July 27, 2000 Comments Welcome

Covariance Risk, Mispricing, and the Cross Section of Security Returns

Kent D. Daniel*

David Hirshleifer**

Avanidhar Subrahmanyam***

We thank three anonymous referees, the editor (René Stulz), Jonathan Berk, Michael Brennan, Bob Jones, Gautam Kaul, Blake LeBaron, Simon Gervais, Rick Green, Terry Odean, Canice Prendergast, Matt Spiegel, Tyler Shumway, Siew Hong Teoh, Sheridan Titman, Ingrid Werner, and seminar participants at Arizona St, Boston College, Carnegie Mellon University, Cornell University, Dartmouth College, Emory University, Hong Kong University of Science and Technology, INSEAD, London School of Economics, MIT, University of North Carolina, NYU, Ohio State University, Princeton University, Stockholm School of Economics, USC, Vanderbilt University, and University of Viriginia for helpful comments and discussions. This paper was presented at the 1998 Econometric Society Meetings in Chicago, the 1998 conference on behavioral finance at Yale, the 1998 conference on efficient markets at UCLA, the 1998 Berkeley Program in Finance at Santa Barbara, the NBER 1998 Asset Pricing Meetings in Chicago, and the 1999 WFA meetings. Hirshleifer thanks the Nippon Telephone and Telegraph Program of Asian Finance and Economics for financial support.

* Finance Department, Kellogg School, Northwestern University, Evanston, IL 60208-2006; (847) 491-4113; *kent@kent.kellogg.nwu.edu*; This working paper is available at *http://kent.kellogg.nwu.edu/*

** Fisher College of Business, The Ohio State University, 740A Fisher Hall, 2100 Neil Avenue, Columbus, OH 43210-1144; (614) 292-5174; *hirshleifer.2@osu.edu*

***Anderson Graduate School of Management, University of California at Los Angeles, Los Angeles, CA 90095-1481; (310) 825-5355; *subra@anderson.ucla.edu*

Covariance Risk, Mispricing, and the Cross Section of Security Returns

- *Abstract* -

This paper offers a model in which asset prices reflect both covariance risk and misperceptions of firms' prospects, and in which arbitrageurs trade against mispricing. In equilibrium, expected returns are linearly related to both risk and mispricing measures (*e.g.*, fundamental/price ratios). With many securities, mispricing of idiosyncratic value components diminishes but systematic mispricing does not. The theory offers untested empirical implications about volume, volatility, fundamental/price ratios, and mean returns, and is consistent with several empirical findings. These include the ability of fundamental/price ratios and market value to forecast returns, and the domination of beta by these variables in some studies.

The classic theory of securities market equilibrium beginning with Sharpe, Lintner, and Black is based on the interaction of fully rational optimizing investors. In recent years, several important studies have explored alternatives to the premise of full rationality. One approach models market misvaluation as a consequence of noise or positive feedback trades. Another approach studies how individuals form mistaken beliefs or optimize incorrectly, and derives the resulting trades and misvaluation.[1]

This paper contributes to the above literature by examining how the cross-section of expected security returns is related to *risk* and *investor misvaluation*. We provide a model of equilibrium asset pricing in which risk averse investors use their information incorrectly in selecting their portfolios. We apply the model to address the ability of risk measures versus mispricing measures to predict security returns; the design tradeoffs among alternative proxies for mispricing; the relation of volume to subsequent volatility; and whether mispricing in equilibrium withstands the activities of smart 'arbitrageurs.'

Many empirical studies show that the cross-section of stock returns can be forecast using not just standard risk measures such as beta, but also market value or fundamental/price ratios such as dividend/price or book/market. The interpretation of these forecasting regressions is controversial, because these price-containing variables can be interpreted as proxies for either risk or misvaluation. So far this debate has been pursued without an explicit theoretical model of what we should expect to see in such regressions if investors misvalue stocks, and also discount for risk. This paper is the first to provide an analysis of how well, in this situation, beta and fundamental/price ratios predict the cross-section of security returns.

Based on extensive psychological evidence[2], our premise is that some or all investors are overconfident about their abilities, and hence overestimate the quality of information signals they have generated about security values. Other individuals exploit the pricing errors introduced by the trading of the informed overconfident individuals, but do not eliminate all mispricing because of risk aversion.[3]

[1]On the first approach, see, *e.g.*, Black (1986), De Long, Shleifer, Summers, and Waldman (1990a, 1990b), Campbell and Kyle (1993)). On the second approach, see, *e.g.* Shiller (1981), De Long, Shleifer, Summers, and Waldmann (1991), Benartzi and Thaler (1995), Hirshleifer, Subrahmanyam, and Titman (1994), Kyle and Wang (1997), Odean (1998), Wang (1998), Benos (1998), Caballé and Sákovics (1996), Barberis, Shleifer, and Vishny (1998), Daniel, Hirshleifer, and Subrahmanyam (1998), Hong and Stein (1999), and Barberis, Huang, and Santos (1999).

[2]See for example, the discussions and references in DeBondt and Thaler (1995), Odean (1998) and Daniel, Hirshleifer, and Subrahmanyam (1998).

[3]Our analysis differs from previous models of investor overconfidence [See, for example, De Long, Shleifer, Summers, and Waldmann (1991), Hirshleifer, Subrahmanyam, and Titman (1994), Kyle and Wang (1997), Odean (1998), Daniel, Hirshleifer, and Subrahmanyam (1998), Gervais and Odean (1999),

This paper examines only static overconfidence in a single period. This makes it tractable to integrate risk aversion, multiple risky securities, and the effects of arbitrageurs within one model. Our focus is therefore on providing a cross-sectional asset pricing model when there is long-run overreaction and correction. The analysis does not address the intertemporal patterns of short-term versus long-term return autocorrelations studied in Daniel, Hirshleifer, and Subrahmanyam (1998), Barberis, Shleifer, and Vishny (1998) and Hong and Stein (1999).[4] The relation of our paper to these dynamic models is discussed further in Section 1.

In addition to offering new empirical implications, the model explains a variety of known cross-sectional empirical findings (see Appendix A), including: (1) value-growth effects, *i.e.*, the ability of fundamental/price ratios (dividend yield, earnings/price, and book/market) to predict cross-sectional differences in future returns, incrementally to market beta, (2) inclusion of fundamental/price variables weakening, and in some tests dominating, the effect of beta on future returns, (3) the ability of firm size to predict future returns when size is measured by market value, but not when measured by non-market proxies such as book value, (4) greater ability of book/market than firm size to predict future returns in both univariate and multivariate studies, and (5) the positive association between aggregate fundamental-scaled price measures and future aggregate stock market returns.

Some recent papers (see Section I) have attempted to explain these patterns with rational asset pricing models. The challenges faced by risk-based explanations are significant (see Appendix A for details). Within the standard asset-pricing framework, the high Sharpe ratios achieved by trading strategies based on these patterns would imply extreme variation in marginal utility, especially given that returns to such strategies seem to have low correlations with plausible risk factors. While we cannot rule out explanations based on risk or market imperfections, it is reasonable to consider alternative

Benos (1998), Wang (1998), Caballé and Sákovics (1996), and Hirshleifer and Luo (1999).] in examining how covariance risk and misvaluation jointly determine the cross-section of expected security returns. Our specification of overconfidence is most similar to those of Kyle and Wang (1997), Odean (1998), and Daniel, Hirshleifer, and Subrahmanyam (1998). The latter paper assumed risk neutrality and a single risky security in order to examine the dynamics of shifts in confidence as a result of biased self-attribution, and the possibility of either over- or under-reaction.

[4]Put differently, we look at overreaction and its correction, but do not model extra dates in which overreaction can temporarily become more severe, and in which overreaction may be slugglishly corrected. Such a dynamic pattern can lead to short-term positive return autocorrelations ('momentum') as well as long term negative autocorrelation ('reversal'). Recently, Jegadeesh and Titman (1999) have provided evidence that momentum, though often interpreted as a simple underreaction, results from a process of continuing overreaction followed by correction.

explanations such as ours, which are based on imperfect rationality.

In our model, investors receive private information about, and misvalue, both systematic factors and firm-specific payoffs. While we assume that investors are overconfident about both types of information, all of the results about the cross-section of security returns follow so long as investors are overconfident about either factor information, residual information, or both. We show that in equilibrium, expected security returns are linearly increasing in the beta of the security with an adjusted market portfolio, as perceived by the overconfident investors. However, expected returns also depend on current mispricing, so returns can be predicted better by conditioning on proxies for misvaluation. A natural ingredient for such a proxy is the security's market price itself, since price reflects misvaluation. For example, following a favorable information signal, investor expectations overreact, so the price is too high. A misvaluation proxy that contains price in the denominator therefore decreases. In this setting, firms with low fundamental/price ratios are overvalued, and vice versa. In consequence, high fundamental/price ratios predict high future returns returns.

The model implies that even when covariance risk is priced, fundamental-scaled price measures can be better forecasters of future returns than covariance risk measures such as beta (Appendix A describes existing evidence). Intuitively, the reason that fundamental/price ratios have incremental power to predict returns is that a high fundamental/price ratio (*e.g.*, high book/market) can arise from high risk and/or overreaction to an adverse signal. If price is low due to a risk premium, on average it rises back to the unconditional expected terminal cash flow. If there is an overreaction to genuine adverse information, then the price will on average recover only part way toward the unconditional expected terminal value. Since high book/market reflects both mispricing and risk, whereas beta reflects only risk, book/market can be a better predictor of returns.[5]

In general, knowing the level of covariance risk (beta) helps disentangle risk and mispricing effects. This is consistent with the findings of several empirical studies (discussed in Appendix A) that beta positively predicts future returns after controlling for fundamental/price ratios or size. Furthermore, the model implies that regressing (or cross-classifying) based on fundamental/price ratios such as book/market weakens the effect of beta. This is also consistent with existing evidence. Interestingly, there is a special case of extreme overconfidence in which risk is priced and beta is a perfect proxy for

[5]Berk (1995) derives an explicit set of statistical conditions under which a price-related variable such as size has incremental power to predict future returns. Here we offer an equilibrium model in order to explore the economic conditions under which this occurs.

risk, yet beta does not have any incremental explanatory power. Thus, such a test can create the appearance that market risk is not priced even if it is fully priced. Subsection 2.3.1 provides a numerical illustration of the basic intuition for these implications.

The positive relation of fundamental/price ratios to future returns is not a general implication of investor misvaluation. Rather, it is a specific consequence of our assumption that individuals are overconfident. If, contrary to psychological evidence, individuals were on the whole underconfident, then they would underreact to adverse private signals— a high price would on average need to rise further, so a low book/market ratio would forecast *high* future returns. Thus, the evidence of a positive relation between fundamental/price ratios to future returns supports theories based on a well-known psychological bias, overconfidence, over theories based on pure underreaction.

The theory has other implications about the ability of alternative misvaluation proxies to predict future returns. Since the market value of the firm reflects misvaluation, firm size as measured by market value predicts future returns but non-market measures of firm size do not. Of course, price (or market value) can vary in the cross-section simply because unconditional expected firm payoffs vary across firms. Scaling prices by fundamental measures (*e.g.*, book value, earnings, or dividends) can improve predictive power by filtering out such irrelevant variation. Thus, a variable such as book/market tends to predict future returns better than size. Nevertheless, if the fundamental proxy measures expected future cash flows with error, market value still has some incremental ability beyond the fundamental/price ratio to predict future returns. In addition, industry normalized measures (*e.g.*, price-earnings ratio relative to industry price-earnings) can filter out industry-wide noise in fundamental measures, at the cost of removing industry-wide misvaluation.

Our analysis also offers empirical implications that are untested or that have received confirmation subsequent to our developing the model. The theory predicts that fundamental/price ratios should better forecast risk-adjusted returns for businesses that are hard to value (*e.g.*, R&D-intensive firms comprised largely of intangible assets). Recent empirical research has provided evidence consistent with this implication (see Section 4). The theory also offers implications about the cross-sectional dispersion in fundamental/price ratios and their power to predict future returns in relation to market-wide levels of fundamental/price ratios.

Further untested empirical implications relate to current volume as a predictor of future market return volatility. High volume indicates extreme signals and strong disagreement between overconfident traders and arbitrageurs. High volume therefore pre-

dicts a larger future correction. This leads to implications regarding the relation between current volume and future market volatility, and how this relation varies over time as confidence shifts.

In our setting, arbitrageurs have an incentive to trade against mispricing. We show that portfolio-based arbitrage strategies have very different consequences for the persistence of idiosyncratic versus systematic mispricing. Risk averse arbitrageurs can profit by investing in value or small-cap portfolios (or funds) and short-selling portfolios with the reverse characteristics. With many securities, arbitrageurs are able to eliminate large idiosyncratic mispricing for all but a few securities, because their arbitrage portfolios remove almost all idiosyncratic risk. In contrast, risk averse arbitrageurs do *not* eliminate the systematic mispricing. Thus, although all the model implications follow so long as there is misvaluation of either residuals or factors, to maintain a large magnitude for the effects on many securities, a non-negligible proportion of investors must be overconfident about their private information concerning systematic factors.[6]

In fact, casual empiricism strongly suggests that investors, rightly or wrongly, do think they have private information about aggregate factors. This is consistent with the existence of an active industry selling macroeconomic forecasts. Consistent with genuine private information about aggregate factors, several studies have provided evidence that aggregate insider trading forecasts future industry and aggregate stock market returns (see, *e.g.*, Lakonishok and Lee (1998)). In addition, there are many market timers who trade based on what they perceive to be information about market aggregates, and investors looking for industry plays such as internet or biotech stocks.

The remainder of the paper is structured as follows. Section 1 describes the relation of this paper to some recent models of overreaction and securities prices. Section 2 presents a pricing model based on investor psychology. Section 3 examines the forecasting of future returns using both mispricing measures and traditional risk-based return measures (such as the market beta), and develops further empirical implications. Section 4 examines further empirical implications relating to variables affecting the degree of overconfidence. Section 5 examines volume and future volatility. Section 6 examines the profitability of trading by arbitrageurs and overconfident individuals.

[6]A further objection to models with imperfect rationality is that if such trading causes wealth to flow from irrational to smart traders, eventually the smart traders may dominate price-setting. In our setting, arbitrageurs exploit the mispricing, but do not earn riskless profits. Furthermore, as in De Long, Shleifer, Summers, and Waldmann (1990a), overconfident individuals invest more heavily in risky assets, and thereby may earn higher or lower expected profits than the arbitrageurs.

