www.cerag.org

Centre d'Etudes et de Recherches Appliquées à la Gestion_ U.M.R. C.N.R.S. 5820

CAHIER DE RECHERCHE n°2009-05 E2

Risk and the Cross-Section of Stock Returns.

BURLACU Radu, FONTAINE Patrice,

JIMENEZ-GARCES Sonia, SEASHOLES Mark S.

Unité Mixte de Recherche CNRS / Université Pierre Mendès France Grenoble 2
150 rue de la Chimie – BP 47 – 38040 GRENOBLE cedex 9
Tél. : 04 76 63 53 81 Fax : 04 76 54 60 68

Sciences sociales & humaines

Risk and the Cross-Section of Stock Returns

Radu Burlacu
Sonia Jimenez-Garcès

Patrice Fontaine
Mark S. Seasholes

This Version: 27-Oct-2008*

Abstract

We derive a proxy for expected returns from a noisy multi-asset rational expectations equilibrium model. A goal/contribution of this paper is to use the same proxy variable for theoretical, numerical, and empirical analyses. Economically, our "*Proxy E[r]*" variable helps differentiate between stocks with precise private information (and low expected returns) and stocks with imprecise/noisy information (and high expected returns). Empirically, the proxy variable is easily estimated with recent stock returns. Using CRSP data starting in 1965, we show stocks with a *Proxy E[r]* measure one standard deviation above and below the average have returns that differ by 0.36% the following month (4.44% per annum). The results are statistically significant at the 1%-level. Our results remain economically and statistically significant after controlling for variables such as stock market capitalization, book-to-market ratio, liquidity, and the probability of information-based trading.

JEL Classification: D8, G1, G10

Keywords: Risk Premia, Cross-Sectional Asset Pricing, REE Models

*We thank Flavio Bazzana, Denis Gromb, Soeren Hvidkjaer, and Joel Peress for helpful comments as well as seminar participants at City University of Hong Kong, HKUST, NTU Singapore, Toulouse University, Universite Catholique de Louvain, Univerisity of Paris Dauphine, Econometric Society 2009 North American Summer Meetings, and AFFI Lyon 2008 Conference. Burlacu is at the University of Nancy and Eurofidai. Fontaine are at the University Pierre Mendès-France and Eurofidai. Jimenez-Garcès is at the University of Lyon and Eurofidai. Seasholes is at Hong Kong University of Science and Technology. Questions and comments can be emailed to "Sonia.Jimenez@wanadoo.fr" or to "Mark.Seasholes@emailias.com".

Risk and the Cross-Section of Stock Returns

This Version: 27-Oct-2008*

Abstract

We derive a proxy for expected returns from a noisy multi-asset rational expectations equilibrium model. A goal/contribution of this paper is to use the same proxy variable for theoretical, numerical, and empirical analyses. Economically, our "*Proxy E[r]*" variable helps differentiate between stocks with precise private information (and low expected returns) and stocks with imprecise/noisy information (and high expected returns). Empirically, the proxy variable is easily estimated with recent stock returns. Using CRSP data starting in 1965, we show stocks with a *Proxy E[r]* measure one standard deviation above and below the average have returns that differ by 0.36% the following month (4.44% per annum). The results are statistically significant at the 1%-level. Our results remain economically and statistically significant after controlling for variables such as stock market capitalization, book-to-market ratio, liquidity, and the probability of information-based trading.

JEL Classification: D8, G1, G10

Keywords: Risk Premia, Cross-Sectional Asset Pricing, REE Models

*

1 Introduction

Why do some stocks have high average returns while others have low average returns? Answering this question fuels much debate and research in the field of financial economics. Many argue that high (or low) returns are compensation for bearing high (or low) levels of risk. Empirically, however, there is a failure to link theoretical risk measures such as the Capital Asset Pricing Model (CAPM) beta to average returns. Does the failure stem from incorrectly specifying risks—i.e., is the CAPM the correct model? Or, as others argue, do stocks temporarily experience short-horizon and medium-horizon mispricings that overwhelm predictions of risk-based models? Part of the debate surrounding average returns stems from each side claiming the same variable supports its case. For example, a stock's book-to-market ratio can be viewed as both a risk proxy and as a measure of mispricing.

The overall contribution of this paper rests in our integrated approach to studying average returns. We derive a single variable that is correlated with expected returns and can be used in theoretical, numerical, and empirical analyses. We start with a multi-asset equilibrium model of a market with a large number of investors who possess diverse and asymmetric pieces of private information.[1] We next derive a proxy variable that is positively related to a stock's expected return. This positive relation can be shown with both closed-form equations and numerical analysis. The model and derivation explicitly state the economic quantities and assumptions behind the risk/return relations documented in this paper. Our proxy variable is easily estimated with recently observed stock returns. We end by showing that our proxy variable is an economically and statistically significant predictor of stock return dispersion. The empirical results are based on standard return data from the Center for Research in Security Prices (or "CRSP").

Information, uncertainty, and portfolio concerns determine an investor's demand for a given asset. Investors with precise information about an asset's future dividends tend to have high demands for the asset. The investors may, however, refrain from investing too heavily in the asset due to portfolio and diversification considerations. Conversely, investors with imprecise information about the asset's dividends tend to have low demands. These investors may be willing to increase their demands for the asset if they are able to glean dividend information from market-clearing (observable) prices.

[1]The theoretical analysis in our paper starts with the Admati (1985) model. As in the original paper, we assume investors' information is unbiased on average. In a large market, the assumption implies prices are a function of information precision but not realizations. Other frameworks, such as Grossman and Stiglitz (1980), give similar results. The Admati (1985) model has one period and two dates. One of our goals is to push this framework to derive a *single* predictor variable for use in theoretical, numerical, and empirical analyses. This paper does not attempt to model a multi-period equilibrium and such work is left for future research.

In a noisy market (e.g., one with supply shocks and/or noise trading), equilibrium prices partially reveal private information.[2] Stock prices are typically below expected future values and these price discounts represent premia investors expect to earn for holding risky assets. The premia are complicated functions of quantities such as the dividends' covariance structure, precision of investors' information, and supply uncertainty. On average, stocks with little private information (low precisions) and/or high degrees of supply uncertainty have low prices and high expected returns. In other words, such stocks are viewed as risky by investors who must be compensated for holding them.

Testing the insights of a noisy rational expectations equilibrium (REE) model can be challenging. The precisions of investors' information and the degrees of supply uncertainty are difficult, if not impossible, to observe. Equally important, equilibrium prices are complicated functions of information and noise variables. The effects of information and noise on expected returns cannot easily be disentangled.[3] Identifying non-informational events such as S&P 500 additions is one approach to studying the effect that a supply shock (noise) has on stock prices. Unfortunately, the event study methodology is better suited to measuring temporary price movements than giving insights into expected returns. As an alternative approach, some financial economists claim there exist certain variables (such as the number of analysts following a stock) that are correlated with information precisions but are uncorrelated with supply uncertainties. The fact that information precisions and supply uncertainties are unobservable makes testing the veracity of such claims impossible.

The first specific contribution of this paper is to derive a proxy for expected returns that is straightforward to calculate and requires only recent return data and industry classification codes. Our proxy variable is defined as the logistic transformation of the fit (time-series R^2) from a multi-variate, time-series regression of stock i's return on its own price and the prices of other stocks in the market.[4] Our proxy variable is negatively

[2] In noisy rational expectations equilibrium models the terms "supply shocks" and "noise trading" are often used interchangeably as they refer to the same quantities. We use the term "supply uncertainty" to denote the variance of the supply shocks.

[3] Expressions showing the interdependence of information and noise variables, as well as derivations are in Section 2 of this paper. In the absence of noise in the market, Grossman (1976) and Radner (1979) propose fully revealing REE models. The Grossman and Stiglitz (1980) paradox points out that such an equilibrium is not possible if information is costly. These authors show, however, that private information is not completely revealed in prices if there is noise trading.

[4] We acknowledge the regression may initially appear non-standard. Theoretically, our model has no impediments to calculating a time-series R^2. Empirically, we use normalized prices to address econometric issues. Please see Section 3 for more descriptions. Note the basic data requirements imply our proxy variable can be calculated using stock market data from around the world—an added benefit. This paper focuses on CRSP data. Finally, our use of a time-series R^2 should not be confused with R^2 measures used in other papers. The literature review at the end of this section and Appendix D highlight differences.

correlated with the precision of investors' information. We show such a correlation implies our proxy variable is positively correlated with expected returns. Therefore, we refer to the variable as "*Proxy E[r]*". At the same time, our proxy variable is positively correlated with the degree of supply uncertainty (and thus positively correlated with expected returns). We are therefore able to use *Proxy E[r]* to explain the observed dispersion in monthly stock returns.

To understand the economics behind our proxy variable, it is helpful to first consider the standard Sharpe-Lintner-Mossin capital asset pricing model (CAPM). In this one-period, two-date 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. In the standard CAPM framework, the covariance between a stock's price and its return is zero due to the fact that prices are deterministic—i.e., the time-series $R^2 = 0$ from a time-series regression of a stock's return on its own price.[5]

In a noisy rational expectations equilibrium framework, future dividends are random as in the standard CAPM. Importantly, a stock's price is a random variable because it is 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 cases, $R^2 \neq 0$ for a time-series regression of stock i's return on its price.

To gain more intuition about our proxy 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.[6] When the private information signals are imprecise and noisy, today's price is 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 (time-series R^2).

[5]The zero covariance (in the CAPM time-series) should not be confused with the cross-sectional relations (where there is a negative relation between stocks' expected returns and prices.) This negative cross-sectional relation exists regardless of whether prices are deterministic (as in the CAPM) or random (as in noisy REE models). Berk (1995) discusses that financial economists *should expect* to find such cross-sectional relations.

