Risk and the Cross-Section of Stock Returns

Radu Burlacu
Université de Nancy 2, CEREFIGE, Eurofidai
Patrice Fontaine
Université de Grenoble 2, CNRS, Eurofidai, Cerag
Sonia Jimenez-Garcès
Université de Lyon 2, CoActiS, Eurofidai
Mark S. Seasholes
HKUST

This Version: 20-Jul-2011*

Abstract

This paper mathematically transforms unobservable REE model parameters (information precision and supply uncertainty) into a single variable that is both correlated with expected returns and that can be estimated with recently observed data. Our variable can be used to explain the cross-section of returns in theoretical, numerical, and empirical analyses. Using CRSP data, we show a -1σ to +1σ change in our variable is associated with a 0.31% difference in average returns the following month (equaling 3.78% per annum). The results are statistically significant at the 1%-level. Our results remain economically and statistically significant after controlling for stocks' market capitalizations, book-to-market ratios, liquidities, and the probabilities of information-based trading.

JEL Codes: D8, G1, G10
Keywords: Risk Premiums, Cross-Sectional Asset Pricing, REE Models
Note: All referenced appendices are in this document's associated Internet Appendix. Please see: *http://dl.dropbox.com/u/6555606/RiskXsectionInternetAppendix.pdf*

*We thank Flavio Bazzana, Bruno Biais, Denis Gromb, Soeren Hvidkjaer, Michael Lemmon, Joel Peress and Bruno Solnik for helpful comments as well as seminar participants at Cerag, City University of Hong Kong, HKUST, NTU Singapore, Toulouse University, U.C. Berkeley Haas School of Business, Université Catholique de Louvain, University of New South Wales, University of Paris Dauphine, the AFFI Lille 2008 Conference and the Econometric Society 2009 North American Summer Meetings. Burlacu, Fontaine, and Jimenez-Garcès acknowledge support from CNRS and Région Rhône-Alpes. Seasholes acknowledges support from the Hong Kong RGC. Corresponding author: Mark S. Seasholes, HKUST, Dept of Finance (Rm 2413), Clear Water Bay, Kowloon, HONG KONG. Tel: +852 2358-7666. Fax: +852 2358-1749. Email: "Mark.Seasholes@gmail.com".

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. Mapping equations from a general equilibrium asset pricing model to an empirically validated measure of risk is difficult. For example, the beta from the frictionless Sharpe-Lintner-Mossin Capital Asset Pricing Model (CAPM) is eloquently defined but it has not been very successful in empirically explaining the cross-section of returns.

We study the relations between expected returns and risk in an existing multi-asset rational expectations equilibrium (REE) model.[1] This class of models introduces frictions. One such friction assumes investors receive diverse and asymmetric pieces of information. There are also supply (noise) shocks. The frictions introduce sources of risk that investors take into account when setting equilibrium prices. Note that the risks are in addition to the dividend uncertainty modeled in the frictionless CAPM.

Our first contribution is to mathematically transform unobservable REE model parameters (information precision and supply uncertainty) into a variable that is both correlated with expected returns and that can be estimated with recently observed data. The transformation makes use of two key economic facts inherent in REE models. First, an investor forms opinions about a given stock's future dividends based on his own private information <u>and</u> from observing the market-clearing mechanism to (imperfectly) glean others' information.[2] Second, and in a multi-asset world, an investor considers the degree of correlation between stock i and other stocks in the market. He considers both his private signals about other stocks' dividends and the market-clearing mechanisms associated with these other stocks.

[1] Our paper starts with the Admati (1985) model. Other REE frameworks, such as Grossman and Stiglitz (1980), lead to similar results. The Admati (1985) model has one period and two dates. Extending our analysis to a multi-period equilibrium is left for future research. In noisy rational expectations equilibrium models the terms "supply shocks" and "noise trading" are often used interchangeably. We use the term "supply uncertainty" to denote the variance of the supply shocks.

[2] 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.

Our second contribution is to show our variable is an economically and statistically significant predictor of cross-sectional average returns. Using Center for Research in Security Prices (CRSP) data starting in 1965, we show a -1σ to +1σ change in our variable is associated with a 0.31% difference in average returns the following month. This difference equals 3.78% per annum and is statistically significant at the 1%-level. Our variable 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 probability of information-based trading (*PIN*) measure of Easley, Hvidkjaer, and O'Hara (2002), the firm-specific return variation (*FSRV*) measure of Durnev, Morck, and Yeung (2004), and the delay (*Delay1*) measure of Hou and Moskowitz (2005).

Our variable helps explain the cross-section of average returns by empirically estimating the degree to which the investors rely on the market-clearing mechanism (vs. private information) when setting the prices of stocks. Consider a stock i for which investors receive very precise information about future dividends. In equilibrium, stock i's price is likely to be high and its expected return low.[3] Most importantly, investors do not need to rely heavily on the market-clearing mechanism when setting stock i's price. For stocks like i, our estimated variable has a low value coinciding with low expected returns.

