Four Essays in Financial Economics

By
Li Jin

B.A. World Economy
Fudan University, 1992

Submitted to the Alfred P. Sloan School of Management In Partial Fulfillment for the Requirement for the Degree of

Doctor of Philosophy in Management

At the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June 2001

Signature of Author: ______________________
Sloan School of Management
May 2, 2001

Certified by: ______________________
Andrew W. Lo
Harris & Harris Group Professor of Management
Thesis Supervisor

Accepted by: ______________________
Birger Wernerfelt
Director, Ph.D. Program, Sloan School of Management

Four Essays in Financial Economics

By
Li Jin

Submitted to the Alfred P. Sloan School of Management In Partial Fulfillment for the Requirement for the Degree of
Doctor of Philosophy in Management

ABSTRACT

This dissertation is composed of four pieces of independent but related works. The first is on optimal risk sharing and CEO compensation. Using a principal agent model, I addressed theoretically the optimal tradeoff between the inefficient risk sharing and a high incentive, to conclude that the optimal CEO pay performance level should be negatively related to the level of idiosyncratic (undiversifiable) risk in the firm. I empirically tested the hypothesis and found strong support.

The second paper is on the intermediation of labor income risk. From the perspective of optimal risk sharing, it looks at the recent trend of the rise intermediated labor force and its relationship with the riskness of human capital. I built an overlapping generations model to address the following issues: 1) what types of laborers get to work in conventional firms versus in labor intermediation firms? 2) how will the labor intermediation business be endogenously determined in equilibrium? 3) what effect does labor intermediation have on the ex ante decision of laborers to acquire specialized human capital?

The third paper is a joint work with Stewart Myers. Empirical researches document that more developed financial systems seem to have higher level of idiosyncratic risk. Also, in the United States, idiosyncratic risk level goes up as the level of economic activity increases. This paper uses the framework developed in Myers (2000 JF) to explain such phenomena.

The fourth paper is a joint work with Andrew Lo. It is on data-mining, and its role in explaining the financial anomalies documented in the literature. We developed a series of statistical and econometric methodologies to differentiate spurious statistical artifact from a real anomaly. We also give an extensive literature review of the landscape of the financial anomalies.

Thesis Committee:

Andrew W. Lo (Chairman)	Title: Harris and Harris Group Professor of Management
Robert C. Merton	Title: John and Natty McArthur University Professor, Harvard University
Lisa Meulbroek	Title: Associate Professor of Finance at Harvard University
Stewart C. Myers	Title: Gordon Young Billiard Professor of Management
David Scharfstein	Title: Dai-Ichi Kangyo Bank Professor of Management

Acknowledgements

I am deeply indebted to my advisors. I am grateful to my committee chair, Andrew Lo, for his stimulating guidance and unwavering support through all stages of my dissertation. I am also deeply indebted to Robert Merton and Stewart Myers, whose insightful comments and suggestions inspired my research in many ways. I was extremely fortunate to work on two papers with Professors Lo and Myers, respectively, and to benefit tremendously from extensive discussions with Professor Merton on another one (the risk sharing of labor market). Their ways of thinking about finance and about research in general have had a profound impact on my own conception of the field. I would like to especially thank Lisa Meulbroek, for many detailed discussions with me about my job market paper. Her detailed advices help to add much richness to the paper, while at the same time weeding out many unnecessary details. I also want to thank David Scharfstein for giving his time to serve on my thesis committee, and advising me on my job market paper.

I also would like to thank all faculty members in the finance program at MIT, as well as many other faculty members at Sloan, especially Robert Gibbons, Bengt Holmstrom and S.P. Kothari, who generously gave me their time and attention during the writing of my dissertation, and gave me valuable advice and suggestions.

Finally, and most importantly, I would like to thank my parents and parent-in-laws for their constant love and encouragement. I would like to thank my beloved wife, Fiona Wang, in particular, for her love and patience during the long time period that this thesis is written. Without her sacrifice and devotion, this paper would not have been possible.

Contents

1 CEO Compensation, Diversification and Incentives **7**
1.1 Introduction and related literature 7
1.2 Theory models 10
1.2.1 Model I: when the CEOs cannot trade the market portfolio 12
1.2.2 Model II: When the CEOs can trade the market portfolio 15
1.2.3 Extensions 17
1.3 Empirical analysis 18
1.3.1 Data and measurement of variables 19
1.3.2 Testing the role of Total risk and its components 23
1.3.3 The two sides of the incentive-diversification tradeoff 27
1.4 Conclusion 31

2 Risk Sharing and Labor Intermediation **48**
2.1 Introduction and literature review 48
2.2 The Model of Labor Intermediation 56
2.2.1 The model setup 56
2.2.2 Solution to the model 64
2.2.3 Effect of labor intermediation firm on labor income risk 73
2.3 Risk sharing and human capital specialization 73
2.3.1 The model 75
2.3.2 Solution 77
2.3.3 Properties 78
2.4 Conclusion and extension 80

3 Explaining The Cross Sectional Variation of Idiosyncratic Risk In International Stock Markets (Joint with Stewart C. Myers) **83**
3.1 Introduction 83
3.2 The model 84
3.3 Conclusion 98

4 How Anomalous Are Financial Anomalies? (Joint with Andrew W. Lo) **99**
4.1 Introduction and Literature Review 99
4.2 Citation Count Analysis 100
4.3 Statistical Inference For Anomalies 102
4.3.1 Bootstrap Methods 102
4.3.2 Order Statistics 113

4.3.3 Bayesian Analysis 119
4.3.4 Empirical Assessment of the Goodness-of-Fit of Conditional Predictive Models 127

A appendices **165**
A.1 Appendices of Chapter 1 on the theory part 165
A.1.1 The Prendergast (2000) critique 165
A.1.2 Proof of proposition 1 167
A.1.3 Proof of proposition 2 169
A.1.4 When principal holds a diversified portfolio of stocks 171
A.1.5 Linear contract on both the market and firm-specific movements . . . 176
A.1.6 The effect of monitoring 179
A.2 Appendices of chapter 1 on the empirical part 183
A.2.1 Measuring performance 183
A.2.2 Measuring pay and pay-performance sensitivity 184
A.2.3 Constructing risk measures through the market model regression . . . 188
A.2.4 Campbell et al. (2000) method 190
A.2.5 Proxies for the productivity of CEO effort 190
A.2.6 Robustness of the results 192
A.3 Appendices of chapter 2 200
A.3.1 Proof of Lemma 1 200
A.3.2 Proof of Lemma 2 201
A.3.3 Proof of Theorem 1 202
A.3.4 Proof of Theorem 2 202
A.4 Appendices of Chapter 3 205
A.4.1 Discussion 205
A.4.2 Proofs 207
A.4.3 Discussion about idiosyncratic risk versus market risk 208
A.4.4 Calibration studies 211

Chapter 1

CEO Compensation, Diversification and Incentives

1.1 Introduction and related literature

This paper examines the impact of imperfect diversification on CEO compensation. CEO pay is linked to the stock market performance of the firm, largely for incentive reasons. However, there is a tradeoff between risk sharing and providing incentives. This is well studied in the principal agent theory. Empirically, Aggarwal and Samwick (1999a), Kraft and Niederprum (1999) and Garen (1994) document that executives at riskier firms do have lower incentive levels.

Most of the existing literature treats both systematic and non-systematic risks in the same manner. Empirical work supporting the principal agent model focuses on the total risk rather than its components. Intuitively, one would expect that non-systematic risk and systematic risk matter differently to incentives: First, it is costly for *both* shareholders and CEOs to bear the systematic risk, but the shareholders have a clear cost advantage in bearing the non-systematic risk, because shareholders can diversify, while CEOs typically hold a large undiversified position in the firm. Second, CEOs can potentially trade the market portfolio to adjust their exposures to the market risk, but for incentive reasons they need to maintain the firm-specific risk. We test in this paper whether the two types of risk have different impacts on CEO incentives, and find empirical support for it.

We start by modeling explicitly the different impacts of systematic and non-systematic risks on incentive levels. We assume that the shareholders and the CEOs are both risk averse.

The incentive-based compensation leaves the CEOs unable to diversify their exposures to systematic risk, but shareholders are assumed to hold diversified portfolios. As a result, while shareholders find it is costly to bear systematic risk but not non-systematic risk, the un-diversified CEO finds it costly to bear both systematic risk and non-systematic risk. We demonstrate that in this case, the optimal incentive level decreases with firm-specific risk, but is ambiguously related to market risk. We also model the cases in which CEOs can adjust their exposures to market risk while retaining the firm-specific risk, and find that while the optimal incentive level decreases with firm specific risk, it does not change with market risk.

We empirically investigate whether decomposing risks into their market and firm-specific components can better explain the observed incentives. We find that, indeed, firm-specific risk, but not market risk, is the driving force of the previously documented relations between risk and incentives. When we control for the level of market risk, we see that firm-specific risk level is negatively related to incentive level; when we control for the level of firm-specific risk, we don't observe a robust relation between market risk and incentive level. This finding is robust to different regression methods, different ways to decompose risk into systematic risk and firm-specific risk, and adding control variables that are known to affect incentive levels.

In implementation, we *simultaneously* study how incentive levels should change with the productivity of CEO effort. The latter captures the benefit of giving CEO incentives. As pointed out by Prendergast (2000), risk might be positively related to the benefit of giving incentives. Ignoring this link will likely bias the estimated effects of risk on incentives, because of the omitted-variable problem. At the least, this will render the estimated effect of risk on incentives unreliable, and it could even eliminate or reverse the observed relation between risk and incentive. This suggests that in order to understand the cost of giving incentives, we should simultaneously address the benefit of giving incentives. We find that, consistent with the principal agent theory, when we simultaneously study the cost and benefit of giving incentives, we see that incentives will decrease with firm-specific risk, and increase with proxies for the productivity of CEO effort, all other things being fixed.

This paper is organized as follows. In the remaining part of this section, we review the literature. In Section 2, we formally lay out theoretical models to illustrate the optimal tradeoff between giving incentives and loss of diversification. These models formalize the intuition and generate testable implications that will be the focus of the paper. In Section 3, we explain the data set we use and empirically test the optimal tradeoff between diversification and incentives in CEO compensation, demonstrating that firm-specific risk is significantly negatively related to incentive level, whereas market risk is ambiguously related to incentive level. We also show that incentive level increases with the productivity of CEO effort, when we control for the risks. Section 4 provides concluding remarks. Proofs and some technical details are put in the appendices to make the formal text short.

There is an extensive and active literature on executive compensation in finance, accounting, industrial organization and labor economics.[1] The classical principal agent models typically assume a risk-neutral principal. A simple argument might be that a principal with sufficient wealth can eliminate all risks. But in finance, we know that even at the aggregate economy level, risk does not all cancel out. If the financial market as a whole becomes truly risk-neutral, it should demand the same expected return from risky and riskless investments. In reality, there is a positive and large market risk premium. The fact that this market risk premium is not arbitraged away suggests that the market participants are still substantially risk averse. Thus, a reasonable modification of the principal agent model assumption in finance would be that principals are also risk averse to the market risk, but if we assume that they hold a diversified position, then they are risk-neutral to firm-specific risks.

Empirical research studying incentives in executive compensation largely focuses on pay-performance sensitivity, which is defined by Jensen and Murphy (1990a) as the change in CEO wealth for every $1,000 increase in shareholder value. This measure captures the proportion of firm value increase that goes to the CEO. Jensen and Murphy (1990a) report that CEO wealth on average increases by $3.25 for every $1,000 increase in shareholder value.

[1]Early works in principal agent models include studies by Ross (1973), Holmstrom (1979, 1982), Jensen and Meckling (1976), Grossman and Hart (1983) and Holmstrom and Milgrom (1987, 1991). Coughlan and Schmidt (1985), Demsetz and Lehn (1985), Murphy (1985, 1986), Jensen and Murphy (1990a, 1990b) and Gibbons and Murphy (1990, 1992) did the pioneering work in CEO compensation studies. Recent works on CEO compensation include papers by Aggarwal and Samwick (1999a, 1999b, 1999c), Beatty and Zajac (1994), Bertrand and Mullainathan (2000a), Bloom and Milkovich (1998), Carpenter (1998, 2000), Core and Guay (1999, 2000), Garen (1994), Garvey (1997), Hall and Liebman (1998), Himmelberg et al. (1999), Holmstrom (1999), Kole (1997), Hubbard and Palia (1995), Meulbroek (2000a, 2000b), Prendergast (2000), Sloan (1993), Smith and Watts (1992) and Yermack (1995). Murphy (1999) conducts an excellent survey on executive compensation.

Using more recent data, Hall and Liebman (1998) and Aggarwal and Samwick (1999a) report a higher pay-performance sensitivity. Aggarwal and Samwick (1999a), Kraft and Niederprum (1999) and Garen (1994) study the cross-sectional variation in pay-performance sensitivity using the risk sharing-incentive tradeoff in principal agent models and find that the incentive levels are negatively related to risks. However, they consider total risk as the cost of giving incentives. This paper differs by recognizing that, compared to other shareholders who can freely invest in any asset (including the market portfolio), the CEOs are forced to take large positions in their own firms, and thus lose diversification. The real cost of giving CEOs incentive is the lost diversification through inefficient bearing of the firm-specific risk, rather than the bearing of total firm risk. Accordingly, firm-specific risk level, rather than total risk level, should affect the pay-performance sensitivity. As shown by Meulbroek (2000a) and several other papers, the cost to the executives of lost diversification can be quite large.

Existing work also finds higher pay-performance sensitivities for CEOs after deregulation or when firms have larger investment opportunity sets, which is consistent with the hypothesis that a higher incentive level is needed when CEO has a potentially large impact on firm value.[2] However, that literature has developed largely independent of the literature on the negative impact of risk on incentives. Most existing studies analyze either the cost or the benefit of giving incentives, but not both at the same time[3]. As Prendergast (2000) argues, the same variables might affect both the benefit and cost of giving incentives. Therefore, a partial study of either side of the tradeoff will likely introduce an omitted variable bias, and the estimated results might be un-reliable.

1.2 Theory models

In this section, I use a principal agent model to analyze the tradeoff between giving CEOs high incentives and letting them bear (inefficiently) the firm-specific risk.[4] I assume that both

[2]See, for example, Hubbard and Palia (1995) and Smith and Watts (1992). Finkelstein and Boyd (1998) did a comprehensive study on the impact of CEO discretion on pay-performance sensitivity.

[3]An exception is Himmelberg, Hubbard and Palia (1999).

[4]Maintaining incentives is just one reason among many others why CEOs are holding disproportionally large amounts of stocks and options in their own firms. See Gibbons (1998) and the works cited there. But no matter what the ultimate reason is for CEOs to hold the un-diversified position, the cost of lost diversification can always be measured by the level of firm-specific risk. For the remaining theory and empirical analysis, I assume that the reason for CEO to hold undiversified position is to maintain incentive. Certainly that can

the CEOs and the outside shareholders are risk averse, although with potentially different levels of risk aversion. Furthermore, the outside shareholders can diversify across different stock holdings and thus only care about the market risk. The CEOs, on the other hand, care about both the market and the firm-specific risks, because they are not diversified. I further assume that the realization of the firm-specific risk is beyond the control of the CEOs.[5]

Following the classical papers of Holmstrom and Milgrom (1987, 1991), I assume that the compensation contract takes the form of a linear contract.[6] Such a contract rewards both market and idiosyncratic performance, but I show that the *sensitivity* of CEO pay to firm performance are affected differently by firm-specific risk and market risk. In particular, while a higher firm-specific risk makes the ex-ante compensation contract more performance-dependent, a higher market risk may not. I study two models. In the first model, I assume that CEO cannot trade on the financial market and thus has to hold whatever risk he gets from his compensation package. In the second model, I further assume that there is a financial market, and CEO can trade the market portfolio with no restrictions, but to maintain incentives, CEO cannot hedge the firm-specific risk, either explicitly or implicitly.[7]

be generalized.

[5]In an appendix I extend the analysis to a case where CEOs can control or know some of the realization of the firm-specific risk before the market knows it, and show that we can still get qualitatively the same results by filtering out the part of firm-specific risk that is not exogenous to the CEOs.

[6]The optimality of linear sharing rule depends critically on the assumption of constant absolute risk aversion utility function, of which the mean-variance optimization can be thought of as a reduced form. When we have more general preference, linear contract may not be the optimal contract in this setting. See the paper by Franke et al. (1998) for a discussion about the non-optimality of linear sharing rule with background risk. However, linear sharing rule provides an easier ground for the empirical tests, and in reality the sharing rule is almost linear. The convexity induced by CEO holdings of options is negligible to the first order: for a return change one standard deviation above or below the median level, the slope of incentive contract changes about one to two percent. In the empirical analysis, the assumption of constant absolute risk aversion preference also has the feature that CEOs' out-of-the-firm wealth, which is unobservable, does not affect the incentive level. Given the difficulty of collecting data on CEO's out-of-firm wealth, this is as far as we can go, without having to resort on some arbitrary assumption about CEOs' wealth out of the firm.

[7]The last assumption, although seemingly natural, can be violated in practice. For example, Ofek and Yermack (2000) document that following reward of new equity-based compensations, executives sometimes sell off some stocks in their firm to diversify. Also, even if CEOs cannot short their own firm's stocks, they might be able to short a "tracking portfolio" of highly correlated stocks. Use of a "tracking" portfolio solves some securities law issues that otherwise are raised by a firm-specific hedge. However, even a tracking portfolio raises some very complicated tax issues. In practice, few, if any, are willing to bear these tax burdens. For detailed discussion, see Schizer (2000). In recent years, firms also become more aggressively engaged in performing internal risk management by awarding employees "collars", which has the effect of reducing the firm-specific risk (thus incentives) of the employees. Adding these more realistic assumptions will reduce the magnitude of the loss of diversification, but generally should not totally eliminate the effect we are studying. For simplicity I assume away such complications, although I want to point out that the

1.2.1 Model I: when the CEOs cannot trade the market portfolio

In this subsection, I assume that CEOs cannot trade the market portfolio, thus the only source of risks for CEOs is from their compensation packages. The shareholders, however, can trade freely on the financial market. As a result, they hold a diversified portfolio, and they also hold the optimal level of market risk.[8]

The total dollar value of a firm next period has both a systematic component and a firm-specific component:

$$\tilde{Z} = \overline{Z} + \eta(\widetilde{r_m} - \overline{r_m}) + \tilde{\delta} \tag{1.2.1}$$

Where $\eta(\widetilde{r_m} - \overline{r_m})$ is the dollar amount of market risk, $\tilde{\delta}$ is the dollar amount of firm-specific risk. $\tilde{r}_m$ and $\bar{r}_m$ are the realized and expected return of the market portfolio, respectively. $Var(\widetilde{r_m} - \overline{r_m}) = \sigma_m^2$ is the variance of the market portfolio's percentage return, and $Var(\tilde{\delta}) = \sigma_{\tilde{\delta}}^2$ is the variance of the dollar firm-specific risk.

$\overline{Z}$ is the mean firm value, which we assume can be affected by CEO effort level e in a linear way: $\overline{Z} = Z_0 + ke$, where Z_0 is the mean firm value when CEO exerts zero effort. k is the productivity of CEO effort.

I further assume that exerting effort costs the agent(CEO) an unobservable cost, which can be expressed in monetary terms as $f(e)$. $f(e)$ is increasing and convex, so that as the effort level increases, the total cost increases more than linearly. An example of the cost function is $f(e) = e^2$.[9]

The principal gives the following linear contract to the agent: $\pi_A = a\tilde{Z} + b$, where π_A is the payoff to the agent, a and b are constants. Principal's payoff is thus $\pi_P = (1-a)\tilde{Z} - b$. I

increased use of risk management, both at firm level and individual level, might be one reason why the *observed* pay-performance sensitivity has been increasing over time.

[8]The model represented here is just a "reduced form" representation. For the detailed model and proof, interested readers can see appendix A.1.4.

[9]The assumption of linear impact and convex cost of effort can be thought of as a "reduced form" of the more general assumption that cost of effort is convex and increasing (increasing marginal cost of effort), while effect of effort is concave and increasing (diminishing marginal benefit of effort). Under that assumption, we could re-scale the effort unit to make it linear in output, and the resulting cost function of effort will be convex. A simple proof of this can be obtained from the author. In more general setting, where for example the impact of CEO effort has an increasing but non-concave functional form, then we typically cannot conclude that the "transformed cost function" will be convex once we linealize the impact function of effort. In that case, the optimization problem of principal might not have a unique interior solution.

assume that the principal and the agent both have mean variance optimization preferences:

$$U_A(w) = E(w) - \frac{1}{2}\gamma_A Var(w) \tag{1.2.2}$$

and,

$$U_P(w) = E(w) - \frac{1}{2}\gamma_P Var(w) \tag{1.2.3}$$

Where γ_P and γ_A could differ.

With these assumptions, we can write out the certainty equivalent payoffs of both the principal and the agent, and lay out the principal's problem:

$$max_{a,b,e}[(1-a)\overline{Z} - b] - \frac{1}{2}(1-a)^2\gamma_P\eta^2\sigma_m^2 \tag{1.2.4}$$

$$st.\ [a\overline{Z} + b - f(e)] - \frac{1}{2}a^2\gamma_A(\eta^2\sigma_m^2 + \sigma_\delta^2) \geq \overline{CE_A} \tag{1.2.5}$$

and

$$[a\overline{Z}(e) + b - f(e)] - \frac{1}{2}a^2\gamma_A(\eta^2\sigma_m^2 + \sigma_\delta^2) \geq [a\overline{Z}(e') + b - f(e')] - \frac{1}{2}a^2\gamma_A(\eta^2\sigma_m^2 + \sigma_\delta^2), \forall e' \tag{1.2.6}$$

where (1.2.5) is the individual rationality constraint with $\overline{CE_A}$ being the reservation level of payoff, and (1.2.6) is the incentive compatibility constraint for e to be the optimal choice of effort level.

When effort level is a continuous variable, equation (1.2.6) can be replaced by the first order condition: $f'(e) = ak$.

Proposition 1 *The above problem has a solution:*

$$a^* = \frac{\gamma_P\eta^2\sigma_m^2 + \frac{k^2}{f"(e)}}{\gamma_P\eta^2\sigma_m^2 + \gamma_A(\eta^2\sigma_m^2 + \sigma_\delta^2) + \frac{k^2}{f"(e)}} \tag{1.2.7}$$

$$e^* = f'^{(-1)}(a^* k) \tag{1.2.8}$$

and

$$b^* = \overline{CE_A} - [a^* \overline{Z} - f(e^*)] - \frac{1}{2} a^{*2} \gamma_A (\eta^2 \sigma_m^2 + \sigma_\delta^2) \tag{1.2.9}$$

Proof: see appendix.

Interpretation

The result has clean interpretations:

1. Pay-performance sensitivity decreases in firm-specific risk σ_δ^2. Higher firm-specific risk increases the cost of giving incentive, through increased cost of lost diversification. Thus, other things equal, the optimal incentive level should decrease with firm-specific risk level.

2. The relation between market risk and pay-performance sensitivity is ambiguous. When $\eta^2 \sigma_m^2$ is high, a^* could be either high or low, depending on the other parameters. Intuitively, the principal's marginal cost of bearing market risk is determined by the market price of (market) risk, $\bar{r}_m - r_f$, but the agent's marginal cost of bearing market risk is determined solely by his market risk exposure in the compensation contract. There could be cases when agent has a lower marginal cost of bearing the market risk, thus when firm's market risk level increases, it could be optimal to *increase* the incentive, a^*.

3. There is a positive relation between the productivity of CEO effort, k, and the pay-performance sensitivity, a^*. Higher CEO effort productivity makes the effort more valuable to the firm, thus increasing the shadow price of the incentive compatibility constraint, therefore, the incentive level should optimally increase.