1 Some Recent Models of Security Return Predictability

Why size and fundamental/price ratios forecast returns, and why systematic risk fails to do so consistently, remains a matter of debate. Rational asset pricing theory provides a straightforward motivation for value/growth effects. Since, holding constant expected payoff, price is inversely related to security risk, a fundamental/price ratio is an inverse measure of risk. If empirical beta is an imperfect measure of risk, the fundamental/price ratio will have incremental power to predict returns (see, *e.g.*, Miller and Scholes (1982), Berk (1995)). Thus, Fama and French (1993) argue that the size- and value-premia are rational risk premia. Investors are willing to pay a premium for growth stocks (and earn correspondingly low returns) because they allow investors, for example, to hedge changes in the investment opportunity set (Merton (1973)). However, such a hypothesis suggests that the returns of these portfolios should comove with aggregate economic variables, which does not appear to be the case. Other rational models of value/growth effects are provided by Berk, Green, and Naik (1999) based on real options, and by Jones and Slezak (1999) based on information asymmetry. It is not clear whether these approaches address the high Sharpe ratios attainable by value/growth strategies (MacKinlay (1995)).

Lewellen and Shanken (2000) find that owing to rational learning, high dividend yields should be associated with high subsequent aggregate market returns in *ex post* data. In addition, if learning about variances is imperfectly rational, these effects can persist in the long run. They also derive the possibility that such an effect can occur cross-sectionally, but depending on investor priors value stocks could be associated with either relatively high or low subsequent returns.

Several recent models have examined underreaction, overreaction, and correction in intertemporal settings to derive implications for short-run versus long-run autocorrelations in individual security returns. Intuitively, a pattern of long run negative autocorrelation for individual securities will tend to induce a cross-sectional value-growth effect at a given time across stocks. So the insights of these models suggest a cross-sectional relation between fundamental/price ratios and subsequent returns.

In Daniel, Hirshleifer, and Subrahmanyam (1998), individuals who are overconfident about private signals overreact to those signals. As they update their confidence over time, this overreaction temporarily becomes more severe before correcting. As a result, there is long-run overreaction and correction. Barberis, Shleifer, and Vishny (1998) is based on the representativeness heuristic and conservatism rather than overconfidence.

Investors who see only a few quarters of good earnings underreact to this good news, but those who see many quarters of good news overreact to it. This overreaction leads to subsequent low returns in the correction. Hong and Stein (1999) focus on the behavior of newswatchers who underreact to private information, and to momentum traders who condition on a subset of past prices. Momentum traders buy a rising stock, causing it to overreact. Again, this overreaction leads to subsequent low long-run returns.

The above papers, like the present paper, considered investors who form erroneous expectations of asset values or do not use all available information in forming such expectations. In contrast, Barberis and Huang (2000) focus on alternative preference assumptions. In their model, the combination of asset-by-asset mental accounting (see Thaler (1980)) and loss aversion of investors results in high equity returns, and in cross-sectional size and value effects.

A common feature of these papers is that they derive implications of investor misvaluation, but do not analyze how risk pricing interacts with mispricing in the cross-section. Our paper differs in examining how measures of misvaluation and systematic risk jointly determine the cross-section of expected future returns.[7]

2 The Model

2.1 The Economic Setting

In the introduction, we argued that the psychological basis for overconfidence is that people overvalue their own expertise. A signal that only a subset of individuals receive presumably reflects special expertise on the part of the recipients. This suggests that people will tend to be overconfident about private signals. We therefore examine a setting in which some traders possess private information and some do not. A trader who possesses a private information signal is *overconfident* about that signal: he overestimates its precision. A trader who does not possess that signal has no personal reason to be overconfident about its precision.[8]

[7] An alternative approach to securities pricing is offered by Shefrin and Statman (1994), who analyze the effect of mistaken beliefs on equilibrium in stock, option and bond markets. Their model allows for general beliefs, and therefore for a wide range of possible patterns. However, their focus is not on empirically predicting the direction of pricing errors or addressing evidence on the cross-section of security returns. In a contemporaneous paper, Shumway (1998) examines the effects of loss-aversion on securities prices. He does not, however, examine whether this approach can explain the known patterns in the cross-section of securities prices.

[8] A purely rational trader would disagree with the overconfident investors as to posterior payoff variances. This suggests that there may be profit opportunities for trading in options markets. If

The analysis has two other equivalent interpretations. First, the class of investors that are not overconfident could instead be viewed as a set of fully rational uninformed investors. These traders can also be viewed as being fully rational *informed* arbitrageurs. All three interpretations lead to identical results. We refer to the signals the informed individuals receive as 'private'.[9] Individuals who receive a private signal about a factor or about a security's idiosyncratic payoff component are referred to as the overconfident informed with respect to that signal. Individuals who do not receive a given signal are referred to as arbitrageurs with respect to that signal.[10]

2.1.1 Timing

A set of identical risk averse individuals who are each endowed with baskets containing shares of $N+K$ risky securities and of a riskfree consumption claim with terminal (date 2) payoff of 1. Prior to trade at date 1, individuals hold identical prior beliefs about the risky security payoffs. At date 1 some, but not all, individuals receive noisy private signals about the risky security payoffs. Whether or not an individual receives a signal affects his belief about the precision of that signal. Individuals then trade securities based on their beliefs. At date 2, conclusive public information arrives, the $N+K$ securities pay liquidating dividends of $\boldsymbol{\theta} = (\theta_1, ...\theta_{N+K})'$, the risk-free security pays 1, and all consumption takes place.[11]

2.1.2 Individuals and the Portfolio Problem

All individuals have identical preferences. Individual j selects his portfolio to maximize $E_j[-exp(-A\tilde{c}_j)]$, where $\tilde{c}_j$, date 2 consumption, is equal to his portfolio payoff. The j subscript here denotes that the expectations are taken using individual j's beliefs,

the model were extended to continuous time using the stylized assumptions of arbitrage-based option pricing (smooth diffusion of information, non-stochastic volatility), then rational traders would be able to obtain large risk-free profits by forming hedge portfolios of options, stocks and bonds. However, as options professionals are well aware, information arrives in discrete chunks such as earnings reports, and volatility evolves stochastically. Thus, even a trader who has a better assessment of volatility cannot make riskfree profits. In other words, a reasonable dynamic extension of the model would provide risky profit opportunities, but not arbitrage opportunities, to rational agents.

[9] An overconfident investor recognizes that those other investors who receive the same signal he does perceive a similarly high precision for it. Since this perception is shared, the investor does not regard the others as overconfident about this signal. The investor does recognize overconfidence in others about signals which they receive, if he does not receive that signal himself.

[10] We therefore allow for the possibility that an individual is overconfident with respect to one signal, but acts rationally to arbitrage mispricing arising from a different signal.

[11] Incorporating a non-zero risk-free rate would increase notational complexity but would not alter the central insights offered here.

conditional on all information available to j as of date 1. Let $\mathbf{P}$ denote the date 1 vector of prices of each security relative to the riskfree security, $\mathbf{x}_j$ denote the vector of risky security demands by individual j, and let $\bar{\mathbf{x}}_j$ be the vector of individual j's security endowment. Let $\boldsymbol{\mu}_j \equiv E_j[\boldsymbol{\theta}]$ denote the vector of expected payoffs, and $\boldsymbol{\Omega}_j \equiv E_j[(\boldsymbol{\theta} - E_j[\boldsymbol{\theta}])(\boldsymbol{\theta} - E_j[\boldsymbol{\theta}])']$ denote the covariance matrix of security payoffs.

Since all asset payoffs are normally distributed, individual j solves:

$$\max_{\mathbf{x}_j} \ \mathbf{x}_j'\boldsymbol{\mu}_j - \frac{A}{2}\mathbf{x}_j'\boldsymbol{\Omega}_j\mathbf{x}_j \quad \text{subject to} \quad \mathbf{x}_j'\mathbf{P} = \bar{\mathbf{x}}_j'\mathbf{P}.$$

All individuals act as price takers. Differentiating the Lagrangian with respect to $\mathbf{x}_j'$ gives the first order condition:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}_j'} = \boldsymbol{\mu}_j - A\boldsymbol{\Omega}_j\mathbf{x}_j - L\mathbf{P} = \mathbf{0}.$$

The condition that the price of the riskfree security in terms of itself is 1 implies that the Lagrangian multiplier $L = 1$, so

$$\mathbf{P} = \boldsymbol{\mu}_j - A\boldsymbol{\Omega}_j\mathbf{x}_j. \tag{1}$$

2.1.3 Risky Security Payoffs - The Factor Structure

Before any information signals are received, the distribution of security payoffs at date 2 are described by the following K-factor structure:

$$\theta_i = \bar{\theta}_i + \sum_{k=1}^{K} \beta_{ik} f_k + \epsilon_i, \tag{2}$$

where β_{ik} is the loading of the ith security on the kth factors, f_k is realization of the kth factor, and ϵ_i is the ith residual. As is standard with factor models, we specify w.l.o.g. that $E[f_k] = 0$, $E[f_k^2] = 1$, $E[f_j f_k] = 0 \ \forall i \neq j$, $E[\epsilon_i] = 0$, $E[\epsilon_i f_k] = 0 \ \forall i, k$. The values of $\bar{\theta}_i$ and β_{ik} are common knowledge, but the realizations of f_k and ϵ_i are not revealed until date 2. Let V_i^ϵ denote $Var(\epsilon_i)$.

With many securities, K mimicking portfolios can be formed that correlate arbitrarily closely with the K factors and diversify away the idiosyncratic risk. As a convenient approximation, we assume that each of the first K securities is a factor-mimicking portfolios for factor K, and therefore that each of these assets has zero residual variance, has a loading of 1 on factor k, and zero on the other $K - 1$ factors.

2.1.4 An Equivalent Maximization Problem

Since an individual can, by means of the K factor portfolios, hedge out the factor risk of any individual asset, he can construct a portfolio with arbitrary weights on the K factors and N residuals. Therefore, the individual's utility maximization problem is equivalent to one in which the investor directly chooses his portfolio's loadings on the K factors and N residuals, and his holdings of the risk-free asset. This can be viewed as the problem that arises when the risky securities are replaced with a set of $N+K$ uncorrelated risky porfolios each of which has a expected payoff (at date 0) of zero and a loading of one on the relevant factor or residual and zero on all others. That is, the k'th factor portfolio ($k = 1, \ldots, K$) has a date 2 payoff of f_k, and the n'th residual portfolio ($n = 1, \ldots, N$) has a date 2 payoff of ϵ_n.

Since this set of portfolios spans the same space as the original set of securities, optimizing the weights on these portfolios generates the same overall consumption portfolio as that formed by taking optimal positions in the individual securities. We solve for the market prices of these portfolios, and then for the market prices of the original securities.

One unit of the i'th original security (as described in equation (2)) can be reproduced by holding $\bar{\theta}_i$ of the risk-free asset, one unit of the i'th residual portfolio, and $\beta_{i,k}$ units of each of the $k = 1, ..., K$ factor portfolios. At any date, the price of any security is the sum of the prices of these components. From this point on we number assets so that the first K risky assets ($i = 1, \ldots, K$) in the equivalent setting are the K factor portfolios, and the remaining N ($i = K+1, \ldots, K+N$) are the N residual portfolios.

2.1.5 Date 1 Signals

Some individuals receive signals at date 1 about the K factors and N residuals. We assume that it is common knowledge that a fraction $\phi_i, i = 1, ..., K+N$ of the population receives a signal about the payoff of the ith asset. For $i = 1, ..., K$ the signal is about a factor realization and for $i = K+1, ..., K+N$ it is about a residual.

We assume that all individuals who receive a signal about a factor or residual receive precisely the same signal.[12] The noisy signals about the payoff of the k'th factor portfolio

[12] Our assumption that all individuals receive exactly the same signal is not crucial for the results, but signal noise terms must be correlated. Some previous models with common private signals include Grossman and Stiglitz (1980), Admati and Pfleiderer (1988), and Hirshleifer, Subrahmanyam, and Titman (1994). If, as is true in practice, some groups of analysts and investors use related information sources to assess security values, and interpret them in similar ways, the errors in their signals will be correlated.

and i'th residual portfolio take the form

$$s_k^f = f_k + e_k^f \quad \text{and} \quad s_i^\epsilon = \epsilon_i + e_i^\epsilon.$$

The true variance of the signal noise terms e_k^f and e_i^ϵ are V_k^{Rf} and $V_i^{R\epsilon}$, respectively (R denotes "Rational"), but because the informed investors are overconfident (C for over$C$onfident), they mistakenly believes the variance to be lower: $V_k^{Cf} < V_k^{Rf}$, and $V_i^{C\epsilon} < V_i^{R\epsilon}$. In much of the analysis, it will be more convenient to use the precision, $\nu \equiv 1/V$. Thus we define $\nu_k^{Cf} \equiv 1/V_k^{Cf}$, $\nu_k^{Rf} \equiv 1/V_k^{Rf}$, $\nu_i^{C\epsilon} \equiv 1/V_i^{C\epsilon}$, and $\nu_i^{R\epsilon} \equiv 1/V_i^{R\epsilon}$.

Finally, we assume independence of signal errors, *i.e.*, $cov(e_i^\epsilon, e_{i'}^\epsilon) = 0$ for $i \neq i'$, $cov(e_k^f, e_{k'}^f) = 0$ for $k \neq k'$, and $cov(e_i^\epsilon, e_k^f) = 0$ for all i, k. This set of assumptions makes the model tractable and is without loss of generality.