Next, consider a stock j for which investors receive imprecise/noisy information about future dividends. In equilibrium, stock j's price is likely to be low and its expected return high. In such cases, investors rely heavily on the market-clearing mechanism when setting stock j's price. For stocks like j, our estimated variable has a high value coinciding with high expected returns. In summary, there are instances when investors tend to rely on the market-clearing mechanism. It is in these instances that both our variable and expected returns are high. For readers who are interested, Appendix A provides an overview of this paper's research approach. We outline both theoretical and empirical steps associated with deriving

[3] We are discussing stocks in general. There may exist some individual stocks for which these relations between information precision and expected return do not hold. However, such situations are anomalous and studied in Admati (1985).

and using our variable. Appendix B provides a number of additional economic insights into our variable. All additional materials can be found in this paper's Internet Appendix and can be downloaded at: *http://dl.dropbox.com/u/6555606/RiskXsectionInternetAppendix.pdf*

A multi-asset framework complicates the reasoning in the previous paragraph. In particular, investors may no longer need to rely on a given stock's market-clearing mechanism if they receive imprecise information about the stock. Instead, they consider whether the stock's dividends have high absolute correlations with dividends of other stocks (i.e., they consider information related to substitutes and hedges). When calculating our theoretical variable for a given stock, we take these possible substitutes and hedges into consideration by considering the prices of all stocks in the market. To empirically implement our procedure for a given stock, we consider groups of other stocks. Groups are chosen in such a manner as to have decreasing levels of economic correlation with the stock in question.

Transforming the REE model's parameters into a variable that can be estimated with recently observed data is complicated. Section 2 of the paper starts with an expected return expression derived from equations in the Admati model. A two-page overview of that REE model is given in Appendix C. In this paper, Section 2 then focuses on our main derivation results. Appendix D provides the three pages of proofs necessary to derive and simplify our variable.

Like some existing papers, regression fits (R^2 measures) are an integral part of our predictor variable. However, the regressions that generate our fits are significantly different from regressions in most existing papers. In particular, our regression equation is directly linked to the theoretical derivation of our variable in an equilibrium model. Regardless, the use of fits can cause confusion when comparing papers. To address these issues, the model precisely defines all quantities used in this paper. We also include appendices that compare and contrast our variable with other variables in the literature.

The final contribution of this paper is our integrated approach to studying the relations between expected returns and risks faced by investors. Theoretically, we link expected returns to our variable. The theory section directly leads to the empirical variable used to explain cross-sectional dispersion in average returns. Likewise, readers who want to better understand why our empirical measure works, can refer to the theory section. Additionally, the associated Internet Appendix contains a numerical analysis of our model. In the analysis, we generate a multi-asset market. We then show, cross-sectionally, that our variable is positively related to expected returns. A goal of this paper is to create a single variable that can explain cross-sectional return dispersion in a theoretical REE framework, numerical analysis, and empirical analysis. For reference, Appendix A provides an overview of this paper's research approach.

Related literature

Our paper is related to both recent empirical and theoretical work. On the empirical side, three papers propose information/friction variables to help explain the cross-section of average returns. Easley, Hvidkjaer, and O'Hara (2002) calculate the probability of information-based trading (*PIN*), Durnev, Morck, and Yeung (2004) propose a measure of firm-specific return variation (*FSRV*), and Hou and Moskowitz (2005) estimate the degree to which frictions cause information to be incorporated into prices with delay (*Delay1*). *PIN* is based on a single-stock microstructure model while our variable is derived from a multi-asset REE model.[4] Both *FSRV* and *Delay1* are derived from linear regression fits (R^2 measures). As mentioned in the introduction, the linear regressions used to estimate these fits are very different from the regression used to create our variable. Because the three measures are often confused with our derivation, we provide comparisons and contrasts with each variable in Appendices H, I, and J.

Empirically, we follow Biais, Bossaerts, and Spatt (2010) by regressing returns on prices—our contribution, however, rests in how we carry out the regression and what we do with the regression results. We measure regression fits and generate a variable that explains the cross-section of all stock returns. The earlier paper uses six test assets to construct a portfolio

[4] In a recent paper, Duarte and Young (2009) decompose the *PIN* variable into an informational asymmetry component and an illiquidity related one. The authors show that the only dimension that is priced by the market is the illiquidity component.

that outperforms the market. In this way, the two papers provide complementary insights about the role of REE prices in two areas of financial economics (portfolio choice vs. the cross-section of expected returns). Finally, Kumar, Sorescu, Boehme, and Danielsen (2008) study information and returns. The authors focus on parameter uncertainty rather than on a REE model. The paper's empirical analysis uses a proxy variable for "the innovation in market volatility" which helps appraise estimation risk.

On the theory side, 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. *PIN*'s relations with expected returns are motivated by the 2004 paper, but *PIN* itself is not derived directly from the 2004 model's results. Both Biais, Bossaerts, and Spatt (2010) and our paper use REE frameworks to study the degree to which prices incorporate information. While our framework is static, our contribution is a derivation that directly links regression fits to expected returns.

Our paper proceeds as follows. Section 2 derives our variable from an existing multi-asset rational expectations equilibrium model. Section 3 describes how to empirically estimate our variable and describes the data used in the paper. Section 4 presents the empirical results. We provide numerous robustness checks and alternative specifications. Section 5 concludes.