[6]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, the term "price informedness" is problematic and may be avoided when comparing our paper with non-REE papers.

When investors receive precise information about future dividends, today's price is high, the risk premium is low, and the market-clearing mechanism conveys relatively little information to investors.[7] 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 (time-series $R^2 \rightarrow 0$).

The reasoning in the above two paragraphs represents the key economic insight of this paper. When private information is imprecise, expected returns tend to be high, and the market-clearing mechanism conveys relatively large amounts of information to investors. The result is a strong (negative) time-series relation between a stock's return and its own price—which financial econometricians can estimate using the fit from a time-series regression. The same time-series relation does not exist when prices are deterministic as in a traditional CAPM model. In effect, our paper is asking how much information about expected returns is conveyed by the market-clearing mechanism? Estimation of our proxy variable identifies stocks in which the market-clearing mechanism conveys relatively more (or less) information about expected returns. This happens when private information is imprecise (or precise) and/or when supply uncertainty is high (or low). During such times, expected returns are also high (or low).[8]

The second specific contribution of this paper comes from studying a multi-asset equilibrium model. The price of a given stock i reflects investors' information about the dividends of *all* stocks in the market place. In other words, due to various correlation structures, investors may learn about stock i's dividend by receiving information about other stocks' dividends. As an example, investors with precise information about stock i's dividends may have relatively low demands for stock i because: 1) The same investors may have more precise information about another stock that is a close substitute of i; 2) Investors may prefer other stocks for portfolio diversification reasons; 3) There may be a high degree of supply uncertainty associated with stock i; or 4) There may be high degrees of supply uncertainty in stocks that serve as hedges. The multivariate approach

[7] The example in this paragraph is focused on the precision of the private signals. The model also includes supply uncertainty. Thus, readers can think about the precision of signal-to-noise.

[8] To further understand our paper, we can make a loose comparison: The CAPM produces stocks' betas. Sorting stocks by their betas should allow an econometrician to predict cross-sectional dispersion in returns. Our paper produces *Proxy E[r]*. Sorting stocks by *Proxy E[r]* can be used to predict cross-sectional dispersion in returns. Empirically, our variable does a much better job predicting return dispersion than the CAPM beta does at predicting dispersion.

sets our work apart from papers that study information asymmetries for a single stock at a time.

In a multi-asset REE framework, the economic intuition discussed earlier is augmented by the ability of investors to glean information about stock i's dividends from the market clearing mechanisms of other stocks. Equilibrium prices are functions of information and uncertainty about each stock's dividend as well as supply uncertainties of all stocks in the market. The correlation structures within and between these variables is, of course, important. Imprecise information about stock i can be overcome if investors have precise information about the dividends of a close substitute (stock j). In such a case, stock i's expected return is low and this is reflected in the relatively weak (negative) time-series relation between j's price and i's return.

We use a multi-variate regression of stock i's returns on the prices of all stocks in order to assess the strength of multiple return–price relations. In the theoretical derivation section of this paper, we "regress" stock i's returns on the prices of stocks i, j, k, etc. Measuring the fit from the regression allows us to represent the strength of the relations in a single number. Both the derivation of our proxy variable and an analysis of computer generated data confirm that the regression fit is positively correlated with stock i's expected returns.

An additional economic insight comes from understanding the economic friction modeled in this paper. Private information cannot be traded. Investors worry about the part of stocks' dividends that cannot be diversified away (as in the CAPM.) Investors *also* worry about private information that is not fully incorporated into prices. Of particular importance is how the private information is correlated with the dividends. In other words, having imprecise information about an asset that can easily be hedged is different from having imprecise information about an asset with no close substitutes. Finally, investors also care about supply uncertainties and how it is correlated with the dividends and information.

The third contribution of this paper is empirical. We first show how to calculate *Proxy E[r]* with recent stock returns and industry classification codes. We then show that our *Proxy E[r]* variable is an economically and statistically significant predictor of cross-sectional average returns. Stocks with *Proxy E[r]* one standard deviation above and below the average have returns that differ by 0.36% the following month. The difference equals 4.44% per annum and is statistically significant at the 1%-level.

The fourth contribution of our paper is to show that *Proxy E[r]* remains an economically and statistically significant predictor of future returns after including additional predictor variables such as an estimate of a stock's beta, market capitalization, and book-to-market ratio. Results also remain significant after including the firm-specific risk variation (*FSRV*) measure of Durnev, Morck, and Yeung (2004), the delay measure of Hou and Moskowitz (2005), and the probability of information-based trading (*PIN*) measure of Easley, Hvidkjaer, and O'Hara (2002). 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 end the paper by showing our empirical results are robust to including lagged stock returns, the standard deviation of returns, turnover measures, the Amihud (2002) illiquidity measure, and the reciprocal of price ($1/P$). We also estimate predictive regressions using portfolios of stocks. The portfolios are formed by sorting stocks into industry groups, *Proxy E[r]* deciles, stock beta deciles, and stock market capitalization deciles. As with the individual stock results, *Proxy E[r]* continues to predict significant cross-sectional differences in returns. We find the ability of *Proxy E[r]* to predict cross-sectional return differences is concentrated in the bottom three size deciles of stocks using NYSE breakpoints. This result compares favorably to existing cross-sectional pricing measures that typically work best for the smallest stocks.

Related Literature

Our paper is related to both recent empirical and theoretical work. On the empirical side, Easley, Hvidkjaer, and O'Hara (2002) study the probability of information-based trading. The authors create a variable to measure this probability called "*PIN*." 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.

Like Easley, Hvidkjaer, and O'Hara (2002), our model is also motivated by the relation between informed investors, asset prices, and returns. A difference between the papers is that our *Proxy E[r]* variable is constructed using only return data and industry classification codes. Therefore, it is possible to construct *Proxy E[r]* for stocks in many markets around the world. In addition, *Proxy E[r]* for stock i uses data from stock i and data from other stocks in the same industry as i. We believe our multi-variate approach explains why both *Proxy E[r]* and *PIN* predict future returns when included in the same regression.

Durnev, Morck, and Yeung (2004) extend arguments put forth in Roll (1988) and propose a measure of firm-specific risk variation ($FSRV$).[9] The authors 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 more informative.

A recent paper by Hou and Moskowitz (2005) proposes 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. A difference between our paper and the earlier work is that our *Proxy E[r]* variable is derived directly from a multi-asset equilibrium model. Including the delay measure in cross-sectional regressions does not affect our results.

Kumar, Sorescu, Boehme, and Danielsen (2008) study information and returns. Their paper is both different from, and complementary to, ours. Their model is not inspired by REE models (like Admati (1985), Grossman and Stiglitz (1980), Brennan and Cao (1997), etc.) Instead, the authors focus on parameter uncertainty. Investors do not know a stock's average return nor do they know the variance of its returns. The author propose a new CAPM with two risk premiums. There is an "intrinsic systematic risk premium" and a premium due to "estimation risk". The paper carries out a cross sectional analysis of stock returns with proxies for "the innovation in market volatility" (appraising estimation risk). The paper does not propose a new proxy variable for expected returns as we do with *Proxy E[r]*.

On the theory side, our *Proxy E[r]* variable is motivated by a recent paper by Biais, Bossaerts, and Spatt (2009). The authors analyze a partially revealing dynamic rational expectations model. They empirically investigate price-contingent portfolio strategies as their model predicts prices contain some of the informed investors' information. In fact, their portfolio allocation is based on"projections of a month's returns onto the vector of relative prices at the beginning of the month." We, too, project returns on prices. Thus, the papers provide complementary insights about the role of equilibrium prices.

Finally, Easley and O'Hara (2004) 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

[9]Roll (1988) shows 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.

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. As mentioned in Footnote 1, our paper could obtain similar results using a Grossman and Stiglitz (1980) framework.

The paper proceeds as follows. Section 2 derives our proxy variable from a multi-asset rational expectations equilibrium model. Section 3 describes how to empirically esimate *Proxy E[r]*. We also describe the data used in the paper. Section 4 presents the empirical results. We provide numerous robustness checks. Section 5 concludes.

2 Theoretical Derivation

We derive a cross-sectional asset pricing measure from the Admati (1985) noisy rational expectations model. A key feature of many noisy REE models is that investors receive private signals about assets' future dividends. Investors can also observe market-clearing prices and are thus able to (imperfectly) infer other investors' information. We measure the relative importance of (public) price signals versus private signals in determining an asset's equilibrium price. The publicly observed signals (prices) play a relatively larger role in determining an asset's equilibrium price when the asset's expected return is high. Appendix A outlines the set-up of the model.