One thing is worth noting. Although the agent cannot trade on the financial market, the principal can. This implies that not only will the principal hold a well diversified portfolio (and thus does not care about idiosyncratic risk), but also he will have the *optimal loading* in the market portfolio, such that his marginal cost of bearing the market risk is equal to the market price of the market risk. We can allow this feature in our model by assuming that the principal will invest an amount I_P in the market portfolio, in addition to his holding of the stock in this firm. Under that case, all the above intuition carries through, although the principal's *personal* cost of bearing the market risk is replaced by the market's cost of bearing the market risk, which is the market risk premium. The proof is not included here for the sake of brevity, but it can be obtained from the author.

1.2.2 Model II: When the CEOs can trade the market portfolio

In this subsection, I explicitly consider the case when both the CEOs and the outside shareholders can trade the market portfolio. In this case, they will be able to augment their holdings of the firm by investments in the market portfolio.

I assume exactly the same model as before, except that now both parties can invest in the market portfolio. Denote by I_A and I_P the investments in the market portfolio by the agent and the principal, respectively.

The total payoffs after accounting for the investment in market portfolio are:

$$\begin{aligned} \pi'_A &= a\widetilde{Z} + b + I_A(\overline{r_m} - r_f) + I_A(\widetilde{r_m} - \overline{r_m}) &\quad (1.2.10)\\ &= a\overline{Z} + (a\eta + I_A)(\widetilde{r_m} - \overline{r_m}) + a\widetilde{\delta} + I_A(\overline{r_m} - r_f) + b &\quad (1.2.11)\end{aligned}$$

and

$$\begin{aligned} \pi'_P &= (1-a)\widetilde{Z} - b + I_P(\overline{r_m} - r_f) + I_P(\widetilde{r_m} - \overline{r_m}) &\quad (1.2.12)\\ &= (1-a)\overline{Z} + [(1-a)\eta + I_P](\widetilde{r_m} - \overline{r_m}) + (1-a)\widetilde{\delta} + I_P(\overline{r_m} - r_f) - b &\quad (1.2.13)\end{aligned}$$

I solve for the optimal level of the market investment by the agent and the principal to get: $I_A = \frac{\overline{r_m} - r_f}{\gamma_A \sigma_m^2} - a\eta$, and $I_P = \frac{\overline{r_m} - r_f}{\gamma_P \sigma_m^2} - (1-a)\eta$.

The principal's problem can now be written as:

$$max_{a,b,e}[(1-a)\overline{Z} - b] + \frac{1}{2}\frac{(\overline{r_m} - r_f)^2}{\gamma_P \sigma_m^2} - (1-a)\eta(\overline{r_m} - r_f) \quad (1.2.14)$$

$$st.\ a[\overline{Z} - \eta(\overline{r_m} - r_f)] + b - f(e) + \frac{1}{2}\frac{(\overline{r_m} - r_f)^2}{\gamma_A \sigma_m^2} - \frac{1}{2}\gamma_A a^2 \sigma_\delta^2 \geq \overline{CE_A} \tag{1.2.15}$$

and

$$g(e) \geq g(e'), \forall e' \tag{1.2.16}$$

where

$$g(e) = a[\overline{Z}(e) - \eta(\overline{r_m} - r_f)] + b - f(e) + \frac{1}{2}\frac{(\overline{r_m} - r_f)^2}{\gamma_A \sigma_m^2} - \frac{1}{2}\gamma_A a^2 \sigma_\delta^2 \tag{1.2.17}$$

When effort level is a continuous variable, (1.2.16) can be replaced by the first order condition, $f'(e) = ak$.

Proposition 2 *The above problem has a solution:*

$$a^* = \frac{k^2}{\gamma_A \sigma_\delta^2 f''(e) + k^2} \tag{1.2.18}$$

$$e^* = f'^{(-1)}(a^* k) \tag{1.2.19}$$

and

$$b^* = \overline{CE_A} - [a\overline{Z} - f(e) - a\eta(\overline{r_m} - r_f) + \frac{1}{2}\frac{(\overline{r_m} - r_f)^2}{\gamma_A \sigma_m^2} - \frac{1}{2}\gamma_A a^{*2} \sigma_\delta^2] \tag{1.2.20}$$

Proof: see appendix.

Interpretation

The result has clean interpretations:

1. pay-performance sensitivity decreases in firm-specific risk, σ_δ^2.
2. The exposure to market risk, η, does not affect a. The intuition is that since both principal and agent can trade the market, they equalize their marginal cost of bearing the market risk to the market price of market risk. Thus, the division of market risk

through incentive plan does not matter, because any party with more or less than their optimal loading of market risk can simply adjust it through the financial market.

3. The higher the productivity of effort (k), the higher the pay-performance sensitivity.

The results here resembles that of the standard pay-performance sensitivity in a model where the principle is risk neutral and the volatility of performance is σ_δ^2. What this result says is that even when the principal is risk averse, the results will still hold, although the appropriate risk measure is the idiosyncratic risk. That is to say, when the two parties can independently invest in the market, the optimal contract will be strictly an incentive contract, which involves the tradeoff between giving incentives and the forced loss of diversification of agent. In contrast, equation (1.2.7) reflects both incentives contract and the optimal risk sharing between both parties. The incentive contract component can be isolated by considering the case of $\gamma_p = 0$, while the market risk sharing component can be isolated by considering $k = 0$.

1.2.3 Extensions

I consider two extensions in the appendix.

Contracting separately on the market and firm-specific performances

The assumption that CEO can trade the market portfolio is not the only way to get the result that the market risk is irrelevant to incentive level. In the appendix, we study a case where the CEO cannot trade the market, but the firm can contract *separately* on the firm-specific and the market performance, and show that this will give a functionally equivalent result as in section 2.2.

Monitoring CEO effort

The paper so far focuses on output-based contract, in which CEO compensation is linked to the firm value. Alternatively, we could have input based contract, where CEO's compensation could be based on a direct monitoring of the CEO's effort. In practice, firms typically combine the two: while CEO compensation does depend on the stock price performance, typically it also depends on the direct signals about the CEO effort, such as accounting measures of performance[10]. Such arrangement is optimal in light of the theoretical argument by

[10]See, for example, Sloan (1993)

Holmstrom (1979), in which firms should use other signals to complement the output-based compensation contract. More recently, Prendergast (2000) argues that, when risk increases, direct monitoring of CEO effort would be more difficult. For example, the accounting numbers become less meaningful. Thus firms might rely more on the incentive plans to align the CEOs' incentive. This shows that it is important to analyze how monitoring could be used to modify the incentive contract. In the appendix, we demonstrate that under some circumstances incentive contract can be thought of as a substitute of monitoring: incentive is in more use when monitoring is less effective.

1.3 Empirical analysis

Aggarwal and Samwick (1999a) find that the total risk is negatively correlated with the pay-performance sensitivity under various regression specifications. This section extends their findings in several ways.

1. I decompose the total risk and study whether firm-specific risk, market risk or both contribute to the cross-sectional variation in pay-performance sensitivity.

2. I employ a more extensive robustness check and sub-sample study.

3. I also include the other side of the tradeoff to determine how productivity of CEO effort is related to incentive level. As Prendergast (2000) argues, risk might also increase the benefit of giving incentive[11]. Thus, ignoring the other side of the tradeoff will introduce an omitted variable problem, thus making our estimate of the true effect of risk on incentive unreliable.

We first replicate the Aggarwal and Samwick's work.

Then, we show that the addition of idiosyncratic risk will change the pattern: it is significant.

Then, we show that the formal decomposition of total risk into systematic and non-systematic risks reveals that the idiosyncratic risk matters.

[11] See appendix for the detailed illustration of Prendergast's point

Then, we change the measure to PPS from total wealth, and show that the formal decomposition of total risk into systematic and non-systematic risks reveals that the idiosyncratic risk matters.

The sub-sample study of breaking the sample into CEOs with more equity holdings and CEOs with less equity holdings.

Then, the last thing is to show that when idiosyncratic risk alone is included, we have that PPS does down with idiosyncratic risk.

To deal with outliers, the existing literature on CEO compensation proposes the use of median regression and robust regression, as well as OLS regression controlling for year effects.[12] I perform all three types of regressions[13]. They generally give qualitatively similar results. In addition, I also truncate the data at 1% and 99% to remove the outliers.

1.3.1 Data and measurement of variables

Our primary data set for executive compensation is the ExecuComp database maintained by Standard and Poors, which contains data from 1992 to 1998 for the S&P 500, the S&P mid cap 400 and the S&P small cap 600 firms. This data set reports the annual compensation of the five most highly compensated executives for each firm, which includes the CEO. It contains data on the level of salary and bonus and the value of other types of grants in each year, as well as CEOs' holdings of the firms' stocks and stock options. From this data set, we can construct our measure of CEO compensation. Altogether we have 2,018 unique firms and 9714 CEO-year observations.

I obtain the stock price information from the CRSP daily and monthly stock files, and the accounting variables from COMPUSTAT. These are used to construct the risk measures and the control variables.

[12] Median regression minimizes the sum of absolute deviations rather than the sum of squared deviations, so that the precise value of the dependent variable in a median regression matters only in determining whether the observation has a positive or negative residual. If the residual is positive or negative, then the dependent variable could increase toward infinity (minus infinity) without affecting the estimated parameters. Koenker and Bassett (1982) discusses median regression, and quantile regression in general. Robust regression uses Huber weight iterations followed by biweight iterations to lower the weight on observations with large residuals, to make the estimated coefficients less influenced by outliers. See Hamilton (1991) for detailed discussion about it.

[13] I also performed regression controlling for firm or industry fixed effects, the results do not change much. To save space, I am not reporting them here.

To conduct regression analysis, I need to measure the dependent variable (pay-performance sensitivity) and the independent variables (risks, productivity of CEO effort, etc). I explain briefly the measurement of some variables. In the appendix, I explain in greater detail the measurements of variables.

Measuring firm performance

Performance can be measured in both dollar terms and percentage terms. As pointed out by many researchers like Murphy (1999) and Aggarwal and Samwick (1999a), the dollar measure of performance and risk makes more economic sense. Another good thing about expressing risk measures in dollar terms is that this measure will be roughly invariant to changes in firm leverage: if we assume debt is riskless, then the total dollar risk of firm will remain constant, whereas the percentage risk of the firm equity will change with firm leverage. For these reasons, this paper will focus on dollar performance and dollar risks. To address the issue that the dollar pay-performance sensitivity decreases with firm size, I explicitly put in control for firm size. See appendix for a detailed discussion[14].

Another issue is whether we should measure the performance relative to the market or industry peers or the absolute performance. As shown by Murphy (1985), it does not seem to matter in our context, as long as the two are not used simultaneously. I focus on total real performance where the performance is the inflation adjusted total dollar change of firm value. For a robustness check, I also analyze the relative performance where performance is adjusted using a market return benchmark. The results are very similar. Table 1 reports sample statistics of the performance measures in year 1995.

Measuring pay and pay-performance sensitivity

As discussed in previous literature, two components of CEO wealth can be influenced by firm performance — human capital and financial capital. These are captured respectively by the total direct compensation change and the re-evaluation of stock and stock options.

Two measures of CEO pay change are of interest in the regression analysis later on. The first one is as suggested by Jensen and Murphy (1990a), the hypothetical change in CEO wealth, should the firm value increase by $1000; the other is the realized change in CEO wealth in each year.

Table 2 reports the sample statistics of CEO's pay and pay-performance sensitivity mea-

[14]See the paper by Baker and Hall (1999) for the justification for the specification using percentage terms.

sures.

In this section, we mainly focus on the PPS from financial wealth. This way, we can use all the actual numbers for the PPS, rather than having to go through a first stage regression analysis to get the PPS from human capital part, which necessarily increases the measurement error. Granted, we might miss some action in bonuses, but as several researchers have documented, the PPS from financial wealth accounts for nearly all the PPS. For robustness checks, we also include results from PPS from the total wealth, and we will see that the results don't change much.

Measuring risks

I adopt three different approaches to estimate the market risk and the firm-specific risk. The first is a market model regression, the second is a regression-free approach as proposed by Campbell et al. (2000), the third is a Fama-French three factor model decomposition of risks.

In the market model regression, firm-specific risk is the mean squared error, while market risk is beta squared multiplied by the variance of market return. Our basic measure of risks comes from the market model regression using up to 60 monthly observations immediately before the current year. For robustness check, we also estimate risks using market model regression of up to one year and three years of weekly data.

Two considerations motivate the use of industry average measure of firm-specific risk to replace the individual firm firm-specific risk measure. First, if there are errors in estimating the individual firm's risk measures through the market model regression, as long as the errors are not correlated with industry, the industry average risk measures will be more accurately estimated.[15] Second, CEOs can potentially control the level of their own firms' risks. In that case, a more robust measure of the risk would be the industry's average level of firm-specific risk, which the CEOs are less likely to affect.[16]

[15] See Fama and MacBeth (1973) and Fama and French (1992) for similar approach in the empirical asset pricing literature. I require that at least five firms are in the industry (two digit SIC code) in any given year, before I include the industry in the later regression analysis.

[16] As a robustness check, I also calculate the risk decomposition of the value-weighted industry portfolio directly, and assign these systematic and firm-specific risk measures to each firms within the same industry. This approach gives similar results. One drawback of this approach is, if the firms within a industry have a large amount of firm level firm-specific risk, but they cancel out when we form the industry portfolio, then the industry portfolio risk measure will not be able to reflect the average firm-specific risk of the firms within the industry.

One alternative specification is to use the individual firm's risk measures as regressors, but use the industry average risk measures as an instrumental variable. This way, we could use more of the individual firm level information, and still handle the measurement error problem with using the individual firm's risk measures. In addition, instrumental variable approach is a more sensible approach when there is a possibility of endogeneity of the risk measures: CEOs might choose their own firm's risk measures in response to the incentive schemes they face.[17]

Campbell et al. (2000) offers a method for estimating the industry average level of risk without estimating firm-specific parameters in a market model. As a robustness check, we calculate two additional risk measures according to the Campbell et al. (2000) approach and assess how our results depend on the measurement of risk. These two measures are calculated using 5 years of monthly data and 2 years of monthly data, respectively. Some detailed procedures for getting the Campbell et al. (2000) measures are included in an appendix.

To be consistent with the dollar performance in the models, I convert the percentage risks into dollar risks. The variance of dollar risk is calculated as the product of the industry average percentage risk measures and the square of each individual firm's beginning-of-the-period market cap.

The summary statistics of various risk measures and their correlation structure are in Table 3.

To be consistent with the existing literature and to guard against outliers, I construct the rank of the risk measures among all firms in the ExecuComp database. I perform regression analysis using both the raw risk measures and the rank measures. The results are qualitatively similar, but the rank measures have the advantage of easier economic interpretations. I only report the regression results using the rank measures of risk.

Proxies for the productivity of CEO effort

One goal of the present research is to empirically test for the other side of the diversification-incentive tradeoff, and determine how productivity of CEO effort is related to the pay-performance sensitivity, once we consider the effect of risk on incentives. Largely following

[17] For the detailed discussion of the use of instrumental variable approach to deal with measurement errors and endogeneity of variables, see, for example, Greene (1993).

the existing literature, I used the following proxies for CEO effort productivity.

- Investment

  Productivity of CEO effort could be positively related to the amount of investment the firm is making. If a firm is undertaking a great amount of investments, the CEO will likely have major influence on the firms' value through making important investment decisions.

- Firm age

  It can be argued that younger firms might have more growth potential, and thus their CEOs are more important in developing the firm through exerting effort.

- Tobin's Q.

  Tobin's Q can be related to future growth opportunity, thus a firm with a higher Tobin's Q will likely be affected more by CEO effort.

1.3.2 Testing the role of Total risk and its components

Aggarwal and Samwick (1999a) demonstrate that the total risk of a firm is negatively related to the pay-performance sensitivity. I want to determine whether firm-specific or market risk or both contribute to this effect.

I first generate the measure of PPS from CEO's holdings of stock and option in the firm. Then, using these as the left hand side variable, I replicate the main results in Aggarwal and Samwick, adding other control variables that are used in later studies. These control variables include: the log of sales, and the square of log of sales (proxies for the size and the non-linear effect of size), the ratio of capital over sales, the ratio of R&D expense over capital and advertising expense over capital, the dummy variables for the missing observations of these, and investment expense over capital. I also add the year dummies.

The first regression I run is:

$$PPS = \alpha + \beta_1 TOT\_RISK + \beta_2 CONTROL + \beta_3 YEAR\_DUMMY + +\epsilon \qquad (1.3.21)$$

Table 4 reports the regression result. I report three different regression methods, where the risks are measured using up to 60 monthly observations of the stock prices [18].

The results supports the basic findings of Aggarwal and Samwick (1999a), that there is an apparent negative relationship between PPS and total risk measure, when we control for many other variables.

Next, we add the measure of idiosyncratic risk to the first equation. The new equation I run becomes:

$$PPS = \alpha + \beta_{11}TOT\_RISK + \beta_{12}ID\_RISK + \beta_2 CONTROL + \beta_3 YEAR\_DUMMY + +\epsilon \tag{1.3.22}$$

Table 5 reports the regression results for OLS, median and robust regressions. Here, we notice that although the idiosyncratic risk measures is in general negative and significant, the total risk measure, *after the idiosyncratic risk measures are included*, becomes insignificant in general.

As a direct test of the hypothesis in the theory section, we would also like to know what happens when we break down the total risk into its systematic and idiosyncratic risk components, and put both in the regression. We do this in Table 6. The regression we run is:

$$PPS = \alpha + \beta_{11}SYS\_RISK + \beta_{12}ID\_RISK + \beta_2 CONTROL + \beta_3 YEAR\_DUMMY + +\epsilon \tag{1.3.23}$$

Table 6 reports results for OLS, median, robust and instrumental variable regression results. We can see that idiosyncratic risk measures remains significant both economically and statistically across various specifications, while systematic risk is in general insignificant.

In Table 7, we re-do the exercise in Table 6, but instead of using PPS from CEOs' stock and options holding only, we now add to it the PPS coming from the change in CEOs' human capital.

[18] Throughout the analysis, risks measured using one year and three-year weekly stock returns give roughly the same picture. For brevity these are not included.

All of these results reveal that firm-specific and market risks have different impacts on the pay-performance sensitivity. While firm-specific risk is clearly negatively related to pay-performance sensitivity even after controlling for market risk, the effect of market risk is ambiguous when controlling for firm-specific risk, consistent with the prediction of model I in Section 2. Furthermore, for many specifications, the coefficient on market risk is insignificant after controlling for firm-specific risk, providing some support for the prediction of model II.

One concern that people might have about the above argument is whether the market model regression correctly characterizes the priced risks. In empirical work, we know that the market model regression, although intuitively appealing, is not strongly supported by data. Empirically, the Fama-French 3 factor model explains the asset returns better. For this reason, we also decompose risks into the idiosyncratic and systematic components using the Fama-French three factor model. The results are qualitatively similar. It is not reported to save space.

I also directly test two related hypothesis:

- Whether firm-specific risk and market risk have the same effects on pay-performance sensitivity. We test the hypothesis: $H_0 : \beta_1 = \beta_2$ in (1.3.21)

  24 regression models are checked: 3 regression methods (median, robust and OLS regressions), 3 different raw measures of risk estimated using the market model regression and one raw measure of risk estimated using the Fama-French model regression, with PPS from total wealth change and from financial wealth change, respectively. In all regressions, I control for size and productivity of CEO effort. Among these, 18 reject H_0 at the 1% significance level, 22 reject H_0 at the 5% level.

- After controlling for rank of total risk, do we still gain extra explanatory power by adding the term of the rank of the firm-specific risk. We test the hypothesis:

  $H_0 : \beta_2 = 0$ in the regression:

$$PPS = \alpha + \beta_1 TOTAL\_RISK + \beta_2 ID\_RISK + \beta_3 CONTROL + \epsilon \qquad (1.3.24)$$

  Of the 24 regression models that I check, 22 reject H_0 at the 1% significance level, the remaining two reject H_0 at the 5% level.

The difference in predictions of my model and the existing model will be most significant in a setting where the stocks have similar total risk levels, but different proportion of firm-specific risk. In such a setting, the existing models will predict that the pay-performance sensitivity will be relatively stable across firms, but my model will predict that the pay-performance sensitivity will be lower for those stocks with a higher level of firm-specific risks, even though their total risk levels are comparable.

I do a test to directly see if the data support my predictions in such a setting. I divide the whole sample into 10 equal-numbered groups each year according to their total dollar risk levels. Then, within each group, I test whether the level of firm-specific risk is correlated with the pay-performance sensitivity level. I find that within each group, there is a significant and negative relationship between firm-specific risk level and the pay-performance sensitivity, using median regression, robust regression and OLS regression controlling for year and firm fixed effects, and the effect of firm-specific risk on pay-performance sensitivities is economically significant, such that it cannot be explained by the small variation of the total risk rank within each group.

sub-sample study

One practical concern that people might have is: CEOs might only have limited capacity in optimizing on the market risk exposure. In the case that CEOs have a very large exposure in the firm risk, they might not be able to completely offset the risks through shorting the market portfolio. This might imply that systematic risk will have a different impact on the PPS of CEOs with large and small holdings in the firm's equity: while for CEOs with little holdings in firm equity the systematic risk should not affect the PPS, for CEOs with larger holdings in firm equity, we might see that PPS decreases with systematic risk of firm also.[19].

I divide the whole sample into two sub-samples: one with above-median equity holdings in the firm (in terms of total dollar amount), and the other with below-median equity holdings in the firm. I run the same regression as in Table 6 for the two sub-samples.

The results are reported in Table 8. We again report three different regression methods (OLS, median and robust). We could see that across various regression methods and both sub samples, the idiosyncratic risk measure is always negatively related to PPS, while the systematic risk measure have different patterns for the two sub sample. For CEOs with above-

[19] I am grateful to an anonymous referee for suggesting this test.

the-median holdings in the firm's equity, systematic risk is actually significantly negative for both median and robust regression, and negative but not significant for the OLS regression. Thus, there seems to be some support that for CEOs with large holdings in the firm's equity, systematic risk could also negatively affect the PPS.

We also re-do the above sub-sample study, using the exposure to "market risk" rather than the exposure to the firm equity. This gives very similar results as reported in Table 8.

Some other robustness checks are performed and the results does not seem to change with these robustness checks. The detailed results are reported in the appendix.

In the subsequent analysis, the focus will be on the firm-specific risk measure to study how the optimal compensation contract will trade-off diversification with incentive.

1.3.3 The two sides of the incentive-diversification tradeoff

Having demonstrated that only firm-specific risk should be considered the cost of giving incentive, we now refine the test to capture the real tradeoff between diversification and incentive. We show that, as predicted by theory, pay-performance sensitivity decreases with firm-specific risk level when we control for productivity of CEO effort, and increases with productivity of CEO effort when we control for firm-specific risk.

The role of firm-specific risk on incentive

Two stage regression

I first perform the two stage regression explained before. I estimate the relation between total pay performance sensitivity and firm-specific risk measures. The specification that I test is:

$$PPS = \alpha + \beta_1 ID\_RISK + \beta_2 CONTROL + \beta_3 YEAR\_DUMMY + \epsilon \qquad (1.3.25)$$

Interpreting the results

The results are summarized in Tables 9 and 10. In Table 9, median regression results are reported. In Table 10, OLS regression controlling for year and firm fixed effects are reported. In each of these tables, I report the coefficients and t-statistics for regressions using the three risk measures estimated from the market model regression and the two risk measures generated through the Campbell et al. (2000) approach.[20] In these regressions, I control for CEO productivity, firm size and CEO tenure.

The risk measures are in percentage rank measures from 0 to 100. Thus, the lowest firm-specific risk stock will have a rank of 0, while the highest ranked firm-specific risk stock will have a rank of 100.