2. Theoretical derivation

Using Corollary 3.5 of Admati (1985), one can define an expression for expected returns within a REE framework. As the original model uses a CARA-normal setting, returns are defined as the difference between date-1 payoffs and date-0 prices. The original model contains n assets, thus $\mathbb{E}[r]$ is a $n \times 1$ vector.[5]

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

Throughout this section, and without loss of generality, we simplify notation by setting the risk-free rate to zero. If the risk-free rate were not zero, Eq.(1) would be $\mathbb{E}[r] \equiv \bar{F} - (1 + r_f)\bar{P}$ where r_f is the risk-free rate. All results in the main text would remain, albeit with more complicated notation.

In the first equation above, $\bar{F}$ is a $n \times 1$ vector of expected future dividends and $\bar{P}$ is the $n \times 1$ mean of the date-0 equilibrium price vector. In the second equation, $\bar{\rho}$ is the investors' average risk tolerance, $\mathbf{V}$ is the covariance matrix of future dividends (payoffs), $\mathbf{Q}$ is the precision matrix of investors' information signals about future dividends, and $\bar{Z}$ and $\mathbf{U}$ are the mean and covariance matrix of per capita supply (noise). We assume stocks are in positive net supply ($\bar{Z} > 0$).

In this model, $\mathbf{Q}$, $\mathbf{U}$, $\mathbf{V}$, $\bar{\rho}$, and $\bar{Z}$ are parameters as opposed to random or endogenous variables. In fact, the first three parameters are covariance/precision matrices of random variables. In this paper, as is normal with REE models, we assume there are no particular relations between the parameters. Put differently, the parameters are assumed not to be functions of one another.[6] We do, however, consider a market with many different stocks. Some stocks may be associated with high (low) information precision, some with high (low) supply uncertainty, and some with high (low) variance of future dividends.

[5] Our derivation comes with all the caveats associated with a CARA-normal framework. While the main text defines returns as price differences, we carry out robustness checks using approximations developed by Hayya, Armstrong, and Gressis (1975) for dealing with ratios of normally distributed variables. In Appendix E.2, we re-derive our variable when returns are defined as price ratios.

[6] While potentially interesting, assuming relations between model parameters go far beyond the scope of this paper. Different assumptions might involve modeling quantities outside the market considered in this paper. They might involve modeling investor behavior in ways beyond the Admati (1985) framework. Therefore, such work is left for future research.

Our goal is to transform unobservable REE model parameters (information precision and supply uncertainty) into a variable that is correlated with expected returns and that can be estimated with recently observed data. To accomplish our goal we project stock i's returns on the prices of all n stocks in the market. The fit from this projection is given by the expression below. Put differently, we measure the time-series fit (denoted R_i^2) from a multi-variate regression of stock i's returns over the t=0 to t=1 interval on prices of all stocks at t=0.

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

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. Although this paper focuses on CRSP data, the basic data requirements imply our variable can be calculated using stock market data from around the world—an added benefit. Our use of a time-series R^2 should not be confused with R^2 measures used in other papers (see Appendices I and J). Our time-series regression results are different from the negative cross-sectional relations discussed in Berk (1995).

To understand the link between stock i's R_i^2, information precisions ($\mathbf{Q}$), supply uncertainties ($\mathbf{U}$), and its expected return ($\mathbb{E}[r_i]$) we consider two cases. First, we study a simpler case with n uncorrelated assets. This simpler case provides closed-form solutions and necessary intuition about how our variable works. Proofs are provided in Appendix D. Second, we study a more complex case with correlated signals, supply uncertainties, and/or dividends. Appendix F provides the equation related to the more complex case, while a numerical analysis used to fully study the more complex case is presented in Appendix G.

For the simpler case, assume the matrices $\mathbf{Q}$, $\mathbf{U}$, and $\mathbf{V}$ are diagonal. Eq.(2) can be simplified to the expression below. Please see Appendix D for details.

$$R_i^2 = \left(\frac{U_i^2}{Q_i^2 V_i + U_i}\right)\left(\frac{U_i V_i}{V_i U_i^2 + 2\bar{\rho} Q_i U_i V_i + \bar{\rho}^2 U_i + \bar{\rho}^2 Q_i^2 V_i}\right). \tag{3}$$

The above expression helps compare and contrast stocks with different levels of information precisions and supply uncertainties. First we consider Q_i and assume U_i and V_i are constant across stocks. When investors' precision about stock i's dividend is higher (lower) than average, our R_i^2 is lower (higher) than its cross-sectional average. If Q_i is very large, our R_i^2 is close to zero.

Second we consider U_i and assume Q_i and V_i are constant across stocks. If a given stock i's supply uncertainty is higher (lower) than its cross-sectional average, our R_i^2 is higher (lower) than average. If U_i is very large, our R_i^2 is close to one. Appendix D provides additional expressions to help clarify the relations between Q_i, U_i, and our R_i^2.

We combine results from Eq.(1) and Eq.(3) to get an expression relating a stock's expected return and the fit (R_i^2) from a time series regression of returns on prices. First, note that Q_i, U_i, V_i, $\bar{Z}_i$, and R_i^2 are all positive (the first three are precision/variance terms). Because fit is bounded by the $[0,1]$ interval, we ultimately define our empirical variable as the logistic transformation of R_i^2. If V_i/U_i is fixed, then an initial inspection shows that if a stock's R_i^2 is higher (lower) than the cross-sectional average, the stock's expected return is higher (lower) than average. If our R_i^2 is close to zero, stock i's expected return is also close to zero.[7] If our R_i^2 is close to one, stock i's expected return is very large.