2.1.6 Expectations and Variances of Portfolio Payoffs

To solve for price in terms of exogenous parameters, we first calculate the expectation of portfolio i's terminal value given all of the signals. For convenience we will now slightly abuse the notation by letting the variable μ refer to means for the factor and residual portfolios instead of the original assets, and x the number of shares of the factor and residual portfolios. Since all variables are jointly normally distributed, the posterior distributions for f_k and ϵ_i are also normal. Let a $\bullet$ denote C or R. Except where otherwise noted, all investor expectations, covariances, and variances are conditioned on all signals available to the individual. The posterior mean and variance of payoffs of factor and residual portfolios are

$$\mu_i^\bullet \equiv E^\bullet[\delta_i] = \frac{\nu_i^\bullet s_i}{\nu_i + \nu_i^\bullet}; \quad E^\bullet\left[(\delta_i - \mu_i^\bullet)^2\right] = \frac{1}{\nu_i + \nu_i^\bullet} \quad \text{for } i = 1, \ldots, N+K \tag{3}$$

where δ_i is the payoff of the ith asset in the equivalent setting, $\delta_k = f_k$ for $k = 1, \ldots, K$, $\delta_i = \epsilon_{i-K}$ for $i = K+1, \ldots, N+K$, and where ν_i denotes the prior precision of the portfolio (*i.e.*, $1/V_i$, where $V_i = V_i^\epsilon$ for a residual, and $V_i = 1$ for a factor). Since the precision of the prior on f is 1 by assumption, $\nu_k = 1$ for $k = 1, \ldots, K$.

2.1.7 Prices and Portfolio Holdings

Because the payoffs of the K factor portfolios and N residual portfolios are uncorrelated, the covariance matrix $\mathbf{\Omega}$ in equation (1) is diagonal, and we can rewrite equation (1) on an element-by-element basis as

$$P_i = \mu_i^\bullet - \left(\frac{A}{\nu_i + \nu_i^\bullet}\right) x_i^\bullet,$$

or

$$x_i^\bullet = \frac{1}{A}(\nu_i + \nu_i^\bullet)\,(\mu_i^\bullet - P_i) = \frac{\nu_i + \nu_i^\bullet}{A} E^\bullet[R_i] \qquad \text{for } i = 1, \ldots, N+K, \tag{4}$$

where $x_i^\bullet$ denotes the number of shares of portfolio i an individual would hold. Also, since individuals have constant absolute risk aversion, it is convenient to define the date 1-2 'return' of portfolio i as the terminal payoff minus the price, $R_i \equiv \delta_i - P_i$.

In this setting there is no noise trading or shock to security supply. In consequence, uninformed individuals can infer all the signals perfectly from market prices. The uninformed end up with the same information as the informed traders, but use it differently as they are not overconfident about these signals (see the discussion in Subsection 2.1.5).

We impose the market clearing condition that that the average holdings of each asset equal the number of endowed shares per individual, ξ_i, of each factor or residual portfolio (this is the number of shares that would be required to construct the market portfolio using just the $N+K$ factor and residual portfolios, divided by the number of individuals). Recall that ϕ_i denotes the fraction of the population which receives information about, and is overconfident about, portfolio i. Thus, by equation (4),

$$\begin{aligned} \xi_i &= \phi_i x_i^C + (1-\phi_i) x_i^R \\ &= \frac{1}{A}\left[\phi_i(\nu_i + \nu_i^C)(\mu_i^C - P_i) + (1-\phi_i)(\nu_i + \nu_i^R)(\mu_i^R - P_i)\right]. \end{aligned}$$

Using the expression for $\mu_i^\bullet$ in (3), the above equation yields:

$$P_i = \left[\frac{\nu_i^R + \phi_i(\nu_i^C - \nu_i^R)}{\nu_i + \nu_i^R + \phi_i(\nu_i^C - \nu_i^R)}\right] s_i - \left[\frac{A}{\nu_i + \nu_i^R + \phi_i(\nu_i^C - \nu_i^R)}\right] \xi_i. \tag{5}$$

Let ν_i^A be the *consensus precision* (A for Average), or, more formally, the population-weighted average assessment of signal precision for signal i:

$$\nu_i^A \equiv \phi_i \nu_i^C + (1-\phi_i)\nu_i^R. \tag{6}$$

Then (5) can be rewritten as

$$P_i = \left(\frac{\nu_i^A}{\nu_i + \nu_i^A}\right) s_i - \left(\frac{A}{\nu_i + \nu_i^A}\right) \xi_i. \tag{7}$$

This expression shows that prices are set as if all agents were identically overconfident and assessed the signal precision to be ν^A. We further define

$$\lambda_i \equiv \frac{\nu_i^A}{\nu_i + \nu_i^A}, \quad \lambda_i^R \equiv \frac{\nu_i^R}{\nu_i + \nu_i^R} \quad \text{and} \quad \lambda_i^C \equiv \frac{\nu_i^C}{\nu_i + \nu_i^C}, \qquad \text{for } i = 1, \ldots, N+K. \tag{8}$$

λ_i is the actual response of the market price of asset i to a unit increase in the signal s_i; λ_i^R is what the response would be if all individuals in the population behaved rationally, and λ_i^C is what the response would be were all individuals overconfident.

If there is a mixture of arbitrageurs and overconfident informed individuals in the population, *i.e.*, if $0 < \phi < 1$, then $\lambda_i^C > \lambda_i > \lambda_i^R$. Thus, prices respond too strongly to private signals, but not as strongly as they would were there no arbitrageurs to trade against the overconfidence-induced mispricing. The higher the fraction of overconfident informed agents ϕ_i, the greater the amount of overreaction to the private signal.

Substituting the definitions of the λ's from (8) into (7), we can calculate the price and expected return of asset i (expectation at date 1 of the date 1-2 price change), conditional on the signal.

$$P_i = \lambda_i^R s_i + (\lambda_i - \lambda_i^R) s_i - \left(\frac{A}{\nu_i + \nu_i^A}\right) \xi_i \tag{9}$$

$$E^R[R_i] = \mu_i^R - P_i = -(\lambda_i - \lambda_i^R) s_i + \left(\frac{A}{\nu_i + \nu_i^A}\right) \xi_i, \quad \text{for } i = 1, \ldots, N+K. \tag{10}$$

The price equation has three terms. The first, $\lambda_i^R s_i = E^R[\delta_i]$, is the expected payoff of the security from the perspective of a rational investor. The second term, $(\lambda_i - \lambda_i^R) s_i$, is the extra price reaction to the signal s_i due to overconfidence. The last term of the equation is the price-discount for risk. The (rationally assessed) expected return on portfolio i depends only on the i-signal. Intuitively, with constant absolute risk aversion news about other components of wealth does not affect the premium individuals demand for trading in the ith portfolio.

From equation (10), it can be seen that the expected return consists of two terms: the correction of the extra price reaction to the signal, and a risk premium that compensates for the risk of the portfolio. Recall that an overconfident informed individual always thinks that the security is less risky than it really is. Hence, the greater the fraction of overconfident individuals in the population (the greater ϕ), and the greater their overconfidence, the lower its risk premium.

Equation (10) gives the expected return, as assessed by a rational arbitrageur. The more general expression that gives the expected return as assessed by either overconfident informed traders or arbitrageurs is

$$E^\bullet[R_i] = \mu_i^\bullet - P_i = (\lambda_i^\bullet - \lambda_i) s_i + \left(\frac{A}{\nu_i + \nu_i^A}\right) \xi_i, \quad \text{for } i = 1, \ldots, N+K. \tag{11}$$

Since under overconfidence, $\lambda_i^C > \lambda_i > \lambda_i^R$, the first term of this equation shows that the arbitrageurs and the informed overconfident traders disagree on whether securities are

over- or under-priced. Ignoring the risk-premium (the last term), if s_i is positive, then an arbitrageur thinks that the price is too high by $(\lambda_i - \lambda_i^R)s_i$, and an overconfident investor thinks that the price is too low by approximately $(\lambda_i^C - \lambda_i)s_i$. Because of these differing beliefs, the arbitrageurs and informed overconfident traders, whose holdings are given by equation (4), take opposing positions following a signal.[13]

The expressions for the price and expected return can be expressed more compactly with the following rescaling:

$$S_i \equiv \lambda_i^R s_i, \quad \text{and} \quad \omega_i \equiv \frac{\lambda_i - \lambda_i^R}{\lambda_i^R} \quad \text{for } i = 1, \ldots, N+K. \tag{12}$$

Here S_i is rescaled so that a unit increase in the signal would cause a unit increase in the price, were all agents rational. However, with overconfident traders there is excess sensitivity to the signal: ω_i denotes the fractional excess sensitivity for the ith signal. Given these definitions, equations (9) and (10) become:

$$P_i = (1+\omega_i)S_i - \frac{A}{\nu_i + \nu_i^A}\xi_i \tag{13}$$

$$E^R[R_i] = -\omega_i S_i + \frac{A}{\nu_i + \nu_i^A}\xi_i, \qquad \text{for } i = 1, \ldots, N+K. \tag{14}$$

2.1.8 The Adjusted Market Portfolio

Using equation (14), we can write the returns on each of the $N+K$ portfolios as:

$$R_i = E^R[R_i] + u_i \qquad \text{for } i = 1, \ldots, N+K,$$

where by rationality of the true expectation $E^R[u_i] = 0$ and $E^R[R_i u_i] = 0$. Also, from (3), $E^R[u_i^2] = 1/(\nu_i + \nu_i^R)$. And, as discussed previously $E^\bullet[u_i u_j] = 0$ for $i \neq j$ and $\bullet = \{C, R\}$, since rational and overconfident agree that the $N+K$ assets are uncorrelated with one another.

Let the true (*per capita*) market return be the return on the porfolio with security weights equal to the endowed number of shares of each security per individual (*i.e.*, the weights are the total market portfolio weights divided by the population size):

$$R_m = \sum_{i=1}^{N} \xi_i R_i,$$

[13]The overconfident also think that the security is less risky than do the arbitrageurs. Hence, they are willing to hold a larger position. For a favorable signal, the return and risk effects are reinforcing, but for an adverse signal they are opposing.

and let the *adjusted market portfolio* M be the portfolio with weights

$$\xi_i^{'} = \xi_i \left(\frac{\nu_i + \nu_i^R}{\nu_i + \nu_i^A} \right), \quad i = 1, \ldots, N + K. \tag{15}$$

The adjustment factor in parentheses is the ratio of the asset's consensus variance $1/[\nu_i + \nu_i^A]$ to the true variance $1/(\nu_i + \nu_i^R)$. The rationally assessed covariance between the asset i return and the adjusted market return is $cov^R(R_i, R_M) = \xi_i/(\nu_i + \nu_i^A)$. Substituting this into equation (14) gives

$$E^R[R_i] = -\omega_i S_i + A\ cov^R(R_i, R_M) \qquad \text{for } i = 1, \ldots, N + K. \tag{16}$$

Thus, the expected return is the sum of a mispricing component and a risk component which is based on the covariance with the adjusted market portfolio.

2.2 Pricing Relationships

The previous subsection derived expressions for prices and expected returns for the factor and residual portfolios. We now derive pricing relationships for the original securities. We now let ω_i^ϵ and S_i^ϵ, for $i = 1, ..., N$, and ω_k^f and S_k^f, for $k = 1, ..., K$, denote the fractional overreaction (ω_i) and scaled signal (S_i) for the N residuals and K factors. For the original $N + K$ assets, equation (16) implies the following asset pricing relationship.

Proposition 1 *If risk averse investors with exponential utility are overconfident about the signals they receive regarding K factors and about the idiosyncratic payoff components of N securities, then securities obey the following relationships:*

$$P_i = \bar{\theta}_i - \alpha\beta_{iM} + (1 + \omega_i^\epsilon)S_i^\epsilon + \sum_{k=1}^{K} \beta_{ik}(1 + \omega_k^f)S_k^f \tag{17}$$

$$E^R[R_i] = \alpha\beta_{iM} - \omega_i^\epsilon S_i^\epsilon - \sum_{k=1}^{K} \beta_{ik}\omega_k^f S_k^f, \tag{18}$$

for all $i = 1, \ldots, N + K$, where

$$\beta_{iM} \equiv cov(R_i, R_M)/var(R_M).$$

Equation (18) implies that true expected return decomposes additively into a risk premium (the first term) and components arising from mispricing (the next two terms).

Mispricing arises from the informed's overreaction to signals about the factor and the idiosyncratic payoff components. The mispricing due to overreaction to factor information is proportional to the security's sensitivity to that factor. In addition, the securities expected return includes a premium for market risk.[14] If there were no overconfidence ($\lambda_i = \lambda_i^R$ and $\lambda_k = \lambda_k^R$), this equation would be identical to the CAPM with zero riskfree rate.

We now use the fact that the expected value of the signals is zero to derive an expression for expected returns without conditioning on current market prices nor on any other proxy for investors' private signals. The following corollary follows by taking the rational expectation of (18); and then taking a weighted sum of security expected returns to show that $\alpha = E[R_M]$.

Corollary 1 *Conditioning only on β_{iM}, expected security returns obey the pricing relationship:*

$$E[R_i] = E[R_M]\beta_{iM}, \quad i = 1, \ldots, N + K,$$

where $E[R_M]$ is the expected return on the adjusted market portfolio, and β_{iM} is the security's beta with respect to the adjusted market portfolio.

This is identical in form to the CAPM security market line (with zero riskfree rate). However, here M is the adjusted market porfolio. This relationship also holds for the true market portfolio m in the natural benchmark case in which investors are *equally overconfident* about all signals, *i.e.*, the ratio

$$\frac{\nu_i + \nu_i^R}{\nu_i + \nu_i^A} = \frac{\nu_j + \nu_j^R}{\nu_j + \nu_j^A} \tag{19}$$

is equal for all factors and residuals i and j.[15] This can be seen by substituting equation (19) into equation (15).[16]

Although consistent with the univariate evidence that mean returns are increasing with beta, the model implies that there are better ways to predict future returns than the CAPM security market line. Proposition 1 implies that better predictors can be obtained by regressing not just on beta, but on proxies for market misvaluation.

[14]The coefficient β_{iM} is a price-change beta, not the CAPM return beta. That is, β_{iM} is the regression coefficient in $\theta_i - P_i = \alpha_i + \beta_{iM}(\theta_M - P_M)$ where P_i and P_M are known. The CAPM return beta is the coefficient in the regression $(\theta_i - P_i)/P_i = \alpha_i^R + \beta_{iM}^{\text{Ret}}(\theta_M - P_M)/P_M$, and is equal to $(P_M/P_i)\beta_{iM}$.

[15]If there are a very large number of securities, as discussed in Section 6, all that is required is that this ratio be equal across factors, but not necessarily across residuals.

[16]Empirically, as discussed in Appendix A, the evidence is supportive of a positive univariate relation between market beta and future returns, although estimates of the strength and significance of the effect varies across studies.