2.1 Prices, Returns, and Expected Returns

Theorem 3.1 of Admati (1985) provides a closed-form solution for the $n \times 1$ vector of equilibrium prices $\tilde{P}$ at date 0. 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 Equation (1) arises from assuming the market has a continuum of investors which implies investors' signals are unbiased on average.

$$\tilde{P} = A_0 + A_1\tilde{F} - A_2\tilde{Z} \tag{1}$$

The three constant expressions $\{A_0, A_1, A_2\}$ are themselves complicated functions of the model's parameters. Expressions for the constants are shown in Appendix A of this paper. The model's parameters are as follows: r_f is the riskfree rate; $\bar{\rho}$ is investors' average risk tolerance; $\bar{F}$ and V are the mean and covariance matrix of future dividends (payoffs); Q is the precision matrix of investors' information signals about future dividends; $\bar{Z}$ and U

are the mean and covariance matrix of per capita supply (noise); and I_n is a $n \times n$ identity matrix. We follow convention and define the $n \times 1$ vector of excess dollar returns as:

$$\begin{aligned} r &\equiv \tilde{F} - (1+r_f)\tilde{P} \\ &= -(1+r_f)A_0 + (I_n - (1+r_f)A_1)\tilde{F} + (1+r_f)A_2\tilde{Z} \end{aligned} \quad (2)$$

The return vector in Equation (2) is a function of the same two random variables ($\tilde{F}$ and $\tilde{Z}$) found in the price vector—see Equation (1). 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 of Admati (1985) provides an expression for the $n \times 1$ vector of expected excess returns (risk premia) as a function of investors' precisions and supply uncertainty.

$$\begin{aligned} \mathbb{E}[r] &\equiv \bar{F} - (1+r_f)\bar{P} \\ &= \left(\bar{\rho}V^{-1} + Q + \bar{\rho}QU^{-1}Q\right)^{-1}\bar{Z} \end{aligned} \quad (3)$$

In Equation (3), the term in parentheses is positive definite. On average, low levels of investor precisions (Q) are associated with high expected returns (provided $\bar{Z} > 0$ which is true for stocks.) High levels of supply uncertainty (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).

2.2 Our Proxy for Expected Returns

Our goal is to transform unobservable model parameters (information precision and supply uncertainty) into a proxy variable for expected returns that can be estimated with recently observed data. To accomplish our goal we project stock i's returns on the prices of all stocks. The fit from this projection is given by the expression below. Put differently, we measure the time-series fit from a multi-variate time-series regression of stock i's returns on prices of all stocks—denoted R_i^2. Throughout this section, we simplify notation by setting the riskfree rate to zero ($r_f = 0$).

$$R_i^2 = 1 - \frac{Var[\tilde{F}_i - \tilde{P}_i|\tilde{P}]}{Var[\tilde{F}_i - \tilde{P}_i]} \quad (4)$$

Because fit is bounded by the $[0, 1]$ interval, we define *Proxy E[r]$_i$* as the logistic transformation of R_i^2.

$$Proxy\ E[r]_i \equiv ln\left(\frac{R_i^2}{1-R_i^2}\right) \quad (5)$$

To understand the link between a stock's R_i^2, information precisions (Q), supply uncertainties (U), and its expected return ($\mathbb{E}[r_i]$) we consider two cases. Section 2.2.1 studies a simpler case with n uncorrelated assets. Section 2.2.2 studies a more complex case with correlated dividends, signals, and/or supply uncertainties.

2.2.1 Uncorrelated Assets

Assume the matrices V, Q, and U are diagonal. Equation (4) can be simplified to the expression below. Please see Appendix B for details.

$$R_i^2 = \left(\frac{U_i^2}{Q_i^2 V_i + U_i}\right)\left(\frac{U_i V_i}{V_i U_i^2 + 2\bar{\rho} Q_i U_i V_i + \bar{\rho}^2 U_i + \bar{\rho}^2 Q_i^2 V_i}\right) \tag{6}$$

The above expression shows that as investors' precision about stock i's dividend (Q_i) increases, the denominator increases, and R_i^2 decreases. For very large values of Q_i, the R_i^2 goes to zero. Equation (6) also shows that for small values of stock i's supply uncertainty (U_i), an increase in U_i leads to an increase in R_i^2. For very large values of U_i, the R_i^2 goes to one. Appendix B provides additional expressions to help clarify the relations between Q_i, U_i, and R_i^2.

We combine results from Equations (3) and (6) and show there is a positive relation between a stock's expected excess return and the fit from a time series relation of returns on prices (R_i^2). Please see Appendix B for additional details:

$$\mathbb{E}[r_i] = \left(\frac{V_i}{U_i} \cdot \frac{R_i^2}{1 - R_i^2}\right)^{1/2} \bar{Z}_i \tag{7}$$

Higher fit predicts higher expected returns and lower fit predicts lower expected returns. As R_i^2 goes to zero, stock i's expected excess return also goes to zero. As the R_i^2 goes to one, stock i's expected return goes to infinity. Note that R_i^2 is a function of V_i, Q_i, and U_i as shown in Equation (6). All three variables, V_i, Q_i, U_i, are positive variance terms. We conclude that a variable based on R_i^2 can be used as a proxy for expected returns.

2.2.2 Correlated Assets

Assume matrices V, Q, and/or U are not diagonal. Define $K \equiv (\bar{\rho}V^{-1} + Q + \bar{\rho}QU^{-1}Q)^{-1}$. The variable K is part of expected returns shown in Equation (3). Below, i_i is a $n \times 1$ vector of zeros with a "1" in the i^{th} position. Please see Appendix B for details on obtaining the expression below from Equation (4):

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

In broad terms, an increase in Q leads to a decrease in the fit (R_i^2). An increase in U leads to an increase in the fit. The relation between the Q, U, and R_i^2 can be complicated by off-diagonal elements of the V, Q, or U matrices. Unfortunately, there is no closed-form solution for the relation between a stock's expected return and R_i^2. To better understand the relations, we turn to a numerical analysis.

We consider a market with 25 different stocks. Appendix C provides details about the parameters used to generate our computer generated data. Figure 1 shows that expected returns decrease as the precisions of investors' information (Q) increases. The figure also shows that expected returns increase as supply uncertainty (U) increases. Figure 2 shows that our *Proxy E[r]* variable decreases as Q increases and *Proxy E[r]* increases as U increases—this result generalizes intuition from Equation (6). Most importantly, Figure 3 shows that expected returns increase (roughly) linearly with our *Proxy E[r]* variable. The third figure supports the idea that our proxy variable can be used on the right-hand side of a linear, predictive regression in which future returns are on the left-hand side.

3 The Empirical *Proxy E[r]* Variable and Data Description

We create empirical variables in each month t that are correlated with stocks' expected returns. The variable for stock i as of month t is denoted *Proxy E[r]*$_{i,t}$. We use the variables to predict return dispersion in month t+1. The variables are calculated using lagged data from month t-12 to t-1. For a given stock i, the measure is based on the strength of the time-series relationship between stock i's return on day k and stock prices on day k-1. The six steps used to calculate the proxy are:

Step 1: For each stock i, we calculate our own price series over the sample period. The price of stock i is set to one the first day a stock appears in our dataset and then increased or decreased by the daily stock return. The price of stock i on day k is thus: $price_{i,k} = price_{i,k-1} \times (1 + r_{i,k})$.

Step 2: We calculate our own price series for the market portfolio over the sample period. The price is set to one in July 1965 and then increased or decreased using the daily market return. The market price on day k is thus: $price_{m,k} = price_{m,k-1} \times (1 + r_{m,k})$.

Step 3: We define the normalized price of stock i on day k as the daily price of stock i from Step 1 divided by the market's price from Step 2: $P_{i,k}^N = \frac{price_{i,k}}{price_{m,k}}$. This step is motivated by work by Biais, Bossaerts, and Spatt (2009) as discussed in our introduction.

Step 4: Equation (4) calls for projecting stock i's return on the prices of *all* stocks in the market. Using the prices of all stocks as right-hand side variables is not feasible. Therefore, for each stock i, we calculate normalized daily prices of four industry portfolios using value-weighted industry returns in a manner similar to Steps 1, 2, and 3. The first industry portfolio is most related to stock i while the fourth industry portfolio is least related.

The normalized price of the first portfolio, $P^N_{SIC4\backslash i,k}$, is calculated using stocks with the same four-digit SIC code as stock i but excludes stock i. The second portfolio, $P^N_{SIC3\backslash 4,k}$, consists of stocks with the same three-digit SIC code as stock i but excludes stocks used in the first portfolio and excludes stock i. The third portfolio, $P^N_{SIC2\backslash 3,k}$, consists of stocks with the same two-digit SIC code as stock i but excludes stocks used in the first two portfolios and excludes stock i. Finally, the fourth portfolio, $P^N_{SIC1\backslash 2,k}$, consists of stocks with the same one-digit SIC code as stock i but excludes stocks used in the first three portfolios and excludes stock i.

Step 5: We project the returns of stock i from day k on normalized prices from day k-1. For stock i in month t, the multi-variate time-series regression uses daily data from the past year (months t-12 to t-1). We require a minimum of 60 daily returns. We estimate coefficients using ordinary least squares on a stock-by-stock basis.

$$r_{i,k} = \zeta_0 + \zeta_1 P^N_{i,k-1} + \zeta_2 P^N_{SIC4\backslash i,k-1} + \zeta_3 P^N_{SIC3\backslash 4,k-1} + \zeta_4 P^N_{SIC2\backslash 3,k-1} + \zeta_5 P^N_{SIC1\backslash 2,k-1} + \eta_{i,k} \tag{8}$$

We record $R^2_{i,t}$ as the fit from the regression shown in Equation (8). The above regression may look like a momentum or relative strength regression. It is different and we are only interested in measuring fit.

Step 6: Our *Proxy E[r]* variable for stock i in month t is defined as the logistic transformation of the fit ($R^2_{i,t}$) shown in the regression from the previous step, Equation (8). As stated above, the *Proxy E[r]*$_{i,t}$ measure is calculated using lagged (available) data from months t-12 to t-1. Using the $R^2_{i,t}$ measure allows us to quantify the strength of $r_{i,k}$'s covariance with the five right-hand size variables shown in Equation (8).

$$Proxy\ E[r]_{i,t} \equiv ln\left(\frac{R^2_{i,t}}{1-R^2_{i,t}}\right) \tag{9}$$

3.1 Data and Overview Statistics

Our empirical analysis focuses on monthly stock returns from the Center for Research in Security Prices (CRSP). The final sample covers 486 months of data starting July

1965 and ending December 2005. Monthly data from July 1962 to June 1965 are used to estimate stock betas as of July 1965. Daily data are used to estimate *Proxy* $E[r]_{i,t}$. We consider American-listed common stocks with CUSIP numbers ending in digits 10 or 11.

