69	3.42	3.05	3.43
# of Months	486	486	486	486	228	228

Appendix N – Table 2.g
Using a Stock's β_{SMB} and β_{HML} as RHS Variables and Standard Fama-MacBeth Coefficients

This table is similar to Table 2 in the main text except it uses β_{SMB} and β_{HML} as RHS variables. The table reports standard Fama-MacBeth coefficients (i.e., equally-weighted as opposed to precision-weighted coefficients.)

	(2)	(3)	(4)	(5)	(6)
Proxy E[r]	0.17	0.20	0.18	0.13	0.19
(T-stat)	*(3.70)*	*(5.55)*	*(4.36)*	*(2.29)*	*(3.63)*
Beta	-0.07	-0.08	-0.07	-0.12	-0.21
(T-stat)	*(-0.74)*	*(-0.83)*	*(-0.69)*	*(-0.75)*	*(-1.32)*
β_{SMB}	0.02	0.01	0.03	-0.24	-0.18
(T-stat)	*(0.22)*	*(0.17)*	*(0.30)*	*(-2.22)*	*(-1.72)*
β_{HML}	0.16	0.17	0.16	0.15	0.18
(T-stat)	*(2.28)*	*(2.36)*	*(2.29)*	*(1.26)*	*(1.52)*
FSRV		-0.01			-0.16
(T-stat)		*(-0.14)*			*(-3.09)*
Delay(1)			-0.08		-0.30
(T-stat)			*(-0.65)*		*(-2.11)*
PIN				1.28	2.72
(T-stat)				*(1.68)*	*(5.00)*
Adj R^2 (%)	3.80	4.33	4.06	3.20	3.82
# of Months	486	486	486	228	228

Appendix N – Table 2.h
Using Our *Raw* R_i^2 as a RHS Variable and Standard Fama-MacBeth Coefficients

This table is similar to Appendix N – Table 2.c. We do not apply the logistic transformation to regression fits when creating our predictor variable. This table reports standard Fama-MacBeth coefficients (i.e., equally-weighted as opposed to precision-weighted coefficients.)

	(1)	(2)	(3)	(4)	(5)	(6)
Raw R_i^2	2.99	2.08	2.45	2.15	4.08	4.15
(T-stat)	*(3.81)*	*(2.96)*	*(3.83)*	*(3.14)*	*(4.44)*	*(4.50)*
Beta	-0.03	0.01	-0.04	0.00	-0.07	-0.14
(T-stat)	*(-0.37)*	*(0.17)*	*(-0.63)*	*(-0.03)*	*(-0.62)*	*(-1.24)*
ln (*MktCap*)		-0.03	-0.05	-0.04	0.15	0.09
(T-stat)		*(-0.52)*	*(-0.91)*	*(-0.82)*	*(2.23)*	*(1.17)*
ln (*Book-to-Mkt*)		0.24	0.25	0.24	0.10	0.11
(T-stat)		*(4.56)*	*(4.84)*	*(4.49)*	*(1.19)*	*(1.34)*
FSRV			-0.08			-0.14
(T-stat)			*(-1.75)*			*(-2.84)*
Delay(1)				-0.19		-0.21
(T-stat)				*(-1.73)*		*(-1.43)*
PIN					2.63	2.92
(T-stat)					*(3.37)*	*(3.92)*
Adj R^2 (%)	1.02	3.44	3.91	3.63	3.20	3.61
# of Months	486	486	486	486	228	228

O Robustness tests along two dimensions

This table provides a number of robustness checks along two dimensions. All follow the form of Table 2 from the main text. The checks are:

Table 2.(c+d)	Using our raw R_i^2 based on price differences
Table 2.(c+e)	Using our raw R_i^2 based on weekly data

Appendix O – Table 2.(c+d)
Using Our *Raw* R_i^2 based on Price Differences as a RHS Variable

This table combines elements of Appendix N – Table 2.c and Table 2.d. We do not apply the logistic transformation to regression fits when creating our variable. The untransformed variable is referred to "*Raw* R_i^2" and is our R^2 measure from the time series regression of price differences on lagged normalized prices. We estimate our proxy variable using price differences as opposed to using returns.