2.3 Proxies for Mispricing

Since mispricing is induced by signals that are not directly observable by the econometrician, it is important to examine how expected returns are related to observable proxies for mispricing, and with measures of risk. We begin with a simple numerical illustration of the basic reasoning.

2.3.1 A Simple Example

We illustrate here (and formalize later) three points:

1. High fundamental/price ratios (henceforth in this subsection, F/P's) predict high future returns if investors are overconfident.

2. Regressing on β as well as an F/P yields positive coefficients on both β and the F/P.

3. If overconfidence about signals is extreme, even though β is priced by the market, β has no incremental power to predict future returns over an F/P.

To understand parts (1)-(3), suppose for presentational simplicity that there is only information about idiosyncratic risk, and consider a stock that is currently priced at 80. Suppose that its unconditional expected cash flow is known to be 100. The fact that the price is below the unconditional fundamental could reflect a rational premium for factor risk, an adverse signal, or both.

Suppose now that investors are overconfident, and consider the case in which $\beta = 0$. Then the low price ($80 < 100$) must be the result of an adverse signal. Since investors overreacted to this signal, the true conditional expected value is greater than 80— the stock is likely to rise. Thus the above-average F/P (100/80) is associated with a positive abnormal expected return. Of course, the reverse reasoning indicates that a below-average value (e.g, 100/120) predicts a negative expected return. Thus, when investors are overconfident, a high fundamental/price ratio predicts high future returns, consistent with a great deal of existing evidence.[17]

If we allow for differing β's, there is a familiar interfering effect (Berk (1995)): a high beta results in a higher risk discount and hence also results in a high F/P. Thus, even if there is no information signal, a high F/P predicts high returns. This illustrates point

[17] If instead investors were underconfident, then the price of 80 would be an *under*reaction to the signal, so price could be expected to fall further. Thus, the high fundamental/price ratio would predict *low* returns, inconsistent with the evidence.

(1) above. However, the confounding of risk and mispricing effects suggests regressing future returns on beta as well as on F/P.

This confounding leads to point (2). Specifically, if the econometrician knew that the discounting of price from 100 to 80 was purely a risk premium, then the expected terminal cash flow would still be 100. In contrast, if this is a zero-β (zero risk premium) stock, the true conditional expected value (90, say) would lie between 80 and 100: the signal is adverse ($90 < 100$), and investors overreact to it ($80 < 90$). Thus, the true expected return is positive, but not as large as in the case of a pure risk premium. By controlling for β as well as the F/P, the econometrician can disentangle whether the price will rise to 100 or only to 90.

To understand point (3), consider now the extreme case where overconfidence is strong, in the sense that the 'signal' is almost pure noise and investors greatly underestimate this noise variance. In this case, even when the price decreases to 80 purely because of an adverse 'signal,' it will still on average recover to 100. This leads to exactly the same expected return as when the price of 80 is a result of a high beta. Both effects are captured fully and equally by F/P, whereas beta captures only the risk effect. So even though beta is priced, the F/P completely overwhelms beta in a multivariate regression.

Our reasoning is *not* based on the notion that if investors were so overconfident that they thought their signals were perfect, they would perceive risk to be zero, which would cause β to be unpriced. Even if investors are only overconfident about idiosyncratic risk, so that covariance risk is rationally priced, β has no incremental power to predict returns. Thus, the effect described here is not founded on weak pricing of risk.

The Part (3) scenario, while extreme, does offer an explanation for why the incremental beta effect can be weak and therefore hard to detect statistically. The following subsections develop these insights formally, and provide further implications about the usefulness of alternative mispricing proxies.

2.3.2 Noisy Fundamental Proxies

Consider an econometrician who wishes to forecast returns. Since prices reflect misvaluation, it is natural to include a price-related predictor. However, it is hard to disentangle whether a low price arises because $\bar{\theta}_i$, its unconditional expected payoff, is low, or because the security is undervalued. The econometrician can use a fundamental measure as a noisy proxy for the unconditional expected value. We examine here how well scaled price variables can predict returns when the fundamental proxy is the true expected

cash flow plus noise,

$$F_i = \bar{\theta}_i + e_i^F. \tag{20}$$

Here e_i^F is *i.i.d.* normal noise with zero mean, and $V^F \equiv E[(e^F)^2]$ is the variance of the error in the fundamental measure of a randomly selected security.

Suppose that the econometrician randomly draws a security, observes the fundamental-scaled price variable $F_i - P_i$, and uses this to predict the future return. We let variables with i subscripts *omitted* denote random variables whose realizations include a stage in which a security is randomly selected (*i.e.*, there is a random selection of security characteristics). This stage determines security parameters such as $\bar{\theta}$ or ω. Other random variables, such as the price, return and signal variables R, P, and S, require a second stage in which signal and price outcomes are realized.

Security expected payoffs $\bar{\theta}_i$, $i = 1, \ldots N$ are assumed to be distributed normally from the econometrician's perspective,

$$\bar{\theta} \sim \mathcal{N}(\bar{\bar{\theta}}, V_{\bar{\theta}}), \tag{21}$$

where $\bar{\bar{\theta}}$ is the cross-sectional expectation of the unconditional expected values, and $V_{\bar{\theta}}$ is the variance of $\bar{\theta}$.

We denote the moments of the factor loading distribution β_k (the loading of a randomly selected security on factor k) by $E[\beta_k], V_{\beta_k}$; of β_M (the beta of a randomly selected security with the adjusted market price change) by $E[\beta_M], V^{\beta_M}$; of ω^ϵ (the excess sensitivity of price to a unit increase in the signal about idiosyncratic risk for a randomly selected security) by $E[\omega^\epsilon], V^{\epsilon\omega}$; of ω_k^f (the excess sensitivity of prices to a unit increase in the signal about factor k) by $E[\omega^f], V^{f\omega}$ (moments assumed to be independent of k); of S_k^f (the normalized signal about factor k) by V_k^{Sf}; and of S^ϵ (the normalized signal about idiosyncratic risk for a randomly selected security i) by $V^{S\epsilon}$ (by our earlier assumptions, the last two random variables have means of zero). Further, we assume that the choice of firm is independent of the signal realization, so that signal realizations are uncorrelated with the ω's and β.[18]

Now, consider the linear projection of the security return R onto $F - P$,

$$R = a + b_{F-P}(F - P) + e,$$

[18]The restriction on exogenous parameters that achieves this is that the signal realizations be uncorrelated with factor loadings and with variances of signals and noises.

where e is mean-zero independent noise. The slope coefficient value that minimizes error variance is

$$b_{F-P} = \frac{cov(R, F-P)}{var(F-P)}. \tag{22}$$

Let

$$cov_{OC} \equiv -\left(E[\omega^\epsilon] + E[(\omega^\epsilon)^2]\right) E[(S^\epsilon)^2] - \left(E[\omega^f] + E[(\omega^f)^2]\right) \sum_{k=1}^{K} E[\beta_k^2] E[(S_k^f)^2] \tag{23}$$

$$var_{OC} \equiv \left[(1+E[\omega^\epsilon])^2 + V^{\epsilon\omega}\right] E[(S^\epsilon)^2] + \left[(1+E[\omega^f])^2 + V^{f\omega}\right] \sum_{k=1}^{K} E[\beta_k^2] E[(S_k^f)^2]. \tag{24}$$

Appendix B shows that the quantities cov_{OC} and var_{OC} are the respective contributions of the individuals' private information to the covariance and variance in equation (22). The appendix also proves the following proposition.

Proposition 2 *The regression of the return R on the fundamental-scaled price $F-P$ yields the following coefficient:*

$$b_{F-P} = \frac{\alpha^2 V^{\beta_M} - cov_{OC}}{\alpha^2 V^{\beta_M} + var_{OC} + V^F} = \left(\frac{var(\bar{\theta} - P)}{var(F-P)}\right) b_{\bar{\theta}-P}, \tag{25}$$

where

$$b_{\bar{\theta}-P} \equiv \frac{\alpha^2 V^{\beta_M} - cov_{OC}}{\alpha^2 V^{\beta_M} + var_{OC}} \tag{26}$$

is the regression coefficient when the fundamental proxy is perfect ($\bar{\theta}$ instead of F). The coefficient is positive if:

1. *Investors are on average overconfident, $E[\omega^\epsilon] \geq 0$ and $E[\omega^f] \geq 0$, with at least one inequality strict, or*

2. *Investors are rational, and not all security betas are equal, i.e., $E[\omega^\epsilon] = E[\omega^f] = 0$, and $V^{\beta_M} \neq 0$.*

The R^2 of the regression is

$$R^2_{F-P} = \frac{[cov(R, F-P)]^2}{var(R) var(F-P)} = \left(\frac{var(\bar{\theta} - P)}{var(F-P)}\right) R^2_{\bar{\theta}-P}.$$

Intuitively, overconfident investors overreact to their private signals, so a high price (low $F - P$) probably means too high a price, and consequently that expected risk-adjusted returns are negative.[19] Any cross-sectional variation in beta contributes further to the tendency of high price to predict low future returns. High beta implies low current price and a high future expected return.[20]

Our premise of overconfidence has empirically distinct implications from an approach based on underreaction. For example, in the model, if investors were on average *under*-confident then a high fundamental/price ratio would be associated with a low expected future return (if V^{ω} is not too large and if risk effects, reflected in V^{β_M}, are small). Intuitively, if investors are underconfident and hence underreact to their private signals, then a high price means that the price is likely to increase still more. Thus, the direction of fundamental/price effects is consistent with the well-known psychological bias of overconfidence.

The slope coefficient b_{F-P} in (25) is smaller than that in (26) because F_i is a noisy proxy for $\bar{\theta}_i$, so that $var(F - P) > var(\bar{\theta} - P)$; similarly, the regression R^2 is also lower. Any adjustment of the fundamental proxy F that decreases the measurement error variance improves R^2_{F-P}. One method of doing so is to adjust fundamental ratios relative to industry values. In particular, accounting measures of value differ across industry for reasons that do not reflect differences in fundamental value. For example, different businesses have differing importance of intangible assets, which are imperfectly reflected in accounting measures of value. This suggests that industry adjustment, by filtering out measurement noise that is correlated for firms within an industry, can improve the ability of fundamental measures to predict return. This is consistent with the evidence of Cohen and Polk (1995) and of Dreman and Lufkin (1996).

We have assumed that the public information signal is conclusive. Similar results would apply with noisy public information arrival. Basically, a noisy public signal only partly corrects the initial overreaction to the private signal. So high fundamental/price still indicates undervaluation (even though the public signal on average has partly corrected the market price upwards). However, as analyzed in Daniel, Hirshleifer, and Subrahmanyam (1998), in a setting with dynamic overconfidence overreactions can tem-

[19] In equation (26), $var_{OC} > 0$ by its definition in (24). If the $E[\omega]$'s are positive for all signals, then cov_{OC} is negative, so the regression coefficient is positive.

[20] Equation (26) shows that even with no overconfidence, there is still a risk-based relationship between $\bar{\theta} - P$ and R arising from risk effects. The V^{β_M} term in the numerator and denominator reflects cross-sectional variation in risk and risk premia. If there is no overconfidence, then $E[\omega]$ and V^{ω} are zero for all signals, which implies $cov_{OC} = 0$. In standard pricing models such as the CAPM, low price firms are those that are discounted heavily, *i.e.*, high beta/high return firms (see Berk (1995)).

porarily continue before eventually being corrected. If we were to combine such dynamics with our assumptions here of risk aversion and multiple risky securities, since any mis-valuation must eventually be corrected, it is intuitively reasonable to conjecture that the cross-sectional relationship between fundamental/price ratios and future returns in such a setting will still be positive.

2.4 Aggregate Fundamental-Scaled Price Variable Effects

This subsection examines the special case of a single risky security ($N = 0, K = 1$), which we interpret as the aggregate stock market portfolio. The implications for the aggregate market are obtained from the general model by deleting all i subscripts.

The model then predicts that future aggregate market returns will be predicted by value/growth variables of the form F/M, where F is a publicly observable *non*-market measure of expected fundamental value and M is aggregate market value. Four examples are aggregate dividend yield, aggregate earnings/price ratio, the aggregate book/market ratio, and the reciprocal of market value (where the numerator $F \equiv 1$ is a constant). Noisy public information may moderate, but does not eliminate this effect. Thus, market mispricing will gradually be corrected, though this correction may be slow. The model therefore explains the empirical finding of a dividend yield effect, and predicts aggregate earnings yield and book/market effects as well. It follows that the model is consistent with the profitable use of asset allocation strategies wherein arbitrageurs tilt their portfolios either toward or away from the stock market depending upon whether variables such as the market dividend/earnings yield are high or low.

3 Risk Measures versus Fundamental-Scaled Price Ratios

3.1 Size, Book/Market, and Returns

Often tests of return predictability look simultaneously at standard risk measures and possible measures of mispricing. These regressions usually involve the market β, and variables such as market value, or a fundamental-scaled variable such as book-to-market (see, *e.g.*, Fama and French (1992) and Jagannathan and Wang (1996)). Firm 'size' or market value (P) is a special case of a fundamental scaled measure in which the

fundamental proxy is a constant.[21]

In our setting, if the expected fundamental value is measured with noise, as in Section 2.3.2, the fundamental-scaled variable (or value/growth measure) is an imperfect proxy for the private signal, and both size (P) and a fundamental-scaled variable ($F_i - P_i$), in addition to the risk measure β, predicts future returns. To see this, consider the linear projection of R onto β_M, $F - P$ and P:

$$R = a + b^{\dagger}_{\beta}\beta_M + b^{\dagger}_{F-P}(F - P) + b^{\dagger}_{P}P + e. \tag{27}$$

The optimal coefficients come from the standard matrix equation (see Appendix B).

Proposition 3 *The regression of the return R on β_M, the price-scaled fundamental $P - F$, and the price P yields the following set of coefficients:*

$$b^{\dagger}_{\beta} = \alpha\left(\frac{var_{OC} + cov_{OC} + K_1}{var_{OC} + K_1}\right) \tag{28}$$

$$b^{\dagger}_{F-P} = -\frac{V_{\bar{\theta}}}{V_{\bar{\theta}} + V^F}\left(\frac{cov_{OC}}{var_{OC} + K_1}\right) \tag{29}$$

$$b^{\dagger}_{P} = \frac{V^F}{V_{\bar{\theta}} + V^F}\left(\frac{cov_{OC}}{var_{OC} + K_1}\right) \tag{30}$$

where

$$K_1 \equiv \frac{V_{\bar{\theta}}V^F}{V_{\bar{\theta}} + V^F}$$

is half the harmonic mean of $V_{\bar{\theta}}$ and V^F, and $K_1 = 0$ if either $V^F = 0$ or $V_{\bar{\theta}} = 0$. Under overconfidence, the coefficient on the fundamental/price ratio is positive and on the price is negative.