Table 1, Panel A gives overview statistics for the variables used in this paper. Each month we calculate a variable's cross-sectional mean, standard deviation, and percentiles. The table presents time series averages of these statistics. The values of *Proxy* $E[r]_{i,t}$ are always negative due to the logistic transformation and the fact that $R^2_{i,t}$ is bounded between zero and one. The mean value is -2.640 with an intra-quartile range of [-3.059, -2.210].

[Insert Table 1 About Here]

The table also presents time-series averages of cross-sectional statistics for the other variables used in the paper. For example, the average monthly excess return of stocks over the riskfree rate is 0.90% per month with an intra-quartile range of [-0.060, +0.063]. The natural log of equity market capitalization has an intra-quartile range of [+9.964, +12.742] while the natural log of the book-to-market ratio has an intra-quartile range of [-0.822, +0.294].

Table 1 includes six other variables that have been shown to explain the cross-section of returns: 1) The firm-specific risk variation, *FSRV*, measure of Durnev, Morck, and Yeung (2004) is estimated with *contemporaneous* returns as right-hand side variables while our *Proxy E[r]* measure is estimated with *lagged* normalized prices as right-hand side variables; 2) The *Delay(1)* measure from Hou and Moskowitz (2005) use contemporaneous and *lagged market returns*, while our measure is based on *lagged normalized prices* as right-hand side variables; 3) Monthly values of the *PIN* measure from Easley, Hvidkjaer, and O'Hara (2002) are downloaded from Soeren Hvidkjaer's website; 4) The natural log of turnover; 5) The Amihud (2002) measure of illiquidity; and 6) The reciprocal of price, 1/P. Appendix D has notes on calculating FSRV and *Delay(1)*.

Table 1, Panel B shows the correlation of our *Proxy E[r]* variable with other variables. For each of the 13,993 stocks in our sample, we first calculate the time-series average of each variable. We then correlate these values across stocks. The table shows that stocks with high average *Proxy E[r]* variable are likely to have higher than average volatility of excess returns (ρ=+0.340), smaller than average market capitalization (ρ=-0.571), and larger than average book-to-market ratios (ρ=+0.311).

Not surprisingly, our *Proxy E[r]* variable has a +0.448 correlation with the *FSRV* measure, a +0.213 correlation with the $Delay(1)$ measure, and a +0.427 correlation with the *PIN* measure. Interestingly, *PIN* and *FSRV* have a correlation of 0.631 which is higher than *Proxy E[r]*'s correlation with any of the variables.

4 Empirical Results

We test whether our empirical *Proxy E[r]* variable helps explain the cross-section of stock returns using monthly Fama-MacBeth regressions. The left hand side variable is the excess return of stock i in month t+1. Right hand side variables use measures from month t including *Proxy E[r]*$_{i,t}$, an estimate of stock i's beta, the natural log of the stock's market capitalization, etc. The main regression equation is thus:

$$r_{i,t+1} - r_{f,t+1} = \gamma_0 + \gamma_1 Proxy\ E[r]_{i,t} + \gamma_2 \beta_{i,t} + \gamma_3 ln(MktCap_{i,t}) + \ldots + \varepsilon_{i,t} \qquad (10)$$

Table 2 presents results at the individual stock level. We report weighted average (through time) coefficients based on the reciprocal of regression standard errors—see Litzenberger and Ramaswamy (1979). Weighted averages help address issues related to time-varying volatility. T-statistics shown in the table are based on the time-series standard deviation of coefficient estimates.[10] All reported coefficients have been multiplied by 100.

[Insert Table 2 About Here]

Table 2, Regression 1 shows that *Proxy E[r]* is a statistically significant predictor of future returns. The regression coefficient is 0.21 with a 3.90 t-statistic. We discuss the economic significance of these results in Section 4.1. A stock's estimated beta is not positively correlated with next period's returns. The coefficient on estimated beta is -0.14 with a -1.91 t-statistic.

In Table 2, Regression 2 we include the natural log of a stock's market capitalization and book-to-market ratio as predictor variables. Only book-to-market is a significant predictor of cross-sectional differences in returns. The coefficient on *ln(Book-to-Mkt)* is 0.27 with a 5.35 t-statistic.

Regressions 3, 4, and 5 test whether *FSRV*, *Delay(1)*, and *PIN* predict future returns in addition to the variables already tested. Regression 5 represents the main results of the

[10] Errors in variables issues are addressed in two ways. First, we apply the Shanken (1992) corrections to our regression results in this section and find similar results (available from authors upon request). Second, in Section (4.3), we estimate regressions similar to Equation (10) using portfolios of stocks rather than individual securities.

paper and we see both *Proxy E[r]* and *PIN* are statistically significant predictors of cross-sectional differences in returns. The coefficient on *Proxy E[r]* is 0.25 with a 4.26 t-statistic and the coefficient on *PIN* is 3.53 with a 4.49 t-statistic. Interestingly, including *PIN* in the predictive regression drives out the significance of *ln(Book-to-Market)* as a predictor variable. Notice *PIN* is available for the 1983 to 2001 time period or 228 months and the fit of Regression 5 is 3.06%. 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. We now turn to evaluating the economic significance of the results shown in Table 2, Regression 5.

4.1 Economic Significance

We calculate the economic significance of our regression results. To do this, we calculate the average return of stocks when a predictor variable is one standard deviation above and below its average. Multiplying two times the standard deviation by the regression coefficient gives an estimate of the monthly return dispersion predicted by the variable.

Table 3, Column 1 reports the coefficients from Table 2, Regression 5. Column 2 reports the unweighted average coefficients which are simply the time series averages over the 228 months used in the regression. Column 3 shows each variable's cross-sectional standard deviation (again, averaged over the 228 months). Multiplying two times Column 3 by Column 2 gives a rough estimate of the monthly differences in returns—see Column 4.

[Insert Table 3 About Here]

Column 5 provides a more accurate estimate of economic significance. Each month we multiply two times the specific month's standard deviation by the specific month's regression coefficient. We then take the time series average of the 228 monthly values. Column 6 annualizes the monthly values.

Stocks with a *Proxy E[r]* measure one standard deviation above the mean have returns that are 4.44% higher than stocks with a measure one standard deviation below the mean. We see similar levels of economic significance from market capitalization (4.32%), book-to-market ratios (4.11%), and the *PIN* measure (3.63%).

4.2 Sort Results

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 in Table 4. 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.

[Insert Table 4 About Here]

If we form a portfolio that is long stocks when *PIN* and *Proxy E[r]* are both "*Hi*" and short 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.

4.3 Robustness Checks

We test whether our results are robust to different specifications. Table 5 includes a number of additional predictor variables in regressions similar to Equation (10). Table 5, Regressions 1, 2, and 3 include past returns. Returns from months t-3 to t-2, from months t-6 to t-4, and from months t-12 to t-7 all predict future returns. Including these variables does not affect the predictive power of our *Proxy E[r]* measure. In fact, the coefficient on *Proxy E[r]* increases from 0.15 to 0.16 to 0.18 across the first three regressions. T-statistics increase as well and Regression 3 has a fit of 6.02%.

[Insert Table 5 About Here]

Table 5, Regressions 4, 5, 6, and 7 include four additional predictor variables. We separately try the standard deviation of a stock's excess returns, turnover, the Amihud (2002) illiquidity measure, and the natural log of one over a stock's price. The *Proxy E[r]* measure remains a significant predictor of cross-sectional return differences.

Although not reported, we calculate the *AR(1)* coefficient of returns for each stock i and month t using lagged data and denote it ϕ_1. We then include ϕ_1 as a right-hand side variable in a predictive regression of similar form as Table 5, Regression 7. The cross-sectional coefficient on ϕ_1 is -0.0061 with a -2.74 t-statistic. The negative coefficient implies stocks with highly negative autocorrelations have higher expected returns. In the regression, our *Proxy E[r]* measure remains a significant predictor of cross-sectional return differences with a 0.09 coefficient and 2.20 t-statistic. Also not reported, we test our *Proxy E[r]* measure in conjunction with the Hou and Moskowitz (2005) *Delay(1)* measure. Our measure continues to be a statistical and economic predictor of return dispersion. Adding the *Delay(1)* measure has no apparent effect on our reported results.

A number of additional robustness checks are available from the authors upon request. We modify the left-hand side of Equation (10) in two ways. First, we calculate *Proxy E[r]* using weekly data. Second, we use the raw time-series fit (R^2) without taking the logit transformation. On the right-hand side of Equation (10), we include a stock's β_{smb} and β_{hml} as predictor variables. Finally, we conduct three analyses of our proxy variable: i) the time-series volatility of *Proxy E[r]*$_i$; ii) stock characteristics after sorting by *Proxy E[r]*$_i$; and iii) histograms of *Proxy E[r]*$_i$. The conclusions of this paper are not changed.

Tests Using Portfolios of Stocks: We test whether our *Proxy E[r]* measure helps predict future returns for portfolios of stocks. Using portfolios of stocks addresses errors in variables issues that might arise from using estimated quantities as predictor variables. Table 6, Regressions 1a and 1b use industry portfolios created at the 3-digit SIC level. We form 450 such portfolios each of which exists for 486 months. We see that our *Proxy E[r]* measure is statistically significant when a stock's beta, market capitalization, and book-to-market are included.