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

Notice that Eq.(4) can be difficult to interpret because R_i^2 is a function of the unobservable REE parameters—see Appendix D and Eq.(3) for additional details. To better understand the relations between R_i^2 and $\mathbb{E}\left[r_i\right]$ we carry out three levels of analyses.

First, we solve for conditions under which $Cov\left[\mathbb{E}\left[r_i\right], \frac{R_i^2}{1-R_i^2}\right] > 0$. Appendix E shows that when there is sufficient cross-sectional dispersion in uncertainty about stocks' payoffs/prices, we have positive relations between $R_i^2/(1-R_i^2)$ and expected returns. In other words, stocks

[7] As mentioned near Eq.(1), the derivation assumes the risk-free rate is zero. If the risk-free rate were not zero, then if R_i^2 is close to zero, stock i's excess return is also close to zero.

must have sufficient dispersion in investors' information about future payoffs (the Q_i's), or the supply uncertainties (the U_i's), or even the payoffs themselves (V_i's). Mathematically, we cannot say the relations between $\mathbb{E}[r_i]$ and $R_i^2/(1-R_i^2)$ are unambiguously positive in all circumstances. Therefore, we turn to a numerical analysis.

Our second approach to understanding the relations in Eq.(4) involves a numerical analysis. In the analysis, we model a market with 25 different stocks. We consider stocks with different levels of Q_i and U_i. We then show that, indeed, expected returns are higher (lower) than average when our R_i^2 variable is higher (lower) than average. Appendix G describes the numerical analysis and contains Fig.G.1, G.2, and G.3. The figures depict the relations between expected returns, our R_i^2, Q_i, and U_i. Fig.G.3, in particular, provides support for using our measure in a predictive linear regression as the cross-sectional relations between our fit variable and expected returns are approximately linear.

The numerical analysis also allows us to study the more complicated case with correlated signals, supply uncertainties, and/or dividends. Results in this more complicated case are qualitatively similar to the case of uncorrelated assets. That is, stocks with high (low) information precisions have low (high) R_i^2 and low (high) $\mathbb{E}[r_i]$. Also, stocks with high (low) supply uncertainty have high (low) R_i^2 and high (low) $\mathbb{E}[r_i]$. Fig.G.4, G.5, and G.6 are related to the more complicated case.

Fig.G.6 again provides support for using our measure in a predictive linear regression. From the figure, we note that the cross-sectional relations between our fit variable and expected returns are approximately linear. In general, we see that stocks with higher (lower) R_i^2 measures have higher (lower) expected returns. However, for a given pair of two randomly-picked stocks, the relations may or may not hold.

The third, and final approach to understanding the relations in Eq.(4) involves an empirical analysis. Starting in the next section, we construct an empirical version of our R_i^2 measure by following steps described in this section as closely as possible. We then use our R_i^2 as a predictor variable in cross-sectional regressions.

3. Our empirical variable and data description

For each stock i and each month t, we create an empirical variable that helps predict return dispersion in month t+1. The variable is calculated using lagged daily data from the month [t-12, t-1] interval. For a given stock i, the measure is based on the strength of the time-series relations between stock i's return on day k and stock prices on day k-1. The six steps used 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 data set and then increased or decreased by the daily stock return. The price of stock i on day k is thus: $price_{i,k} = price_{i,k-1} \times (1 + r_{i,k})$. Here, $r_{i,k}$ is stock i's daily adjusted return obtained from CRSP.

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's price on day k is thus: $price_{m,k} = price_{m,k-1} \times (1 + r_{m,k})$. Here, $r_{m,k}$ is the daily return of the CRSP value-weighted market index.

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 price from Step 2 such that: $P^N_{i,k} = \frac{price_{i,k}}{price_{m,k}}$. Economically, the normalization gives prices in units proportional to fractions of the economy. The normalization is done to ensure regressor variables in Step 5 are stationary.[8]

Step 4: Eq.(2) calls for projecting stock i's return on the prices of *all* stocks in the market. Using separate prices of all stocks as right-hand side variables is not feasible. Therefore, for each stock i, we calculate normalized daily prices of four portfolios—each with a decreasing economic relations to stock i. We use value-weighted returns in a manner similar to Steps 1,

[8] The idea of extracting expected returns from REE prices is a relatively new approach in empirical finance. We see a benefit in using a normalization method that is promoted by Biais, Bossaerts, and Spatt (2010). An alternative and common method of normalizing involves taking differences. However, if we were to take first differences, the Stage A regression (Step 5 of this section) would involve a regression of returns on lagged returns—a type of AR(1) regression. The theory section, however, points us to studying the link between returns and beginning of period prices and not the link between returns and lagged returns.

2, and 3. The first portfolio is most related to stock i while the fourth 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 stock i's return 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). Stock i's return on day k comes directly from CRSP. We require a minimum of 60 daily returns. We estimate coefficients using ordinary least squares on a stock-by-stock basis. We sometimes refer to Eq.(5) below as our "Stage A Regression" to differentiate it from later cross-sectional regressions.[9]

$$r_{i,k} = \zeta_0 + \zeta_1 P^N_{i,k-1} + \zeta_2 P^N_{SIC4\backslash i,k-1} + \zeta_3 P^N_{SIC3\backslash 4,k-1} + \zeta_4 P^N_{SIC2\backslash 3,k-1} + \zeta_5 P^N_{SIC1\backslash 2,k-1} + \eta_{i,k}. \tag{5}$$

We record $R^2_{i,t}$ as the fit from the regression shown in Eq.(5). The above regression may look like a momentum or relative strength regression. It is different and we are only interested in measuring fit. Using the Stage A regression fit allows us to quantify the strength of $r_{i,k}$'s covariance with the five right-hand side variables in Eq.(5).