This proposition provides a theoretical motivation for the use of book/market ratios in the well-known regressions (and cross-classifications) of Fama and French (1992) and Jagannathan and Wang (1996).

If individuals are on average overconfident, $E[\omega^f] > 0$, then equation (23) shows that cov_{OC} is negative. Also, provided the fundamental measure is not perfect (*i.e.*, $V^F > 0$), these equations show that: (1) the regression coefficient on β is positive but less than

[21] P can be interpreted as either a per-share price or a total firm market value. Here we assume that θ_i is proportional to the total value of the firm. However, the analysis is equally valid on a per-share basis, and is therefore consistent with the empirical evidence that, cross-sectionally, share price is negatively correlated with future returns. A fuller analysis of this topic would include the number of shares relative to total firm value as a source of cross-sectional noise, so that firm value versus share price could have different degrees of predictive power for future returns.

α (the CAPM 'market price of risk'); (2) the coefficient on size (P) is negative; and (3) the coefficient on a $F - P$ variable (such as book/market) is positive. These results are thus consistent with the evidence of Fama and French (1992).

If the fundamental proxy is a noiseless indicator of unconditional expected returns, $V^F = 0$, then the coefficient on P is zero, because $F - P$ captures mispricing perfectly whereas P reflects not just mispricing but scale variability. Alternatively, if there is no variability in fundamental values across firms (if $V_{\bar{\theta}} = 0$), then P_i is a perfect proxy for the signal s_i, and the fundamental-scaled price variable has no additional explanatory power for future returns. Further, if a multiple regression is run with any number of fundamental-scaled price variables, such as book/market and price/earnings ratios, and if the errors of the different fundamental proxies are imperfectly correlated, so that each proxy adds some extra information about $\bar{\theta}_i$, the coefficient on each variable is non-zero. If the errors are independent, then the coefficient on price is negative and the coefficient on each regressor that contains price inversely (such as book/market) is positive.

As the variability in unconditional expected cash flows across securities becomes large, $V_{\bar{\theta}} \to \infty$, $b^{\dagger}_{P} \to 0$ and $b^{\dagger}_{F-P}$ does not. Thus, if the variance of expected cash flows across securities is large relative to the noise in the fundamental proxy, the size variable is dominated by the fundamental/price variable in the multiple regression. More broadly, when $V_{\bar{\theta}}$ is large, the coefficient on size will be less significant than the coefficient on a fundamental/price variable such as book market. Intuitively, if securities have very different expected cash flows (as is surely the case), it becomes very important to find a proxy to filter out scale variation in order to locate mispricing effects. Two recent studies find that in a multivariate regression or cross-classification, book/market is more significant than size (Fama and French (1992), Brennan, Chordia, and Subrahmanyam (1998)); see also Davis, Fama, and French (2000).

The model implies that future returns should be related to market value, but not to non-price measures of size. This is because such measures, exemplified by the number of employees or book value, are unrelated to the error in the informed's signal e_i, and are therefore also unrelated to the future return on the security. This is consistent with the empirical findings of Berk (2000).

The analysis suggests that the relationship between book/market and future returns is a valid one rather than an *ex post* relationship arising from data-snooping. However, there is no implication that there exists any meaningful book/market factor, nor that sensitivities with respect to a factor constructed from book/market portfolios can be used to price assets. In this regard, our analysis is consistent with the evidence of Daniel

and Titman (1997) that the book/market effect is associated with the book/market characteristic, not with the risk of an underlying factor (distinct from the market return). Specifically, our analysis suggests that book/market captures a combination of market risk and mispricing.[22]

More broadly, Fama and French (1993) offer evidence which they, and many subsequent papers, have interpreted as supporting a rational asset pricing explanation for the cross-section of security returns over a psychological interpretation. Fama and French's argument is based on three pieces of evidence. First, there is comovement among value and growth stocks, in the sense that the HML portfolio (a zero-investment portfolio that is long in high book/market stocks and short in low book/market stocks) has a variance considerably greater than zero. This contrasts with the close-to-zero variance that would obtain for a zero investment portfolio of randomly picked stocks, because such a portfolio would be well-diversified and have approximately zero loadings on economic factors. Second, the HML and SMB portfolios load on factors other than the market: when the Fama/French HML and SMB returns are regressed on the market portfolio, the R^2's are not close to one. Third, using a Gibbons, Ross, and Shanken (1989) regression method, Fama and French find that a three factor model consisting of the market, HML, and SMB does a fairly good job of explaining the cross-section of expected securities returns.

These empirical results are consistent with standard rational asset pricing theory,[23] but they are also consistent with explanations based on imperfect rationality. In our single period model, there is a risk premium only for market factor risk; all other factors have zero net supply and hence rationally command no risk premium. Nevertheless, as we explain below, our model is entirely consistent with Fama and French's three empirical observations. Thus, the Fama and French tests cannot discriminate between a risk premium theory of value and size effects and a mispricing theory, such as the model explored here.

Intuitively, as pointed out by Daniel and Titman (1997), the fact that HML and SMB portfolios seem to capture factors other than the market in itself says nothing about whether these factors are rationally priced: in our setting, stocks with extreme book/market ratios (value or growth stocks) tend to be stocks that have extreme loadings

[22]To see this, consider the special case in which there is just a single factor (essentially the market) and in which there is no information about the factor, only about idiosyncratic security return components. Clearly there is no book/market factor as distinct from the market factor. Nevertheless, a price-scaled fundamental (such as book/market) predicts future returns, and may dominate beta.

[23]In a rational single-period CAPM setting such zero-net-supply factors would be unpriced. Fama and French give an intertemporal hedging interpretation for the factors.

on those factors about which people have received extreme signal values. Thus, zero investment portfolios formed by investing equal long and short amounts in stocks with high and low book/market, respectively ("HML" portfolios) have systematic risk because such portfolios tend to pick out securities which load heavily on mispriced factors. In our model, there is no risk premium for these non-market factors, because they do not contribute to aggregate risk. Nevertheless, HML loadings predict future returns, because a security's loading on HML is a proxy for mispricing. Thus, the Fama/French tests cannot distinguish between the null hypothesis of rational asset pricing and the alternative hypothesis of inefficient asset prices.

Daniel, Hirshleifer, and Subrahmanyam (2000) derive the above implications formally, and examine regressions of future returns on both book/market and on the HML loadings (as examined empirically in Daniel and Titman (1997)). They show that if there is cross-sectional correlation in the factor loadings of different stocks on different factors, then the book/market characteristic will predict future returns better than HML loadings,[24] but under other conditions HML loadings may be better forecasters of future returns.

The arguments developed in Daniel, Hirshleifer, and Subrahmanyam (2000) suggest that whether characteristics or B/M better forecasts future returns may be related to issues other than whether assets are rationally priced. In the end the only definitive test discriminating between the two stories may be to see whether HML's realized returns covary with variables that proxy for investors' marginal utilities (as must be the case in the rational risk premium story). As discussed in the appendix, so far there is little evidence supporting such a link.

3.2 Biases in the Estimation of Beta Pricing

Proposition 3 implies not only that size and fundamental/price predict returns, but that the inclusion of these variables in a return-predicting regression weakens the effect of beta. To see this, note by equation (18) of Proposition 3 that α is the coefficient in the projection of the expected return on market beta. So by equation (28), overconfidence attenuates the effect of beta in a cross-sectional return regression that includes fundamental/price ratios or size. However, the low coefficient on beta does *not* mean that risk is weakly priced. Variables such as market value or F/P are proxies for misvaluation

[24] For example, auto firms may load positively on an auto industry factor and negatively on oil price and steel price factors. Such similarity across firms in their loadings illustrates correlatedness of factor loadings.

as well as risk. Because of misvaluation, high F/P predicts high returns. The resulting positive coefficient on F/P causes this variable to capture some of the risk effect. This biases downward the coefficient on beta. Thus, the correlation of the regressors causes regressions to underestimate the effect of risk.

This spurious weakening of the beta effects is most intense when overconfidence is high, as illustrated by the following Corollary.

Corollary 2 *If the fundamental measure is noiseless, $V^F = 0$, if the expected tendency to overreact approaches infinity, $E[\omega^e], E[\omega^f] \to \infty$, and variability in overconfidence (as reflected in $V^{e\omega}$ and $V^{f\omega}$) remains finite then the regression coefficient on β, $b^\dagger_\beta$ in Proposition 3 approaches zero, whereas the coefficient on the fundamental-scaled price approaches unity.*

The limiting case of infinite tendency to overreact occurs with λ^C constant and $\lambda^R \to 0$. In other words, signals become close to pure noise ($V^R \to \infty$) and investors drastically underestimate this noise.[25] Alternatively, λ^C could approach infinity, but this would lead to infinitely volatile prices. In the limit, trading strategies can be viewed as noise trading, because nontrivial price revisions are triggered by very little information. As this occurs, the coefficient on β approaches zero, and the coefficient on $F - P$ approaches 1.

To understand the intuition for the corollary, consider first the intuition for why book/market and beta jointly predict returns in the completely general case. When the fundamental/price ratio is high (low value of market price relative to unconditional expected terminal cash flow) price should on average rise (either because overreaction is reversed, or because of a risk premium). However, the amount of the expected rise differs depending on whether the low price arose from adverse private information, or from high risk. In particular, if price is low because of a risk premium, on average it rises to the unconditional expected terminal value. If, on the other hand, price is low because overconfident traders overreacted to adverse information, then the conditional expected value is below the unconditional expected value. As pointed out earlier, regressing on both beta and fundamental/price ratios helps disentangle these two cases.

When overconfident individuals trade based on pure noise, however, the conditional expected value of θ is equal to the unconditional value, so there is no predictive power to

[25] As signals become close to pure noise, the rational reaction to a signal becomes close to zero. By equations (8) and (9), if the consensus signal precision approaches zero at the rate of the square root of the rational precision, price volatility is asympotically proportional to 1 (*i.e.*, it approaches neither zero nor infinity).

be gained by disentangling the cases. The fundamental/price ratio completely captures both risk and mispricing, but beta only captures risk. Thus, in this case risk measures such as β provide no incremental explanatory power for future returns.

The above discussion implies that if investors' private signals are very noisy, the statistical relationship between fundamental/price ratios and expected return will be strong, and the relation between β and expected return will be weak. Thus, the theory is consistent with the differing findings of several studies regarding the existence of a cross-sectional relation between return and β after controlling for book-to-market or for market-value (see Appendix A).

4 Variations in Confidence: Empirical Implications

There is evidence that individuals tend to be more overconfident in settings where feedback on their information or decisions is slow or inconclusive than where the feedback is clear and rapid (Einhorn (1980)). Thus, mispricing should be stronger for businesses which require more judgment to evaluate, and where the feedback on the quality of this judgment is ambiguous in the short run. This line of reasoning suggests that fundamental/price effects should be stronger for businesses that are difficult to value, as with high-tech industries (as measured by high R&D expenditures) or industries (*e.g.* service industries) with high intangible assets.[26] Subsequent to our developing this prediction, Chan, Lakonishok, and Sougiannis (1999) investigated this question, and have reported evidence consistent with this empirical implication of the model.[27]

A further implication of our approach is based on the notion that market overconfidence can vary over time. To develop this implication, we begin by stating the following corollary (proved in Appendix B).[28]

[26]A low book/market ratio is itself an indicator of high intangible assets, but can also be low for other reasons such as a risk premium or market misvaluation. Thus, conditioning on other intangible measures provides a test of how intangible measures affect the misvaluation-induced relation between fundamental-price ratios and future returns.

[27]Their Table 3 sorts firms into five groups based on R&D expenditures relative to sales. They then sort each of these portfolios into high and low sales/market firms. Sales/market is a fundamental/price variable here. The average return differential between the high and low sales/market firms for the low R&D firms is 3.54% per year over the three post-formation years. For the high R&D firms, this differential is 10.17% per year. This evidence indicates that the fundamental/price ratio effect is far higher for the high R&D firms then for the low R&D firms.

[28]We derive this corollary for the conditions under which Corollary 2 holds, but the intuition seems more general.

Corollary 3 *If the conditions in Corollary 2 are satisfied, then as the average level of confidence ($E[\omega^e]$ and $E[\omega^f]$) increases, (1) the cross-sectional variance in fundamental/price ratios increases without bound, and (2) the R^2 of the regression coefficient of returns on fundamental/price ratios approaches 1.*

Treating shifts in confidence as exogenous, periods of high overconfidence tend to be associated with extreme values for aggregate stock market fundamental/price ratios (dividend yield, book/market, and so on). Corollary 3 suggests that such periods of high confidence (very high or very low fundamental/price ratios) should be associated with greater cross-sectional dispersion in both fundamental/price ratios and in return, and with a stronger relation between the two.

5 Market Volume and Future Volatility

We now examine the relationship between the market's current trading volume and its future return volatility. Consider the special case of model in which $N = 0, K = 1$. We call this single security the market portfolio; it has no idiosyncratic risk. Let

$$R = \gamma s + \eta$$

be the theoretical regression of the future (date 1-2) market return on the signal about the market at date 1. Under overconfidence, $\gamma < 0$, so $-\gamma$ is an index of overconfidence.