[Insert Table 6 About Here]

Regressions 2a and 2b use 100 portfolios formed by sorting stocks into *Proxy E[r]* deciles and Beta deciles. Again, our *Proxy E[r]* measure continues to be statistically significant. However, in Regression 2b, neither market capitalization nor book-to-market predict future returns. Finally, Regressions 3a and 3b use 100 portfolios formed by sorting stocks into *Proxy E[r]* deciles and size deciles. The size portfolios use the same NYSE break points as used in Fama and French (1993).

Additional Robustness Checks: We carry out a final series of tests to ensure our results are robust to different sample definitions. Table 7, Regression 1 only includes stock-months with ten or more days of data. Regressions 2 and 3 split the sample roughly in half. The first half runs from 1965 to 1985. The second half runs from 1986 to 2005. Regression 4 excludes Nasdaq stocks and only uses NYSE and Amex stocks. Regression 5 uses only Nasdaq stocks. Our *Proxy E[r]* measure is statistically significant at the 5%-level in each of the first five regressions except Regression 2. Note the *Proxy E[r]* coefficient is larger for Nasdaq stocks than NYSE/Amex stocks (0.17 > 0.13) though less statistically significant.

[Insert Table 7 About Here]

We check whether the *Proxy E[r]* measure is more effective at predicting return dispersion of large or small stocks. Each month we sort our sample into three market-capitalization groups based on NYSE breakpoints. The first group consists of stocks in deciles 1, 2, and 3. The second group consists of stocks in deciles 4 to 7. The third group consists of stocks in deciles 8, 9, and 10.

Table 7, Regressions 6, 7, and 8 report coefficients from the three groups. The *Proxy E[r]* is an economic and statistically significant predictor of return dispersion for stocks in the bottom three deciles. The *Proxy E[r]* coefficient is 0.27 with a 4.46 t-statistic. We exclude the natural log of market capitalization from these last three regressions as stocks have already been sorted by this variable.

We conclude that the *Proxy E[r]* measure is an economically and statistically significant predictor of cross-sectional return differences for stocks in the bottom three NYSE size deciles.

5 Conclusion

This paper derives a new proxy variable to explain the cross-section of stock returns. Our proxy variable comes from a multi-asset rational expectations equilibrium model and is straightforward to calculate. Both theoretical and numerical analyses predict higher values of our measure (denoted *Proxy E[r]*) are associated with higher expected returns.

In multi-asset markets, expected returns (price discounts) depend on complicated correlations of information, dividends, and noise. Our proposed measure for a given stock i is derived from the prices of many stocks. This multivariate approach sets our work apart from most papers which focus primarily on estimating information asymmetries for a single stock at a time.

We show our *Proxy E[r]* measure economically and statistically predicts cross-sectional dispersion in stock returns. Empirically, stocks with a measure one standard deviation above and below the average have returns that differ by 0.36% the following month. The difference equals 4.44% per annum and is statistically significant at the 1%-level. Our results hold after controlling for many predictor variables.

Interestingly, we find both our *Proxy E[r]* measure and the *PIN* measure are significant predictors of cross-sectional return differences. The two measures appear to pick up different effects. Our measure is motivated by a multi-asset model and the prices of many stocks go into its construction. The *PIN* measure relies on analyzing trades only of one stock at a time.

We begin with the question: "Why do some stocks have high average returns, while others have low average returns?" To answer the question, this paper adopts a three-pronged approach: 1) We start with a multi-asset equilibrium model that explicitly states all our assumptions. For example, investors possess diverse and asymmetric pieces of information. Also, there is a friction in our market as information cannot be freely traded. We show a theoretical link between the "fit" from a time-series regression and expected returns. In our model, expected returns differ across stocks due to risk/portfolio considerations. 2) We next show the link between the regression fit and expected returns using numerical analysis—this is especially helpful when dividends, information signals, and/or supply shocks are correlated across assets. 3) Finally, we estimate the same time-series fit using recently observed CRSP data. We show that our empirical proxy for expected returns helps explain dispersion in future average returns. We conclude that risks faced by investors are priced into stocks. These risks are complicated functions of uncertainty about future dividends, private information about the dividends, and supply uncertainty. Our three-pronged approach provides expressions that show the sources of the risks and how an empirical proxy variable for expected returns can be constructed—see Equations (3) and (7) for examples.

There are a number of directions for future research. First, one could test the ability of *Proxy E[r]* to explain returns in other markets around the world. Second, one could work to theoretically disentangle the effects of information risk from the effects of supply uncertainty. Third, it may be possible to derive observable/empirical measures that separate the precision of information (Q) from supply risk (U). Fourth, we could develop a dynamic model that allows testing of time-varying risk factors.

References

[1] Admati, Anat R. 1985. "A Noisy Rational Expectations Equilibrium for Multi-Asset Securities Markets." *Econometrica*, 53(3): 629-58.

[2] Amihud, Yakov. 2002. "Illiquidity and Stock Returns: Cross-Section and Time-Series Effects." *Journal of Financial Markets*, 5(1): 31-56.

[3] Avramov, Doron, Tarun Chordia, and Amit Goyal. 2006. "Liquidity and Autocorrelations in Individual Stock Returns." *The Journal of Finance*, 61(5): 2365-94.

[4] Ball, Ray, S. P. Kothari, and Jay Shanken. 1995. "Problems in Measuring Portfolio Performance: An Application to Contrarian Investment Strategies." *Journal of Financial Economics*, 38(1): 79-107.

[5] Berk, Jonathan B. 1995. "A Critique of Size-Related Anomalies." *The Review of Financial Studies*, 8(2): 275-86.

[6] Bhagat, Sanjai, Wayne M. Marr, and Rodney G. Thompson. 1985. "The Rule 415 Experiment: Equity Markets." *Journal of Finance*, 40(5): 1385-401.

[7] Biais, Bruno, Peter Bossaerts, and Chester Spatt. 2009. "Equilibrium Asset Pricing and Portfolio Choice under Heterogeneous Information." *Review of Financial Studies*, in press.

[8] Brennan, Michael J., Tarun Chordia, and Avanindhar Subrahmanyam. 1998. "Alternative Factor Specifications, Security Characteristics, and the Cross-Section of Expected Stock Returns." *Journal of Financial Economics*, 49(3): 345-73.

[9] Campbell, John Y., Martin Lettau, Burton G. Malkiel, and Yexiao Xu. 2001. "Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk." *The Journal of Finance*, 54(1): 1-43.

[10] Carhart, Mark M. 1997. "On Persistence in Mutual Funds Performance." *The Journal of Finance*, 52(1): 57-82.

[11] Cochrane, John H. 2005. Asset Pricing. Princton Univeristy Press.

[12] Chordia, Tarun, Avanidhar Subrahmanyam, and Ravi V. Anshuman. 2001. "Trading Activity and Expected Stock Returns." *Journal of Financial Economics*, 59(1): 3-32.

[13] Collins, Daniel W., S. P. Kothari, and Judy D. Rayburn. 1987. "Firm Size and the Information Content of Prices with Respect to Earnings." *Journal of Accounting and Economics*, 9(2): 111-38.

[14] Core, John, and Wayne Guay. 1999. "The Use of Equity Grants to Manage Optimal Equity Incentive Levels." *Journal of Accounting and Economics*, 28 (2): 151-84.

[15] Core, John, and Wayne Guay. 2001. "Stock Option Plans for Non-Executive Employees." *Journal of Financial Economics*, 61(2): 253-87.

[16] Datar, Vinay, Narayan Naik, and Robert Radcliffe. 1998. "Liquidity and Stock Returns: An Alternative Test." *Journal of Financial Markets*, 1(2): 203-19.

[17] Demsetz, Harold, and Kenneth Lehn. 1985. "The Structure of Corporate Ownership: Causes and Consequences." *Journal of Political Economy*, 93(6): 1155-77.

[18] Durnev, Art, Randall Morck, and Bernard Yeung. 2004. "Value-Enhancing Capital Budgeting and Firm-Specific Stock Return Variation." *Journal of Finance*, 59(1): 65-105.

[19] Easley, David, Soeren Hvidkjaer, and Maureen O'Hara. 2002. "Is Information Risk a Determinant of Asset Returns." *The Journal of Finance*, 52(5): 2185-221.

[20] Easley, David and Maureen O'Hara. 2004. "Information and the Cost of Capital." *The Journal of Finance* 59(4), 1553-1583.

[21] Fama, Eugene F., and Kenneth R. French. 1992. "The Cross-Section of Expected Stock Returns." *Journal of Finance*, 47(2): 427-65.

[22] Fama, Eugene F., and Kenneth R. French. 1993. "Common Risk Factors in the Returns of Stocks and Bonds." *Journal of Financial Economics*, 33(1): 3-56.

[23] Fama, Eugene F., and James D. Macbeth. 1973. "Risk, Return, and Equilibrium: Empirical Tests." *The Journal of Political Economy*, 81(3): 607-36.

[24] Firth, Michael. 1976. "The Impact of Earnings Announcements on Share Price Behaviour of Similar Type Firms." *The Economic Journal*, 86(342): 296-306.

[25] Foster, George. 1980. "Externalities and Financial Reporting." *Journal of Finance*, 35(2): 521-33.

[26] Foster, George. 1981. "Intra-Industry Information Transfers Associated with Earnings Releases." *Journal of Accounting and Economics*, 3(3): 201-32.

[27] Goyal, Amit, and Pedro Santa-Clara. 2003. "Idiosyncratic Risk Matters!" *Journal of Finance*, 54(3): 1199-219.

[28] Grossman, Sanford J. 1976. "On the Efficiency of Competitive Stock Markets Where Traders Have Diverse Information." *Journal of Finance* 31: 573-585.