The coefficient on firm-specific risk is statistically significant for both the median regression and the OLS regression controlling for year and firm fixed effects and using the heteroskedasticity-robust standard errors.

In the median regression, for all five regressions, the coefficients on the firm-specific risk measures are close to -0.2. The constant is about 21. For the year 1998, where all the year dummies will have value of zero, the average pay-performance sensitivity at the median firm-specific risk level (rank of firm-specific risk equal to 50) is about \$10.7 for a \$1,000 increase in shareholder value. However, the average pay-performance sensitivity at the 75% firm-specific risk level will drop to about \$5.6, a decrease of about 46%. Thus, the firm-specific risk indeed has a big impact on the level of pay-performance sensitivity.

In the OLS regression, the coefficients on the firm-specific risk measures are close to -0.5, while the constant is about 40. Thus, for the year 1998, a movement of the firm-specific risk from the median level to the 75th percentile level will reduce the pay-performance sensitivity level from about \$17 per \$1000 increase in shareholder value to about \$4 per \$1000 increase in shareholder value, a reduction of about 75%.

As a robustness check, I test the most important component of compensation, the change in financial wealth for a \$1,000 change in CEO wealth. I find that the firm-specific risk measure exerts a similar effect on the pay-performance sensitivity in CEO financial wealth. To save space, the results are not reported here.

Interaction regression

I also employ the interaction regression to cross-validate the previous results. Instead of going through the middle step of constructing the pay-performance sensitivity measures, the

[20]The Fama-French decomposition of risk is not reported, but the results there are qualitatively similar. The results are available from the author.

interaction regression will directly test the following specification:

$$\Delta CEO\_wealth = \beta_1 Z + \beta_2 Z * CDF(variance) + \beta_3 CDF(variance) + constant + \epsilon \quad (1.3.26)$$

Where Z is the performance of the firm, CDF(variance) is the cumulative density function on firm variance, ranging from 0 to 1. This method is adopted by Aggarwal and Samwick (1999a) and Bertrand and Mullainathan (2000a). Under this specification, the pay-performance sensitivity is captured by the term $\beta_1 + \beta_2 CDF(variance)$. When we have the lowest firm-specific risk, CDF(variance) = 0, the pay-performance sensitivity will be β_1; when we have the highest firm-specific risk, CDF(variance) = 1, the pay-performance sensitivity will be $\beta_1 + \beta_2$. Thus the effect of firm-specific risk on pay-performance sensitivity is captured by the coefficient β_2.

Tables 11 and 12 report the results of the interaction regression. Table 11 is the result of the median regression, while Table 12 is the results of the OLS regression controlling for year and firm fixed effects. Although the magnitudes of the coefficient β_2 are different for these two different regression methods, the results are qualitatively similar. Furthermore, a move of the firm-specific risk from the median level to the 75th percentile level will decrease the pay-performance sensitivity by about 45% under both regression methods. I also tried to control for size related heterogeneity and CEO effort productivity by adding log of firm size and proxies for the CEO effort productivity. The results are qualitatively similar. To save space, they are not reported.

The role of productivity of effort on incentive

In this subsection, we test various proxies of the productivity of CEO effort.

All three proxies work as the theory predicts. The median regression and robust regression give similar results and adding the control variables about CEO tenure and market cap does not alter the qualitative results. The results are reported in Tables 13. We summarize the findings of Table 13, which controls for CEO tenure and market cap.

Tobin's Q as a proxy for CEO effort productivity

At the median level of firm-specific risk, the average pay-performance sensitivity is about \$13.5 at the median productivity of CEO effort. It increases to about \$14.2 at the 75th percentile of CEO effort productivity, an increase of about 6% for the 25 percentile increase in CEO effort productivity.

Investment as a proxy for CEO effort productivity

At the median level of firm-specific risk, the average pay-performance sensitivity is about \$13.2 at the median productivity of CEO effort. It increases to about \$14.4 at the 75th percentile of CEO effort productivity, an increase of about 9% for the 25 percentile increase in CEO effort productivity.

Firm age as a proxy for CEO effort productivity

At the median level of firm-specific risk, the average pay-performance sensitivity is about \$14.6 at the median productivity of CEO effort. It increases to about \$16.8 at the 75th percentile of CEO effort productivity (which is the 25th percentile of firm age), an increase of about 14% for the 25 percentile increase in CEO effort productivity.

One competing theory would be an intuitive story relating firm performance and incentives. Firms that are doing well could afford to reward their CEOs more with a higher pay-performance sensitivity, or they might expect to continue to do well in the future. Thus they are willing to give their CEOs a higher pay-performance sensitivity. It might be that firms tend to reward CEOs more on the upside, and punish them less on the downside. Some of the proxies that I use for the productivity of CEO effort might be correlated with good performance, for example, Tobin's Q.

To differentiate between the principal agent theory and this competing story, I do a robustness check to compare the pay-performance sensitivity of distressed firms with that of the whole sample. Since distressed firms should heavily rely on their CEOs' effort, under the prediction of principal agent model, we would expect to see a higher pay-performance sensitivity for this sub sample than for the rest of firms. I find that indeed distressed firms (here defined as firms with the bottom decile returns each year) have a pay-performance sensitivity that is about twice as large as the rest of the firms in the sample. The detailed results are in the appendix.

Some robustness checks are performed and the results do not seem to change with these robustness checks. The detailed results are reported in the appendix.

1.4 Conclusion

I extend the basic principal agent model to incorporate two important aspects of finance: first, shareholders are risk averse to the market risk, and second, CEO can trade the market portfolio. I show that incorporating these two considerations affects the design of optimal compensation contracts. When we don't allow CEOs to trade on the market portfolio, a higher firm-specific risk should always lower the pay performance sensitivity, while a higher market risk is ambiguously related to incentive. Furthermore, when CEOs can trade the market portfolio without restrictions, market risk is irrelevant to the optimal level of incentive.

I design empirical tests to investigate the predictions of these models. Using a variety of risk measures and regression methods, I demonstrate that there is a robust pattern between the idiosyncratic risk of firm and the incentive level. The relation is both economically and statistically significant. There is no such robust relation between market risk of firm and incentive level.

In doing the empirical tests of the principal agent theory, I incorporate the critique by Prendergast (2000), and address simultaneously both the cost and the benefit of giving CEO incentives. This also works to link two largely independent literature, on estimating the impact of either side of the tradeoff alone, and arrive at a balanced and complete picture of how principal agent theory could explain the way firms set incentives.

Table 1

Performance Measures for the year 1995

			Dollar Performance ($Mn)		
			Performance Benchmarks Used		
	Market Cap ($Mn)	Nominal Return (%)	Zero	Inflation	Market return
Number of Obs.	1550	1580	1545	1545	1545
Mean	2538.99	29.86	925.36	860.96	79.19
Standard Deviation	6306.53	49.93	3014.05	2876.49	1590.36
Skewness	6.80	2.51	6.92	6.90	2.95
Kurtosis	63.07	14.82	61.02	60.78	63.55
Maximum	31053.29	208.642	15163.34	14544.93	6650.94
Minimum	59.76	-58.49	-556.71	-645.71	-2743.45
Percentiles					
P5	111.48	-35.50	-194.82	-239.56	-1044.99
P10	153.71	-21.60	-80.21	-93.51	-622.09
P25	271.54	1.99	8.00	-1.96	-198.78
P40	452.65	16.79	85.03	71.55	-83.13
P50	668.77	26.05	145.15	133.45	-38.70
P60	1063.35	34.49	255.34	230.74	4.04
P75	2078.69	48.80	658.70	607.56	111.63
P90	5587.00	76.16	2019.68	1926.73	612.82
P95	10291.09	101.79	4035.01	3826.20	1281.83

Note: Market capitalization and the dollar performances are in millions of 1995 dollars. Nominal returns are in percentages. All measures are truncated at 1% and 99% to remove outliers.

Table 2

A: Measures of sensitivity of CEO's financial wealth change to performance

	Real change	Hypothetical change			
	unit: ($000)	HL1 ($000)	HL2	HL3	HL4 ($)
Number of Obs.	5031	5132	5107	5107	5128
Mean	3798.35	4973.40	96%	4.065	25.889
Standard Dev.	19987.95	9317.80	46%	1.932	45.702
Skewness	2.98	4.41	0.37	0.367	3.604
Kurtosis	16.82	22.47	-0.642	-0.642	14.779
Maximum	124590.26	65471.38	209%	8.835	286.129
Minimum	-57504.68	103.06	14%	0.593	0.271
Percentiles					
P5	-14724.18	279.76	29%	1.237	0.669
P10	-6359.32	433.28	40%	1.681	1.283
P25	-1155.63	898.76	61%	2.557	3.496
P40	82.79	1487.84	78%	3.311	6.797
P50	728.23	2000.09	91%	3.846	9.871
P60	1713.98	2691.97	104%	4.407	14.331
P75	4587.16	4788.89	130%	5.471	25.350
P90	15035.82	10969.94	162%	6.848	63.469
P95	30987.33	18845.00	176%	7.435	110.829

Note:

1. The real changes in CEO's firm-specific financial wealth due to performance are calculated as the realized real value change of CEO's beginning-of-year stock and option portfolio.
2. The hypothetical changes in CEO's firm-specific financial wealth are calculated in accordance with Hall and Liebman (1998), table6. Four measures of change in CEO's firm-specific financial wealth are reported, for a hypothetical change of firm performance from 50th to 75th percentile: (1) HL1: dollar change; (2) HL2: percentage change; (3) HL3: elasticity of CEO wealth change to firm value change; (4) HL4: Jensen and Murphy statistics, which measures how much CEO wealth changes for a $1000 increase in shareholder value.
3. All measures are truncated at the 1% and 99% to remove outliers.
4. All dollars are 1994 constant dollars.

Table 2

B: Measures of Pay-performance sensitivity from CEO's human capital change

	Real change	Hypothetical changes				
	unit:($000)	JM11 ($)	JM21 ($)	JM22 ($)	JM23 ($)	JM24 ($)
Number of Obs.	7301	9432	9018	9018	8587	8582
Mean	2095.56	2.44	2.53	4.20	2.50	4.55
Standard Dev.	2684.65	0.86	1.18	2.22	1.13	2.51
Skewness	3.20	-0.55	0.25	0.98	0.48	0.30
Kurtosis	12.05	0.92	-0.27	1.36	0.82	-0.40
Maximum	16747.37	4.43	5.62	11.44	6.00	11.14
Minimum	139.38	0.18	0.18	0.20	0.13	0.22
Percentiles						
P5	270.44	0.40	0.49	0.77	0.54	0.74
P10	364.07	1.38	1.16	1.90	1.16	1.55
P25	630.08	2.07	1.67	2.77	1.87	2.33
P40	931.18	2.54	2.05	3.52	2.22	3.81
P50	1189.90	2.54	2.43	3.95	2.46	4.85
P60	1522.95	2.54	2.75	4.15	2.59	5.30
P75	2409.64	2.83	3.28	4.82	2.92	5.97
P90	4673.25	3.48	4.07	7.33	3.94	7.92
P95	7187.80	3.94	4.44	8.86	4.59	9.03

Note: The real change in human capital is measured by total direct compensations including salary, bonus and the value of other grants that the CEO receives each year. The hypothetical changes in human capital are calculated according to Jensen and Murphy (1990a). We use Jensen and Murphy (1990a) equation (2) to estimate the sensitivity of pay to current and lagged performance, assuming that CEO retires at 70 and real riskless rate is 3% in calculating the PV of future pay change. JM11(inflation rate benchmark, same regression coefficients for all years); JM21 (inflation rate benchmark, different regression coefficients for different years), JM22 (market return benchmark, different regression coefficients for different years), JM23 and JM24 are same as JM21 and JM22, respectively, except using a "mid-year convension", where if a stock has fiscal yearend before June, then the end-of-year compensation numbers will be treated as last year's end-of-year compensation numbers. All measures are truncated at 1% and 99% to remove outliers. All dollars are 1994 constant dollars.

Table 3

A. Measures of idiosyncratic risks

	Market model regression measures			Campbell et al. (2000) measures	
	monthly	weekly1	weekly2	monthly1	monthly2
No. of Obs.	9431	9432	9432	9430	9431
Mean	0.00747	0.00182	0.00185	0.00924	0.00735
Standard Dev.	0.00344	0.00086	0.00085	0.00512	0.00420
Skewness	0.57550	0.64039	0.73023	0.81494	0.82627
Kurtosis	-0.44166	-0.25128	0.32503	-0.09229	-0.08127
Maximum	0.01718	0.00427	0.00474	0.02286	0.01857
Minimum	0.00240	0.00058	0.00059	0.00121	0.00093
Percentiles					
P5	0.00284	0.00072	0.00072	0.00327	0.00243
P10	0.00332	0.00079	0.00080	0.00351	0.00294
P25	0.00472	0.00109	0.00120	0.00534	0.00430
P40	0.00585	0.00137	0.00144	0.00673	0.00522
P50	0.00679	0.00170	0.00173	0.00826	0.00598
P60	0.00800	0.00200	0.00201	0.00963	0.00754
P75	0.01019	0.00241	0.00247	0.01213	0.01018
P90	0.01231	0.00294	0.00296	0.01734	0.01339
P95	0.01295	0.00349	0.00326	0.02001	0.01524

B. Correlations among different measures of idiosyncratic risks

		Market model regression measures			Campbell et al. (2000) measures	
		monthly	weekly1	weekly2	monthly1	monthly2
Market model regression measures	monthly	1	0.90012	0.95188	0.83745	0.84872
	weekly1	0.90012	1	0.94739	0.82613	0.89321
	weekly2	0.95188	0.94739	1	0.84178	0.86879
Campbell et al. measures	monthly1	0.83745	0.82613	0.84178	1	0.90005
	monthly2	0.84872	0.89321	0.86879	0.90005	1

Notes:

1. Three risks measures are calculated using the market model: monthly (5 year of monthly data), weekly1 (1 year of weekly data), weekly2 (3 years of weekly data). Two idiosyncratic risk measures are calculated using the Campbell et al. (2000) regress-free method: monthly1 (5 years of monthly data) and monthly2 (2 years of monthly data).
2. All risks are in rates of returns, and not annualized.

Table 4

Regression of PPS from financial wealth on the total risk measure and other control variables

Variable	OLS Regression	Median Regression	Robust Regression
Intercept	73.87446	60.50553	33.57616
	24.40	2.991231	3.286244
rank of total volatility	-0.5050095	-0.1637194	-0.1134609
	0.08	0.0073425	0.0080603
ln(sales)	-4.498671	-8.804751	-2.838473
	6.07	0.772266	0.8489837
$\ln(\text{sales})^2$	0.2553527	0.4713237	0.108721
	0.33	0.0487379	0.0536183
capital/sales	-5.945588	-0.9967084	-1.847046
	3.16	0.5531135	0.6077672
capital/sales^2	0.1638715	-0.0211367	0.3485562
	0.90	0.1877925	0.206882
RND/capital	2.876127	-1.144745	-0.9769489
	3.01	0.5071909	0.5570812
missing RND	1.022898	-0.5676998	-0.8823418
	1.98	0.2571163	0.2822676
Advertising/Capital	24.07639	-0.4460267	-2.795533
	10.49	0.6630383	0.7550523
missing Advertising	4.419811	-0.9789651	-1.561945
	2.49	0.3362533	0.371787
Invest/capital	-2.619656	7.901633	7.72488
	6.55	0.9466254	1.03809

Notes:

1. Heteroskedasticity-robust standard errors are reported for the OLS regression.
2. Year effets are included in the regressions but not reported to save space.

Table 5

Regression of PPS from financial wealth on the total risk, idiosyncratic risk and other control variables

Variable	OLS Regression	Median Regression	Robust Regression
Intercept	69.00574	59.76021	32.32657
	20.79378	2.939683	3.260863
rank(total risk)	0.2966054	0.0000362	-0.0100445
	0.1939375	0.0274151	0.0304131
rank(idiosyncratic risk)	-0.843233	-0.173266	-0.1087961
	0.1968473	0.0278066	0.0308694
ln(sales)	-3.086045	-8.466618	-2.545144
	5.374061	0.7598282	0.8427556
$\ln(sales)^2$	0.2064962	0.4547045	0.0968453
	0.3389557	0.047913	0.0531547
capital/sales	-5.582224	-0.5305221	-1.848127
	3.840849	0.5427489	0.6023186
$capital/sales^2$	0.1887695	-0.1291893	0.3712036
	1.307107	0.1843128	0.2049793
RND/capital	3.36361	-1.030212	-0.8865772
	3.521514	0.4974803	0.5522408
missing RND	2.091043	-0.4432846	-0.702613
	1.800733	0.2544394	0.2823893
Advertising/Capital	23.20137	-0.5525144	-2.879358
	4.77484	0.6501776	0.7487862
missing Advertising	4.118899	-1.211557	-1.532746
	2.350024	0.3291887	0.3685287
Invest/capital	-1.59705	8.105055	7.789786
	6.563059	0.9285428	1.029213

Notes:

1. Heteroskedasticity-robust standard errors are reported for the OLS regression.
2. Year effets are included in the regressions but not reported to save space.

Table 6

Regression of PPS from financial wealth on the systematic risk, idiosyncratic risk and other control variables

Variable	OLS Regression	Median Regression	Robust Regression	IV Regression
Intercept	68.06396	60.82449	32.58505	77.36087
	24.89615	3.000864	3.264422	21.16975
rank(systematic risk)	-0.0431238	0.0168805	0.0144416	0.249806
	0.1885482	0.0229872	0.0252334	0.3237859
rank(idiosyncratic risk)	-0.5114312	-0.1897143	-0.1321208	-0.8056384
	0.1920031	0.0231938	0.025496	0.3089941
ln(sales)	-2.957638	-8.678798	-2.605798	-6.032265
	6.199981	0.7752502	0.8434125	5.462805
$\ln(\text{sales})^2$	0.2071027	0.4660485	0.0993563	0.411665
	0.3359489	0.0488417	0.0531675	0.3484217
capital/sales	-5.428665	-0.580123	-1.851999	-2.115216
	3.211189	0.5526355	0.602245	4.42381
capital/sales^2	0.1354692	-0.1332298	0.3759105	-0.550227
	0.9090386	0.1877562	0.2050265	1.376118
RND/capital	3.649972	-1.067448	-0.9185735	6.592128
	3.094053	0.5084653	0.5524636	3.72762
missing RND	1.62393	-0.4493896	-0.6665545	1.541289
	2.049499	0.2584719	0.2812732	1.833515
Advertising/Capital	23.61575	-0.5831138	-2.914769	24.90858
	10.32735	0.6629299	0.7484514	4.998672
missing Advertising	4.218051	-1.208122	-1.526529	4.197627
	2.46622	0.3362358	0.3684654	2.440617
Invest/capital	-1.4758	7.976339	7.790351	3.355519
	6.438593	0.9471534	1.029477	6.882693

Notes:

1. Heteroskedasticity-robust standard errors are reported for the OLS regression.
2. Year effets are included in the regressions but not reported to save space.

Table 7

Regression of PPS from total wealth on the systematic risk, idiosyncratic risk and other control variables

Variable	OLS Regression	Median Regression	Robust Regression	IV Regression
Intercept	70.31949	61.2763	37.95609	79.71406
	24.77527	3.761007	3.256549	21.10974
rank(systematic risk)	-0.0560621	-0.0172634	0.0106458	0.22623
	0.1882111	0.0290944	0.0251763	0.3230145
rank(idiosyncratic risk)	-0.5010367	-0.1589905	-0.1313672	-0.7859569
	0.1913549	0.0293926	0.0254371	0.3082424
ln(sales)	-2.937808	-8.272447	-3.30576	-6.046603
	6.170356	0.9714309	0.8413659	5.447042
$\ln(\text{sales})^2$	0.2056885	0.446142	0.1398343	0.4136232
	0.3342099	0.0612243	0.0530397	0.3474189
capital/sales	-5.368218	-1.062846	-1.715486	-2.151907
	3.21925	0.6957563	0.6025024	4.425694
capital/sales^2	0.1076516	-0.0085765	0.2932974	-0.5652881
	0.9111255	0.2361596	0.2049295	1.375328
RND/capital	4.115247	-0.2641357	-0.3580329	7.03455
	3.087655	0.6333972	0.5511711	3.717381
missing RND	1.710568	-0.2887766	-0.5143405	1.647535
	2.046153	0.3246933	0.2808338	1.829746
Advertising/Capital	23.46092	-0.8215324	-3.199208	24.67328
	10.33681	0.8649248	0.7468215	4.985663
missing Advertising	3.937719	-1.536558	-1.853034	3.889224
	2.47063	0.4255702	0.3678614	2.43557
Invest/capital	-1.039697	8.246951	8.030941	3.789707
	6.422959	1.188991	1.027091	6.864002

Notes:

1. Heteroskedasticity-robust standard errors are reported for the OLS regression.
2. Year effets are included in the regressions but not reported to save space.

Table 8

Regression of PPS from total wealth on the systematic risk, idiosyncratic risk and other control variables

Panel A: CEOs with above median holdings in the firm's equity

Variable	OLS Regression	Median Regression	Robust Regression
Intercept	131.1001	147.5695	116.6688
	42.95256	8.482522	5.966669
rank(systematic risk)	-0.4602097	-0.2947548	-0.1827378
	0.4126144	0.0741069	0.0521004
rank(idiosyncratic risk)	-0.9799442	-0.2916141	-0.1856427
	0.4229715	0.0749721	0.0527142
ln(sales)	-3.770556	-20.74755	-16.55022
	10.87519	2.151605	1.514737
$\ln(\text{sales})^2$	0.484239	1.195184	0.9114668
	0.5858349	0.1321255	0.0930519
capital/sales	-0.6511653	1.006847	0.2946651
	6.163234	1.623994	1.152209
capital/sales^2	-1.638244	-1.069438	-0.527204
	1.678108	0.5362002	0.3818449
RND/capital	7.598095	1.34074	0.4781045
	4.299648	1.1936	0.8380491
missing RND	5.069485	0.6296332	-0.1827058
	3.147653	0.6918661	0.4866633
Advertising/Capital	48.31666	11.41384	-0.5391934
	10.43704	1.762909	1.235894
missing Advertising	14.62554	2.012055	-0.8436097
	3.077934	0.8387133	0.5899357
Invest/capital	-19.7909	-1.434598	0.7374459
	9.984443	2.529932	1.787803

Table 8

Panel B: CEOs with below median holdings in the firm's equity

Variable	OLS Regression	Median Regression	Robust Regression
Intercept	74.51506	47.06342	32.41411
	9.71533	3.288759	2.83588
rank(systematic risk)	-0.0045402	-0.017708	-0.0170508
	0.0403287	0.0222678	0.0190456
rank(idiosyncratic risk)	-0.3130597	-0.1589422	-0.1322195
	0.0550276	0.0223925	0.0191542
ln(sales)	-11.55878	-7.449999	-4.206305
	2.440157	0.8689557	0.7511841
$\ln(\text{sales})^2$	0.7902289	0.4806269	0.268878
	0.1519293	0.0570952	0.0494934
capital/sales	-2.483638	-1.695532	-1.819136
	0.8711568	0.5545865	0.4782057
capital/sales^2	0.4025245	0.4323499	0.3768851
	0.3420231	0.1920038	0.1657807
RND/capital	-1.209301	-0.0368376	0.3303668
	1.350432	0.6895349	0.6105056
missing RND	-1.676021	-0.3466045	-0.3251754
	0.8838124	0.28586	0.2452508
Advertising/Capital	-2.949456	-1.63927	-0.790724
	1.132647	0.7935014	0.6840853
missing Advertising	-1.729891	-0.3870345	0.1270823
	0.8113998	0.415253	0.3560102
Invest/capital	-1.581737	4.782568	4.62093
	3.109912	1.02784	0.8927042

Notes:

1. Heteroskedasticity-robust standard errors are reported for the OLS regression.
2. Year effets are included in the regressions but not reported to save space.