Step 6: Our predictor variable for stock i in month t is defined as the logistic transformation of the fit ($R^2_{i,t}$) from the regression in Eq.(5). The measure follows from Eq.(4) in Section 2, hence we call the variable "*Proxy E[r]*$_{i,t}$". As stated above in Step 5, all month t

[9] Eq.(5) uses CRSP daily returns on the left-hand side. We also consider a similar regression with price differences on the left-hand side. Such an approach follows from the derivation in Section 2 and the regression result shown in Eq.(2). Please see results in Appendix N and O.

variables are calculated using lagged daily (available) data from months t-12 to t-1.

$$Proxy\ E[r]_{i,t} \equiv ln\left(\frac{R^2_{i,t}}{1 - R^2_{i,t}}\right). \tag{6}$$

3.1. Data and overview statistics

Our cross-sectional empirical analysis focuses on monthly stock returns from 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. 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 risk-free rate is 0.009 per month with an intra-quartile range of [-0.060, +0.063]. The natural log of our equity market capitalization has an average value of 11.407 and an intra-quartile range of [+9.964, +12.742] while the natural log of the book-to-market ratio has an average value of -0.270 and 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 return 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 *Delay1* measure from Hou and Moskowitz (2005) uses 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 for years 1983 to 2001; 4) The natural log of turnover; 5) The Amihud (2002) measure of illiquidity; and 6) The natural log of the reciprocal of price, 1/P. Appendices H, I, and J have notes on calculating *PIN*, *FSRV*, and *Delay1.*

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 (i.e., monthly OLS cross-sectional 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. We call Eq.(7) below our "Stage B Regression" to differentiate it from the regression we used to initially calculate the *Proxy* $E[r]_{i,t}$ variable.[10]

$$r_{i,t+1} - r_{f,t+1} = \gamma_0 + \gamma_1 Proxy\ E[r]_{i,t} + \gamma_2 \beta_{i,t} + \gamma_3 ln(MktCap_{i,t}) + \ldots + \varepsilon_{i,t}. \quad (7)$$

Table 2 presents results at the individual stock level. We report equal-weighted average (through time) coefficients which is standard with the Fama-MacBeth methodology.[11] T-statistics shown in the table are based on the time-series standard deviation of coefficient estimates. All reported coefficients have been multiplied by 100. Note that Table 2 represents the paper's main empirical results. As such, most alternative specifications are based on this table.

[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 0.19 with a 3.39 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 -0.04 with a -0.49 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. Book-to-market is a significant predictor of cross-sectional differences in returns. The coefficient on *ln(Book-to-Mkt)* is 0.24 with a 4.59 t-statistic.

Regressions 3, 4, and 5 test whether *FSRV*, *Delay1*, 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, and in Section 4.3, we estimate regressions similar to those in Eq.(7) but we use portfolios of stocks rather than individual securities. Second, we apply the Shanken (1992) corrections to our regression results in this section and find similar results (Please see Table 2.f in Appendix N).

[11] In Appendix L, we replicate all of this paper's main tables using precision-weighted average coefficients. For each coefficient-month, we use the reciprocal of its OLS standard error (squared) to calculate the weight. In Appendix M, we again replicate all tables using the reciprocal of the heteroskedastic-consistent (White) standard errors (squared). Weighted averages help address issues related to time-varying volatility and down-weight months for which the cross-sectional regression produces noisy estimates.

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 0.22 with a 4.10 t-statistic and the coefficient on *PIN* is 2.63 with a 3.38 t-statistic. Interestingly, including *PIN* in the predictive regression drives out much of the significance of *ln(Book-to-Market)* as a predictor variable. Notice that *PIN* is available for the 1983 to 2001 time period or 228 months and the fit of Regression 5 is 3.16%. One explanation is that 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. A second explanation is that *PIN* captures illiquidity effects as described by Duarte and Young (2009).

Table 2, Regression 6 tests all variables together for completeness. Results remain qualitatively unchanged. 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 shows each variable's cross-sectional standard deviation (again, averaged over the 228 months). Multiplying two times Column 2 by Column 1 gives a rough estimate of the monthly differences in returns—see Column 3.

[Insert Table 3 About Here]

Column 4 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 5 annualizes the monthly values.

Stocks with a *Proxy E[r]* measure one standard deviation above the mean have returns that are 3.78% higher than stocks with a measure one standard deviation below the mean. We see higher levels of economic significance for market capitalization (7.20%); slightly lower economic significance for book-to-market ratios (2.39%); and slightly higher significance for the *PIN* measure (4.49%).

4.2. Robustness checks with individual stocks

We test whether the results shown in Table 2 are robust to different specifications. Table 4 includes a number of additional predictor variables in the "Stage B Regression". Regressions 1, 2, and 3 include past returns. Cumulative 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 statistical significance of our *Proxy E[r]* measure. In fact, across the first three regressions, t-statistics range from 2.92 to 3.83 and Regression 3's fit is 4.95%.