[29] Grossman, Sanford J., and Joseph E. Stiglitz. 1980. "On the Impossibility of Informationally Efficient Markets." *The American Economic Review*, 70(3): 393-408.

[30] Hellwig, Martin F. 1980. "On the Aggregation of Information in Competitive Markets." *Journal of Economic Theory*, 22(3): 477-98.

[31] Hou, Kewei, and Tobias J. Moskowitz. 2005. "Market Frictions, Price Delay, and the Cross-Section of Expected Returns." *Review of Financial Studies*, 18(3): 981-1020.

[32] Jegadeesh, Narasimhan, and Sheridan Titman. 1993. "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency." *Journal of Finance*, 48(1): 65-91.

[33] Kelly, Patrick J. 2005. "Information Efficiency and Firm-Specific Return Variation." Working paper, Arizona State University: 55 p.

[34] Kumar, Praveen, Sorin M. Sorescu, Rodney D. Boehme, and Bartly R. Danielson. 2008. "Estimation Risk, Information, and the Conditional CAPM: Theory and Evidence." *Review of Financial Studies*, 21(3): 1037-1075.

[35] Lehmann, Bruce N. 1990. "Residual Risk Revisited." *Journal of Econometrics*, 45(1-2): 71-97.

[36] Lintner, John. 1965. "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets." *Review of Economics and Statistics*, 47(February): 13-37.

[37] Jin, Li, and Stewart Myers, C. 2006. "R2 around the World: New Theory and New Tests." *Journal of Finacial Economics*, 79(2): 257-92.

[38] Lintner, John. 1965. "Security Prices, Risk, and Maximal Gains from Diversification." *Journal of Finance*, 20(4): 587-615.

[39] Litzenberger, Robert H., and Krishna Ramaswamy. 1979. "The Effect of Personal Taxes and Dividends on Capital Asset Prices: Theory and Empirical Evidence." *Journal of Financial Economics*, 7(2): 163-96.

[40] Merton, Miller H., and Myron Scholes, S. 1982. "Dividends and Taxes: Some Empirical Evidence." *Journal of Political Economy*, 90(6): 1118-41.

[41] Morck, Randall, Bernard Yeung, and Wayne Yu. 2000. "The Information Content of Stock Markets: Why Do Emerging Markets Have Synchronous Stock Price Movements?" *Journal of Financial Economics*, 58(1-2): 215-38.

[42] Mossin, Jan. 1967. "Equilibrium in a Capital Asset Market." *Econometrica*, 34(October): 768-783.

[43] Pastor, Lubos, and Robert Stambaugh. 2003. "Liquidity Risk and Expected Stock Returns." *Journal of Political Economy*, 111(3): 642-85.

[44] Radner, Roy. 1979. "Rational Expectations Equilibrium: Generic Existence and the Information Revealed by Prices." *Econometrica*, 47: 655-678.

[45] Roll, Richard. 1988. "R2." *The Journal of Finance*, 43(3): 541-66.

[46] Shanken, Jay. 1992. "On the Estimation of Beta-Pricing Models." *The Review of Financial Studies*, 5(1): 1-33.

[47] Sharpe, William F. 1964. "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk." *Journal of Finance*, 19(September): 425-442.

[48] Szewczyk, Samuel H. 1992. "The Intra-Industry Transfer of Information Inferred from Announcements of Corporate Security Offerings." *The Journal of Finance*, 17(5): 1935-45.

[49] Tinic, Seha M., and Richard R. West. 1986. "Risk, Return and Equilibrium: A Revisit." *Journal of Political Economy*, 94(1): 126-47.

[50] Van Ness, Bonnie F., Robert A. Van Ness, and Richard S. Warr. 2001. "How Well Do Adverse Selection Components Measure Adverse Selection?" *Financial Management*, 30(3): 77-98.

[51] Wang, Jiang. 1993. "A Model of Intertemporal Asset Prices under Asymmetric Information." *The Review of Economic Studies*, 60(2): 249-82.

[52] Wang, Jiang. 1994. "A Model of Competitive Stock Trading Volume." *The Journal of Political Economy*, 102(1): 127-68.

[53] Wurgler, Jeffrey A. 2000. "Financial Markets and the Allocation of Capital." *Journal of Financial Economics*, 58(1-2): 187-214.

Appendix A: Theory Set-Up

Following Admati (1985), there is 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 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 $V = Var[\tilde{F}]$. 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 also receives an independent signal about future asset 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 S_a. The weighted average of the signal precision matrices is: $Q = \int \rho_a S_a^{-1} da$.

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 $U = Var[\tilde{Z}]$.

The closed-form solution for prices at date 0 is given by Theorem 3.1 on page 637 of Admati (1985).

$$\tilde{P} = A_0 + A_1 \tilde{F} - A_2 \tilde{Z}$$

The three constants are:

$$\begin{aligned}
A_0 &= \frac{\bar{\rho}}{1+r_f} \left(\bar{\rho}\bar{V}^{-1} + \bar{\rho} Q U^{-1} Q + Q\right)^{-1} \left(\bar{V}^{-1}\bar{F} + Q U^{-1} \bar{Z}\right) \\
A_1 &= \frac{1}{1+r_f} \left(\bar{\rho}\bar{V}^{-1} + \bar{\rho} Q U^{-1} Q + Q\right)^{-1} \left(Q + \bar{\rho} Q U^{-1} Q\right) \\
A_2 &= \frac{1}{1+r_f} \left(\bar{\rho}\bar{V}^{-1} + \bar{\rho} Q U^{-1} Q + Q\right)^{-1} \left(I_n + \bar{\rho} Q U^{-1}\right)
\end{aligned}$$

Following convention, excess returns are defined to be the change in prices between date 0 and date 1 and given by the $n \times 1$ vector: $r \equiv \tilde{F} - (1 + r_f)\tilde{P}$. The return of stock i is the i^{th} element of this vector.

The ex-ante relation between expected returns and model parameters (i.e., before realizations of random variables), is given by Corollary 3.5 on page 640 with $\bar{P} = \mathbb{E}[\tilde{P}]$.

Appendix B: Expression for R_i^2

We start with the fit from a multi-variate, time-series regression of stock i's return on the prices of all stocks. Note, we have set the riskfree rate to zero (r_f=0) to reduce notation. Below, we repeat Equation (4):

$$R_i^2 = 1 - \frac{Var[\tilde{F}_i - \tilde{P}_i|\tilde{P}]}{Var[\tilde{F}_i - \tilde{P}_i]}$$

We re-write the above expression in matrix-vector form and use i_i to denote a $n \times 1$ vector of zeros with a one (1) in the i^{th} position.

$$R_i^2 = 1 - \frac{i_i \cdot Var[\tilde{F} - \tilde{P}|\tilde{P}] \cdot i_i'}{i_i \cdot Var[\tilde{F} - \tilde{P}] \cdot i_i'}$$

The conditional variance matrix in the numerator can be re-written as below. There are three unique terms, each of which can be simplified. The first of these terms appears in the denominator.

$$Var[\tilde{F} - \tilde{P}|\tilde{P}] = Var\left[\tilde{F} - \tilde{P}\right] - Cov\left[\tilde{F} - \tilde{P}, \tilde{P}\right] \cdot Var^{-1}\left[\tilde{P}\right] \cdot Cov\left[\tilde{F} - \tilde{P}, \tilde{P}\right]'$$

Before simplifying the three unique terms above, note that from Theorem 3.1, page 637 of Admati (1985), we have: $\tilde{F} - \tilde{P} = -A_0 + (I_n - A_1)\tilde{F} + A_2\tilde{(Z)}$. The expression contains three constant expressions $\{A_0, A_1, A_2\}$ defined in the original paper and shown in Appendix A of our paper. We also make the following substitution: $K \equiv \left(\bar{\rho}V^{-1} + \bar{\rho}QU^{-1}Q + Q\right)^{-1}$.

Uncorrelated Assets

When the V, Q, and U matrices are diagonal, the n regression fits (R_i^2) can be found on the main diagonal of the matrix R^2 (shown below). To solve for the R^2 matrix, we start with the definition of R_i^2, again expand the conditional variance to four terms, and simplify.

$$\begin{aligned} R^2 &= I_n - Var\left[\tilde{F} - \tilde{P}|\tilde{P}\right] \cdot Var^{-1}\left[\tilde{F} - \tilde{P}\right] \\ &= K\left(U^{-1}QVQU^{-1} + U^{-1}\right)^{-1}\left(U + \bar{\rho}Q + \bar{\rho}K^{-1}\right)^{-1}K^{-1} \end{aligned}$$

For a single stock in the above expression, the R_i^2 shown in Equation (6) is obtained as follows:

$$\begin{aligned} R_i^2 &= K_i \left(U_i^{-1}Q_iV_iQ_iU_i^{-1} + U_i^{-1}\right)^{-1} \left(U_i + \bar{\rho}Q_i + \bar{\rho}K_i^{-1}\right)^{-1} K_i^{-1} \\ &= \left(\frac{Q_i^2V_i}{U_i^2} + \frac{1}{U_i}\right)^{-1} \left(\frac{U_i^2V_i}{U_iV_i} + \frac{\bar{\rho}Q_iU_iV_i}{U_iV_i} + \frac{\bar{\rho}^2U_i + \bar{\rho}^2Q_i^2V_i + \bar{\rho}Q_iU_iV_i}{U_iV_i}\right)^{-1} \\ &= \left(\frac{U_i^2}{Q_i^2V_i + U_i}\right)\left(\frac{U_iV_i}{V_iU_i^2 + 2\bar{\rho}Q_iU_iV_i + \bar{\rho}^2U_i + \bar{\rho}^2Q_i^2V_i}\right) \end{aligned}$$

To more easily see the relation between Q_i and R_i^2, Equation (6) can be factored. From the expression below, we can easily show that when Q_i increases, the denominator increases, and the R_i^2 decreases. For large values of Q_i, the R_i^2 goes to zero. As Q_i goes to zero, the expression below converges 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, Equation (6) can again be factored. For small values of U_i, an increase in U_i leads to an increase in R_i^2. For very large values of U_i, the R_i^2 goes to one. As U_i goes to zero, the R_i^2 goes to zero.