To calculate γ, use (5) (but with mean $\bar{\theta}$ restored), giving the market's price as

$$P = \bar{\theta} + \Pi s - \frac{A}{\nu + (1-\phi)\nu^R + \phi\nu^C},$$

where

$$\Pi \equiv \frac{(1-\phi)\nu^R + \phi\nu^C}{\nu + (1-\phi)\nu^R + \phi\nu^C}.$$

The signal is $s = \bar{\theta} + f + e$. For convenience and without effect on the results, we set the market's sensitivity to the factor to be one, so $R = \bar{\theta} + f - P$, and the regression coefficient γ is

$$\begin{aligned} \gamma &= \frac{cov(R,s)}{var(s)} \\ &= \frac{\phi(\nu^R - \nu^C)}{(1+\nu^R)[1+\nu^R+\phi(\nu^C-\nu^R)]}. \end{aligned}$$

Assume that all individuals begin with identical endowments. We refer to the overconfident/informed as those who observe the signal s. Let the signed turnover be defined as twice the difference between the total position of the overconfident/informed traders and their endowment ξ:

$$\text{Signed Turnover} = 2\phi(x^C - \xi).$$

By (4),

$$\begin{aligned} 2\phi(x^C - \xi) &= \frac{2\phi}{A}(\nu + \nu^C)(\mu^C - P) \\ &= \kappa(s + A\xi), \end{aligned} \tag{31}$$

where

$$\kappa \equiv -\frac{2(1-\phi)(\nu + \nu^R)\gamma}{A\nu}.$$

The effect we wish to focus on is signal-induced volume. Because overconfidence causes underestimation of risk, the overconfident/informed trade with the arbitrageurs in the mistaken belief that this improves risk-sharing. This effect is nonstochastic. We therefore deduct from turnover the risk-shifting induced component $\kappa A\xi$ on the RHS of (31), the aggregate pure risk-shifting position the overconfident informed would have taken if $s = 0$. This gives the *signal-induced residual turnover*,

$$2\phi(x^C - \xi) - A\kappa\xi = \kappa s.$$

We define the *absolute signal-induced turnover*, or *volume*, as

$$X \equiv \kappa|s|.$$

Then by symmetry,

$$E[R|X] = \frac{A}{\nu + (1-\phi)\nu^R + \phi\nu^C},$$

which is independent of X and the same as the unconditional expected return. Intuitively, each value of X corresponds to two equiprobable signal values, one of which is associated with high future return and the other with low.

Conditional return volatility is

$$\begin{aligned} \text{var}(R|X) &= E[R^2|X] \\ &= E[(\gamma s + \eta)^2|X] \\ &= \left(\frac{\gamma}{\kappa}\right)^2 X^2 + \sigma^2(\eta). \end{aligned}$$

Thus, conditional volatility increases quadratically in absolute turnover X.

The above arguments lead to the following proposition:

Proposition 4 *1. The market's future volatility is an increasing quadratic function of its current volume.*

2. A greater mass of overconfident/informed individuals uniformly increases the absolute slope of this relationship.

3. A strong relation between volatility and volume is associated with a strong relation between the market fundamental/price ratio and future returns.

Part 2 of the proposition follows because the greater is the mass of the overconfident/informed, the greater is the ratio γ/κ. Thus the greater the mass of overconfident agents, the stronger is the convex relationship between future volatility and current volume.[29]

Part 3 follows from Proposition 2 applied to the special case of a time-series regression for a single security whose β with the market is known with certainty (= 1), and for which overconfidence is a constant. This eliminates the V^{β_M} terms in equation (26), and sets $K = 1$, $\omega^\epsilon \equiv 0$, $V^{\epsilon\omega} = 0$, $E[\omega^f] = \omega$ (a constant), $E[\omega^2] = \omega^2$, and $V^{f\omega} = 0$ in equations (23) and (24). It follows from the proposition that the coefficient of the fundamental price ratio is increasing in ω. From the definition of ω (see equation (12), λ (see equation (8)) and ν^A (see equation (6)), ω is increasing in ϕ, the fraction of informed/overconfident investors. Thus, Part 3 obtains. Treating shifts in confidence as exogenous, this suggests that during time periods where a stronger convex relationship between volatility and volume obtains, the effect of fundamental/price ratios on price should be particularly strong.

Most past empirical studies of volume and volatility focus on relatively shorter horizons of days or weeks. However, empirical evidence such as fundamental-scaled price variable effects and return reversals suggest that these effects are important at horizons of several years. Thus, the volatility and volume effects that we describe here are predicted to occur at these longer horizons.[30]

[29] Our analysis focuses on volume arising from traders receiving information and taking positions based on this information. In a dynamic setting, it would be important to take into account volume generated by unwinding of trades. Intuitively, such volume will be unrelated to future volatility, because it is not indicative of any *current* market price overreaction. This will make volume a noisier predictor of future volatility, but will not reverse the effect identified here.

[30] Numerous theoretical papers have analyzed determinants of volume and volatility. However, most

Reasoning very similar to the above demonstrates that Proposition 4 also extends to the relation between securities' idiosyncratic volume and their idiosyncratic volatility. This relationship holds even when the number of securities becomes large. However, with many securities, the fraction of idiosyncratically informed traders per security becomes very small, so that information-induced volatility becomes extremely small (see Section 6 below). The result for individual securities is therefore not very interesting in our setting. A broader interpretation of our approach is that market imperfections limit arbitrage of idiosyncractic mispricing. This suggests that there could be a substantial fraction of idiosyncratically informed investors who substantially influence price. In such a broader setting we conjecture a result similar to that in Proposition 4 will apply, relating idiosyncratic volume to idiosyncratic volatility for individual securities.

6 Profitability of Trading, Mispricing, and Diversified Arbitrage Strategies

This section analyzes the profitability of trading for the overconfident informed, and the extent to which arbitrageurs can eliminate mispricing through diversified trading strategies.[31,32] To allow for diversified trading strategies, we analyze what occurs in the limit as the number of available securities becomes large.

An argument sometimes put forth in favor of efficient markets is that if a large number of securities were mispriced, a portfolio that is long on underpriced stocks and short on overpriced stocks would diversify risk and thus achieve very high Sharpe ratios. This would imply very high volume of trade and a very large flow of wealth from imperfectly

focus on a contemporaneous relation between the two. Odean (1998) and Daniel, Hirshleifer, and Subrahmanyam (1998) theoretically examined the relation of overconfidence to volatility and/or volume. However, in Odean's single-period setting volatility is non-stochastic; and Daniel, Hirshleifer, and Subrahmanyam (1998) did not focus on volume. In contrast, here we examine how current volume is correlated with stochastic shifts in subsequent long-horizon volatility. Empirical researchers have identified a contemporaneous volume/volatility relation (see, *e.g.*, Daigler and Wiley (1999)). We are not aware of any empirical work relating variations in volume to subsequent volatility at a long horizon.

[31]Several previous papers have argued that there are limits to the degree that arbitrage reduces market inefficiencies, and that imperfectly rational or overconfident traders can earn higher expected profits than fully rational traders and therefore can be influential in the long run; see, *e.g.*, De Long, Shleifer, Summers, and Waldmann (1990a), Kyle and Wang (1997), and Shleifer and Vishny (1997). Also, even if overconfident traders make less money, it may also be the case that those traders who make the most money become more overconfident (see Daniel, Hirshleifer, and Subrahmanyam (1998) and Gervais and Odean (1999)).

[32]As described earlier, these arbitrageurs can be viewed either as fully rational uninformed traders, or as fully rational informed traders, or as overconfident uninformed traders, without affecting the analysis.

rational traders to arbitrageurs. Thus, it is argued that such mispricing should not exist in equilibrium. This conclusion does not follow in our setting. Based on a reasonable specification of information arrival (see below), we show that the mispricing of most residuals approaches zero, but factor mispricing does not.

In order to analyze mispricing a limiting economy with a very large number of securities, we make the following assumptions. First, we assume that each individual receives a finite number of signals about firm-specific payoffs. Implicitly this reflects the notion that it is costly to obtain information about a very large number of firms. Second, each individual receives at least one factor signal, with no factor shunned by all individuals. This reflects the intuitive notion that even if individuals were to study only individual firms, since the aggregate market is the sum of its constituent firms, such study provides information about the fixed set of market factors. Third, we assume that who observes what factor is independent of the number of securities. Assumptions 2 and 3 imply that as the number of securities N grows, a fixed positive fraction of the population continues to receive a signal about any given factor. Assumption 3 is not important for the result, but allows simple presentation. However, assumption 1 is important, because it leads to a finite volume of trade even though there is disagreement between investors who are overconfident about their information about residuals and arbitrageurs who trade against them.

Under the above specification, although all securities are mispriced, and trading against mispricing is profitable, there are no riskfree arbitrage opportunities, and volume of trade remains finite. Indeed, for the reasons analyzed by De Long,Shleifer, Summers, and Waldman (1990a, 1991) (DSSW), the overconfident informed in our setting may make as much or greater expected profits than the arbitrageurs. Intuitively, if people believe they have information about a 'new economy' factor (for example), they may misprice an entire industry that loads upon this factor. Playing an arbitrage game based on such a mispriced factor could be profitable, but is certainly risky.

Using the equivalent setting presented in Subsection 2.1.4, we now formalize the above arguments. We analyze on a security-by-security basis how much an arbitrageur gains as a result of the overconfidence of others in relation to the relative numbers of arbitrageurs versus overconfident traders. We then examine overall portfolio profitability, and examine how profitability changes as the number of available securities becomes large.

We compare the expected return of an individual who is overconfident about every security to that of an arbitrageur whose beliefs about all securities are rational. The

expressions derived generalize easily to give the expected return of an individual who receives (and is overconfident about) private information on only a subset of securities. The true expected returns (based on rational beliefs) of an investor's optimal portfolio can be derived from (4) by substituting for expected returns from equation (11), taking expectations over s_i, and simplifying:

$$E^R[R_p^R] = \sum_{i=1}^{N+K} x_i^R E^R[R_i] = \frac{1}{A}\sum_{i=1}^{N+K}\left[\phi_i^2 C_i + \frac{(\nu_i+\nu_i^R)}{(\nu_i+\nu_i^A)^2}(A\xi_i)^2\right] \tag{32}$$

$$E^R[R_p^C] = \sum_{i=1}^{N+K} x_i^C E^R[R_i] = \frac{1}{A}\sum_{i=1}^{N+K}\left[-\phi_i(1-\phi_i)C_i + \frac{(\nu_i+\nu_i^C)}{(\nu_i+\nu_i^A)^2}(A\xi_i)^2\right], \tag{33}$$

where

$$C_i \equiv \frac{(\nu_i^C-\nu_i^R)^2\nu_i^2}{\nu_i^R(\nu_i+\nu_i^R)(\nu_i+\nu_i^A)^2}. \tag{34}$$

The first term in the brackets in equations (32) and (33) is the expected return gain (or loss) that results from the trading on the overconfidence-induced mispricing. The population-weighted sum of the mispricing terms (with weights ϕ_i and $1-\phi_i$) is zero, *i.e.*, the mispricing results in a wealth transfer from the overconfident/informed to the arbitrageurs.

The second term in the brackets above is larger for the overconfident/informed than for the arbitrageurs (because $\nu_i^C > \nu_i^R$). Since an overconfident informed individual underestimates risk, he holds a larger position in riskier assets than does an arbitrageur, and thereby captures a greater risk premium (the right-hand-side term in brackets); see DSSW. Thus, whether the overconfident make more or less money than would a rational 'arbitrageur' depends on a balance of effects. There is no presumption that rational traders will drive out overconfident ones.

Mathematically, our assumption that each individual observes only a finite number of signals implies that for all but a finite number of the N securities, the fraction of individuals who are informed about the residual component is $O(1/N)$ (*i.e.*, the fraction approaches zero is proportional to $1/N$). This does not preclude a finite number of firms receiving a great deal of attention; for example, even if there are many internet stocks, a few like Amazon or Yahoo may garner attention from a substantial portion of the public. In contrast, the assumption that everyone observes a factor signal (and no factor is shunned) in a fixed observation structure implies that as N varies the fraction of individuals informed about any given factor is a positive constant.

Now, equation (10) can be rewritten as:

$$E^R[R_i] = -\phi_i \left(\frac{\nu_i^C - \nu_i^R}{\nu_i + \nu_i^R} \right) \left(\frac{\nu_i}{\nu_i + \nu_i^A} \right) s_i + \left(\frac{A}{\nu_i + \nu_i^A} \right) \xi_i$$

giving the expected return for each factor or residual portfolio. The first term is the extra return from mispricing, and is proportional to ϕ_i. When the fraction of overconfident informed investors is small, the rational investors compete to drive away almost all mispricing. This removes the component of return deriving from correction of mispricing (the term multiplied by s_i). When the number of securities N is large, and since for residuals, $\phi_i^\epsilon = O(1/N)$, the mispricing term tends to zero at the rate $1/N$. In contrast, factor mispricing is *not* arbitraged away; for factors, $\phi_i^f = O(1)$, *i.e.*, ϕ_i^f does not approach zero as N becomes large.

Consider how much an arbitrageur gains as a result of mispricing. Rewriting equation (32) in terms of factors and residuals gives

$$E^R[R_p^R] = \frac{1}{A} \sum_{i=1}^{K} (\phi_i^f)^2 C_i + \frac{1}{A} \sum_{i=K+1}^{N+K} (\phi_i^\epsilon)^2 C_i + \frac{1}{A} \sum_{i=1}^{N+K} \frac{(\nu_i + \nu_i^R)}{(\nu_i + \nu_i^A)^2} (A\xi_i)^2. \tag{35}$$

The last term is the risk premium. The second term is the portfolio return gain for the arbitrageur resulting from mispricing of residuals, and is proportional to $(\phi_i^\epsilon)^2$. Since $\phi_i^\epsilon = O(1/N)$, and there are N residuals, this term approaches zero at the rate $1/N$ (*i.e.*, it is $O(1/N)$) so long as the cross-sectional variation in C_i is not too large (*e.g.*, the C_i's are bounded above by a finite number):

$$\frac{1}{A} \sum_{i=K+1}^{N+K} (\phi_i^\epsilon)^2 C_i = O\left(\frac{1}{N}\right).$$

The arbitrageur's portfolio variance remains above zero, even in the limit as $N \to \infty$, so the Sharpe ratios of the arbitrageurs' portfolios do not explode, even for large N.

Despite disagreement, *per capita* volume of trade in each security remains finite. Under the assumption of Section 5 that the date 0 endowments of all agents are identical (and therefore equal to ξ_i), an arbitrageur's expected date 1 trade in security i is

$$E[x_i^R - \xi_i] = -\phi_i \left(\frac{\nu_i^C - \nu_i^R}{\nu_i + \nu_i^A} \right) \xi_i.$$

This is negative because, on average, an overconfident individual underestimates the risk of a security he has information about, and therefore tends to purchase it.