[Insert Table 4 About Here]

Table 4, Regressions 4, 5, 6, and 7 include four additional predictor variables. We separately try the standard deviation of a stock's excess returns, the natural log of 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 in all but Regression 6. Table 4, Regression 8 tests all variables together for completeness. Results remain qualitatively the same.

In addition to the robustness checks in Appendix L and Appendix M (see footnote 15), we provide a number of alternative specifications. In Appendix N, we: a) Include an estimate of a stock's *AR(1)* coefficient and $|AR(1)|$ as predictor variables to control for possible illiquidity and/or slow price adjustment. b) Include a stock's β_{smb} and β_{hml} as predictor variables. c) Use the raw time-series fit (R^2) without taking the logistic transformation. d) Use price difference as the left-hand side variable in the Stage A regression to estimate a variant of our

Proxy E[r] variable. Using price differences follows literally from the Admati (1985) framework. e) Calculate *Proxy E[r]* using weekly data. f) Apply the Shanken (1992) correction to the standard errors. When we compare results in Appendix L with the weekly results in Appendix N, we notice a drop in statistical significance. We believe calculating Proxy E[r] with different frequencies of data can lead to different results based on the rate at which information is incorporated into stock prices.

In Appendix O, specifications modify regressions along two dimensions. For example, we use the raw time-series fit (R^2) and use price difference as the left-hand side variable in the Stage A regression. Finally, Appendix P analyzes *i*) Stock characteristics after sorting by *Proxy E[r]*; *ii*) Stock characteristics after sorting by 1-digit SIC codes; and *iii*) Histograms of *Proxy E[r]*.

The conclusions of this paper are not changed by the alternative specifications nor by the analysis of stock characteristics.

4.3. Robustness checks with 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 from the Stage A regression as predictor variables in the Stage B regression. Table 5, 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 in regressions that also include a portfolio's beta, market capitalization, and book-to-market.

[Insert Table 5 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 in Regression 2a. However, in Regression 2b, only ln(Book-to-Mkt) remains statistically significant at the 10%-level. 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). Our *Proxy E[r]* variable remains statistically significant at the 10%-level in Regression 3a and at the 5%-level in Regression 3b.

4.4. Robustness checks with different samples of individual stocks

We carry out a final series of tests in the main paper to ensure our results are robust to different sample definitions. Table 6, 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. Our *Proxy E[r]* is statistically significant at the 5%-level in Regressions 1 and 3. We note that in the first half of the sample, the natural log of a stock's book-to-market ratio has strong predictive power (which may lower the predictive power of our *Proxy E[r]*). The 1986 to 2005 period has almost twice as many listed stocks with which to estimate our *Proxy E[r]* measure. Finally, we note that when reporting Litzenberger and Ramaswamy precision weighted coefficients, our *Proxy E[r]* measure has a 1.86 t-statistic using data from 1965 to 1985 (see results in Appendix L - Table 6.)

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 to 3. The second group consists of stocks in deciles 4 to 7. The third group consists of stocks in deciles 8 to 10.

[Insert Table 6 About Here]

Table 6, Regressions 4, 5, and 6 report coefficients from the three size 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 0.24 with a 4.04 t-statistic. We exclude the natural log of market capitalization from these three regressions as stocks have already been sorted by this variable.

To end, Regression 7 uses only NYSE and Amex stocks. Regression 8 uses only Nasdaq stocks. Our *Proxy E[r]* measure is statistically significant at the 5%-level in Regression 7 and at the 10%-level in Regression 8. Note the *Proxy E[r]* coefficient is larger for Nasdaq stocks than NYSE/Amex stocks (0.14 > 0.11) though less statistically significant.

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. Conclusions

This paper derives a new variable that helps explain the cross-section of stock returns. Our variable comes from an existing multi-asset rational expectations equilibrium model and is straightforward to calculate. Both theoretical and numerical analyses show that there are relations between our measure and expected returns. Our empirical analysis demonstrates that these relations exist and are positive.

In multi-asset markets, expected returns (price discounts to expected future dividends) depend on complicated correlations of information 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-related variables for a single stock at a time.

We show our variable 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 0.31% the following month. The difference equals 3.78% per annum and is statistically significant at the 1%-level. Our results hold after controlling for many predictor variables.

Interestingly, we find both our variable and *PIN* 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 of one stock at a time.

We begin this paper 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 an existing 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 friction/portfolio considerations. 2) We next show the link between the regression fit (our variable) and expected returns using numerical analysis—this is especially helpful when information signals, supply shocks and/or dividends are correlated across assets. 3) Finally, we estimate the same variable (also based on the time-series fit) using recently observed CRSP data. We show that our variable helps explain dispersion in future average returns. We conclude that certain frictions lead to risks that are priced into stocks. These risks are complicated functions of the precisions of private information and supply uncertainties. Our three-pronged approach provides expressions that show the sources of theses risks in a REE framework. We also show how an empirical predictor variable for expected returns can be constructed. Eq.(1) and Eq.(4) succinctly summarize the derivation of our variable.

There are a number of directions for future research. First, one could test the ability of our variable to explain average 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 empirical measures that separate the precision of information from supply risk.

References

[1] Admati, A., 1985. A noisy rational expectations equilibrium for multi-asset securities markets. Econometrica 53(3), 629-658.