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

When assets are uncorrelated, Equation (3) can be written as:

$$\mathbb{E}[r_i] = \frac{\bar{Z}_i}{\bar{\rho}V_i^{-1} + Q_i + \bar{\rho}Q_i^2U_i^{-1}}$$

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

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

Using Equation (6), 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:

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

Correlated Assets

Step 1: Simplify the expression $Var[\tilde{F}-\tilde{P}]$

$$
\begin{aligned}
Var[\tilde{F}-\tilde{P}] &= Var\left[(I_n - A_1)\tilde{F} + A_2\tilde{Z}\right] \\
&= (I_n - A_1)V(I_n - A_1)' + A_2 U A_2' \\
&= K\bar{\rho}V^{-1}V(K\bar{\rho}V^{-1})' + A_2 U A_2' \\
&= K\left(U + \bar{\rho}Q + \bar{\rho}K^{-1}\right)K'
\end{aligned}
$$

Step 2: Simplify the expression $Cov[\tilde{F}-\tilde{P},\tilde{P}]$

$$
\begin{aligned}
Cov[\tilde{F}-\tilde{P},\tilde{P}] &= Cov[(I_n - A_1)\tilde{F} + A_2\tilde{Z}, A_1\tilde{F} - A_2\tilde{(Z)}] \\
&= (I_n - A_1)VA_1' - A_2UA_2' \\
&= (I_n - K(Q + \bar{\rho}QU^{-1}Q))V(K(Q + \bar{\rho}QU^{-1}Q))' \\
&\quad -(K(I_n + \bar{\rho}QU^{-1}))U(K(I_n + \bar{\rho}QU^{-1}))' \\
&= -K\left(U + \bar{\rho}Q\right)K'
\end{aligned}
$$

Note that the matrices K, U, and Q are all positive definite. Therefore, the covariance matrix is negative definite. Economically, lower prices today imply higher returns (on average).

Step 3: Simplify the expression $Var^{-1}[\tilde{P}]$

$$
\begin{aligned}
Var[\tilde{P}] &= Var\left[A_1\tilde{F} - A_2\tilde{Z}\right] \\
&= A_1VA_1' + A_2UA_2' \\
&= \left(K(U + \bar{\rho}Q)U^{-1}Q\right)V\left(K(U + \bar{\rho}Q)U^{-1}Q\right)' \\
&\quad +K(U + \bar{\rho}Q)U^{-1}U(K(U + \bar{\rho}Q)U^{-1})' \\
&= K(U + \bar{\rho}Q)\cdot(U^{-1}QVQU^{-1} + U^{-1})\cdot(U\bar{\rho}Q)K'
\end{aligned}
$$

We combine the simplified expressions show in Step 1, 2, and 3 to get the final equation for R_i^2:

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

Appendix C: 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 $\bar{\rho} = 0.15$. The average precisions of investors' private information signals is:

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

The variance-covariance of supply shocks is:

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

The variance of asset dividends (V) 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 to find the correlation matric as shown below:

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

The chart below provides a partial overview of the stocks used in the numerical analysis. The precision of investors' information is shown in Column 2. The level of supply uncertainty 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_i	U_i	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%

Figure 1 graphs the relation between two model parameters and expected returns. Expected returns are lower when the precision of investors' information (Q) is higher. Expected returns are higher when the supply uncertainty (U) is higher.

Figure 2 graphs the relation between two model parameters and our *Proxy E[r]* measure. The proxy is lower when the precision of investors' information (Q) is higher. The proxy is higher when the supply uncertainty (U) is higher.

Figure 3 graphs the relation between expected returns and our *Proxy E[r]* measure. A "best fit" or regression line is included in the scatter plot and shows the relation is approximately linear. Thus our regressions in Section 4 are well specified in this regard.

Appendix D: Other Variables in the Empirical Analysis

The FSRV Measure: The firm-specific risk variation (FSRV) measure of Durnev, Morck, and Yeung (2004) 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). We require a stock to have a minimum of 60 days of data in order to estimate *FSRV.*

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

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

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

The Delay Measure: The Delay(1) 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.

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

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

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

Table 1
Overview Statistics

This table provides overview statistics of the data used in this paper. Data start July 1965, end December 2005, and cover 486 months. There are 13,993 ordinary common stocks, an average of 3,167 stocks per month, and a total of 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. We include the natural log of stocks' market value of equity and the natural log of stocks' book to market ratio. *FSRV* is a measure of firm-specific risk variation. *Delay(1)* is a measure of a stock's delayed price reaction. *PIN* is a stock's probability of information-based trading.

Panel A: Cross-Sectional Distributions

This panel presents time-series averages of cross-sectional statistics. Each month, we calculate the cross-sectional mean, standard deviation, and percentiles for each of twelve variables. We then present time series means of each cross-sectional statistic.

Variable	Mean	Stdev	25%	50%	75%	Average # of Observations per Month
Proxy E[r]	-2.640	0.681	-3.059	-2.629	-2.210	3,168
Excess Rets (Ri – Rf)	0.009	0.139	-0.060	-0.001	0.063	3,168
Std (Ri – Rf)	0.038	0.026	0.021	0.031	0.046	3,168
Beta	0.970	0.603	0.602	0.940	1.311	3,157
ln (Mkt Cap)	11.407	1.938	9.964	11.260	12.742	3,168
ln (B-to-M)	-0.270	1.101	-0.822	-0.237	0.294	2,886
FSRV	2.676	1.395	1.688	2.510	3.492	2,773
Delay(1)	0.491	0.290	0.242	0.464	0.735	3,115
PIN	0.207	0.080	0.151	0.192	0.246	1,774
ln (Turnover)	-6.437	0.905	-6.999	-6.388	-5.819	3,003
Amihud Illiquid	7.619	58.790	0.048	0.312	2.183	2,903
ln (1/P)	-2.541	1.060	-3.304	-2.720	-1.901	3,168

Panel B: Correlations of Variables

This panel presents cross-sectional correlations of time series means. For each stock in our sample, we first calculate the time series average for each of the twelve variables. We then correlate the average values across stocks.

	Proxy E[r]	Ri - Rf	Std(Ri - Rf)	Beta	ln(MktCap)	log(B-to-M)	*FSRV*	*Delay(1)*	*PIN*	ln(Turn)	Illiquid	ln(1/P)
Proxy E[r]	1.0											
Ri - Rf	-0.002	1.0										
Std(Ri - Rf)	0.340	-0.188	1.0									
Beta	-0.216	-0.005	-0.018	1.0								
ln(MktCap)	-0.571	0.117	-0.555	0.185	1.0							
log(B-to-M)	0.311	-0.087	0.098	-0.095	-0.399	1.0						
FSRV	0.448	-0.105	0.442	-0.288	-0.649	0.198	1.0					
Delay(1)	0.213	0.006	0.254	-0.219	-0.442	0.167	0.499	1.0				
PIN	0.427	-0.035	0.325	-0.200	-0.723	0.348	0.631	0.335	1.0			
ln(Turn)	-0.471	-0.026	0.124	0.276	0.283	-0.334	-0.242	-0.228	-0.448	1.0		
Illiquid	0.051	0.048	0.216	-0.017	-0.080	0.105	0.061	0.145	0.187	-0.027	1.0	
ln(1/P)	0.319	-0.285	0.801	-0.008	-0.684	0.156	0.521	0.342	0.460	0.082	0.118	1.0

Table 2
Return Regressions Using Individual Stocks

This table presents time-series average coefficients from Fama-MacBeth regressions of monthly excess stock returns on lagged stock characteristics. All coefficients have been multiplied by 100 and are calculated as weighted averages (through time) based on the reciprocal of regression standard errors. Data start July 1965 and end December 2005 for a total 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 risk variation. *Delay(1)* is a measure of a stock's delayed price reaction. *PIN* is a stock's probability of information-based trading. T-statistics, shown in parentheses, are based on the time-series standard deviations of coefficient estimates.

	(1)	(2)	(3)	(4)	(5)
Proxy E[r]	0.21	0.15	0.17	0.15	0.25
(T-stat)	*(3.90)*	*(3.01)*	*(3.83)*	*(3.19)*	*(4.26)*
Beta	-0.14	-0.13	-0.14	-0.13	-0.18
(T-stat)	*(-1.91)*	*(-1.70)*	*(-2.07)*	*(-1.72)*	*(-1.58)*
ln (Mkt Equity)		0.01	0.01	-0.01	0.18
(T-stat)		*(0.18)*	*(0.26)*	*(-0.16)*	*(2.76)*
ln (Book-to-Mkt)		0.27	0.29	0.27	0.14
(T-stat)		*(5.35)*	*(5.58)*	*(5.47)*	*(1.64)*
FSRV			-0.01		
(T-stat)			*(-0.33)*		
Delay(1)				-0.07	
(T-stat)				*(-0.65)*	
PIN					3.53
(T-stat)					*(4.49)*
Adj R^2 (%)	0.89	3.81	3.76	3.42	3.06
# of Months	486	486	486	486	228

Table 3
Economic Significance of Predictor Variables

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.