Table 9

Impact of idiosyncratic risk: controlling for firm size, CEO effectiveness, and CEO tenure, median regression result

Independent Variable	Dependent Variable: Pay-Performance Sensitivity				
	(1)	(2)	(3)	(4)	(5)
Intercept	20.766	20.678	20.764	20.336	20.192
id_risk_monthly	-0.203 (-37.341)				
id_risk_weekly1		-0.200 (-38.371)			
id_risk_weekly2			-0.201 (-39.159)		
id_risk_CP_monthly1				-0.197 (-35.595)	
id_risk_CP_monthly2					-0.196 (-39.744)
PPS at median risk level	10.593	10.682	10.707	10.488	10.379
PPS at 75th percentile risk level	5.506	5.684	5.678	5.564	5.472
Percentage reduction in PPS	48.021	46.787	46.964	46.950	47.275
Number of Obs	2929	2929	2929	2929	2929

1. The regression being run is: $PPS=\alpha+\beta_1 ID\_RISK+\beta_2 YEAR+\beta_3 MARKET+\beta_4 INVEST+\beta_5 TENURE+\varepsilon$.

2. PPS is defined as the dollar increase of CEO's total wealth, for a $1000 increase in shareholder value.

3. Id_risk is the percentage rank of dollar idiosyncratic risk estimated using market model regressions; id_risk_CP is the percentage rank of dollar idiosyncratic risk estimated using Campbell et al. (2000) method.
4. t-statistics are reported under the coefficients.

5. The coefficients for control variables and year dummies are not reported in the table.

Table 10

Impact of idiosyncratic risk: controlling for firm size, CEO effectiveness, and CEO tenure, OLS regression result

Independent Variable	Dependent Variable: Pay-Performance Sensitivity				
	(1)	(2)	(3)	(4)	(5)
Intercept	42.372	42.407	42.167	40.061	41.093
id_risk_monthly	-0.512				
	(-6.834)				
id_risk_weekly1		-0.510			
		(-6.895)			
id_risk_weekly2			-0.510		
			(-6.844)		
id_risk_CP_monthly1				-0.485	
				(-6.865)	
id_risk_CP_monthly2					-0.490
					(-6.848)
PPS at median risk level	16.761	16.909	16.666	15.818	16.598
PPS at 75th percentile risk level	3.955	4.161	3.916	3.696	4.350
Percentage reduction in PPS	76.404	75.395	76.506	76.633	73.790
R^2	0.1934	0.1926	0.1930	0.1859	0.1872
Number of Obs	2929	2929	2929	2929	2929

1. The regression being run is: $PPS=\alpha+\beta_1 ID\_RISK+\beta_2 YEAR+\beta_3 MARKET+\beta_4 INVEST+\beta_5 TENURE+\varepsilon$.

2. Each regression controls for firm fixed effects. t-statistics using heteroskedasticity-robust standard errors are reported below the coefficients.

3. id_risk is the percentage rank of dollar idiosyncratic risk estimated using market model regressions; id_risk_CP is the percentage rank of dollar idiosyncratic risk estimated using Campbell et al. (2000) method.

4. t-statistics are reported under the coefficients.

5. The coefficients for control variables and year dummies are not reported in the table.

Table 11

Impact of idiosyncratic risk: Interaction regression, Median regression result

Independent Variables	Dependent Variable: total CEO wealth change		
	(1)	(2)	(3)
Performance (Z)	19.373	18.996	19.323
	117.101	104.579	121.415
Z X CDF(id_risk_monthly)	-18.611		
	-111.394		
CDF(id_risk_monthly)	13.118		
	11.093		
Z X CDF(id_risk_weekly1)		-18.211	
		-99.262	
CDF(id_risk_weekly1)		12.596	
		9.599	
Z X CDF(id_risk_weekly2)			-18.560
			-115.458
CDF(id_risk_weekly2)			12.665
			11.055
	Estimated Pay-performance sensitivity		
PPS at the median level of idiosyncratic risk	10.067	9.891	10.044
PPS at the 75th percentile level of idiosyncratic risk	5.414	5.338	5.404
Percentage reduction in PPS	46.22%	46.03%	46.20%

Note:

1. The regression being run is: CEO_wealth_change=$\alpha+\beta_1 Z+\beta_2 F(\sigma^2)Z+\beta_3 F(\sigma^2)+\beta_4$Year_dummy$+\varepsilon$, where Z is the dollar return to shareholders, $F(\sigma^2)$ is the cumulative distribution function (CDF) of the dollar idiosyncratic risk in the firm, ranging from 0 to 1.
2. t-statistics are underneath each coefficient estimate.

3. Three measures of risk are used: 5 year monthly data (1), 1 year weekly data (2) and 3 year weekly data (3).

Table 12
Impact of idiosyncratic risk: Interaction regression, OLS regression result

Independent Variables	Dependent Variable: total CEO wealth change		
	(1)	(2)	(3)
Performance (Z)	49.674	48.605	49.999
	8.247	8.381	8.227
Z X CDF(id_risk_monthly)	-47.127		
	-7.670		
CDF(id_risk_monthly)	-26.951		
	-0.718		
Z X CDF(id_risk_weekly1)		-46.049	
		-7.805	
CDF(id_risk_weekly1)		-29.416	
		-0.776	
Z X CDF(id_risk_weekly2)			-47.462
			-7.795
CDF(id_risk_weekly2)			-29.107
			-0.758
	Estimated Pay-performance sensitivity		
PPS at the median level of idiosyncratic risk	26.110	25.581	26.269
PPS at the 75th percentile level of idiosyncratic risk	14.328	14.068	14.403
Percentage reduction in PPS	45.12%	45.00%	45.17%

Note:

1. The regression being run is: CEO_wealth_change=$\alpha+\beta_1 Z+\beta_2 F(\sigma^2)Z+\beta_3 F(\sigma^2)+\beta_4$Year_dummy+$\varepsilon$, where Z is the dollar return to shareholders, $F(\sigma^2)$ is the cumulative distribution function (CDF) of the dollar idiosyncratic risk in the firm, ranging from 0 to 1.

2. Each regression controls for firm fixed effects. t-statistics using heteroskedasticity-robust standard errors are reported below the coefficients.
3. Three measures of risk are used: 5 year monthly data (1), 1 year weekly data (2) and 3 year weekly data (3).

Table 13

Testing proxies of effectiveness of CEO effort, controlling for CEO tenure and market cap, median regression

	Dependent Variable: PPS					
Independent Variable	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	22.677	22.529	22.766	20.867	20.863	21.032
Idiosyncratic risk						
id_risk_monthly	-0.220			-0.201		
	-34.274			-35.293		
id_risk_weekly1		-0.217			-0.199	
		-35.145			-35.234	
id_risk_weekly2			-0.221			-0.203
			-38.648			-39.660
Proxies of effectiveness of CEO effort (k)						
Rank_q	0.033	0.033	0.034			
	5.987	6.358	6.932			
Rank_invest				0.047	0.047	0.048
				9.440	9.589	10.699
Number of Obs	3177	3177	3177	2929	2929	2929
PPS for Year 1998, at median idiosyncratic risk level						
Median k level	13.322	13.360	13.421	13.155	13.245	13.251
75th percent k level	14.142	14.197	14.267	14.320	14.419	14.442
Percentage Change	6.15%	6.26%	6.30%	8.86%	8.87%	8.99%

Note:

1. The regression being run is: $PPS=\alpha+\beta_1 ID\_RISK+ \beta_2 PROXY\_EFFECT+\beta_3 TENURE+\beta_4 MARKET+\beta_5 YEAR+\varepsilon$.

2. PPS is defined as the dollar amount of increase of CEO's total wealth, for a $1000 increase in shareholder value.

3. Rank_q is the percentage rank of Tobin's q, from 0 to 100; Rank_invest is the percentage rank of investment measure, from 0 to 100.
4. t-statistics are reported below the coefficients.
5. Coefficients on control variables are not reported.

Table 13 (continued)

Testing proxies of effectiveness of CEO effort, controlling for CEO tenure and market cap, median regression

Independent Variable	Dependent Variable: PPS (1)	(2)	(3)	(4)	(5)	(6)
Intercept	30.496	30.389	30.224	23.151	23.100	23.082
Idiosyncratic risk						
id_risk_monthly	-0.232			-0.194		
	-33.801			-27.486		
id_risk_weekly1		-0.228			-0.190	
		-33.739			-27.706	
id_risk_weekly2			-0.229			-0.195
			-34.716			-33.599
Proxies of effectiveness of CEO effort (k)						
Rank_q				0.011	0.011	0.012
				1.808	1.884	2.446
Rank_invest				0.038	0.038	0.038
				6.459	6.632	7.709
Rank_age	-0.083	-0.087	-0.084	-0.045	-0.047	-0.043
	-12.364	-13.393	-13.135	-7.200	-7.765	-8.517
Number of Obs	3177	3177	3177	2929	2929	2929
PPS for Year 1998, at median idiosyncratic risk level						
Median k level	14.748	14.632	14.581			
75th percent k level	16.812	16.816	16.681			
Percentage Change	14.0%	14.9%	14.4%			

Note:

1. The regression being run is: $PPS=\alpha+\beta_1 ID\_RISK+ \beta_2 PROXY\_EFFECT+\beta_3 TENURE+\beta_4 MARKET+\beta_5 YEAR+\varepsilon$.

2. PPS is defined as the dollar amount of increase of CEO's total wealth, for a $1000 increase in shareholder value.

3. Rank_q is the percentage rank of Tobin's q, from 0 to 100; Rank_invest is the percentage rank of investment measure, from 0 to 100; Rank_age is the percentage rank of firm age, from 0 to 100.
4. t-statistics are reported below the coefficients.
5. Coefficients on control variables are not reported.

6. Columns (4) to (6) reports regression results where all three proxies of the effectiveness of CEO effort are included.

Chapter 2

Risk Sharing and Labor Intermediation

2.1 Introduction and literature review

SYNOPSIS: This section motivates our study. We first argue that labor income/human capital is very risky, and specialization further increases the riskness of human capital. Then we list several institutional arrangements that could potentially reduce the labor risk, among them labor intermediation. We then define labor intermediation formally and give some examples (temporary staffing companies and consulting firms) and document the rapid development of this business as a result of technological innovation and changing demand. Last we list the objectives for the rest of the paper.

This paper studies labor intermediation and its impact on labor risk sharing and human capital specialization. Before we go into more detail, we first briefly explore the special characteristics of the human capital that make it particularly risky (as compared to financial capital), and then discuss several institutional arrangements to share labor risk.

labor income risk, specialization, and risk management in labor market

Human capital is a significant ingredient of personal and national wealth[1]. The following

[1] Jorgenson and Fraumeni (1989) have estimated that for the period 1948 to 1984, the share of human capital in the aggregate wealth of the United States has been around 93%. A more recent account by Krueger (1999) has that labor's share in National Income has been ranging between 76% to 79% from 1988 to 1998, and slightly decreasing.

special characteristics of human capital make it particularly risky:

1. Non-tradability: cannot be sold off, also difficult to use as collateral. The main reason for this is the moral hazard problem[2][3].

2. Irreversibility: it takes a long time to build up human capital, and once established, one cannot "decrease human capital" and get money back or switch to some other skill[4].

3. Indivisibility: the deployment of human capital is often indivisible: you don't work 25% as a secretary, 50% as a manager and 25% as a janitor[5]. Unlike in financial market, where diversification often requires a small weight on each single asset, when it comes to human capital, choices are predominantly either zero or one. Indivisibility means the loss of diversification.

4. The cost of moving workers across production firms is high[6]

5. Many of the time, the labor income shocks cannot be easily hedged by *laborers themselves* on the financial market, because doing so sometimes requires borrowing and/or short selling certain financial assets, which might be difficult for some individuals who

---

[2] Although we typically talk of the moral hazard problem here as if it is one-sided, it is actually two sided: firm can also exert some effort that affect the product of labor, and therefore there is also a moral hazard problem on the side of the firm.

[3] The view that human capital is difficult to use as collateral is not without debate. Diaz-Gimenez et al (1992) report that at the end of 1986 the total market value of equities held by household was 0.80 GNP. On the other hand, the outstanding stock of mortgage (0.60 GNP), consumer credit (0.16 GNP) and bank loans to the household sector (0.04 GNP) also add up to 0.80 GNP. Although many of these credit have some form of primary collateral, human capital can be thought of as a secondary collateral here.

[4] We don't want to take this point too far, as, obviously, when the first best usage of the human capital is unavailable for some reason, there are often other ways to apply this accumulated human capital to some second-best usage. Thus, the accumulated human capital might still have a high "scrap value", therefore the investment therein is not completely wasted.

[5] To some degree this indivisibility is endogenous: you can choose to spread your human capital over 1000 professions or work simultaneously in 1000 types of jobs if you really want, but economies of scale makes such arrangement extremely inefficient.

[6] This is mostly due to the asymmetric information on labor market. To show the hiring and firing cost, here is some back of the envelope calculation: suppose that human capital is discounted at 20% per annum (roughly the discount rate of equity), and fired worker is given 10% of the annual salary as compensation, then the transaction cost as a percentage of the Net Present Value of the stream of labor income will be equal to is $\frac{10\% W_{annual}}{\frac{W_{annual}}{20\%}} = 2\%$, where W_{annual} is the annual wage. This does not include the cost of hiring a replacement, and any possible legal bills or loss in the morale of remaining workers, and it is already much higher than the transaction cost of most financial asset.

have only limited access to the financial markets. Furthermore, there does not always exist assets that are highly correlated with labor income[7].

In addition, recent technological development is propelling the demand for specialization of human capital:

1. Technological development is constantly refining functions. Most professionals cannot master all the skills out there and have to be more and more specialized: for example, computer programmers in the 80s could be generalists, today most programmers are specialized according the operation systems: UNIX, Windows, MAC, etc;

2. Technological development increases the ability to *standardize the function across firms*, so that a core function can be performed at different firms by the same laborer who process a certain (possibly very specialized) skills[8];

Thus, given that at one time, at most one or two skills of the laborer is used, it is more efficient from the society's point of view to let the laborer to be close to fully specialized in a particular function[9].

Specialization means to spend a lot of effort to develop certain particular skills and expertise, which, once used, can generate a higher productivity. However, more specialized skills also means a lower chance of finding a good match in any particular firm. For example, a jack-of-all-trade financial analyst in a manufacturing firm might not be very productive

[7]see Bodie (1975) and Shiller (1993) for ideas about issuing securities highly correlated with macroeconomic factors (including labor income) for risk sharing purposes, and more generally the security design literature as in Allen and Gale (1994) and Duffie and Rahi (1995).

[8]Note this and the first are not necessarily contradicting each other. Job functions can be very specialized (instead of having N types of jobs 10 years ago, we could have 2N types today), but the same type can be used across more firms today than 10 years ago, because the standardization, such as in the software industry, allows people to easily transform to another firm and do basically the same job. Thus, a temporary secretary, once mastering Word or Excel, can work in almost every firm

[9]The above applies in a developed country where risk sharing system is relatively sophisticated. In a developing country, due to the highly uncertain demand for labor in a society, there need to be a lot of flexibility in the training of labor force. If the society is equipped with a good risk-sharing mechanism, then not all individual workers need to maintain the flexibility even if the society as a whole has some uncertain demand regarding type of labor (the society only needs to assign a small group of workers to training that are comprehensive enough to maintain the optionality at the society level, and put the rest of the workers into highly specialized training). However, in developing countries the risk sharing mechanisms are typically imperfect, thus most likely the uncertainty in the aggregate demand of labor is taken by more than just the necessary amount of the laborers in the economy. This is just a different representation of the point made by Samantha Merton (1992).

in dealing with an Initial Public Offering than a person who specializes in Initial Public Offering, but the later will find himself idle (or doing something not requiring his specialized knowledge) most of the time in the manufacturing firm. We assume that the more specialized the human capital is, the higher the average labor productivity could be, but also the higher the risk of not finding a good match is. The jack-of-all-trade type of human capital is on average the least productive, but it is also the least risky. Therefore, when a risk averse laborer decides on the degree of human capital specialization, he will have to take into account the tradeoff between high return and high risk. Individual laborers might find it sub-optimal to be highly specialized, if they were to act alone. With labor risk sharing arrangements, laborers will have more incentive to further specialize, and this is in general value enhancing.

In the real world we observe several solutions to reduce the risk associated with human capital:

- Firms often share some of the labor productivity risk by issuing a long term labor contract at a relatively fixed wage. In doing so, the owners of the firms (for privately held firm) or equity holders (for publicly held firm) take on some of the labor productivity shock, so as to enable a better spreading of the labor income risk across the society[10]. This could have direct implication on the empirical observation that, in contrast to the real side of the economy, equity prices seem to be much more volatile[11]. We refer interested readers to the implicit contracting literature, and to Weitzman (1984) on the sharing economy for further discussion on profit sharing.

- Government is playing a role in the use of social security system, unemployment benefit,

[10]Furthermore, if we assume that equity holders and owners of private firm are less likely to be constrained on financial market than laborers themselves, then it makes sense to shift some of the labor income risk to owners and equity holders, so that the risk can be further traded on the financial market and born by the lowest cost bearer.

[11]Campbell (1997) documented that the stock market is much more volatile than the real dividend process or the real consumption process. Over the period 1947.2 to 1993.4, the annualized standard deviation of real stock return is 15.8%, much greater than the annual standard deviation of 7.3% in the real dividend process, which in turn is much greater than the standard deviation of the seasonally adjusted annual growth rate of real consumption at 1.1%. Of course, we should read these estimates with a grain of salt: first, these estimates are taken from data observed with different frequencies, and they might not be directly comparable; second, it could be that although annual data shows that stock price is more volatile than real variables, in a much longer time horizon that is commensurate with the life span of a representative agent of the economy, such as 20 years, this effect could be less strong or completely disappear.

health care and other benefits. Recent years have seen a trend of government shifting more of these functions to the more efficient private institutional forms.

- Labor union has been and still is an important form to share some of the labor risk. It sets minimum wage for its members, negotiates with employers on wage and benefits, intervenes in the hiring decisions, sometimes move member workers across firms according to seniority (need to get more evidence on these), and sometimes manage the pension for its members. There is a common wisdom that unionized workers have on average a higher level of job security[12] [13].

- Some firms, especially large conglomerates, might be able to internally shift laborers from one position to another, thus smoothing the demand shock to some extent[14].

- Under some circumstances, the cost of using spot market can be greatly reduced and thus enabling laborer to move across firms and better handle the risk arising from demand shock[15].

Labor intermediation is another form of labor income risk sharing, and it is becoming increasingly important in today's economy. We define labor intermediation as the practice

[12]From the later discussion of the labor intermediation firm, we think that union is performing functions that are very similar to a labor intermediation firm, therefore, in a broader sense, we could think of union as one form of labor intermediation.

[13]If we observe non-unionized workers being paid a lower wage than the unionized workers, does that mean that non-unionized workers are exploited by the employers? Not necessarily. First, employers might willingly pay the unionized workers a higher wage if that means the union is taking over part of the responsibilities of sharing the labor income risk also, so that firms will be relieved of some of that burden. Given that union often have the ability to move workers across firms, it could be a more efficient arrangement for union to take over the "labor risk management" task and get compensated for that; second, it could well be that people with higher human capital choose to be unionized, so their wage is not because of union but because of skill.

[14]Fay and Medoff (1985) reported that during off-peak season, manufacturing plants often assign cleaning and maintenance work to production workers who otherwise would have been less than fully occupied. Abraham and Taylor (1996) confirmed this in their study.

[15]For example, in Silicon Valley, programmers develop a good knowledge of the ability of programmers in other companies, which drastically reduce the asymmetric information. As a result, they can relatively easily move between firms to take advantage of changing demand, without incurring a high hiring and firing cost. Also, people in the trading departments of different investment banks often know each other very well, so that moving from one bank to another is not rare and the cost for doing so is not prohibitive. There are also evidence, for example, in Von Hippel (1986), that scientists and engineers in rival corporations often develop a surprising degree of information sharing. One result is that, when needed, a move between firms can be achieved with relatively low cost. All of these seem to suggest that, *under certain circumstances*, society has developed natural mechanisms to accommodate workers' move across firm, so as to ameliorate the labor productivity shock resulting from shocks to the demand of a particular firm.

where a firm pays its employees a fixed wage, and then assigns them to work for *other* firms[16]

Labor intermediation includes temporary staffing agencies, contracting job firms, management consulting firms, and to some extent other specialized service provision firms such as investment banks, legal consulting firms, accounting agencies, etc.[17]

labor intermediation firm and its role in sharing labor income risk

The recent years have witnessed rapid development in the labor intermediation business.

The temporary staffing industry in the United States has experienced rapid development in the 90s. According to the Staffing Services Annual Update published in May 1999 by the National Association of Temporary and Staffing Services (available online at: http://www.natss.org/staffingservicesupdate.html), from 1992 to 1995, temporary help employment grew at an annual rate of 17%. Since 1995, growth has slowed to a still healthy annual rate of 9%. In 1998, the growth rate is 9.8%, which is more than twice the growth in US GDP. All of these happens with an already tight labor market, where the unemployment rate is at its 28-year low of 4.3% in December 1998. Currently, help supply services employs about 2% of non-farm employment.

The management consulting business was at its infancy in the early 1980s, with about 18,000 practicing management consultants worldwide, and the industry as a whole had revenues of $1.2 billion in the United States, and worldwide perhaps $2 billion. In 1997, the industry has grown to around $35 billion globally, with an annual growth rate of more than 20%. Today, there are approximately 140,000 consultants worldwide. In another perspec-

[16] Here we different between a principle type of intermediation, whereas the intermediation takes in the labor, versus an agent type, whereas the intermediation just facilitates the transaction but does not have a position. From now on when we talk about labor intermediation firm we refer to only the principle type, which has the feature of risk sharing. Examples of the agent type include head hunters and employment agency.

[17] The difference between labor intermediation and other forms of labor risk sharing are:

1. while a firm intermediated labor or a laborer in an internal labor market within a conglomerate will most likely be limited by the scope of the firm they work with, laborers in the labor intermediation firm can be assigned to the position that best fits his specialized skills;
2. labor intermediation allows labor to move across firms without cost of hiring and firing, differ from spot market;
3. by issuing its employees a relatively fixed contract, labor intermediation firm is taking on some of the labor income risk from its laborers, this differs from the agency type.

tive, today there are approximately 70,000 management consultants in the United States, compared to around 150,000 executives in comparable positions. That is, for each executive there are 0.5 management consultants who advice full time. In 1980, this ratio was approximately 0.1. Furthermore, management consulting firms today employ about 25% of the graduates from leading business schools in the United States.[18][19][20].

In this paper, we want to focus on labor intermediation firms' function of sharing labor risk[21].

Technically this paper has the following characteristics:

1. This paper develops a general equilibrium model to characterize the endogenously developed labor intermediation. It also shows a "participation externality". Some interesting comparative static can be derived out of this model;

2. To demonstrate the role of "picking the best match" by labor intermediation firm, it

[18]Source: Staffan Canback, 1998: The logic of Management Consulting, Part One, in *Journal of Management Consulting*, 10, 3-11

[19]To the question: "Could it be that the increase in number of consultant are due to increase in independent consultant rather than growth in large consulting firms", *Management Consulting*, published by Harvard Business School, 1999, reports that three out of four of all consultants work for firms with over 100 professionals.

[20]Another source, cited from *Management Consulting: a complete guide to the industry*, by Saguta Biswas and Daryl Twitchel and published by John Wiley & Sons, Inc., NY 1999, has even bigger figures. I decide to take the more conservative estimates from the first source.