[2] Amihud, Y., 2002. Illiquidity and stock returns: Cross-section and time-series effects. Journal of Financial Markets 5(1), 31-56.

[3] Berk, J., 1995. A critique of size-related anomalies. Review of Financial Studies 8(2), 275-286.

[4] Biais, B., Bossaerts P., Spatt, C., 2010. Equilibrium asset pricing and portfolio choice under heterogeneous information. Review of Financial Studies 23(4), 1503-1543.

[5] Dasgupta, S., Gan, J., Gao, N., 2010. Transparency, price informativeness, and stock return synchronicity: Theory and evidence. Journal of Financial and Quantitative Analysis 45(5), 1189-1220.

[6] Duarte, J., Young, L., 2009. Why is *PIN* priced? Journal of Financial Economics 91, 119-138.

[7] Durnev, A., Morck, R., Yeung, B., 2004. Value-enhancing capital budgeting and firm-specific stock return variation. Journal of Finance 59(1), 65-105.

[8] Durnev, A., Morck, R., Yeung, B., Zarowin, P., 2003. Does firm-specific return variation mean more or less informed stock pricing? Journal of Accounting Research 41(5), 797-836.

[9] Easley, D., Hvidkjaer, S., O'Hara, M., 2002. Is information risk a determinant of asset returns? Journal of Finance 52(5), 2185-2221.

[10] Easley, D., O'Hara, M., 2004. Information and the cost of capital. Journal of Finance 59(4), 1553-1583.

[11] Fama, E., French, K., 1993. Common risk factors in the returns of stocks and bonds. Journal of Financial Economics 33(1), 3-56.

[12] Fama, E., MacBeth, J., 1973. Risk, return, and equilibrium: Empirical tests. Journal of Political Economy 81(3), 607-636.

[13] Grossman, S., Stiglitz, J., 1980. On the impossibility of informationally efficient markets. American Economic Review 70(3), 393-408.

[14] Hayya, J., Armstrong, D., Gressis, N., 1975. A note on the ratio of two normally distributed variables. Management Science 21(11), 1338-1341.

[15] Hou, K., Moskowitz, T., 2005. Market frictions, price delay, and the cross-section of expected returns. Review of Financial Studies 18(3), 981-1020.

[16] Kumar, P., Sorescu, S., Boehme, R., Danielson, B., 2008. Estimation risk, information, and the conditional CAPM: Theory and evidence. Review of Financial Studies 21(3), 1037-1075.

[17] Litzenberger, R., Ramaswamy, K., 1979. The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence. Journal of Financial Economics 7(2), 163-196.

[18] Morck, R., Yeung, B., Yu, W., 2000. The information content of stock markets: Why do emerging markets have synchronous stock price movements? Journal of Financial Economics 58(1-2), 215-238.

[19] Roll, R., 1988. R2. Journal of Finance 43(3), 541-566.

[20] Shanken, J., 1992. On the estimation of beta-pricing models. Review of Financial Studies 5(1), 1-33.

[21] White, H., 1980. A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. Econometrica 48(4), 817-838.

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 3,168 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 beta, the natural log of stocks' market value of equity in thousands of dollars and the natural log of stocks' book-to-market ratio. *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. We also include the natural log of stocks' turnover, the Amihud (2002) illiquidity measure, and the natural log of the inverse of stock prices.

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 ($R_i - R_f$)	0.009	0.139	-0.060	-0.001	0.063	3,168
Std ($R_i - R_f$)	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 (*MktCap*)	11.407	1.938	9.964	11.260	12.742	3,168
ln (*Book-to-Mkt*)	-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
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(*Turnover*)	*Illiquid*	ln(*1/P*)
Proxy E[r]	1.0											
$R_i - R_f$	-0.002	1.0										
Std(R_i - R_f)	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(*Turnover*)	-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 cross-sectional regressions of monthly excess stock returns on lagged stock characteristics. 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. We report equal-weighted means of coefficients. 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	0.12	0.14	0.12	0.22	0.22
(T-stat)	*(3.39)*	*(2.50)*	*(3.36)*	*(2.69)*	*(4.10)*	*(4.19)*
Beta	-0.04	0.00	-0.05	-0.01	-0.07	-0.14
(T-stat)	*(-0.49)*	*(0.06)*	*(-0.71)*	*(-0.14)*	*(-0.62)*	*(-1.23)*
ln (*MktCap*)		-0.03	-0.05	-0.04	0.15	0.08
(T-stat)		*(-0.57)*	*(-0.95)*	*(-0.86)*	*(2.16)*	*(1.11)*
ln (*Book-to-Mkt*)		0.24	0.26	0.24	0.10	0.11
(T-stat)		*(4.59)*	*(4.86)*	*(4.52)*	*(1.20)*	*(1.35)*
FSRV			-0.07			-0.14
(T-stat)			*(-1.70)*			*(-2.80)*
Delay(1)				-0.19		-0.21
(T-stat)				*(-1.69)*		*(-1.41)*
PIN					2.63	2.93
(T-stat)					*(3.38)*	*(3.95)*
Adj R^2 (%)	0.98	3.41	3.88	3.60	3.16	3.57
# of Months	486	486	486	486	228	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 (see the $2\times\sigma$ 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 Table 2, Reg 5 (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	0.716	0.32%	0.31%	3.78%
Beta	-0.07	0.644	-0.09%	-0.07%	-0.80%
ln (*MktCap*)	0.15	2.053	0.62%	0.58%	7.20%
ln (*Book-to-Mkt*)	0.10	0.999	0.20%	0.20%	2.39%
PIN	2.63	0.080	0.42%	0.37%	4.49%