	(1)	(2)	(3)	(4)	(5)	(6)
Raw R_i^2	2.41	2.32	2.56	2.34	3.03	3.33
(T-stat)	*(3.81)*	*(4.47)*	*(5.09)*	*(4.57)*	*(3.37)*	*(3.56)*
Beta	-0.23	-0.23	-0.22	-0.22	-0.22	-0.24
(T-stat)	*(-3.32)*	*(-3.22)*	*(-3.47)*	*(-3.26)*	*(-2.09)*	*(-2.46)*
ln (*MktCap*)		0.08	0.11	0.08	0.23	0.19
(T-stat)		*(1.82)*	*(2.18)*	*(1.77)*	*(3.98)*	*(3.04)*
ln (*Book-to-Mkt*)		0.28	0.29	0.28	0.14	0.14
(T-stat)		*(6.69)*	*(6.63)*	*(6.65)*	*(1.78)*	*(1.74)*
FSRV			0.04			-0.07
(T-stat)			*(1.01)*			*(-1.61)*
Delay(1)				0.07		-0.13
(T-stat)				*(0.61)*		*(-0.92)*
PIN					4.29	4.43
(T-stat)					*(6.28)*	*(6.69)*
Adj R^2 (%)	0.99	3.32	3.70	3.51	3.11	3.48
# of Months	486	486	486	486	228	228

Appendix O – Table 2.(c+e)
Using Our *Raw* R_i^2 based on Weekly Data as a RHS Variable

This table combines elements of Appendix N – Table 2.c and Table 2.e. We do not apply the logistic transformation to regression fits when creating our variable. The untransformed variable is referred to "*Raw* R_i^2" because it is our R^2 measure from the time series regression of returns on lagged normalized prices. Weekly data are used in the regressions that create our variable.

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Raw R_i^2	0.97	0.82	0.81	0.80	1.15	1.26
(T-stat)	*(3.88)*	*(3.40)*	*(3.45)*	*(3.40)*	*(3.12)*	*(3.36)*
Beta	-0.25	-0.25	-0.23	-0.24	-0.27	-0.29
(T-stat)	*(-3.50)*	*(-3.37)*	*(-3.58)*	*(-3.48)*	*(-2.54)*	*(-2.89)*
ln (MktCap)		0.06	0.11	0.07	0.21	0.19
(T-stat)		*(1.52)*	*(2.20)*	*(1.55)*	*(3.68)*	*(2.94)*
ln (Book-to-Mkt)		0.30	0.31	0.29	0.21	0.21
(T-stat)		*(6.70)*	*(6.84)*	*(6.65)*	*(2.71)*	*(2.72)*
FSRV			0.07			-0.05
(T-stat)			*(1.83)*			*(-1.11)*
Delay(1)				0.10		-0.12
(T-stat)				*(0.93)*		*(-0.83)*
PIN					4.34	4.41
(T-stat)					*(6.40)*	*(6.70)*
Adj R^2 (%)	0.74	3.29	3.68	3.48	2.91	3.24
# of Months	486	486	486	486	228	228

P Additional overview statistics

We provide a three sets of overview statistics:

i. Presenting stock characteristics after sorting by our *Proxy E[r]* variable

ii. Presenting stock characteristics after sorting by SIC 1 code

iii. Drawing histograms of R_i^2 and *Proxy E[r]*

Appendix P-*i*
Stock Characteristics by Deciles of *Proxy E[r]*

This table provides overview statistics of some data used in this paper. We first sort stocks into deciles based on our *Proxy E[r]* measure. We then report time series averages of cross-sectional averages by decile. "*Proxy E[r]*" is the logistic transformation of the fit (R^2) from a regression of returns on prices and defined in the text. *PIN* is a stock's probability of information-based trading. *FSRV* is a measure of firm-specific return variation. Also included are standard deviation of excess returns, the natural log of turnover, and Amihud's illiquidity measure.