[21]There are several other functions performed by various form of labor intermediation firms, such as:

- providing on-the-job screening, so as to better solve the lemon problem on labor market.
- solving moral hazard problem by closely monitoring and developing expertise in handling labor related problems;
- by being the focal point, can have economies of scale and thus reduce cost of collecting information;
- reducing the fixed cost (of hiring and firing laborers) by specializing in labor market
- an intermediation over the institutional rigidity

  For example, avoidance of unnecessary health care and pension coverage. People in these labor intermediation firms often get paid higher wages but less benefits (like health insurance and pension). But maybe that is exactly because these people don't need that much of these benefits. Often they can get alternative coverage through a family member, or because they are relative young and healthy, they don't value these benefits that much.

  Another example being the concern about equity: some firms, due to union or "equity" considerations, might find it hard to set very different wages for its "regular" employees, and thus in order to pay the fair market price for some jobs it will have to rely on outsourcing. Such view is expressed by, among others, Abraham and Taylor (1996) uses the order statistics result, which I think is unique[22];

3. To show the effect of risk pooling and diversification, it uses the Liapounov Central Limit Theorem for independent but non-identically distributed random variables, which I think is unique in such context.

Literature Review

Some recent papers that are addressing related topics: recent literature in this area includes Acemoglu and Zilibotti (1997), Atje an Jovanovic (1993), Bencivenga and Smith (1991), Greenwood and Jovanovic (1990), Greenwood and Smith (1997), Levine (1997), Obstfeld (1994), Pagano (1993), Rajan and Zingalis (1998) and Saint-Paul (1992).

The organization of the paper

The rest of the paper is organized as followed: Section 2 develops a model of labor intermediation using a three period overlapping generations model. We differentiate between functional specificity and firm specificity, and show that the more functional specific (and the less firm specific) laborers tend to work in a labor intermediation firm. We then characterize an endogenously generated labor intermediation and derive the conditions for equilibrium. Section 3 analyze in a more general setting (by allowing the possibility of firm risk sharing) the effect of labor risk sharing on the ex ante decision of human capital specialization. There we derive an interesting result: that as the risk sharing improves, the level of specialization will increase very quickly, such that the ex post level of labor income risk, *after taking account of the risk sharing*, is even higher than when there is no risk sharing. Section 4 concludes and poses some questions and empirical work to be done in future research. All proofs are in appendix.

[22]The result I use here is the simplest case: when firm's realized productivity shock is independent. We have also worked out some more complicated results for the case when the realized productivity shock is correlated across firms. For brevity the result is not included.

2.2 The Model of Labor Intermediation

In this section we develop a model of labor intermediation. In doing so, we try to accommodate the following real world observation:

- We observe that people have different level of (functional) specialization, and people's decision as to whether to join the labor intermediation firm or not depends on their level of functional specificity. Some people are highly specialized, for example an expert of a particular type of corporate litigation, and these people tend *not* to work in any particular conventional firm, but rather in a firm that provide specialized services to those conventional firms; compared to them, there are other, more general legal service providers who work as an attorney of a particular conventional firm[23];
- there is also heterogeneity across people regarding the level of *firm specific* human capital. Some laborers are only productive within a certain firm (for example people familiar with a software used only by a particular firm), and these people tend to work inside a conventional firm rather than be outsourced into a labor intermediation firm.
- the level of human capital specialization is determined both by a discreet choice of worker and by some factors that people don't control (financial constraints, luck, etc). Therefore in the model our agent will have his human capital determined by his discreet choice as well as luck.
- doing labor intermediation has a fixed cost. The fixed cost makes it only efficient if there are sufficiently large number of people using labor intermediation[24].

2.2.1 The model setup

This is a three period overlapping generations model. During the first period an agent will acquire human capital through "schooling", and no consumption is made[25]; during the

[23]One of the corporate attorney's jobs is to refer the company to expert in certain specialized field when necessary.

[24]Also it means in our simple model that if at all there are labor intermediation firms, there will be only *one*. This is certainly an abstraction from the real world where many labor intermediation firms co-exist, and we think that is explained by some other factors, for example the informational efficiency of a labor intermediation firm might deteriorate with the scale of the firm

[25]Alternatively, we could assume that each "student" receives a fixed amount of consumption good from the government, which they cannot trade or transfer to later periods. Thus they could receive a fixed amount

second period he works and gets paid a wage[26], and make a decision to consume and to save; in the third period he invest his savings in the financial market, which consist of a risky asset and a riskless asset, and at the end of that period he consumes the proceeds from investment. There is no bequest.

Following the paper by Constantinides et al (1998), this model has the feature that young and middle-aged people cannot borrow against their future labor income to invest in risky asset.[27].

preference

The agents all have the same additively separable utility function with discount rate β, and in each period he has a CARA preference with risk aversion parameter α, therefore if consumption is c_t and c_{t+1}, respectively for middle and old age, his total utility is given by:

$$U(c_t, c_{t+1}) = -e^{-\alpha c_t} - \beta e^{-\alpha c_{t+1}} \tag{2.2.1}$$

three components of human capital

We model a laborer's human capital as composed of three components:

1. a common component. This is some general skill that the laborer acquired. It will generate the same amount of marginal revenue product in *any* production firm;
2. a functional specific component. In reality, this corresponds to a certain skill that *may* be useful in the firm where the laborer works, for example, the expertise in corporate litigation cases. At any firm, the marginal revenue product from this functional specific component is random: sometimes the firm will value the special skill very highly and thus the marginal revenue product is very high, at other times the special skill could

of utility from this consumption.

[26]depending on where he works he will get paid differently: in a conventional firm he will be paid the (random) marginal revenue product of his labor, while in a labor intermediation firm he will be paid a constant wage that depends on the observable characteristics of his human capital

[27]The reason is primarily what Diamond and Mireless explained in their paper, "Insurance Aspects of Pensions", that individuals don't have as good a credit as firms, therefore they can not borrow in the financial market for the fear that they might default.

be nearly useless to a particular firm. To simplify our analysis, we assume that people are choosing different functions to be specialized in, so there is no substitutions across workers. If a firm happens to need the skill of a particular worker and that worker is not employed by the firm, then the opportunity to generate income through doing that particular job is lost.[28]

3. a firm specific component. Some people have special skill that are only valuable at certain firms. For example, people familiar with certain types of software are more fit with firms that use these software. In reality we could observe that over time people could develop significant amount of firm specific human capital, and if they no longer work in the particular firm (due to termination of employment relationship, or due to merger/acquisition), this human capital will transpire.

In reality, people's human capital differs in the level of all three components. The variability in the proportion of these three components of human capital across people is the result of both the deliberate choice in acquiring human capital and different constraints and luck that people face. In our model, we accommodate both the discrete choice and luck in the accumulation of human capital, and assume that agents make deliberate choice, and then luck also affect the final composition of human capital.

production firm[29]

We assume that there are N production firms. These firms are identical and independent in their production technologies, which takes as input capital and labor, and produces a single consumption good. We assume that N is large, so that these firms are perfectly competitive.

Capital has a constant gross return of A, namely, for a capital investment of K, the next

[28]That means, firms are generating more jobs than its employees are actually taking. The reason that these firms don't merge into a big conglomerate so as to "internalize" the job vacancy utilization is simply because there might be other constraints that limits the size of firm. In the real world, while we do observe large firms internally transferring their laborers, we also observe many smaller firms, which don't just merge to better solve the labor matching problem internally: there are presumably other functions to be served by these production firms, and those are of higher priority. Therefore, in this model, we take the size of the firm and the number of firms in the economy as given.

[29]Although we call it a production "firm", it might also be appropriate to think of these "firms" as "industries". If this is the case, then in later sections when we talk about firm-specific human capital, it could be understood as industry-specific human capital.

period product of capital is AK, and the capital is completely perishable after one period[30].

The total product of firm n, with K_n units of capital input and laborers represented by Ω_n, is:

$$P_n = AK_n + \sum_{i \in \Omega_n} \tilde{B}_i \tag{2.2.2}$$

Production firm just acts as a venue where the factors of production can be applied to generate output. The productivity of capital is constant, while the productivity of labor is random[31].

Under these conditions, for any agent working in a production firm n, the wage he gets is equal to the marginal revenue product, which is:

$$MPL_{pf,n} = (1 - \phi_{11}\phi_{12})B_0 + \phi_{11}\phi_{12}\tilde{B}_1 + \phi_2 I_k(n) \tag{2.2.3}$$

where,

B_0 is the marginal revenue product of one unit of common skill, and $\tilde{B}_1$ is the (random) marginal revenue product of one unit of specialized skill. We assume that $\tilde{B}_1 \sim N(\bar{B}_1, \sigma_{B1}^2)$. Furthermore, we assume that specialization is on average value-enhancing:

$\bar{B}_1 > B_0$.

Agents choose the parameter ϕ_{11} to decide on their level of functional specialization. The choice, however, is amplified by the result of a natural game, ϕ_{12}, to determine the true

[30]Here we are assuming a riskless production technology for capital. This might be viewed as a reduced form model, where all capital market risk are diversifiable and thus have already been taken care of, thus we can focus on the sharing of labor income risk without worrying about the interaction between labor productivity risk and capital productivity risk.

[31]Other than tractability consideration, we assume this simple linear form of production function because we want to avoid the *externality* between labor and capital, for example, more capital increases or decreases the marginal productivity of labor, or vice versa. Such an externality might suggest that it is optimal to have one firm pooling capital and labor thus increasing economy of scale. We believe that to amply demonstrate firm's function of polling production factors together is going to be an interesting paper by itself. One will have to explain why such polling doesn't result in a *single* firm in the economy by explicitly modeling all the costs associated with making the firm large, and then obtain some proper tradeoff as the firm size increases so as to determine endogenously an optimal size of the firm and thus an optimal number of firms in an economy. Such a task, while interesting, could easily dilute the point we are trying to make here, thus we think it is wise to leave it for future research.

level of functional specialization, $\phi_{11}\phi_{12}$. The random variable ϕ_{12} is distributed uniformly over $(0,K_1)$.[32] Given the limited total resource an agent has, devoting more resource into development of specialized skills will necessarily reduce the resource to be used in developing common skills, thus the relationship between the first term and second term on the right hand side of the equation.

The term ϕ_2 is a random parameter chosen by nature, which determines the level of firm specificity of the human capital. We assume that ϕ_2 is uniformly distributed over $(0,K_2)$.

$I_k(.)$ is an indication function of a "match" between skill and firm: $I_k(n) = 1$ iff n=k, otherwise $I_k(n) = 0$, so that for workers of type k, his firm specific human capital will pay off if and only if he is employed by firm k.[33]

To summarize, in this model, a production firm transforms capital from today to tomorrow at a fixed rate of transformation A. It also generates job opportunities. For a laborer with parameters $\{\phi_{11}, \phi_{12}, \phi_2$ and $I_k\}$, the firm generates job that if employed will pay off $W_{pf,n}$, which is equal to $MPL_{pf,n}$ as defined before. If not exploited, these job opportunities will simply be lost.

labor intermediation firm

There is a maximum of one[34] labor intermediation firm, which is set up to intermediate some of the labor (in particular, those labor with high level of the functional specific human capital and low level of firm specific human capital).

labor intermediation firm serves two functions that are welfare enhancing:

1. Labor intermediation firm is good at matching skills with demand, and we assume for simplicity that it can match the skill of its member to the best use among all the production firms, thus increasing the expected MRP_L;

[32] We can choose different re-normalizations of K_1 without loss of generality. We have in our numerical solution part chosen K_1 to be 2, so as to give the interpretation that agents are choosing the *mean* of their levels of human capital specialization (because $E(\tilde{\phi}_{12}) = 1$), and a lottery determines whether their realized level of human capital specialization is more or less than the mean amount

[33] In reality, things could be more complicated: even if a person does not get to work with the firm that best matches his firm-specific skill, he might be able to find another firm, that allows him to use *part* of his firm-specific human capital, rather than have to totally waste it. For simplicity, we assumed away such complication. This simplification is inconsequential to our model since adding the complication only means we have to re-scale the effect of ϕ_2.

[34] There could be no labor intermediation firm, if the cost of doing labor intermediation is too high.

2. In addition, labor intermediation firm can poll the labor risks of its members and thus achieve diversification (cross-subsidization). Furthermore, by issuing a fixed wage contract with its members, labor intermediation firm (in fact, the owner or equity holder of the firm) will take on any residual risk in labor productivity, thus better spreading risk in society.

There are also some drawbacks of using labor intermediation firm:

- First, setting up such a firm is costly, and we explicitly model the cost as a fixed cost, C, for the operation of the labor intermediation firm every period[35]. This cost is covered by charging members of labor intermediation firm a fixed membership fee of c[36].

- Second, when laborer join the labor intermediation firm and work at different firms all the time, we assume for simplicity that labor intermediation firm does not get to know the function $I_k(.)$, so it does not assign laborer according to their firm specific human capital Therefore, for a worker in the labor intermediation firm, she only has $1/N$ chance to work in the "right" firm and realize the gain in ϕ_2. When N is large, the effect of ϕ_2 is negligible.[37]. Thus the firm specific component of human capital is given up.

Formally, we assume the following:

- the labor intermediation firm writes a contract, that for a fixed fee c (to be paid after wage is received, and subtracted from wage), people can become members of labor intermediation firm. The fee serves two purposes: one is to pay off the fixed cost C, the other is to be used as a risk premium to compensate equity holders of the labor intermediation firm, for holding the risky labor intermediation firm equity.

[35]There is potentially another cost: if a labor intermediation firm employee and a conventional firm employee have same level of caliber, the former might have a lower productivity, because of unfamiliarity with the host firm, and maybe a lower morale. This is also included in C.

[36]In reality c does not take the form of an explicit membership fee. It is rather a reduction of wage

[37]we think this is roughly correct in real world, although it might be possible that labor intermediation firm could exert some effort to match the laborer with firms that they have specific skill in, and some labor intermediation firm might even ask a particular employee to work for a client firm for a long time (presumably due to the firm specific skill the employee developed), it is not an important concern of the labor intermediation firm. Adding complications like this will merely re-scale the effects of ϕ_2, without helping to better understand the intuition.

- labor intermediation firm has equity holders, who are the people at the beginning of their old age in the model. These equity holders have two alternatives to invest their money:

  1. a riskless investment in the equity of the production firm, with constant rate of return A;
  2. invest in the capital in labor intermediation firm. The equilibrium equity price (and return) is determined by a competitive equity market.

- labor intermediation firm has rational expectation about the fraction of the population to be in the intermediated labor sector, and calculate the mean and variance of its cash flow, thus can offer a competitive term on c, so that when the market clears, its equity holders are getting a fair return.

- For a member of the labor intermediation firm with functional specific human capital ϕ_{12} and firm specific human capital ϕ_2,

$$MPL_{lif}(\phi_{11}, \phi_{12}) = (1 - \phi_{11}\phi_{12})B_0 + \phi_{11}\phi_{12} \max_n(\tilde{B}_{1,n}) \tag{2.2.4}$$

  where $\max_n(\tilde{B}_{1,n})$ is the maximum of the realized values of $\tilde{B}_{1,n}$ among all n.

  The labor intermediation firm pays the worker a fixed wage of $W_{lif}(\phi_{11}, \phi_{12})$ that depends on ϕ_{11} and ϕ_{12} (after subtracting the fixed fee c, the member actually takes home $W_{lif}(\phi_{11}, \phi_{12}) - c$).[38]

The time line of the model

1. At the beginning of the first period, young laborers choose parameter ϕ_{11}, to control the level of functional specialization of their human capital

[38]Here I assume that labor intermediation firm cannot observe ϕ_2, otherwise it will want to set the cost c dependent on ϕ_2. The intuition is that, labor intermediation firm wants to have the most people joining to better share the fixed cost C. For lower realization of ϕ_2, a worker is in labor intermediation firm anyway, so can charge him a higher c; for higher ϕ_2, lowering the c might induce more people to join, thus lowering the PER CAPITA fixed cost c for each people.

2. Random shocks ϕ_{12} and ϕ_2 are realized at the end of the first period, whereas ϕ_{12} affect level of functional specialization, and ϕ_2 affects level of firm specific human capital

3. at the beginning of the second period, (up to one) labor intermediation firm is set up, at a fixed cost of C.

4. also at the beginning of the second period, observing their own parameters of ϕ_{12} and ϕ_2, middle aged people choose whether to join the labor intermediation firm or conventional production firm.

5. If a worker work for a production firm, then the firm and the worker sign a contract that gives the worker a wage equal to his realized MRP_L.

6. If the worker work for the labor intermediation firm, he will be charged a fixed membership fee, c, which is to be paid at the end of the production period to cover the fixed cost C, and then he will get a wage equal to the expected MRP_L given ϕ_{12} and ϕ_2.

7. after the contract is signed, the demand of firm, $\tilde{B}_1$, is realized for each production firm and each worker. Observing these realizations, labor intermediation firm can assign its workers to firms that has the best match, i.e., the highest realization of $\tilde{B}_1$ for a given worker.

8. At the end of the second period, production is carried out, and marginal revenue product of labor accrue to the labor (in the case of a production firm) or the labor intermediation firm (in the case of a labor intermediation firm). In the case of labor intermediation firm, wage is paid out as contracted.

9. Also at the end of the second period, middle aged carry out consumption and save the rest for the next period.

10. At the beginning of the third period, old people make investment decision (risky versus riskless investment). At the end of the third period, he consumes the proceeds from his investment. No bequest is left.

2.2.2 Solution to the model

Decision to join labor intermediation firm

We first derive conditions for people to be in the labor intermediation firm. To do that, for a given c, we compare utility within the labor intermediation firm and outside of labor intermediation firm, and derive people's decision to join the labor intermediation firm as a function of ϕ_2 and ϕ_{12}.

We first write out the wage that a laborer gets from production firm and labor intermediation firm, respectively. If a person chooses to work in a production firm, she will choose to work in the one that matches her type of the firm specific human capital. Thus, the wage from working in production firm becomes:

$$W_1 = MRP_L = W_{pf,n} = (1 - \phi_{11}\phi_{12})B_0 + \phi_{11}\phi_{12}\tilde{B}_1 + \phi_2 \tag{2.2.5}$$

Wage in labor intermediation firm is a little bit more tricky. The laborer in labor intermediation firm is paid the expected MRP of labor, *minus* the fixed membership fee, c. However, the MRP differs from the case of production firm by two considerations:

1. The term of ϕ_2. As discussed before, this part is assumed lost.
2. The fact that labor intermediation firm assigns the worker to their most productive use. That is, of N independent realization of $\tilde{B}_{1,n}$'s, labor intermediation firm assigns the worker to the one with $\max_n \tilde{B}_{1,n}$.

Lemma 1 (order statistics) *The maximum of N i.i.d normal random variables, X_i, i=1,...,N, where $X_i \sim N(\mu, \sigma^2)$, has mean*

$$E(X_{N:N}) = \sigma f_1(N) + \mu \tag{2.2.6}$$

where $f_1(N)$ is increasing in N,

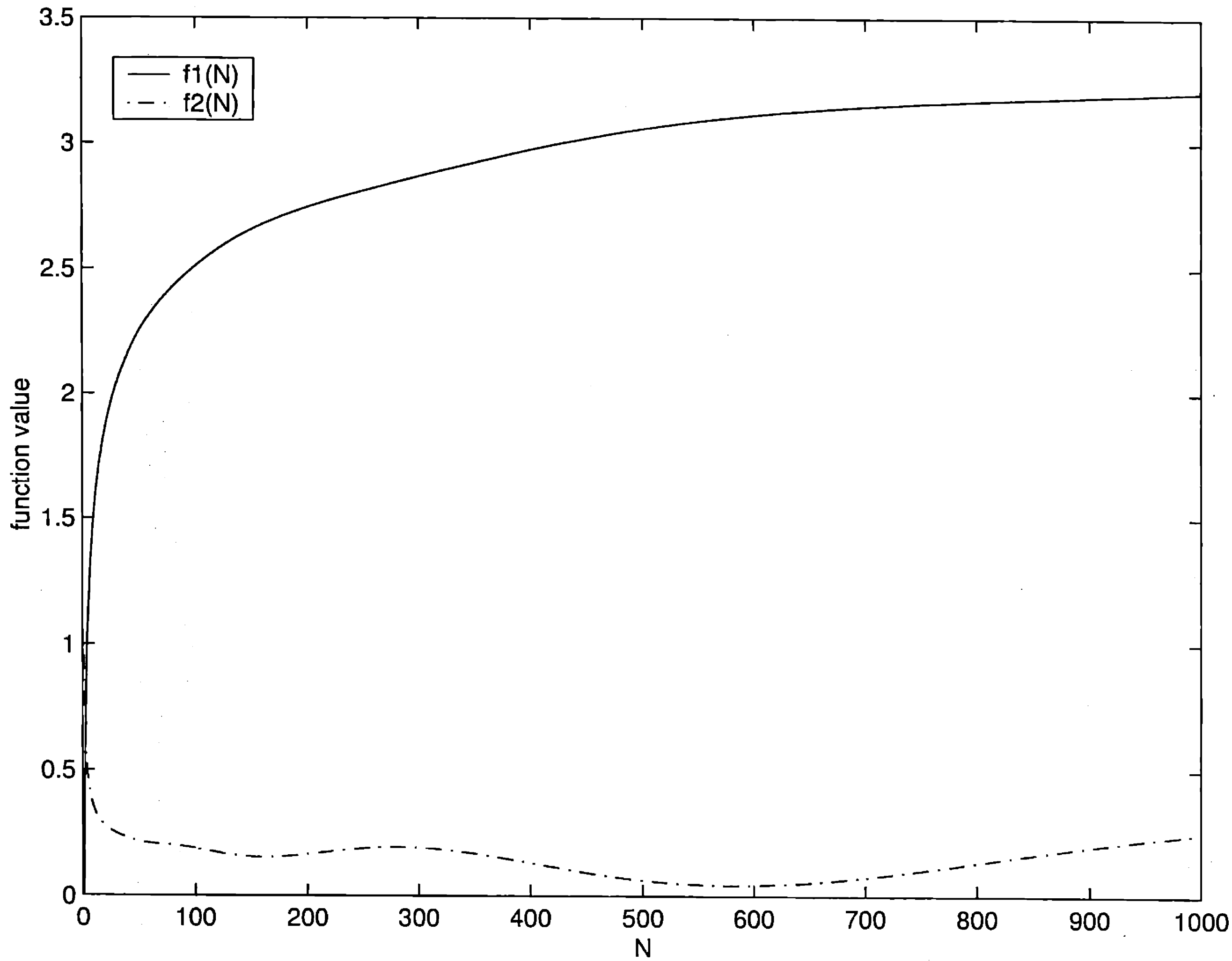
and variance

$$Var(x_{N:N}) = \sigma^2 f_2(N) \tag{2.2.7}$$

Proof: see appendix.

In Figure 2.1, we plot the function $f_1(N)$ and $f_2(N)$.

Figure 2.1: $f_1(N)$ and $f_2(N)$ as functions of N (N from 1 to 1000)



It is important to note that $f_1(N)$ and $f_2(N)$ only depend on N. Therefore, once N is fixed, $f_1(N)$ and $f_2(N)$ are constants.