Beyond this non-stochastic risk-sharing component of volume is the volume induced by the value of the signal realization. The amount of this stochastic volume is measured by the variability of the trade. The variance of the trade is obtained by substituting equations (8) and (11) into equation (4), and taking expectations over s_i:

$$var(x_i^R - \xi_i) = \phi_i^2 \nu_i \left(\frac{\nu_i + \nu_i^R}{\nu_i^R} \right) \left(\frac{\nu_i^R - \nu_i^C}{A(\nu_i + \nu_i^A)} \right)^2 = O\left(\frac{1}{N^2} \right)$$

for residual (but not factor) portfolios. In the limit, the variability of the arbitrageur's trade in each residual portfolio approaches zero. So does the risk-shifting component of volume mentioned earlier. It follows that expected per-capita volume $E(|x_i^R - \xi_i|)$ also approaches zero for residual portfolios. Intuitively, as the number of securities grows large, if each individual only receives information about a finite number of securities, residual mispricing becomes very small so arbitrageurs take vanishingly small bets on each residual. Thus, even with many securities the volume of trade based on residual mispricing remains bounded. We summarize this analysis as follows.

Proposition 5 *Suppose that C_i as defined in (34) is bounded above, that each individual receives a finite number of signals about firm-specific payoffs, that each individual receives at least one factor signal with no factor shunned by all individuals, and that who observes what factor is independent of the number of securities. Then as the number of securities N grows large:*

1. *The idiosyncratic mispricing of all but a finite number of residuals approaches zero.*
2. *The systematic mispricing of each of the K factors is bounded above zero*
3. *Per capita volume in every residual approaches zero.*

It is worth emphasizing that the model does not rule out severe residual mispricing among a small set of assets. The market can make extreme errors about a few particular firms, but arbitrageurs will still have to bear substantial risk to remove this mispricing completely. The removal of large mispricing for all but a finite set of firms when there are many securities is analogous to a similar result of the Arbitrage Pricing Theory (see Ross (1976), Chamberlain and Rothschild (1983)) and is based on essentially the same reasoning.

Going beyond the formal model, transaction costs of trading could limit arbitrage activity enough to allow substantial average idiosyncratic mispricings to persist. For example, if a set of residuals are underpriced, an arbitrageur must buy the underpriced

securities and short the correct amount of each of the relevant factor portfolios, all of which are constructed from many securities. In a dynamic world with evolving factor sensitivities, the weights on securities within each portfolios and the weights placed upon each of these portfolios would have to be readjusted each period. Taking into account the cognitive costs of identifying mispriced securities and of calculating optimal arbitrage strategies would widen the bounds for possible mispricing. Thus, the conclusion that average idiosyncratic mispricing nearly vanishes may be sensitive to our assumptions of perfect markets and near-perfect rationality. In contrast, the conclusion that factor mispricing persists is robust with respect to plausible model variations.

7 Conclusion

This paper offers a multisecurity theory of asset pricing based on investor risk aversion and overconfidence. In the model covariance risk and mispricing jointly determine the cross-section of expected security returns. Several insights derive from this approach:

- The effects of risk and mispricing in our model separate additively into a 'beta' term and a set of 'mispricing' terms, where factor mispricing is inherited by securities according to their factor sensitivities. Thus, in general, beta and price-related misvaluation measures jointly predict future returns. The inclusion of price-related variables weakens the beta effect. Furthermore, when investors are overconfident about pure noise and fundamental measures are imperfect, fundamental/price ratios completely dominate beta– even though covariance risk is priced. These results are consistent with the joint effects found in several empirical studies.

- The usual size and book/market effects do not follow from arbitrary specification of investor irrationality. Rather, they are consistent with overconfidence, an information-processing bias documented by research on the psychology of the individual. This pervasive bias implies stock market overreaction and correction. Underconfidence would reverse the sign of these effects, inconsistent with the empirical evidence.

- The analysis provides a conceptual basis for choosing between alternative measures of mispricing as predictors of future returns. Normalizing price by a fundamental measure such as book value or earnings helps filter out variations in market value that arise from differences in scale rather than mispricing. However, a fundamental

measure such as book value measures scale (unconditional expected payoffs) with error, and thereby introduces its own noise. Adjusting a mispricing measure by examining deviations from industry levels can filter out industry-related noise in the fundamental measure, at the cost of filtering out some industry-level mispricing as well.

In addition, the analysis suggests that a constructed accounting index may be superior to scale proxies such as dividends, earnings or book value if it can provide a better estimate of unconditional expected value. When combined with a mispricing proxy such as market price, such an index may be a better predictor of future returns than noisier variables such as size, price/earnings, or book/market. The findings of Abarbanell and Bushee (1999), Frankel and Lee (1996), and Chang, Chen, and Dong (1999) suggest that such an approach can be effective.

- The theory provides additional empirical implications regarding the strength of fundamental/price (*i.e.*, value/growth) effects in stocks that are difficult to value, about the cross-sectional dispersion and predictive power of fundamental/price ratios as a function of aggregate levels of fundamental/price ratios, and about the relation of volume to subsequent returns. One of these implications has received *ex post* confirmation (see Section 4), but most remain to be tested.

- When there are many securities, owing to the activities of 'risk arbitrageurs', misvaluations of most idiosyncratic components of security payoffs approach zero. In contrast, misvaluation of industry or market-wide factors persists. In our model size and value funds can be built to exploit factor mispricing. Such funds do not provide arbitrage profits because they load on systematic risk factors.

It is likely that the relative numbers of overconfident individuals and fully rational 'arbitrageurs' varies with stock characteristics such as liquidity and firm size. Explicit modeling of market imperfections, such as fixed setup costs of trading in a given security, may offer further implications for the cross-section of expected security returns.

Our focus has been on fundamental/price ratios as proxies for market mispricing, but the pricing model accomodates other possible proxies. For example, publicly disclosed insider trading may be a proxy for market misvaluation if insiders trade against mispricing. Lakonishok and Lee (1998) provide evidence that imitation of insider trades for up to about two years after disclosure is a profitable strategy even after controlling for size and book/market. A further set of possible proxies for market misvaluation in-

volve corporate actions such as aggregate new-issue versus repurchase activity. Indeed, Loughran and Ritter (1995) explicitly propose that managers time new issues in order to exploit market misvaluation.

It would be interesting to extend our approach to address the issues of market segmentation and closed-end fund discounts, in the spirit of De Long, Shleifer, Summers, and Waldmann (1990a), and Lee, Shleifer, and Thaler (1991)). In the pure noise approach, discounts reflect mispricing and therefore forecast future stock returns. In our approach, since the mispricing arises from overreaction to genuine information, changes in fund discounts should predict not just future stock performance, but also future fundamentals such as accounting performance. Swaminathan (1996) finds such predictive power for future fundamentals, which he interprets as tending to support a rational risk premium hypothesis as opposed to a noise/sentiment approach. His evidence at lags of greater than one year is surprising, because high discounts predict both low future accounting profits and high future stock returns. This evidence is consistent with an overconfidence approach, wherein genuine adverse information is associated with large discounts and low future profitability, yet high future stock returns as the market corrects its overreaction.

An important question is whether the misvaluation effects identified in our model should persist in the long run. We mentioned models in which biased learning can cause traders, based on experience, to become more overconfident instead of converging toward rationality. In our model (see Section 6), the overconfident can make greater expected profits than rational traders, a possibility demonstrated in several papers cited earlier.

Stepping beyond the model, suppose initially arbitrageurs are not sure whether there are overconfident traders in the market, and that some sort of noise prevents an arbitrageur from inferring instantly and perfectly the information of overconfident traders. Over time, by statistical analysis of the history of fundamentals and prices, arbitrageurs will learn that other players were in fact overconfident. This encourages more aggressive contrarian trading strategies. (One might interpret these arbitrageurs as 'quants'.) Thus, one interpretation of the high predictability of stock returns over the last several decades is that some investors are overconfident, and this was not fully recognized by other investors who could have exploited this. This interpretation suggests that as arbitrageurs' expectations become more accurate, anomalous predictability of returns should diminish but not vanish. Going one step further, however, arbitrageurs themselves could be overconfident about their abilities to identify statistical patterns, and could be too attached to the patterns they have identified. If so, then mispricing effects

could fluctuate dynamically over time.

More broadly, an investor may become overconfident about a theory of how the economy or the stock market works (*e.g.*, "it is profitable to buy on the dips") rather than about the realization of a signal. We expect such overconfidence if the investor has originated the theory himself or if he thinks that he is among a set of investors with superior analytical ability who were smart enough to adopt the theory. An important avenue for further research is to study what determines how different groups of investors fasten their overconfidence upon one analysis versus another, and how the social process of theory adoption influences security prices. See Shiller (2000) for pioneering work in this field; much remains to be done.

Appendices

A The Literature on Securities Price Patterns

A positive univariate relation of beta with expected returns is found in some studies and not others, depending on the country and the time period. Empirically, an incremental effect of beta after controlling for market value or fundamental/price ratios is found in some studies but not others.

Black, Jensen, and Scholes (1972) and Fama and MacBeth (1973) provide evidence of a significant positive univariate relation between security betas and expected returns. Both studies find significant time variation in this effect. In a more recent sample Fama and French (1992) also find a positive but insignificant unconditional relation between return and market beta. Internationally, Rouwenhorst (1999) finds no significant unconditional relation between average return and beta, relative to a local market index, on common stocks from 20 emerging markets. Heston, Rouwenhorst, and Wessels (1999) find some evidence of an unconditional univariate relation between market beta and future returns across stocks in 12 European countries.

On the incremental importance of conventional risk measures versus fundamental-scaled price variables, Fama and French (1992) find that size and book/market predict future returns, and that when firms are sorted simultaneously by β and size, or by β and book/market, β has no power to explain cross-sectional return differences. However, in contrast to the Fama and French (1992) results, Jagannathan and Wang (1996) find that the incremental effect of beta on future returns is significant when human capital is included in the definition of the market, and conditional rather than unconditional betas are calculated. Knez and Ready (1997) present evidence that market β is priced after controlling for size if robust test statistics are used. Heston, Rouwenhorst, and Wessels (1999) find that size and international market beta are both positively associated with future returns in 12 European countries.

There is strong evidence from numerous studies that firm size as measured by market value predict future returns. This predictive power vanishes when size is measured by book value or other non-market measures (see Berk (2000)).

Fama and French (1993) provide evidence that a three-factor model explains the average returns of stocks sorted on market equity and book/market ratio, which they interpret as a model of equilibrium risk premia. However, Daniel and Titman (1997) argue that the Fama and French (1993) results are also consistent with a 'characteristics' model, and present evidence that, after controlling for the size and book/market ratios, returns are not related to loadings on the Fama and French (1993) factors. Jagannathan,

Kubota, and Takehara (1998) find some evidence that, in Japan, both factor loadings and characteristics determine future returns.

Furthermore, MacKinlay (1995) finds evidence that high Sharpe-ratios (relative to the market) can be achieved with strategies based on fundamental-scaled price variables. As Hansen and Jagannathan (1991) point out, high Sharpe-ratios are only possible in a rational asset pricing model when there is highly variable marginal utility across states. Brennan, Chordia, and Subrahmanyam (1998) show that these strategies produce Sharpe ratios about three times as high as what is achievable with the market, and argue, like MacKinlay, that these are too high relative to the market Sharpe ratio to be plausible within a rational, frictionless asset pricing model. Moreover, as pointed out in Hawawini and Keim (1995), the returns from these strategies have very low correlations across international stock markets, meaning that the achievable Sharpe-ratio with a globally diversified portfolio, and the implied variation in marginal utility, would have been still higher.

The ability of fundamental-scaled price variables to predict cross-sectional differences in future returns is confirmed by numerous studies. Jaffe, Keim, and Westerfield (1989) find that the ratio of earnings to price has predictive power for the future cross-section of returns. Rouwenhorst (1999) finds evidence that firm size and fundamental-scaled price measures predict returns on common stocks from 20 emerging markets. He finds little correlation between book/market- and size-sorted portfolios across 20 countries.

Several earlier studies also find evidence of size and market-to-book effects (see, *e.g.*, DeBondt and Thaler (1987)). Davis (1994) finds that the book/market effect is present in pre-COMPUSTAT US common-stock returns, and Fama and French (1998) find evidence of an international book/market effect in the 1975-95 period. Fama and French find that other fundamental-scaled price variables also have power to forecast the future cross-section of returns, but that these other variables have no predictive power over and above book/market and size.

For the size or the book/market ratio of a firm to be a good proxy for risk, the returns of small and high book/market firms' stocks would have to be negatively correlated with marginal utility, meaning the returns should be particularly high in good times (relative to other stocks) and low in bad times. No such correlation is obvious in the data. [See, for example, Lakonishok, Shleifer, and Vishny (1994).] Also, fundamental-scaled price variables may be related to the liquidity of a firm's stock. However, Daniel and Titman (1997) find that, if anything, the common stocks of firms with higher book/market ratios are more liquid.