Table 4
Additional Return Regressions Using Individual Stocks

This table presents time-series average coefficients from standard Fama-MacBeth cross-sectional regressions of monthly stock excess returns on lagged stock characteristics. All coefficients have been multiplied by 100. 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. 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, Std(*Ret*), the natural log of turnover, ln(*Turnover*), Amihud's illiquidity measure, *Illiquid*, and the natural log of the reciprocal of price, ln(*1/P)*. *PIN* is a stock's probability of information-based trading. Estimated regression constants are not shown. T-statistics are shown in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Proxy E[r]	0.13	0.14	0.15	0.12	0.08	0.06	0.09	0.18
(T-stat)	*(2.92)*	*(3.49)*	*(3.83)*	*(2.62)*	*(2.60)*	*(1.64)*	*(2.12)*	*(2.88)*
Beta	-0.02	-0.04	-0.02	0.04	0.05	0.02	0.04	-0.11
(T-stat)	*(-0.19)*	*(-0.53)*	*(-0.35)*	*(0.61)*	*(0.78)*	*(0.28)*	*(0.49)*	*(-1.15)*
ln (*MktCap*)	-0.02	-0.03	-0.03	-0.11	-0.04	0.25	-0.04	0.04
(T-stat)	*(-0.51)*	*(-0.57)*	*(-0.66)*	*(-3.54)*	*(-0.82)*	*(2.98)*	*(-1.04)*	*(0.41)*
ln (*B-to-M*)	0.23	0.23	0.26	0.21	0.20	0.21	0.20	0.16
(T-stat)	*(4.73)*	*(4.89)*	*(5.78)*	*(4.52)*	*(4.04)*	*(3.45)*	*(3.98)*	*(2.32)*
Ret *t-3* to *t-2*	0.89	0.92	0.88					0.72
(T-stat)	*(3.44)*	*(3.66)*	*(3.56)*					*(2.18)*
Ret *t-6* to *t-4*		1.22	1.18					1.33
(T-stat)		*(5.43)*	*(5.31)*					*(4.78)*
Ret *t-12* to *t-7*			1.11					1.33
(T-stat)			*(8.15)*					*(7.36)*
Std(*Ret*)				-11.04				-6.85
(T-stat)				*(-2.34)*				*(-1.08)*
ln(*Turnover*)					-0.16			-0.32
(T-stat)					*(-1.91)*			*(-2.33)*
Illiquid						0.24		-0.14
(T-stat)						*(4.89)*		*(-1.63)*
ln(*1/P*)							0.00	0.40
(T-stat)							*(-0.04)*	*(3.21)*
PIN								2.16
(T-stat)								*(3.10)*
Adj R^2	3.94	4.46	4.95	4.81	4.76	4.15	4.50	6.19
# of Months	486	486	486	486	486	486	486	228

Table 5
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 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. 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]	0.21	0.20	0.14	0.02	0.13	0.17
(T-stat)	*(2.42)*	*(2.78)*	*(2.12)*	*(0.29)*	*(1.92)*	*(2.71)*
Beta	-0.07	-0.10	0.01	0.01	-0.11	0.22
(T-stat)	*(-0.66)*	*(-0.93)*	*(0.08)*	*(0.06)*	*(-0.41)*	*(1. 18)*
ln (Mkt Equity)		0.05		-0.03		-0.03
(T-stat)		*(1.03)*		*(-0.56)*		*(-0.57)*
ln (Book-to-Mkt)		0.22		0.16		0.16
(T-stat)		*(3.70)*		*(1.89)*		*(1.83)*
Adj R^2 (%)	1.63	5.09	10.54	17.68	8.29	24.34
# of Portfolios	450	450	100	100	100	100
# of Months	486	486	486	486	486	486

Table 6
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 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. 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)** 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]	0.14	0.01	0.22	0.24	-0.03	-0.07	0.11	0.14
(T-stat)	*(3.04)*	*(0.19)*	*(3.13)*	*(4.04)*	*(-0.60)*	*(-1.59)*	*(2.66)*	*(1.81)*
Beta	-0.04	0.01	0.00	-0.01	0.13	-0.03	0.02	-0.04
(T-stat)	*(-0.50)*	*(0.09)*	*(-0.02)*	*(-0.13)*	*(1.09)*	*(-0.23)*	*(0.21)*	*(-0.47)*
ln (Mkt Equity)	-0.06	-0.11	0.06	-	-	-	-0.02	-0.01
(T-stat)	*(-1.28)*	*(-1.56)*	*(0.90)*	-	-	-	*(-0.43)*	*(-0.23)*
ln (Bk-to-Mkt)	0.25	0.25	0.24	0.26	0.24	0.16	0.14	0.30
(T-stat)	*(4.14)*	*(4.23)*	*(2.69)*	*(5.25)*	*(3.47)*	*(2.14)*	*(2.58)*	*(4.51)*
Avg R^2 (%)	3.72	4.32	2.48	1.22	2.80	4.42	3.69	2.10
# of Months	486	246	240	486	486	486	486	486