The MRP of laborers in labor intermediation firm, with $\{\phi_{11}, \phi_{12}, \phi_2\}$, is

$$MPL_{lif} = (1 - \phi_{11}\phi_{12})B_0 + \phi_{11}\phi_{12} \max_n \tilde{B}_{1,n} \tag{2.2.8}$$

Therefore, according to Lemma 1, we have the wage from working in the labor intermediation firm, after subtracting the membership fee c, is:

$$W_2(\phi_{11}, \phi_{12}, \phi_2) = E(\tilde{W}_{lif}) - c = (1 - \phi_{11}\phi_{12})B_0 + \phi_{11}\phi_{12}(\bar{B}_1 + \sigma_{B1} f_1(N)) - c \tag{2.2.9}$$

and,

$$Var(MPL_{lif}) = \phi_{11}^2 \phi_{12}^2 \sigma_{B1}^2 f_2(N) \tag{2.2.10}$$

Next, we determine the utility from getting the wages thus outlined. Note that after getting wage (W_1 or W_2), people have two periods to consume it: middle age and old. Therefore we first need to recursively calculate the derived implicit utility, V(w), which is the result of optimization over the next two period consumption, given wage w.

Lemma 2 *Suppose that wage is W, and people have two investment opportunities: one is a riskless asset with a return of A units one period later for each unit invested now, and the other is a risky asset which has a price of C units of consumption goods now, and which will return $\tilde{X}$ units of consumption goods one period later, where $\tilde{X} \sim N(\bar{X}, \sigma_x^2)$. The agents have additively separable preference over time, with a discount rate of β, and in each period they have a CARA utility function with parameter α.*

Then:

1. *The choice of investment on risky asset, ξ, is independent of wealth level W.*

2. *The indirect utility from wealth of W is of the form:*

$$V(W) = -e^{-\frac{\alpha AW - lnAK}{1+A}} - Ke^{-\frac{\alpha AW + AlnAK}{1+A}} \quad (2.2.11)$$

where,

$K = \beta e^{-\frac{1}{2}\frac{(\bar{X}-AC)^2}{\sigma_x^2}}$

Proof: see appendix.

For the ease of analysis and tractability of the model, from now on, we make the simplifying assumption that $A = K = 1$, This assumption is without much loss of generality given the main point we want to show.

Under these simplification, we have,

$$V_t(W) = -2e^{-\frac{\alpha W}{2}} \quad (2.2.12)$$

We now calculate the utility from staying in conventional production firm versus utility from staying in labor intermediation firm, and determine the condition for people to decide to stay in labor intermediation firm.

The condition under which people are *indifferent* between joining labor intermediation firm and production firm is:

$V_{PF}(\phi_{12}, \phi_2) = V_{LIF}(\phi_{12}, \phi_2)$

$\Leftrightarrow$

$E(-2e^{-\frac{\alpha W_1}{2}}) = E(-2e^{-\frac{\alpha W_2}{2}})$

$\Leftrightarrow$

$-2exp(-\frac{\alpha}{2}E(W_1) + \frac{1}{8}Var(W_1)\alpha^2) = -2exp(-\frac{\alpha}{2}W_2)$

$\Leftrightarrow$

$\frac{\alpha}{2}E(W_1) - \frac{1}{8}\alpha^2 Var(W_1) = \frac{\alpha}{2}W_2$

$\Leftrightarrow$

$E(W_1) - \frac{1}{4}\alpha Var(W_1) - W_2 = 0$

$\Leftrightarrow$

plug in numbers,

$(1-\phi_{11}\phi_{12})B_0+\phi_{11}\phi_{12}\bar{B}_1+\phi_2-\frac{1}{4}\alpha\phi_{11}^2\phi_{12}^2\sigma_{B1}^2-(1-\phi_{11}\phi_{12})B_0-\phi_{11}\phi_{12}(\bar{B}_1+\sigma_{B1}f_1(N))+c=$ 0

$\Leftrightarrow$

$$\phi_2=\frac{1}{4}\alpha\sigma_{B1}^2\phi_{11}^2\phi_{12}^2+\sigma_{B1}f_1(N)\phi_{11}\phi_{12}-c \tag{2.2.13}$$

This gives the boundary condition of joining labor intermediation firm, which has ϕ_2 as a quadratic function of ϕ_{12}. From now on we call this function $\phi_2^*(\phi_{12})$.

(A1) *We assume that K_2 is pretty large, so that the value of $\phi_2(K_1)$ is less than K_2.*

This is a simplifying assumption that guarantees that, even for the highest level of functional specificity, it is still possible to stay in the convention production firm, because firm specific human capital is very high[39] [40].

Under Assumption (A1), a plot can be drawn showing the effect of distribution of ϕ_{12} and ϕ_2. We plot the function of ϕ_2 in ϕ_{12} in Figure 2.2.

The shaded area represents laborer that joins labor intermediation firm. Larger value of c will shift the boundary curve to the lower-right, while smaller values of c will shift the boundary curve to the upper-left. From the graph we can see, people with higher level of functional specific human capital tend to stay in a labor intermediation firm. In addition, for people with the same level of functional specific human capital, those with a lower level of firm specific human capital will more likely join the labor intermediation firm.

The mean and variance of LIF equity, as functions of c

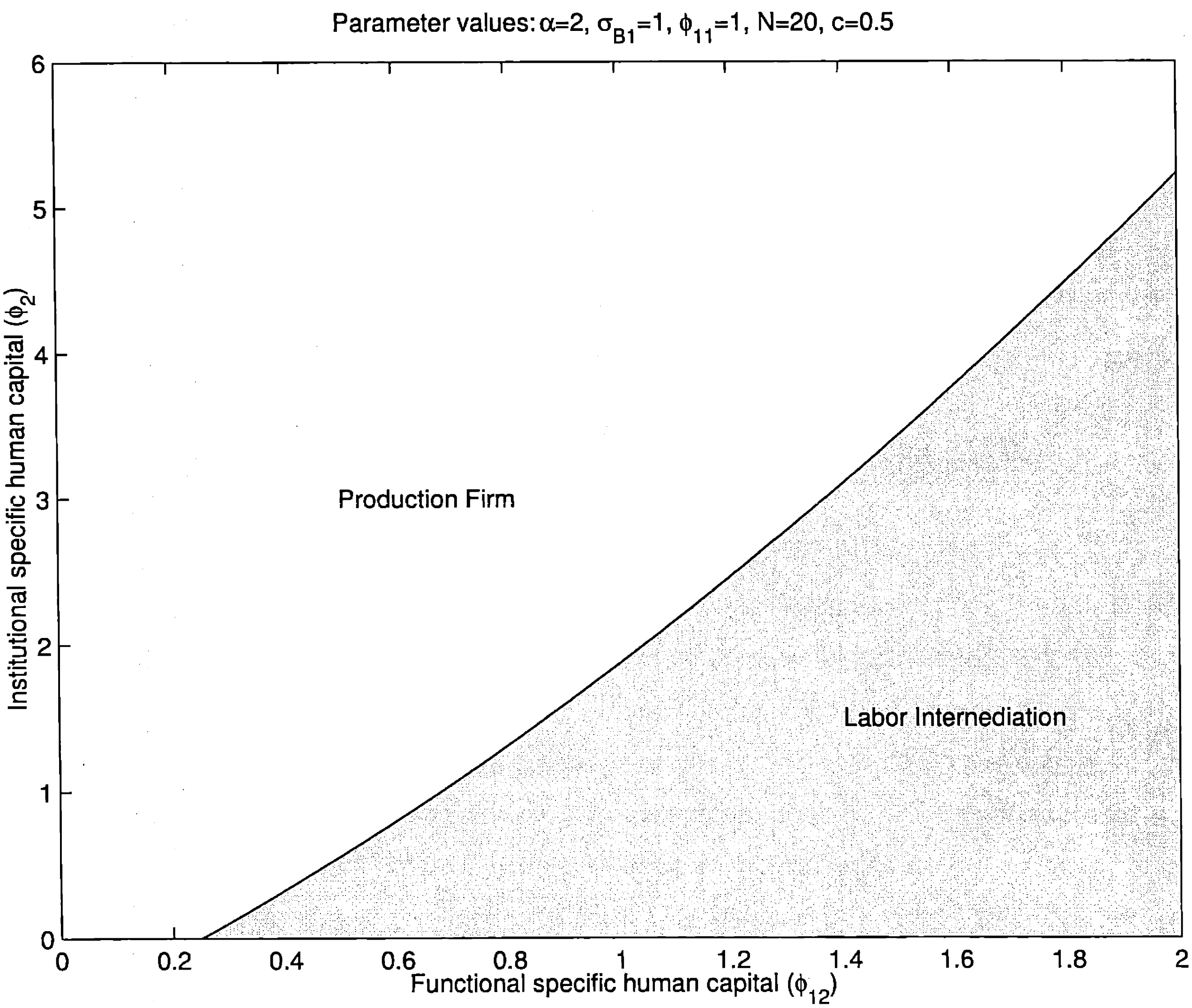
The value of c affects the boundary condition for people's choice to stay in the labor intermediation firm. Large value of c will discourage people from joining labor intermediation

[39]We have in mind the high level corporate executives that are on strategic positions in a firm, such as CEO, CFO, Chief Technology Officer, etc

[40]This assumption implies that there are more people in the conventional firm than there are in the labor intermediation firms, which accords well with the statistics we quoted earlier on the status of temporary staffing industry and management consulting industry.

The decision to join labor intermediation firm is a function of functional specific human capital ϕ_{12} and firm specific human capital ϕ_2. In the figure, for the $\{\phi_{12}, \phi_2\}$ combination that is in the shaded area, a worker will join the labor intermediation firm; otherwise a worker will join the conventional production firm. The boundary has ϕ_2 as a quadratic function in ϕ_{12}.

Figure 2.2: DECISION TO JOIN LABOR INTERMEDIATION



firm, while smaller value of c will induce more people to join, even those with a relatively high level of firm specific human capital ϕ_2. As a result, as c changes, the characteristics of the cash flow (both mean and variance) will change. In Figure 4.2, higher c reduces the area of the shaded area, thus reducing the total "risk" associated with labor intermediation firm equity (because with higher c, less people will join the labor intermediation). The following theorem gives the mean and variance of the labor intermediation firm next period residual cash flow, after labor wages are paid out, as a function of c and other parameters.

Theorem 1 *The cash flow of the labor intermediation firm next one period from now is a normal distribution, whose mean and variance are given, respectively, by:*

$$E(\tilde{X}) = c[\frac{N}{K_1K_2}\frac{1}{12}\alpha\sigma_{B1}^2\phi_{11}^2\phi_{12}^3|_{\phi_{12}^0}^{K_1} + \frac{1}{2}\sigma_{B1}f_1(N)\phi_{11}\phi_{12}^2|_{\phi_{12}^0}^{K_1} - c(K_1 - \phi_{12}^0)] \tag{2.2.14}$$

and,

$$Var(\tilde{X}) = \frac{N}{K_1K_2}\phi_{11}^2\sigma_{B1}^2 f_2(N)[\frac{1}{20}\alpha\sigma_{B1}^2\phi_{11}^2\phi_{12}^5|_{\phi_{12}^0}^{K_1} + \frac{1}{4}\sigma_{B1}f_1(N)\phi_{11}\phi_{12}^4|_{\phi_{12}^0}^{K_1} - \frac{1}{3}c\phi_{12}^3|_{\phi_{12}^0}^{K_1}] \tag{2.2.15}$$

Proof: see appendix.

Some interpretation of the result:

$Var(\tilde{X})$ decreases in c. When c increases, for each ϕ_{12}, ϕ_2^* will be lower (i.e. less likely for people to join the labor intermediation firm), thus there are less terms in the integral to sum up, and less "systematic" risk out of it, after risk sharing has been carried out.

The effect of c on $E(\tilde{X})$ is ambiguous. On the one hand, a higher c will decrease the measure of people in the labor intermediation firm, thus lowering the number of people paying the fixed cost (which is the sole source of the $E(\tilde{X})$). However, a higher c increases the *per capita* payment. The total effect can be ambiguous. Numerical results shows that when c is zero, $E(\tilde{X})$ is zero (since nobody is paying any fees for their membership in labor intermediation firm); as c increases, $E(\tilde{X})$ first goes up, and then eventually goes down. When c is sufficiently large, $E(\tilde{X})$ goes to zero, since nobody is in the labor intermediation firm.

The relationship between $E(\tilde{X})$ and $Var(\tilde{X})$, as linked by c, provides one implicit function to be used to solve for the market equilibrium level of $E(\tilde{X})$ and $Var(\tilde{X})$. The function is highly non-linear, and numerical methods or approximation methods will be necessary to gain further insights into these expressions.

the mean and variance of LIF equity, from market clearing condition

We now turn to another equation linking $E(\tilde{X})$ and $Var(\tilde{X})$, as determined by the market clearing condition of the labor intermediation firm equity.

The holder of ξ units of labor intermediation firm equity (total 1 unit outstanding) will have the following cash flow:

1. at time t, invest ξC
2. at time t+1, gets $\xi\tilde{X}$ in return, where

   $\tilde{X} \sim N(\mu_x, \sigma_x^2)$

Assume agent has initial wealth of W_0 to be invested for next period. He has two choices of investment: a riskless equity in the production firm, or a risky equity in the labor intermediation firm. If he buys ξ share of labor intermediation firm equity, then, the next period, he will have the following cash flow:

1. gets $A(W_0 - \xi C)$ from production firm equity, from initial investment of $W_0 - \xi C$.
2. gets $\xi\tilde{X}$ from labor intermediation firm equity, where

   $\tilde{X} \sim N(\mu_x, \sigma_x^2)$

The optimization problem of investor is:

$$\max_{\xi} E - e^{-a\{AW_0+\xi[\tilde{X}-AC]\}}$$

$\Leftrightarrow$

$$\max_{\xi} E - e^{-a\xi[\tilde{X}-AC]}$$

$\Leftrightarrow$

$\max\limits_{\xi} a(\mu_x - AC)\xi - \frac{1}{2}a^2\xi^2\sigma_x^2$

The first order condition is:

$a(\mu_x - AC) = a^2\xi\sigma_x^2,$

so,

$\xi = \frac{\mu_x - AC}{a\sigma_x^2}$

Note that the demand for risky asset does not depend on initial wealth[41].

The market clearing condition

The total demand of risky LIE equity turns out to be *independent* of initial wealth, and there are a total of measure N investors, so the market demand can be written as:

$1 = N\xi,$

solving this gives:

$$\sigma_x^2 = \frac{N}{a}(\mu_x - AC) \tag{2.2.16}$$

This gives another relationship linking $E(\tilde{X})$ and $Var(\tilde{X})$, through the market clearing condition.

the endogenous equilibrium with labor intermediation

The equilibrium of this economy can be found from the intersection of the loci of the two relationship between $E(\tilde{X})$ and $Var(\tilde{X})$ just identified. We will be able to solve it numerically, to get the equilibrium level of $E(\tilde{X})$ and $Var(\tilde{X})$ and c^*, as shown in Figure 2.3.

What we see is, there are generally three possibilities:

1. no equilibrium involving labor intermediation firm. This could be because the fixed cost C, is too high

[41] We assume that the demand for risky asset turns out to be very small, so that agents will *almost* for sure have enough initial wealth to invest in risky asset, without violating the assumption that agents cannot borrow.

2. one equilibrium involving labor intermediation firm. This is the case if the two loci are tangent to each other.

3. two equilibria involving labor intermediation firm. This will be the case for low enough value of C. In this case, the two equilibria can be Pareto ranked. If a sufficiently low per capita fixed membership fee c is charged, then a large number of people will be in the labor intermediation firm, thus resulting in a higher $E(\tilde{X})$ and $Var(\tilde{X})$, and utility is high; if, on the other hand, the labor intermediation firm expect less people to join it and thus set a high fixed membership fee c to cover the fixed cost C, then it could be that this expectation is self-fulfilling: with less people in the labor intermediation firm, the $E(\tilde{X})$ and $Var(\tilde{X})$ are indeed both lower. This equilibrium has lower utility.[42]

2.2.3 Effect of labor intermediation firm on labor income risk

The effect of labor intermediation firm on labor income risk is obvious. For those laborers with a ϕ_{12}, ϕ_2 combination in the shaded area of Figure 2, the labor productivity risk is totally taken by the equity holders of the labor intermediation firm, or the old generation. The expected labor income risk will depend on the area of the shaded region.

2.3 Risk sharing and human capital specialization

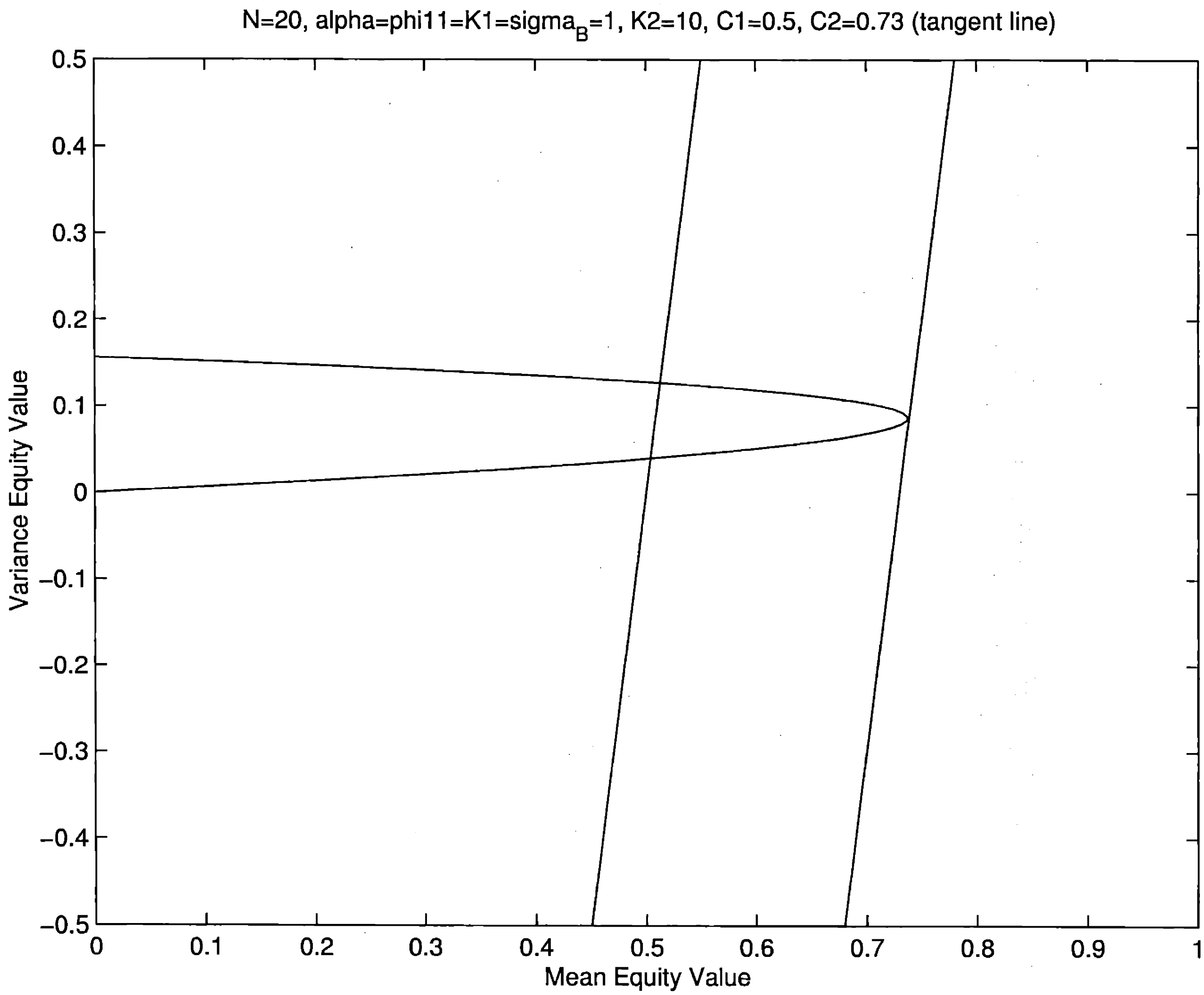
The theme in this section is a trade-off between specialization and diversification, and how that trade-off is being altered with increased risk sharing on the labor market.

The development of human capital has some economies of scale: higher level of specialization can on average have a higher labor product than the jack-of-all-trade. So with risk neutrality people would want to fully specialize. However, the *risk* of labor product also increases with level of specialization. And since labor market cannot perfectly efficiently diversify these risks, the workers will take some of the risks. If workers are risk-averse, they will have a trade-off between the higher average labor product and the higher risks. And often we see the optimal choice is an interior solution: laborers are not jack-of-all-trade, but neither are they fully specialized.

[42]it is pointed out by Professor Peter Diamond that there is a specific literature on this kind of "participation externality".

In Figure 2.3, we plot two possible scenarios: i) there are two equilibria. This is the case when the locus of the market clearing condition cuts through the locus of the mean-variance relationship through c; ii) there are just one equilibrium. This is the case when the two loci are tangent to each other.

Figure 2.3: EQUILIBRIUM WITH LABOR INTERMEDIATION



Labor intermediation as discussed in the last section, together with other forms of sharing the labor productivity risk, are aimed at reducing the risk inherent in the labor income. This will change the risk-return tradeoff for acquiring specialized huamn capital. In this section, we intend to show the effect of improved risk sharing on the decision to accumulate specialized human capital ex ante, and, in particular, we want to demonstrate an interesting phenomenon: that as risk management on labor market increases, the actual level of risk, *after risk management is taken*, might *increase.*

2.3.1 The model

We study a three period overlapping generations model, where we endogenously determine the level of specialization of human capital, and relate it to level of risk sharing of the economy.

We build a model based on the model in the last section. To better make our point, we restrict the previous model in some way, but relax it in some other way.

Since our primary interest is on the effect of improved risk sharing opportunity on the level of human capital specialization rather than determining the criterion to do labor intermediation, we want to isolate the specialization decision made by laborer from the random effects present in the last section. We do so by freezing the random variation across laborers over the amount of firm specific human capital (previously called ϕ_2) and functional specific human capital (previously called ϕ_{12}), so as to focus on the degree of functional specialization chosen by laborers, ϕ (previously called ϕ_{11})

We relax the previous model to accommodate the possibility of the existence of risk sharing by firm. Risk sharing by production firm is an important institutional arrangement to solve the risk sharing problem on the labor market, and we assume that it is *possible* that firm is issuing some contract that partially shares the risk in labor productivity. Then, the wage can be written as:

$$w_t = b\tilde{B}_t + c \tag{2.3.17}$$

where B_t is the labor productivity, $0 \leq b \leq 1$ is the share of the labor productivity risk taken by the laborer himself, and (1-b) is the share of labor productivity risk taken by the firm (or the equity holder of the firm). c is a fixed component which makes the risk sharing contract between laborer and production firm actuarially fair[43].

Consequently, the equity holder of the production firm no longer receives a riskless return of A, but instead, for an investment of K_t, he will receive:

$$q_t = AK_t + (1-b)\tilde{B}_t - c \tag{2.3.18}$$

Note here a simplifying assumption is made that there is no "leakage" of welfare through C, the cost of doing labor intermediation. The reason is two-fold: first, here we want to show that improved risk sharing in the labor market leads to more specialization, while labor intermediation is only one way to improve labor market risk sharing; second, even within labor intermediation, this is a harmless assumption to our goal of showing the effect on specialization, because in this simple setting (where every worker has the same decision to use labor intermediation or not), we could subtract a fixed U_c in utility terms for using the labor intermediation, without changing the propensity to specialization.

The model has the same "three stages" characteristic as in the last section. In the first stage, individuals receive schooling from the government to acquire a human capital for the use in the second stage. They choose the degree of specialization, and the choice will depend on the level of risk sharing in the second stage. There is no income and no consumption for the first stage[44]. The second and third periods are similar to the layout in the previous section.

Individuals choose the degree of human capital specialization by choosing the parameter $\phi > 0$. For a chosen ϕ, the next period's labor productivity will be

$$\tilde{B} = B_0 + \phi\tilde{B}_1, \tag{2.3.19}$$

[43]Note this is an expansion of the setting in the last section: there we discussed the special case of this general model: the case when $b = 1$

[44]same comment as in last section

where $\tilde{B}_1 \sim N(\bar{B}_1, \sigma_L^2)$

Therefore $E\tilde{B} = B_0 + \phi\bar{B}_1$ and $Var(\tilde{B}) = \phi^2\sigma_L^2$.