B Proofs

Proof of Proposition 2 Since the signals in the model are mean zero, by equation (18), $E[R] = E[\bar{\theta} - P] = \alpha E[\beta_M]$. From equations (17) and (18) (and applying the law of iterated expectations),

$$cov(R, \bar{\theta} - P) = E\left[\omega^\epsilon(1+\omega^\epsilon)(S^\epsilon)^2 + \sum_{k=1}^{K} \beta_k^2 \omega_k^f (1+\omega_k^f)(S_k^f)^2 + \alpha^2 \beta_M^2\right] - \alpha^2 E[\beta_M]^2.$$

Now, applying $V^{\epsilon\omega} \equiv var(\omega^\epsilon) = E[(\omega^\epsilon)^2] - (E[\omega^\epsilon])^2$ and $V^{\beta_M} \equiv E\left[(\beta_M)^2\right] - (E[\beta_M])^2$ we have

$$\begin{aligned} cov(R, \bar{\theta} - P) &= \alpha^2 V^{\beta_M} - cov_{OC} && (36) \\ var(\bar{\theta} - P) &= \alpha^2 V^{\beta_M} + var_{OC} && (37) \end{aligned}$$

where cov_{OC} and var_{OC} are as defined in the text. ||

Proof of Proposition 3 Let the vector $X \equiv [\beta_M, F - P, P]$ and define

$$\Sigma_{YX} \equiv [\text{cov}(R, \beta_M),\ \text{cov}(R, F - P),\ \text{cov}(R, P)].$$

Further, let Σ_{XX} denote the variance-covariance matrix of $\beta_M, P - F$ and P. Then the OLS predictor of R is $\Sigma_{YX}\Sigma_{XX}^{-1}X$. The vector of regression coefficients in (27) can therefore be written as

$$[b_\beta^\dagger,\ b_{F-P}^\dagger,\ b_P^\dagger] = \Sigma_{YX}\Sigma_{XX}^{-1}. \tag{38}$$

By (36) and (37), the covariances and variances required to calculate Σ_{YX} and Σ_{XX} are

$$\begin{aligned} cov(R, P) &= -\alpha^2 V^{\beta_M} \\ cov(R, F - P) &= -cov_{OC} + \alpha^2 V^{\beta_M} \\ var(P) &= var_{OC} + \alpha^2 V^{\beta_M} + V_{\bar{\theta}} \\ var(F - P) &= var_{OC} + \alpha^2 V^{\beta_M} + V^F \\ var(\beta_M) &\equiv V^{\beta_M} \\ cov(R, \beta) &= \alpha V^{\beta_M} \\ cov(\beta, F - P) &= \alpha V^{\beta_M}. \end{aligned}$$

Explicitly calculating these coefficients and substituting into (38) yields the expressions in Proposition 3. ||

Proof of Corollary 2: In equation (28), since $V^F = 0$, $K_1 = 0$. The term in parentheses is therefore $1 + (cov_{OC}/var_{OC})$. As $E[\omega^\epsilon], E[\omega^f] \to \infty$ in equations (23) and (24), and noting that $E[(\omega^\epsilon)^2] = V^{\epsilon\omega} + E[\omega^\epsilon]^2$ and $E[(\omega^f)^2] = V^{f\omega} + E[\omega^f]^2$, we see that the terms containing $E[(\omega^f)^2]$ and $E[(\omega^\epsilon)^2]$ dominate constants (1 or V^ω) and the linear terms containing $E[\omega^f]$ or $E[\omega^\epsilon]$. It follows that in the limit, $cov_{OC}/var_{OC} = -1$, so the term in parentheses in (28) approaches zero, proving the result. ||

Proof of Corollary 3: As shown in the Proof of Proposition 3:

$$var(F - P) = var_{OC} + \alpha^2 V^{\beta_M} + V^F.$$

Under the assumptions of Corollary 2, and from the definition of var_{OC} in equation (37), $var_{OC} \to \infty$ as $E[\omega^\epsilon], E[\omega^f] \to \infty$. This proves part 1.

Next, from equations (2), (17), and (20),

$$R_i \equiv \theta_i - P_i = (F_i - P_i) - e_i^F + S_i^\epsilon + \sum_{k=1}^{K} \beta_{ik} S_k^f + (\epsilon_i - E[\epsilon_i]) + \sum_{k=1}^{K} \beta_{ik}(f_k - E[f_k]).$$

Since the expectations in this expression are rational, we have

$$var(R_i - [F_i - P_i]) \leq var\left(-e_i^F + S_i^\epsilon + \sum_{k=1}^{K} \beta_{ik} S_k^f + \epsilon_i + \sum_{k=1}^{K} \beta_{ik} f_k\right).$$

Note that the cross-sectional variance of everything on the RHS of this expression, with the exception of $F - P$, remains finite, but $var(F - P) \to \infty$ as $E[\omega^\epsilon], E[\omega^f] \to \infty$. This implies that $var(R)/var(F - P) \to 1$. Now, the R^2 of the regression of R on $F - P$ is given by

$$R^2_{F-P} = \frac{[cov(R, F - P)]^2}{var(R)var(F - P)},$$

and, as shown in the proof of Corollary 2, in the limit, $cov_{OC}/var_{OC} \to -1$. It follows that $R^2 \to 1$, proving part 2. ||

References

Abarbanell, Jeffery S., and Brian J. Bushee, 1999, Abnormal returns to a fundamental strategy, *Accounting Review* 73, 19–45.

Admati, Anat, and Paul Pfleiderer, 1988, A theory of intraday patterns: Volume and price variability, *Review of Financial Studies* 1, 3–40.

Barberis, Nicholas, and Ming Huang, 2000, Mental accounting, loss aversion, and individual stock returns, University of Chicago Manuscript, November 1999.

———, and Jesus Santos, 1999, Prospect theory and asset prices, forthcoming *Quarterly Journal of Economics*, 2001.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny, 1998, A model of investor sentiment, *Journal of Financial Economics* 49, 307–343.

Benartzi, Shlomo, and Richard H. Thaler, 1995, Myopic loss aversion and the equity premium puzzle, *Quarterly Journal of Economics* 110, 73–92.

Benos, Alexandros, 1998, Aggressiveness and survival of overconfident traders, *Journal of Financial Markets,* Forthcoming.

Berk, Jonathan, 1995, A critique of size related anomalies, *Review of Financial Studies* 8, 275–286.

———, 2000, A view of the current status of the size anomaly, in Donald B. Keim, and William T. Ziemba, ed.: *Security Market Imperfections in Worldwide Equity Markets* . chap. 5, pp. 90–115 (Cambridge University Press: Cambridge).

———, Richard C. Green, and Vasant Naik, 1999, Optimal investment, growth options, and security returns, *Journal of Finance* 54.

Black, Fischer, 1986, Noise, *Journal of Finance* 41, 529–543.

———, Michael Jensen, and Myron Scholes, 1972, The capital asset pricing model: Some empirical tests, in Michael C. Jensen, ed.: *Studies in the Theory of Capital Markets* . pp. 79–121 (Praeger: New York).

Brennan, Michael, Tarun Chordia, and Avanidhar Subrahmanyam, 1998, Alternative factor specifications, security characteristics and the cross-section of expected stock returns, *Journal of Financial Economics* 49, 345–373.

Caballé, Jordi, and József Sákovics, 1996, Overconfident speculation with imperfect competition, Working Paper,Universitat Autonoma de Barcelona.

Campbell, John Y., and Albert S. Kyle, 1993, Smart money, noise trading, and stock price behaviour, *Review of Economic Studies* 60, 1–34.

Chamberlain, Gary, and Michael Rothschild, 1983, Arbitrage, factor structure, and mean-variance analysis on large asset markets, *Econometrica* 51, 1281–1304.

Chan, Louis K.C., Josef Lakonishok, and Theodore Sougiannis, 1999, The stock market valuation of research and development expenditures, NBER Working Paper #7223, July.

Chang, Chiayu, Zhiwu Chen, and Ming Dong, 1999, Investing with a stock valuation model, Working paper, Ohio State University and Yale University.

Cohen, Randolph B., and Christopher K. Polk, 1995, An investigation of the impact of industry factors in asset-pricing tests, University of Chicago working paper, October 1995.

Daigler, Robert T., and Marilyn K. Wiley, 1999, The impact of trader type on the futures volatility-volume relation, *Journal of Finance* 54, 2297–2316.

Daniel, Kent D., David Hirshleifer, and Avanidhar Subrahmanyam, 1998, Investor psychology and security market under- and over-reactions, *Journal of Finance* 53, 1839–1886.

———, 2000, Discriminating between efficient and inefficient markets theories of cross-sectional return predictability, Northwestern University Working Paper.

Daniel, Kent D., and Sheridan Titman, 1997, Evidence on the characteristics of cross-sectional variation in common stock returns, *Journal of Finance* 52, 1–33.

Davis, James, Eugene F. Fama, and Kenneth R. French, 2000, Characteristics, covariances, and average returns: 1929-1997, *Journal of Finance* 55, 389–406.

Davis, James L., 1994, The cross-section of realized stock returns: The pre-COMPUSTAT evidence, *Journal of Finance* 50, 1579–1593.

De Long, J. Bradford, Andrei Shleifer, Lawrence Summers, and Robert J. Waldmann, 1990a, Noise trader risk in financial markets, *Journal of Political Economy* 98, 703–738.

———, 1990b, Positive feedback investment strategies and destabilizing rational speculation, *Journal of Finance* 45, 375–395.

———, 1991, The survival of noise traders in financial markets, *Journal of Business* 64, 1–20.

DeBondt, Werner F. M., and Richard H. Thaler, 1987, Further evidence on investor overreaction and stock market seasonality, *Journal of Finance* 42, 557–581.

———, 1995, Financial decision-making in markets and firms: A behavioral perspective, in Robert A. Jarrow, Voijslav Maksimovic, and William T. Ziemba, ed.: *Finance, Handbooks in Operations Research and Management Science*, vol. 9 . chap. 13, pp. 385–410 (North Holland: Amsterdam).

Dreman, David, and Eric A. Lufkin, 1996, Do contrarian strategies work within industries?, Dreman Value Advisors.

Einhorn, Hillel J., 1980, Overconfidence in judgment, *New Directions for Methodology of Social and Behavioral Science* 4, 1–16.

Fama, Eugene F., and Kenneth R. French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427–465.

——— , 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

——— , 1998, Value versus growth: The international evidence, *Journal of Finance* 53, 1975–1999.

Fama, Eugene F., and James MacBeth, 1973, Risk, return and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.

Frankel, Richard, and Charles Lee, 1996, Accounting valuation, market expectation, and the book-to-market effect, University of Michigan working paper.

Gervais, Simon, and Terance Odean, 1999, Learning to be overconfident, Working paper, Wharton School.

Gibbons, Michael R., Stephen A. Ross, and Jay Shanken, 1989, A test of the efficiency of a given portfolio, *Econometrica* 57, 1121–1152.

Grossman, Sanford J., and Joseph E. Stiglitz, 1980, On the impossibility of informationally efficient markets, *American Economic Review* 70, 393–408.

Hansen, Lars P., and Ravi Jagannathan, 1991, Implications of security market data for models of dynamic economies, *Journal of Political Economy* 99, 225–262.

Hawawini, G, and Donald B. Keim, 1995, On the predictability of common stock returns: World-Wide evidence, in Robert A. Jarrow, Voijslav Maksimovic, and William T. Ziemba, ed.: *Finance, Handbooks in Operations Research and Management Science and Management Science* . chap. 17, pp. 497–544 (North Holland: Amsterdam).

Heston, Steven L., K. Geert Rouwenhorst, and Roberto E. Wessels, 1999, The role of beta and size in the cross-section of european stock returns, *European Financial Management* 4, 4–28.

Hirshleifer, David, and Guo Ying Luo, 1999, On the survival of overconfident traders in a competitive security market, Rutgers University Mimeo.

Hirshleifer, David, Avanidhar Subrahmanyam, and Sheridan Titman, 1994, Security analysis and trading patterns when some investors receive information before others, *Journal of Finance* 49, 1665–1698.

Hong, Harrison, and Jeremy Stein, 1999, A unified theory of underreaction, momentum trading and overreaction in asset markets, *Journal of Finance* 54, 2143–2184.

Jaffe, Jeffrey, Donald B. Keim, and Randolph Westerfield, 1989, Earnings yields, market values, and stock returns, *Journal of Finance* 44, 135–148.

Jagannathan, Ravi, Keiichi Kubota, and Hitoshi Takehara, 1998, Relationship between labor-income risk and average return: Empirical evidence from the Japanese stock market, *Journal of Business* 71, 319–348.

Jagannathan, Ravi, and Zhenyu Wang, 1996, The CAPM is alive and well, *Journal of Finance* 51, 3–53.

Jegadeesh, Narasimhan, and Sheridan Titman, 1999, Profitability of momentum strategies: An evaluation of alternative explanations, NBER working paper W7159.

Jones, Charles, and Stevel L. Slezak, 1999, Security returns in a multi-asset dynamic rational expectations model, Working Paper, Columbia University, New York, NY.

Knez, Peter J., and Mark J. Ready, 1997, On the robustness of size and book-to-market in cross-sectional regressions, *Journal of Finance* 52, 1355–1382.

Kyle, Albert, and F. Albert Wang, 1997, Speculation duopoly with agreement to disagree: Can overconfidence survive the market test?, *Journal of Finance* 52, 2073–2090.

Lakonishok, Josef, and Inmoo Lee, 1998, Are insiders' trades informative?, University of Illinois at Urbana-Champaign manuscript.

Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny, 1994, Contrarian investment, extrapolation and risk, *Journal of Finance* 49, 1541–1578.

Lee, Charles, Andrei Shleifer, and Richard Thaler, 1991, Investors sentiment and the closed-end fund puzzle, *Journal of Finance* 46, 75–109.

Lewellen, Jonathan, and Jay Shanken, 2000, Market efficiency, rational expectations, and estimation risk, Working Paper, University of Rochester, Rochester, NY.

Loughran, Tim, and Jay Ritter, 1995, The new issues puzzle, *The Journal of Finance* 50, 23–52.

MacKinlay, A. Craig, 1995, Multifactor models do not explain deviations from the CAPM, *Journal of Financial Economics* 38, 3–28.

Merton, Robert C., 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867–887.

Miller, Merton H., and Myron Scholes, 1982, Dividends and taxes: Some empirical evidence, *Journal of Political Economy* 90, 1118–1141.

Odean, Terrance, 1998, Volume, volatility, price and profit when all traders are above average, *Journal of Finance* 53, 1887–1934.

Ross, Stephen A., 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341–360.

Rouwenhorst, K. Geert, 1999, Local return factors and turnover in emerging stock markets, *Journal of Finance* 54, 1439–1464.

Shefrin, Hersh, and Meir Statman, 1994, Behavioral capital asset pricing theory, *Journal of Financial and Quantitative Analysis* 29, 323–349.

Shiller, Robert J., 1981, Do stock prices move too much to be justified by subsequent changes in dividends?, *American Economic Review* 71, 421–498.

——— , 2000, Measuring bubble expectations and investor confidence, *Journal of Psychology and Financial Markets* 1, 49–60.

Shleifer, Andrei, and Robert Vishny, 1997, The limits to arbitrage, *Journal of Finance* 52, 35–55.

Shumway, Tyler, 1998, Explaining returns with loss aversion, University of Michigan Business School Mimeo.

Swaminathan, Bhaskaran, 1996, Time-varying expected small firm returns and closed-end fund discounts, *Review of Financial Studies* 9, 845–888.

Thaler, Richard, 1980, Toward a positive theory of consumer choice, *Journal of Economic Behavior and Organization* 1, 39–60.

Wang, F. Albert, 1998, Strategic trading, asymmetric information and heterogeneous prior beliefs, *Journal of Financial Markets* 1, 321–352.