	(1) Precision Weighted Coefficient Estimate from Table 2, Reg 5 (#)	(2) Unweighted Coefficient Estimate (#)	(3) Average Standard Deviation (Monthly) (σ)	(4) Rough Cross-Sectional Estimate $2 \times \sigma \times \gamma$ (#)	(5) Time-Series Average of $2 \times \sigma_t \times \gamma_t$ (#)	(6) Annualized Economic Significance (#)
Proxy E[r]	0.25	0.25	0.716	0.37%	0.36%	4.44%
Beta	-0.18	-0.11	0.644	-0.15%	-0.12%	-1.42%
ln (Mkt Eq)	0.18	0.10	2.052	0.40%	0.34%	4.32%
ln (B-to-M)	0.14	0.16	0.999	0.32%	0.34%	4.11%
PIN	3.53	2.29	0.080	0.37%	0.30%	3.63%

Table 4
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)*
PIN	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)
	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)		

Table 5
Additional Tests Using Individual Stocks

This table presents time-series average coefficients from Fama-MacBeth regressions of monthly stock excess returns on lagged stock characteristics. Data start July 1965 and end December 2005 for a total 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. As control variables we add lagged stock returns from *t-3:t-2*, from *t-6:t-4*, and from *t-12:t-7*. Also included are standard deviation of returns, turnover, Amihud's illiquidity measure, and the reciprocal of price (1/P). T-statistics, shown in parentheses, are based on the time-series standard deviations of coefficient estimates.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Proxy E[r]	0.15	0.16	0.18	0.14	0.08	0.20	0.14
(T-stat)	*(3.29)*	*(3.78)*	*(4.20)*	*(2.93)*	*(2.32)*	*(4.83)*	*(2.99)*
Beta	-0.14	-0.15	-0.14	-0.08	-0.02	-0.15	-0.10
(T-stat)	*(-1.96)*	*(-2.22)*	*(-2.03)*	*(-1.18)*	*(-0.34)*	*(-1.98)*	*(-1.48)*
ln (Mkt Eq)	0.01	0.00	0.00	-0.08	0.00	-0.16	-0.01
(T-stat)	*(0.19)*	*(0.10)*	*(-0.02)*	*(-2.75)*	*(-0.08)*	*(-1.94)*	*(-0.40)*
ln (B-to-M)	0.26	0.25	0.28	0.23	0.23	0.27	0.24
(T-stat)	*(5.53)*	*(5.73)*	*(6.55)*	*(5.38)*	*(4.90)*	*(5.10)*	*(4.99)*
Ret *t-3* to *t-2*	0.78	0.83	0.79				
(T-stat)	*(3.00)*	*(3.29)*	*(3.17)*				
Ret *t-6* to *t-4*		1.17	1.13				
(T-stat)		*(5.72)*	*(5.66)*				
Ret *t-12* to *t-7*			0.88				
(T-stat)			*(7.01)*				
Std(Ret)				-0.12			
(T-stat)				*(-2.74)*			
Turn					-0.29		
(T-stat)					*(-3.49)*		
Illiquid						-0.14	
(T-stat)						*(-3.01)*	
ln(1/P)							-0.10
(T-stat)							*(-0.88)*
Adj R^2	4.56	5.20	6.02	4.82	4.54	4.14	4.42
# of Months	486	486	486	486	486	486	486

Table 6
Return Regressions Using Portfolios of Stocks

This table presents time-series average coefficients from Fama-MacBeth regressions of monthly stock excess returns on lagged stock characteristics. Data start July 1965 and end December 2005 for a total 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. T-statistics, shown in parentheses, are based on the time-series standard deviations of coefficient estimates.

	3-Digit SIC Indus. Portfolios		Portfolios Sorted on *Proxy E[r]* and Beta		Portfolios Sorted on *Proxy E[r]* and Size	
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)
Proxy E[r]	0.22	0.21	0.17	0.10	0.18	0.14
(T-stat)	*(2.45)*	*(2.68)*	*(2.64)*	*(2.02)*	*(2.48)*	*(2.29)*
Beta	-0.12	-0.20	-0.06	-0.04	-0.12	-0.32
(T-stat)	*(-1.91)*	*(-1.51)*	*(-2.18)*	*(-0.29)*	*(-2.07)*	*(-1.51)*
ln (Mkt Equity)		-0.18		0.10		-0.11
(T-stat)		*(-1.88)*		*(0.88)*		*(-0.84)*
ln (Book-to-Mkt)		0.33		0.13		0.21
(T-stat)		*(3.81)*		*(1.15)*		*(1.44)*
Adj R^2 (%)	1.65	5.69	10.29	20.78	7.84	23.59
# of Portfolios	450		100		100	
# of Months	486		486		486	

Table 7
Robustness Checks and Size Results

This table presents time-series average coefficients from Fama-MacBeth regressions of monthly stock excess returns on lagged stock characteristics. Data start July 1965 and end December 2005 for a total 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. T-statistics, shown in parentheses, are based on the time-series standard deviations of coefficient estimates.

	(1) Month Has Min of 10 Trading Days	(2) Data From 1965 – 1985	(3) Data From 1986 – 2005	(4) Stocks From NYSE & Amex	(5) Stocks From Nasdaq	(6) Deciles 1-3 NYSE Breakpoints	(7) Deciles 4-7 NYSE Breakpoints	(8) Deciles 8-10 NYSE Breakpoints
Proxy E[r]	0.16	0.06	0.24	0.13	0.17	0.27	0.02	-0.06
(T-stat)	*(3.43)*	*(0.98)*	*(3.05)*	*(3.21)*	*(1.96)*	*(4.46)*	*(0.34)*	*(-1.56)*
Beta	-0.16	-0.16	-0.10	-0.10	-0.13	-0.12	-0.01	-0.10
(T-stat)	*(-2.04)*	*(-1.43)*	*(-1.00)*	*(-1.22)*	*(-1.59)*	*(-1.73)*	*(-0.13)*	*(-0.75)*
ln (Mkt Equity)	-0.02	-0.03	0.05	0.02	-0.02	N.M.	N.M.	N.M.
(T-stat)	*(-0.50)*	*(-0.45)*	*(0.71)*	*(0.44)*	*(-0.22)*			
ln (Bk-to-Mkt)	0.30	0.22	0.33	0.17	0.38	0.30	0.24	0.16
(T-stat)	*(5.28)*	*(4.56)*	*(3.56)*	*(3.26)*	*(5.56)*	*(5.89)*	*(3.92)*	*(2.53)*
Avg R^2 (%)	3.57	4.31	2.27	3.58	2.33	1.16	2.75	4.40
# of Months	486	246	240	486	486	486	486	486

Figure 1
Expected Returns

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".

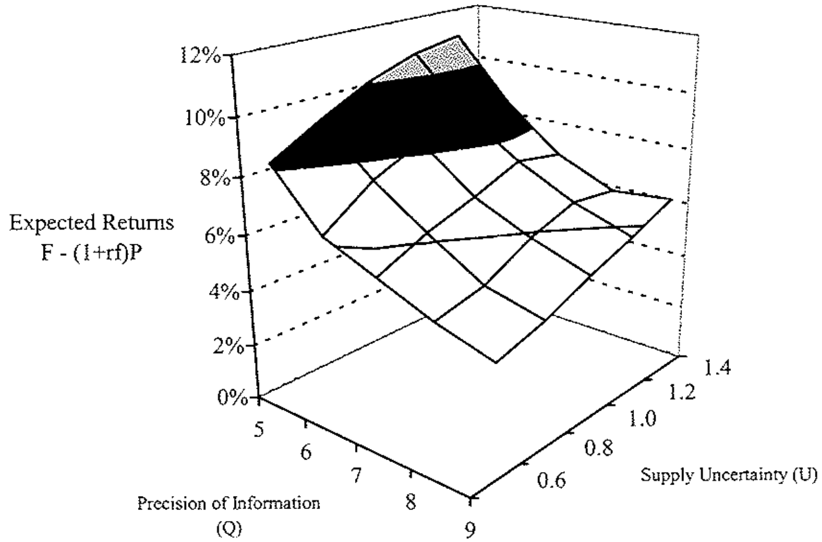

Figure 2
Our Proxy E[r] Variable

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".

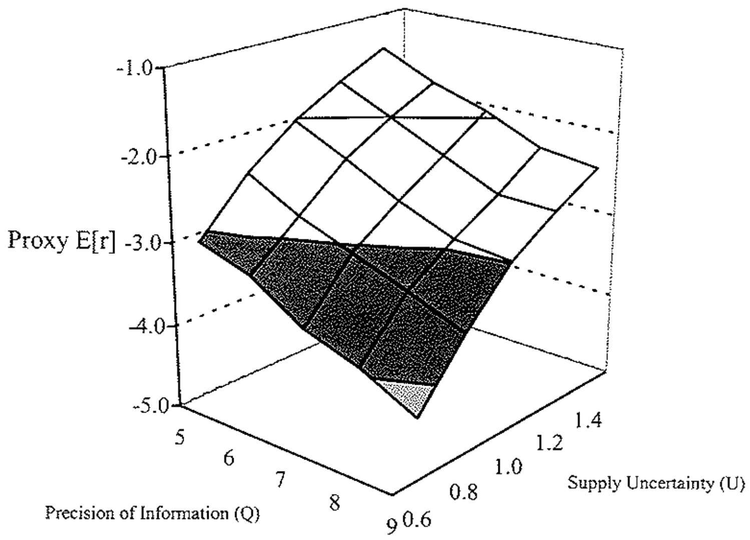

Figure 3
Expected Returns and our *Proxy E[r]* Measure

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 the fit (R^2) from a time-series regression of returns on prices and defined in the text.