As in last section, we further assume that $B_0 < \bar{B}_1$, this says that specialization increases the mean of labor productivity: when $\phi = 0$, human capital is safe, but also not very productive. As ϕ increases, human capital on average becomes more productive, however, there is also higher labor productivity risk.

We study how the choice over ϕ is affected by the existing risk sharing in the economy. The intuition is, improved risk sharing not only enhances utility at every level of specialization, but also have an impact on the chosen level of specialization. As Arrow (1971, P137) put it, "the mere trading of risks, taken as given, is only part of the story and in many respects the less interesting part. The possibility of shifting risks, of insurance in the broadest sense, permits individuals to engage in risky activities that they would not otherwise undertake.". Indeed, we want to show that risk sharing in the economy (represented in the decrease of the σ_L^2, which is the residual risk *after* risk sharing) will **enhance** the agents' appetite in taking on more risk, and, in the end, *improved risk sharing opportunity could lead to increased risk, even after accounting for risk sharing.*

2.3.2 Solution

Assume we have a "symmetric equilibrium", where everybody choose the same level of specialization[45].

Assume that laborers are paid according to a linear sharing contract, where,

Per capita labor income:

$$w_t = b\tilde{B}_t + c \tag{2.3.20}$$

and,

Per capita capital income:

[45]although level of specialization is the same, they could be specializing in very different things. For example, a lawyer or doctor with specialty.

$$q_t = AK_t + (1-b)\tilde{B}_t - c \tag{2.3.21}$$

Theorem 2 *Suppose that $\tilde{B} = B_0 + \phi\tilde{B}_1$, where $\tilde{B}_1 \sim N(\bar{B}_1, \sigma_L^2)$, then the level of specialization chosen by everyone is*

$$\phi^* = \frac{[Ab + (1-b)]\bar{B}_1}{\alpha[(1-b)^2 + \frac{A^2}{1+A}b^2]\sigma_L^2} \tag{2.3.22}$$

and the maximum ex-ante utility is

$$U^* = \max_{\phi} \underset{\tilde{B}_t}{E}\, U = -\theta \exp(-\frac{\alpha}{1+A}\{[Ab + (1-b)]B_0 + (A-1)c + \frac{1}{2}\frac{[Ab+(1-b)]^2\bar{B}_1^2}{\alpha[(1-b)^2 + \frac{A^2}{1+A}b^2]\sigma_L^2}\}) \tag{2.3.23}$$

where, $\theta = [e^{\frac{ln A\beta}{1+A}} + e^{-\frac{A ln A\beta}{1+A}}]$

Proof: see appendix

2.3.3 Properties

It is easy to derive the comparative static for ϕ^* and U^* given the result of Theorem 2. When external risk sharing opportunity increases, leading to a decline in σ_L^2, ϕ^* will increase, meaning specialization of labor will increase, and the ex ante utility of individual also increases.

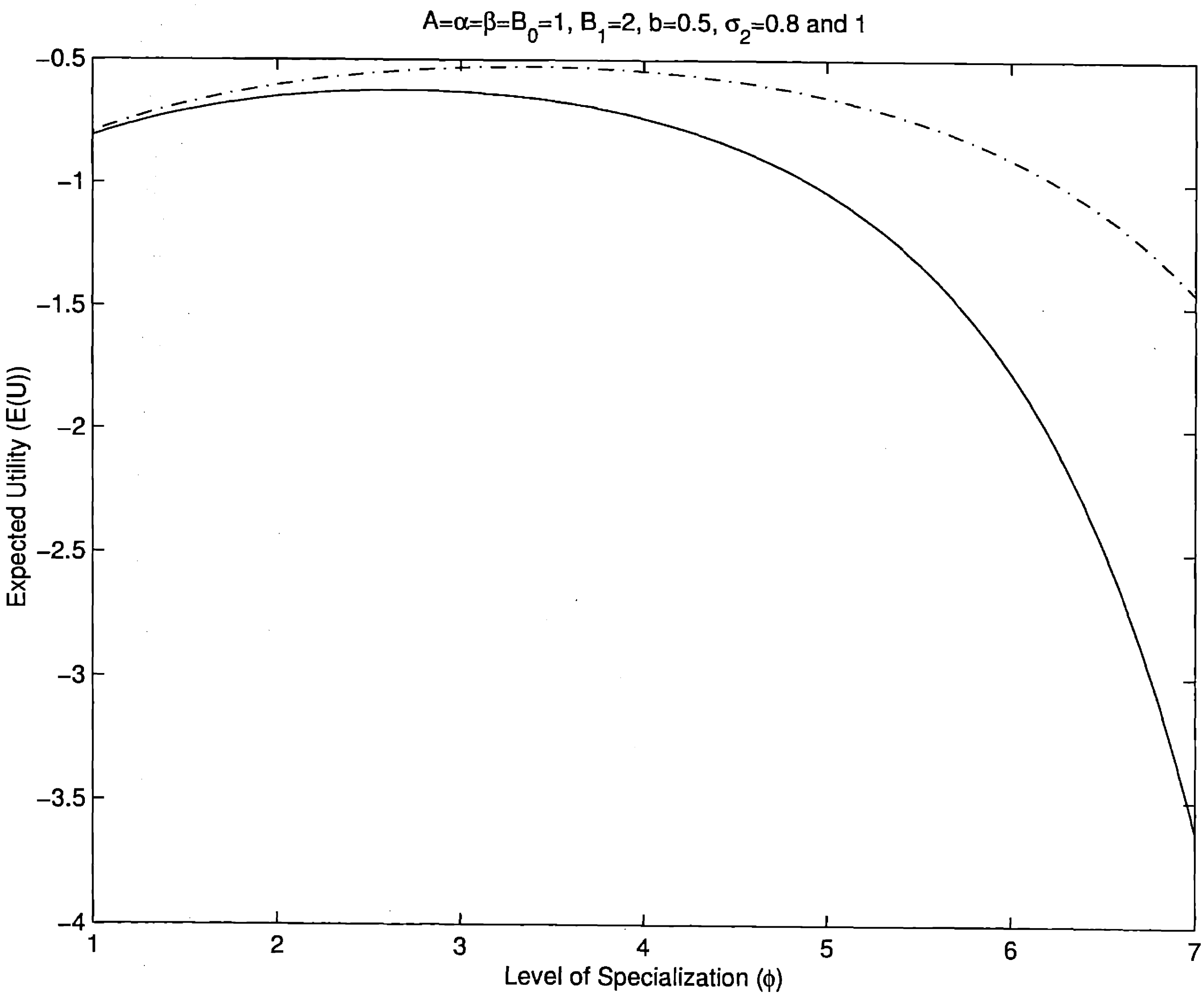
Increased risk-sharing on the labor market has two effects on the total benefits of workers. First, the increased risk sharing results in a higher benefit at *every* level of specialization; second, it increases the optimal level of specialization to *a higher level*, thus further increases total benefit of labor.

The effect is shown in Figure 2.4.

One important characteristic to note is, as risk sharing improves, although σ_L^2 goes down, the total *risk* in the labor income *increases*! This can be seen from the expression:

In Figure 2.4, when the standard deviation of labor income risk decrease from 1 to 0.8 (represented by the move from the solid line to the dashed line), expected utility increases at *every* level of specialization. Furthermore, the level of specialization that maximizes expected utility also increases.

Figure 2.4: EFFECT OF SPECIALIZATION



$$Var^*(\tilde{B}) = \phi^{*2}\sigma_L^2 = \frac{[Ab + (1-b)]^2 \bar{B}_1)^2}{\alpha[(1-b)+\frac{A^2}{1+A}b^2]^2\sigma_L^2} \tag{2.3.24}$$

which is a decreasing function of σ_L^2.

2.4 Conclusion and extension

conclusion

This paper studies the role of labor intermediation firm in sharing the labor productivity risk, and its effect on the ex ante choice of human capital specialization. Our main result is: highly functional specific human capital could join labor intermediation firm to better share the labor income risk, and highly firm specific human capital will want to stay in the traditional employment arrangement. The use of labor intermediation firm will reduce the ex post riskness of the labor income, and thus induce a higher level of human capital specialization ex ante. The increase in specialization could be so high that, even after accounting for the better risk sharing, laborers are taking on more labor income risks than they were when there is no labor risk sharing.

We point to several extensions which, due to space, were not fully exploited here, but could be interesting for future research in this area:

adding layoff and job mobility

One suggestion by Professor Peter Diamond: it would be nice if labor mobility and/or layoff is included. That can be accommodated by assuming a four period OLG model: there are two middle-aged periods: in the first period, people decide to work for a conventional production firm or a labor intermediation firm; in the second period, those previously working in labor intermediation firm will switch to the conventional production firm that has the highest productivity, while those previously working in the conventional labor firm will either stay there, or switch back to labor intermediation firm (after a really low realization of the $\tilde{B}_1$). Such an extension is doable. It also accords well with the empirical observation that, many consultants (also investment bankers), after initially spending some years at the consulting firm (investment bank), will eventually move to a "client firm"[46].

link to development

Early literature on the link between (financial) intermediation and development are Goldsmith (1969), McKinnon (1973) Shaw (1973) and Townsend (1983), recent literature in this area includes Acemoglu and Zilibotti (1997), Atje an Jovanovic (1993), Bencivenga and Smith (1991), Greenwood and Jovanovic (1990), Greenwood and Smith (1997), Levine (1997), Obstfeld (1994), Pagano (1993), Rajan and Zingalis (1998) and Saint-Paul (1992).

We would like to contrast our model with that of Acemoglu and Zilibotti (1997). In their model, economic development leads to increased wealth, which leads to more sectors to be opened given the minimum scale threshold. As more sectors open, the mean return doesn't increase, while the risk decreases because of improved diversification.

In our model, economic development comes in two senses:

First, it means that more firms will be operating in the economy. While each firm may not experience improvement in its production technology (and in our model we assume that indeed it doesn't), the increased number of production firm in the economy increases the potential scale of the production frontier, which is only exploitable with the labor intermediation: there, since labor intermediation firm always assign people to their highest productivity firm, the increased number of firms in the economy means that the expected productivity increases (mathematically, we have $f_1(N)$ increases in N).

Second, economic development, or learning over time, will reduce the cost of doing labor intermediation. Therefore, as economy develops, C goes down. It will be nice if we could characterize the dynamic of $\{\phi_{11}, c, \xi, \tilde{X}, c\}$ as functions of C.

empirical implications

Some empirical work can be carried out to better understand the function of labor intermediation. In particular, with data we could try to answer:

1. why there is variation across industrial countries on the level of labor intermediation?

   For example, compared with United States, France has higher cost of firing work-

[46]Professor Robert Merton pointed out that, as consultants/investment bankers go along their career, their job functions are changing from "getting the job done" to "getting the job". The leave/stay decision is a natural result of the consideration of the comparative advantage that one have. To maintain focus, this point is left for further research to fully exploit

ers. Does that imply a higher level of usage of labor intermediation, because other arrangements of labor risk sharing seems to be less feasible?

2. why do different industries differ in their usage of labor intermediation? This model would predict that, both the volatility of demand and the level of firm specificity of human capital might play a role here. For example, industries with a stable demand might have less labor intermediation, and industries where most jobs require only standard skills (i.e., not much firm specific human capital) might have more labor intermediation. Can we empirically test that?

3. historically, with the change in market demand/technological innovation/government regulation, how does the outsourcing practice in a particular country/industry change? For example, a natural experiment might be, for an industry that suddenly open to foreign competition (and thus might experience highly volatile demand), how does the employment structure change in accordance? For an industry in which suddenly regulation is relaxed and competition is allowed, how does the regular employment/labor intermediation pattern change?

We believe answering these questions will bring very helpful insight into the development of the labor intermediation industry, as well as the development of specialized human capital.

Chapter 3

Explaining The Cross Sectional Variation of Idiosyncratic Risk In International Stock Markets (Joint with Stewart C. Myers)

3.1 Introduction

This paper seeks to explain the negative relation between financial development and the levels of stock market synchronicity. Morck et al. (1999) document that in a cross-sectional setting, stock prices move together more in low-income economies than in high-income economies. They find this pattern is not due to market size difference, nor is it totally explained by the higher fundamentals correlation in low-income economies. Morck et al. (1999) as well as Campbell et al. (2000) also document that for the U.S. stock market, the level of stock market synchronicity (measured as the proportion of stocks moving in the same direction and the average regression R-squared using the market model) has been decreasing.

Morck et al. (1999) tested various explanations for their findings. They suggest that a country's institutionalized respect for property rights appear to explain the observed pattern of stock market synchronicity. In particular, they suggest that the lack of private property rights protection against insider's inter-corporate income shifting makes the firm-specific information less useful to risk arbitrageurs, therefore the market can be more affected by the noise traders. They further suggest that, as in De Long et al (1989, 1990), these noise traders' noise might be systematic in nature, thus the total market risk will be higher in those less developed financial markets because systematic noise are more incorporated in the price. This could lead to the observed higher synchronicity in these less developed markets.

However, upon closer look, it is not convincing that the noise traders' noise should necessarily be systematic. In fact, as argued in an appendix, the crucial assumption that leads to the results in De Long et al. (1989, 1990) is not that all noise traders' noise be systematic, it is rather that all the noise traders have the *same* noise about any particular stock, and these noise traders' *unanimous* beliefs could well *differ* across stocks! Thus, it is not convincing to assume that the lack of risk arbitrageur and the resulting higher noise level will necessarily lead to higher market risk, thus explaining the observed higher synchronicity in less developed market.

In this paper, we offer an alternative explanation through the mechanism as developed in Myers (2000). We argue that informational asymmetry between corporate insider and outsiders could prohibit the incorporation of firm specific information into the stock price. This combined with the less protection of outsider shareholder rights in less developed countries can explain the fact that less developed countries stock all seem to move together.

3.2 The model

In this paper, we assume the firm has fixed capital, no depreciation, no re-investment (thus $k = 0$).

A firm is composed of an inside manager, some outside equity holders and some physical assets. The insider and outside equity holders collectively own the firm's physical assets. The assets generate cash flow each period. The intrinsic value of the firm is defined as the present value of the discounted future cash flows that these assets generates, we discuss two cases:

1)

If both the insider and the existing physical assets work together, the total future cash flow will be C_t for period t, and the present value of these total cash flows are discounted by a constant (for now) rate of return r. We define ex-dividend intrinsic value to be mK_t at time t:

$$E(mK_t|I_t) = PV\{E_t(C_{t+1}), E_t(C_{t+2}), ...; r\} \tag{3.2.1}$$

where I_t is the information set that we condition on to evaluate the value of the firm.

2)

If the outsider kicks out the insider and hire someone else to manage the firm, then the intrinsic value of the firm will be proportionally lower:

$$E(K_t|I_t) = PV\{E_t(C_{t+1}/m), E_t(C_{t+2}/m), ...; r\} \tag{3.2.2}$$

The role of insider is important: he put in his human capital and increase the intrinsic value of the firm by a multiple of $m \geq 1$.

The market value of these assets, which is defined as the value of one unit of the asset to an outside investor at time t, can potentially be different from both mK_t and K_t, because the cash flow that are generated by the firm does NOT all go to the equity holders: some might be intercepted by the insiders.

The ability of the outsider to seize K_t, or a part of it (aK_t in the corporation model of Myers (2000)), is essential in determining the market value of firm. It provides an "outside option", which is the least amount of value that the firm can give to the equity holders. Thus, at any point in time, if the firm is going on with the insider, the market value of the firm must be at least equal to this value reachable by the outsiders (aK_t in the corporation model of Myers (2000)).

The incentive for insider to keep working with the firm, rather than to disband it and start anew, is that he will lose one period of cash flow should he default. Myers (2000) describes the conditions under which the insider fined it optimal to stick to the firm. In this paper, we make the same assumption and use the result there.

Note that the true underlying process is the cash flow process $\{C_t\}$, and the value process

$\{K_t\}$ is a "derivative" process. K_t is stochastic: although the physical asset remains fixed, its intrinsic value, which is defined as its cash generating ability, is changing over time.

In the following, for the most of the time, we refer to K_t as the "value of assets", but sometimes, when we talk about other values of assets, we make special emphasis of the terms to make sure there is no confusion. For simplicity of illustration, for the most part we also implicitly assume that the outsider owns 100% of the firm's physical asset (so that when the firm liquidates they get 100% of the proceeds). This can be generalized to the case when outsider owns only x proportion of the firm, see discussion in Myers (2000).

The following discussion characterizes the scenario that the insider works with the firm.

We assume the total cash flow of the firm at time t is

$$C_{t+1} = C_0 r_{t+1} \tag{3.2.3}$$

Where C_0 is a constant[1] , and r_{t+1} is a random variable that captures the random shock to the cash flow process. We further assume that this shock follows a one-factor model:

$$r_{t+1} = \beta f_{t+1} + \theta_{t+1} \tag{3.2.4}$$

where f_{t+1} is a common factor, and θ_{t+1} is a firm specific factor.

For simplicity, in the following analysis, we assume $\beta = 1$, thus

$$r_{t+1} = f_{t+1} + \theta_{t+1} \tag{3.2.5}$$

The first term captures an innovation of some macroeconomic factor, and the second term captures the firm specific innovation.

AR(1) assumption:

[1] C_0 can be viewed as the amount of physical capital. It is a constant. The cash flow each period can be thought of as generated by this fixed amount of physical capital, at a stochastic productivity r_{t+1}.

We assume that both f_{t+1} and θ_{t+1} processes are AR(1):

$$f_{t+1} = f_0 + \varphi_1 f_t + \epsilon_{t+1} \tag{3.2.6}$$

$$\theta_{t+1} = \theta_0 + \varphi_2 \theta_t + \xi_{t+1} \tag{3.2.7}$$

For simplicity we assume that f and θ have the same φ and that it is both positive and stationary:

$\varphi_1 = \varphi_2 = \varphi$ and $0 < \varphi < 1$

This guarantees that r_t itself is stationary AR(1):[2]

$$r_{t+1} = r_0 + \varphi r_t + \lambda_{t+1} \tag{3.2.8}$$

where $r_0 = f_0 + \theta_0$,and $\lambda_{t+1} = \epsilon_{t+1} + \xi_{t+1}$.

The unconditional mean of θ_t, f_t and r_t are:

$E(\theta_t) = \frac{\theta_0}{1-\varphi}$
$E(f_t) = \frac{f_0}{1-\varphi}$
$E(r_t) = \frac{r_0}{1-\varphi}$

and the unconditional variance of θ_t, f_t and r_t are:

$Var(\theta_t) = \frac{\sigma_\xi^2}{1-\varphi^2}$
$Var(f_t) = \frac{\sigma_\epsilon^2}{1-\varphi^2}$

[2]Stationarity is desirable, as we need to characterize the "steady state" equilibrium, and don't want the discrepancy of insider and outsider valuation of the firm to explode. See discussion at end of paper for more on this issue.

$Var(r_t) = \frac{\sigma_\lambda^2}{1-\varphi^2}$

where $\sigma_\xi^2, \sigma_\epsilon^2$ and σ_λ^2 are the variance of $\xi_t, \epsilon_t$and λ_t, respectively. See, for example, Hamilton (1994) equation 3.4.3 and 3.4.4. We also define the ratio of firm specific variance to the systematic variance:

$\kappa = \frac{Var(\theta_t)}{Var(f_t)} = \frac{\sigma_\xi^2}{\sigma_\epsilon^2}$

With these assumptions, the C_t process also follows AR(1):

$$\begin{aligned} C_{t+1} &= C_0 r_{t+1} \\ &= C_0(r_0 + \varphi r_t + \lambda_{t+1}) \end{aligned}$$

$$C_{t+1} = C_0 r_0 + \varphi C_t + C_0 \lambda_{t+1} \tag{3.2.9}$$

The unconditional mean for C_{t+1} is

$$E(C_t) = \frac{C_0 r_0}{1-\varphi} \tag{3.2.10}$$

And the unconditional variance for C_t is:

$$Var(C_t) = \frac{C_0^2 \sigma_\lambda^2}{1-\varphi^2} \tag{3.2.11}$$

$$\begin{aligned} E(C_{t+k}|C_t) &= E[C_0(r_0 + \varphi r_{t+k-1} + \lambda_{t+k-1})] \\ &= \ldots \\ &= \sum_{i=0}^{k-1} C_0 r_0 \varphi^i + \varphi^k C_0 r_t \end{aligned}$$

$$E(C_{t+k}|C_t) = C_0 r_0 \frac{1-\varphi^k}{1-\varphi} + \varphi^k C_t \tag{3.2.12}$$

In particular,

$$E(C_{t+1}|C_t) = C_0 r_0 + \varphi C_t \tag{3.2.13}$$

This combined with (3.2.1) gives:

$$\begin{aligned} E(mK_t|C_t) &= PV\{E(C_{t+1}|C_t), E(C_{t+2}|C_t), ...; r\} \\ &= \sum_{j=1}^{+\infty} \frac{C_0 r_0 \frac{1-\varphi^j}{1-\varphi} + \varphi^j C_t}{(1+r)^j} \end{aligned}$$

Therefore,

$$E(mK_t|C_t) = \frac{1}{r}\frac{C_0 r_0}{1-\varphi} + \frac{\varphi}{1+r-\varphi}(-\frac{C_0 r_0}{1-\varphi} + C_t) \tag{3.2.14}$$

Thus,

$$E(K_t|C_t) = \frac{1}{m}[\frac{1}{r}\frac{C_0 r_0}{1-\varphi} + \frac{\varphi}{1+r-\varphi}(-\frac{C_0 r_0}{1-\varphi} + C_t)] \tag{3.2.15}$$

Also, by plugging in the unconditional expected value of C_t in (3.2.10), the unconditional expected value of K_t is:

$$E(K_t) = \frac{1}{m}\frac{1}{r}\frac{C_0 r_0}{1-\varphi} = \frac{1}{rm}E(C_t) \tag{3.2.16}$$

This equation has the intuitive interpretation that, on average, the value of the firm, in the eyes of outsiders, is worth the discounted cash flow that can be generated when the outsider operates the firm without the insider, or, the outside option value of the outside investors.

(3.2.16) also gives the K_t in the form of a linear function of C_t :

$$E(K_t|C_t) = a + bC_t \tag{3.2.17}$$

where $a = \frac{1}{m}\frac{C_0 r_0}{1-\varphi}(\frac{1}{r} - \frac{\varphi}{1+r-\varphi})$, and $b = \frac{1}{m}\frac{\varphi}{1+r-\varphi}$.

Thus,

$$\begin{aligned} E(K_{t+1}|C_t) &= a + bE(C_{t+1}|C_t) \\ &= a + b(C_0 r_0 + \varphi C_t) \\ &= a + bC_0 r_0 + b\varphi C_t \\ &= a' + b'C_t \end{aligned}$$

where $a' = a + bC_0 r_0$, and $b' = b\varphi$;

Also we have,

$$\begin{aligned} E(K_{t+1}|K_t) &= a' + b'C_t \\ &= a' + b'(K_t - a)/b \\ &= a + bC_0 r_0 + b\varphi(K_t - a)/b \\ &= a(1-\varphi) + bC_0 r_0 + \varphi K_t \end{aligned}$$

$$E(K_{t+1}|K_t) = \frac{1}{mr}C_0 r_0 + \phi K_t \tag{3.2.18}$$

Thus K_t follows an AR(1) process.