Proxy E[r]	**ln (*MktCap*)**	**ln (*B-to-M*)**	***PIN***	***FSRV***	**Std ($R_i – R_f$)**	**ln (*Turnover*)**	***Illiquid***
1 (low)	12.605	-0.646	0.185	2.160	0.031	-6.005	2.408
2	12.309	-0.594	0.192	2.328	0.033	-6.040	2.551
3	12.106	-0.555	0.197	2.439	0.034	-6.078	2.533
4	11.917	-0.514	0.201	2.549	0.035	-6.121	3.207
5	11.717	-0.464	0.206	2.680	0.036	-6.184	3.734
6	11.523	-0.410	0.212	2.802	0.037	-6.242	5.223
7	11.279	-0.352	0.219	2.955	0.039	-6.307	6.263
8	11.014	-0.267	0.225	3.146	0.042	-6.399	8.915
9	10.618	-0.139	0.239	3.433	0.047	-6.549	15.564
10 (High)	10.016	0.098	0.264	3.865	0.059	-6.900	38.593

Appendix P-*ii*
Stock Characteristics by SIC 1

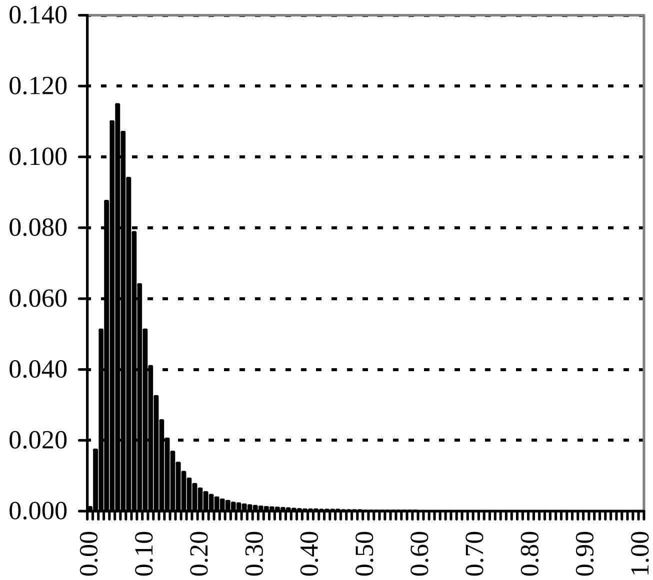
This table provides overview statistics of some data used in this paper. We first sort stocks into deciles based on SIC codes at the one-digit level. We then report time series averages of cross-sectional averages by decile. "*Proxy E[r]*" is the logistic transformation of the fit (R^2) from a regression of returns on prices and defined in the text. *PIN* is a stock's probability of information-based trading. *FSRV* is a measure of firm-specific return variation. Also included are standard deviation of excess returns, the natural log of turnover, and Amihud's illiquidity measure.

SIC 1	*Proxy E[r]*	**ln (*MktCap*)**	**ln (*B-to-M*)**	*PIN*	*FSRV*	**Std ($R_i - R_f$)**	**ln (*Turnover*)**	*Illiquid*
1	-2.808	10.298	0.171	0.256	3.220	0.041	-7.081	9.866
2	-2.573	11.275	-0.473	0.215	2.836	0.048	-6.313	14.688
3	-2.715	11.849	-0.502	0.201	2.681	0.035	-6.419	6.175
4	-2.655	11.194	-0.376	0.217	2.769	0.042	-6.143	8.474
5	-2.642	12.479	-0.300	0.172	2.599	0.027	-6.498	2.508
6	-2.650	11.252	-0.303	0.216	3.021	0.042	-6.200	10.085
7	-2.501	11.725	-0.164	0.206	3.027	0.033	-6.609	8.561
8	-2.704	11.355	-0.735	0.210	2.922	0.051	-5.763	12.217
9	-2.719	11.189	-0.748	0.218	3.250	0.052	-5.755	12.074
10	-2.718	10.846	-0.896	0.237	3.898	0.054	-6.242	9.132

Appendix P-*iii*
Histograms

We show the distribution of our predictor variable. The figure on the left (Panel A) shows the distribution of the regression fits (our R^2). The figure on the right shows the distribution of *Proxy E[r]* which is defined as the logistic transformation of the regression fits.

Panel A: Our R^2



Panel B: *Proxy E[r]*