At this point, it might help to draw the time line of the model:

C_t $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ C_{t+1}

$\downarrow$ $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ $\downarrow$

$\uparrow$ $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\uparrow$

K_t(ex-dividend intrinsic value) $\qquad\qquad\qquad$ K_{t+1}

t $\quad$ t^+ $\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ t+1 $\qquad$ $t+1^+$

At time t, a cash flow C_t is realized. After that, K_t is evaluated using estimated future cash flow series. Then, next period, $C_{t+1}$is realized, etc.

At this point, we discuss the general information structure, and then talk about both the general case and two extreme cases.

We assume that the insider observes both f_t and θ_t at the time these are realized. Outsiders observes the macro factor f_t, too, but they may not be able to observe *all* the information in θ_t. We denote the piece of θ_t that the outsiders do observe as $\theta_{1,t}$, and the piece of θ_t that the outsiders don't observe as $\theta_{2,t}$. Thus we have $\theta_t = \theta_{1,t} + \theta_{2,t}$. Furthermore, we assume these two pieces of information follows the same AR(1) process:

$$\theta_{1,t+1} = \theta_{1,0} + \varphi\theta_{1,t} + \xi_{1,t+1} \tag{3.2.19}$$

and

$$\theta_{2,t+1} = \theta_{2,0} + \varphi\theta_{2,t} + \xi_{2,t+1} \tag{3.2.20}$$

We further assume that the two pieces of firm specific information are independent, thus observing one does not give information about the other. We can define the *transparency (I temporarily use this for the lack of a better name)* of the firm as the ratio of the variance of

$\theta_{1,t}$ to the sum of the variance of $\theta_{1,t}$ and $\theta_{2,t}$[3]:

$$\eta = \frac{Var(\theta_{1,t})}{Var(\theta_{1,t} + \theta_{2,t})} = \frac{Var(\theta_{1,t})}{Var(\theta_{1,t}) + Var(\theta_{2,t})} \tag{3.2.21}$$

The last equation holds because of independency assumption.

The transparency of a country's financial and accounting system can be measured by η, which is between 0 and 1. A high η means most of the firm specific information is revealed to outsiders through accounting reports, a lower η means that not much of the firm specific information is revealed to outsiders. The United States and some western European countries are probably close to the extreme of $\eta = 1$, while some emerging market economies might have an η close to 0.

The general case.

Each period, the outsiders observe both f_t and θ_t, and they form estimation about the cash flow using their available information. Outsiders are interested in knowing the C_t series, not because they can capture it (as in the original Myers (2000) model, the insider has full right to intercept any cash flow for the current period), but because it gives them a reference in estimating the future cash flow, thus the value K_t, which is the basis for their outside option value.

We have that the outsider's estimation of r_t, C_t and mK_t are:

$E(r_t | f_t, \theta_{1,t}) = f_t + \theta_{1,t} + E(\theta_{2,t})$

[3]imperfections of the developing countries financial market, we have considered two alternative ways: First, we could assume that the firm specific information is all revealed, but in less developed countries it is accompanied with a large amount of noise. Thus, although the outsider does not know the exact firm specific information (because it is clouded by a large amount of noise), he will on average have an unbiased estimate of this information. Second, we could assume that some firm specific information is simply non-existent. We think the second approach is probably closer to the truth: the report of firm specific investment and cash flow is often nonexistent, and even when they do exist, it is not sure whether they are unbiased estimate of the truth.

so

$$E(r_t|f_t,\theta_{1,t}) = f_t + \theta_{1,t} + \frac{\theta_{2,0}}{1-\varphi} \tag{3.2.22}$$

$E(C_t|f_t,\theta_{1,t}) = E(C_0 r_t|f_t,\theta_{1,t})$

$$E(C_t|f_t,\theta_{1,t}) = C_0(f_t + \theta_{1,t} + \frac{\theta_{2,0}}{1-\varphi}) \tag{3.2.23}$$

$$\begin{aligned} E(C_{t+k}|f_t,\theta_{1,t}) &= E[E(C_{t+k}|C_t)|f_t,\theta_{1,t}] \\ &= \tfrac{C_0 r_0(1-\varphi^k)}{1-\varphi} + \varphi^k E(C_t|f_t,\theta_{1,t}) \end{aligned}$$

$$E(C_{t+k}|f_t,\theta_{1,t}) = \frac{C_0 r_0(1-\varphi^k)}{1-\varphi} + \varphi^k C_0(f_t + \theta_{1,t} + \frac{\theta_{2,0}}{1-\varphi}) \tag{3.2.24}$$

$$\begin{aligned} E(mK_t|f_t,\theta_{1,t}) &= E[E(mK_t|C_t)|f_t,\theta_{1,t}] \\ &= \tfrac{1}{r}\tfrac{C_0 r_0}{1-\varphi} - \tfrac{\varphi}{1+r-\varphi}\tfrac{C_0 r_0}{1-\varphi} + \tfrac{\varphi}{1+r-\varphi} E(C_t|f_t,\theta_{1,t}) \end{aligned}$$

$$E(mK_t|f_t,\theta_{1,t}) = \frac{1}{r}\frac{C_0 r_0}{1-\varphi} - \frac{\varphi}{1+r-\varphi}\frac{C_0 r_0}{1-\varphi} + \frac{\varphi}{1+r-\varphi} C_0(f_t + \theta_{1,t} + \frac{\theta_{2,0}}{1-\varphi}) \tag{3.2.25}$$

We can check that the *unconditional* mean of mK_t is:

$$E(mK_t) = \frac{1}{r}\frac{C_0 r_0}{1-\varphi} \tag{3.2.26}$$

This is the average value of the firm to the insider. It has the interpretation that the firm value to insider is equal to the mean cash flow, $\bar{C} = \frac{C_0 r_0}{1-\phi}$, discounted at rate of return r.

$$E(mK_{t+1}|f_t,\theta_{1,t}) = E[E(mK_{t+1}|K_t)|f_t,\theta_{1,t}]$$

and from (3.2.18)

$$E(mK_{t+1}|f_t,\theta_{1,t}) = C_0 r_0 \tfrac{1}{r} + \phi E(mK_t|f_t,\theta_{1,t})$$

Thus,

$$E(mK_{t+1}|f_t,\theta_{1,t}) = C_0 r_0 \frac{1}{r} + \phi\left(\frac{1}{r}\frac{C_0 r_0}{1-\varphi} - \frac{\varphi}{1+r-\varphi}\frac{C_0 r_0}{1-\varphi} + \frac{\varphi}{1+r-\varphi} C_0 (f_t + \theta_{1,t} + \frac{\theta_{2,0}}{1-\varphi})\right) \tag{3.2.27}$$

Define

$$V_t^{ex} = \frac{E(Y_{t+1}|f_t,\theta_{1,t}) + E(V_{t+1}^{ex}|f_t,\theta_{1,t})}{1+r} \tag{3.2.28}$$

Where V_t^{ex} is the ex-dividend market value of the outsiders' asset, conditioning on the observation of $\{f_t, \theta_{1,t}\}$.

In equilibrium, this value should be at least equal to the outsider's expectation of the reachable liquidation value of the firm, $E(\alpha K_t|f_t,\theta_{1,t})$, and optimality means that insider will make just enough payment of dividend, to ensure that there is an equality between the two (assume, as in Myers (2000), in the case of indifference, both parties favor an ongoing concern).

Thus we have, at time t,

$$\begin{aligned}
E(\alpha K_t|f_t,\theta_{1,t}) &= V_t^{ex} \\
&= \tfrac{E(Y_{t+1}|f_t,\theta_{1,t}) + E(V_{t+1}^{ex}|f_t,\theta_{1,t})}{1+r} \\
&= \tfrac{E(Y_{t+1}|f_t,\theta_{1,t}) + E(\alpha K_{t+1}|f_t,\theta_{1,t})}{1+r} \\
&= \tfrac{E(Y_{t+1}|f_t,\theta_{1,t}) + \frac{\alpha}{m}[C_0 r_0 \frac{1}{r} + \varphi E(mK_t|f_t,\theta_{1,t})]}{1+r}
\end{aligned}$$

Reorganize to get

$$E(Y_{t+1}|f_t,\theta_{1,t}) = \alpha E(K_t|f_t,\theta_{1,t})(1+r-\varphi) - \frac{\alpha}{m}C_0 r_0 \frac{1}{r} \tag{3.2.29}$$

We can check and verify that the unconditional mean of Y_t is $E(Y_t) = \frac{\alpha}{m}\frac{C_0 r_0}{1-\varphi}$, which is equal to $\alpha r E(K_t)$.

Interpretation: the unconditional expected dividend each period is r (the required rate of return) times the unconditional expected value of the asset reachable by outsiders.

Define

$$Y_t = \alpha\frac{1+r-\varphi}{\varphi}[E(K_t|f_t,\theta_{1,t}) - \frac{1}{r}\frac{1}{m}C_0 r_0] - \alpha\frac{1}{r}\frac{1}{m}C_0 r_0 \tag{3.2.30}$$

Then, easy to verify that Y_t satisfies (3.2.29), thus Y_t will be one solution to the problem.[4] It is easy to verify that this solution defines a Nash Equilibrium: as long as the unobserved firm specific shock is not too bad, the insider will find it optimal to stick to this dividend payment plan. The outsider has no incentive to deviate either. Furthermore, following the discussion in Myers (2000), this solution has the good property that insider will exert the right amount of effort and take all positive NPV projects. Thus this equilibrium will lead to Pareto optimal outcome. Therefore, even though there could exist other types of equilibrium, given the optimality of this particular equilibrium, it is likely that the actual equilibrium will settle on this one.

Expressed in terms of C_t rather than K_t, we get a much simplified expression:

[4] To find this solution, we did the following "backward" search:
(3.2.18) implies that
$K_{t+1} = \frac{1}{r}\frac{1}{m}C_0 r_0 + \varphi K_t + error\_term$,
Thus $K_t = \{K_{t+1} - [\frac{1}{r}\frac{1}{m}C_0 r_0]\}/\varphi + error\_term$,
Plug this (without the error term) into the right hand side of the equation (3.2.29) to get the conjecture of the solution to equation (3.2.29).

$$Y_t = \frac{\alpha}{m} E(C_t | f_t, \theta_{1,t}) \tag{3.2.31}$$

See appendix for the proof of equation (3.2.31). Similarly,

$$E(Y_{t+k} | I_t) = \frac{\alpha}{m} [C_0 r_0 \frac{1 - \phi^k}{1 - \phi} + \phi^k E(C_t | I_t)] \tag{3.2.32}$$

Also, a calculation similar to (3.2.14) gives, $PV(Y_{t+1}, Y_{t+2}, ...; r) = \alpha[a + bC_t | I_t] = \alpha K_t$, as would be expected.

We also calculate the insider's rent each period, which is the total cash flow minus what is given to the outsiders:

$$Z_t = C_t - Y_t = C_t - \frac{\alpha}{m} E(C_t | f_t, \theta_{1,t}) \tag{3.2.33}$$

Note that under extreme circumstances, Z_t could be small, or even negative. These correspond to the cases when the firm's opaque firm specific information is very bad, and insiders have to make up the difference in the insider and outsiders' valuation, using his own money. In the case when Z_t is indeed negative, then the insider will have to take some money out of his own salary for this period, to make up for the difference. Thus, in the realization of a very bad $\theta_{2,t}$, the insider will take a pay cut to try to make up for the dividend to the outsiders.

We can now calculate the total return for the stock.

Define

$$r_{total,t+1} = \frac{V_{t+1}^{ex}(f_{t+1}, \theta_{1,t+1}) + Y_{t+1}(f_{t+1}, \theta_{1,t+1})}{V_t^{ex}(f_t, \theta_{1,t})} - 1 \tag{3.2.34}$$

$r_{total,t+1}$ is the realized return on the stock in period t+1.

In the appendix, we verify that

$$r_{total,t+1} = r + \frac{(1+r)(\epsilon_{t+1} + \xi_{t+1})}{\frac{r_0(1+r)}{r} + \phi(f_t + \theta_{1,t})} \tag{3.2.35}$$

This is as would be expected. The random component in r_{total} is caused by innovations of both f_{t+1} and $\theta_{1,t+1}$. The expected return for the stock next period conditional on f_t and $\theta_{1,t}$ and the unconditional expected return for the stock next period are both r, which is the required discount rate. Thus, although the cash flow process is partially predictable (due to the AR(1) process), it does not result in market inefficiency.

Then, define the "market portfolio" as the portfolio of the whole market. We can think of the return on the market portfolio as the return of a stock that does not have idiosyncratic risk ($\theta = 0$), since the firm specific risk is all washed out by forming the portfolio. Then,

$$r_{m,t+1} = r + \frac{(1+r)(\epsilon_{t+1})}{\frac{r_0(1+r)}{r} + \phi(f_t)} \tag{3.2.36}$$

We can see that, conditioned on the stock prices at time t, the next period stock price can be explained by two parts: a "market movement" ϵ_{t+1}, that is captured by the market return $r_{m,t+1}$, and a firm specific component ξ_{t+1}. The simple linear regression specification is mis-specified, since the derivative of $r_{total,t}$ on $r_{m,t}$ is of course dependent on the realization of the variables f_t and $\theta_{1,t}$, but conditioned on the realization of these two, the proportion of innovation that is explainable by the market movement is fixed:

$$\frac{Var(\epsilon_{t+1})}{Var(\epsilon_{t+1}) + Var(\xi_{t+1})} = \frac{1}{\kappa\eta + 1} \tag{3.2.37}$$

The reason for a less developed country stock market to have lower regression R-squared could be either because of a lower κ, which means the stocks have less idiosyncratic risk or more market risk, or both, or because of a lower η, which means the stocks have a lower *observable* idiosyncratic risk.

We discuss two special cases:

1. when the outsiders observe all firm specific information that the insider knows.

In this case, $\theta_1 = \theta$, and $\theta_2 = 0$. $\eta = 1$, which means for a given κ, the regression R^2 will be the lowest.

2. when the outsiders observe no firm specific information that the insider knows.

In this case, $\theta_1 = 0$, and $\theta_2 = \theta$. $\eta = 0$, which means for a given κ, the regression R^2 will be the highest.

3.3 Conclusion

In this model, insider observes firm specific information about cash flows. But due to the opaqueness of the firm, he has no credible way to convey this information to the outsiders. Since the insider acts as a residual claimer of the firm's value, he effectively has to issue an insurance to outsiders, insuring them against the fluctuation of the firm specific cash flow that is un-observable by the outsiders. That is, insiders absorb the un-observable firm-specific shocks[5].

[5]This argument is supported by the recent researches by Friedman and Johnson (2000) which shows that in countries with weak legal environments, entrepreneurs can engage in both tunneling (which transfers resources out of a firm) and propping (which transfers resources to a firm).

Chapter 4

How Anomalous Are Financial Anomalies? (Joint with Andrew W. Lo)

4.1 Introduction and Literature Review

The issue of financial anomalies have received tremendous attention in the finance literature, generating both empirical works trying to document these anomalies and theoretical works trying to "explain" them. The issue remains, however, whether these anomalies are really existing patterns of the data that warrants future study, or just statistical artifact due to extensive data mining. In a followup of the Lo and MacKinglay (1990, RFS) study, we developed several statistical and econometric methods to study the issue of data mining, to try to differentiate real financial anomalies from ones resulting from datamining. Our work can be thought of as proposing the right way to do data mining.

Following Lo and MacKinlay (1990 a), there has been an increasing realization among researchers that the degree of data-snooping bias in a given field tends to increase with the number of researches done in that field. In making statistical inferences we need to take into consideration the effect of data-snooping. Recently, White (1997) and Sullivan et. al. (1998, 1999) have a series of papers proposing using the "reality check" to guard against the data-snooping effects. White's reality check draws on West (1996) for its theoretical underpinning, and the stationary bootstrap technology developed in Politis and Romano (1994) for implementation.

Recently, Foster et al (1997) documented the potential problem of "choosing the best k out of m regressors" in doing a regression specification search. They studied the setting that the researcher can chooses k out of m candidate regressors to perform the regression. Since there are $\binom{m}{k}$ ways to do the regression, if the researcher reports the best measure of the goodness of fit among the $\binom{m}{k}$ measures, we should not do a test *as if* he only had one fixed set of the k parameter at hand and just did his experiment once. Indeed, such "naive" test statistics will severely underestimate the possibility of data-mining, and overestimate the likelihood of an anomaly. In view of this effect, Foster et al provided simple procedure to adjust the critical R^2 values to account for the data-snooping effect. They went on to use their methodology to re-examine several studies by estimating "the 95 percent confidence limit for m^*", which is "the minimum number of regressors required to achieve an R-squared at least as high as that reported in a number of recent studies. If the researcher had access to more than the listed m^* regressors (either directly, or through the reading of other researcher's papers), then their reported R-squared would occur at least five percent of the time through random chance, according to the test used[1]". They found that, after considering the data-snooping effect, while some effects seem to remain significant, some others are less anomalous.

The literature also see significant development of Bayesian methods (both bayesian decision theory and bayesian estimation method) in the finance literature. Examples of early work include Shanken (1987), Madhavan and Smidt (1991, JFE) and Harvey and Zhou (1990), more recent studies include Kandel and Stambaugh (1996) and Pastor (2000).

4.2 Citation Count Analysis

We did the citation search in two steps.

In the first step, we tried to define the anomalies that we want to study. We searched all the economics and finance articles as published in the ECONLIT database for articles discussing anomalies. We searched using keywords "anomaly", "anomalous", "puzzle", "market efficiency", "efficient market", "market inefficiency", "inefficient market" and the plural form. The search produced 4043 matching articles [2]. Out of these, we did a careful analysis to weed out those irrelevant ones, and identified about 40 anomalies that we think are of significant interest to the finance and economics literature.

[1]Quoted from the heading of Table IV in their paper
[2]as of June 6, 1999

In the second step, we tried to trace the origin of each of these anomalies, and then trace all the citations of the original papers discussing these anomalies. To ascertain the origin of the literature on any particular anomaly, we identified usually two prominent articles discussing the anomaly[3]. We read through these prominent articles, paying special attention to the literature review sections of them, which usually trace the starting point of the literature dealing with the anomaly. From these literature review sections, we identified some of the earliest and/or prominent articles studying the particular anomaly, and we used them as the "seeds" for our citation search. Since these later articles might not cite any one particular "seed", we picked usually three seeds, hoping that the majority of subsequent articles discussing the anomaly will cite at least one of these seed articles, in order to have any impact on the existing literature discussing that anomaly.

The database we used to do the citation search is the online version of "the Web of Science", published by the Institute of Scientific Information, version 4.02 (URL: http://webofscience.com). This database has three sub-categories:

- Science Citation Index Expanded (SCI-EXPANDED)–1983-present
- Social Sciences Citation Index (SSCI)–1983-present
- Arts & Humanities Citation Index (A&HCI)–1983-present

For our purpose of identifying anomalies in finance, we find it suffice to use the Social Sciences Citation Index.

For each of the anomalies that we identified, we searched through the Web of Science database for any articles that cited those "seed" articles that we assigned for the anomaly. If possible, we carefully sift through the abstract section of the citation search results, to make sure that they are indeed discussing the anomaly in some way. Such a double-check may not always be possible for some of the articles published in earlier years, since the database we use does not give complete abstract for some articles published in earlier years.

[3]these articles are not necessarily the earliest ones discussing the anomaly, but they are typically published in major journals, such as Journal of Finance, Journal of Financial Economics, Journal of Financial and Quantitative Analysis, American Economic Review, Journal of Political Economy. Our hope is, these articles will mention in the literature review section what these authors thought were the origin of the literature dealing with the anomaly under discussion

We tabulate the citation search results according to the years that these articles are published, in order to get a pattern of the relative interest of the literature to each of these anomalies.

4.3 Statistical Inference For Anomalies

4.3.1 Bootstrap Methods

intuition of White's reality check

Following Lo and MacKinlay (1990), there has been an increasing realization among researchers that the degree of data-snooping bias in a given field tends to increase with the number of researches done in that field. In making statistical inferences we need to take into consideration the effect of data-snooping. Recently, White (1997) and Sullivan et. al. (1998, 1999) have a series of papers proposing using the "reality check" to guard against the data-snooping effects. In this subsection, we modify and extend the analysis in these papers. To better compare our extension with their initial method, we choose to focus on the same examples used in the paper, "Dangers of Data-Driven Inference: The Case of Calendar Effects in Stock Returns", by Sullivan et al. (1998).

White's reality check draws on West (1996) for its theoretical underpinning, and the stationary bootstrap technology developed in Politis and Romano (1994) for implementation.

The implementation of the White's reality check is explained in detail in White (1997), pp.11 to pp.13. The particular application testing the mean performance of the calendar strategies, as done in Sullivan et al (1998), can be summarized as followed:

1. choose the number of bootstrap re-samples, N, and the smoothing parameter for the Politis and Ramano (P&R) stationary bootstrap, q. In Sullivan et al (1998), N is chosen to be 500, and q is chosen to be 0.1.

2. apply the stationary bootstrap of P&R N times, to generate N sets of random observation indexes, each of length P, for the bootstrap resamples. The ith random observation index is a random function, $\theta_i(.)$, mapping from {R, ..., T} to {R, ..., T}, where R and T are the starting and ending dates of the sample of time series variables. This

same set of N random observation indexes, once created, will be used for generating the bootstrapped values for all of the calendar rules under consideration.

3. The *performance* of the kth calendar rule on date t+1 is defined as the difference between log returns of the calendar rule under consideration and a benchmark, or

$$f_{k,t+1}(\hat{\beta}_t) = ln[1 + y_{t+1}S_k(.)] - ln[1 + y_{t+1}S_0(.)], k = 1, ..., l$$

where $S_k(.)$ is the "signal" of the kth calendar strategy by using the data. It can take on values of {0, -1, 1}. A signal of 1 means longing the stock market, a signal of -1 means shorting the stock market, and a signal of 0 means a neutral position in the stock market. $S_0(.)$ is a benchmark, which in Sullivan et al (1998) is always equal to 1 (the strategy of always investing in the stock market). The performance is a function of $\hat{\beta}_t$, which in a more general setting is the estimate of some parameters under the specification being considered[4]. In the case of calendar effects, $\hat{\beta}_t$ can be viewed as some function of the signal $S_k(.)$.

Also, define the sample mean of performance as:

$$\bar{f}_k = P^{-1} \sum_{t=R}^{T} f_{k,t+1}(\hat{\beta}_t)$$

4. Using the observation indexes derived earlier, we can construct the bootstrapped values:

$$\bar{f}^*_{k,i} = P^{-1} \sum_{t=R}^{T} f_{k,\theta_i(t)+1}(\hat{\beta}_{\theta_i(t)}), i = 1, ..., N$$

where $\bar{f}^*_{k,i}$ is the performance of kth calendar rule, in the ith iteration of the bootstrap.

[4]For example, a set of regression coefficients. See White (1997) for examples.

5. From the $\bar{f}_k$'s (the sample mean) and $\bar{f}^*_{k,i}$'s (the bootstrapped value) thus constructed, we can do the reality check *recursively*:

   First, define

   $\bar{V}_1 = P^{1/2}\bar{f}_1$, and

   $\bar{V}^*_{1,i} = P^{1/2}(\bar{f}^*_{1,i} - \bar{f}_1)$, i=1, ..., N.

   At this stage, if we want, we can compare the value of $\bar{V}_1$ to the percentiles of $\bar{V}^*_{1,i}$, i=1, ..., N, following the next step to get the p-value to the test of the hypothesis that the first calendar rule does not outperform.

   Then, we can recursively define

   $\bar{V}_{k+1} = max\{P^{1/2}\bar{f}_{k+1}, \bar{V}_k\}$, and,

   $\bar{V}^*_{k+1,i} = max\{P^{1/2}(\bar{f}^*_{k+1,i} - \bar{f}_{k+1}), \bar{V}^*_{k,i}\}, i = 1, ..., N$

   After the $\bar{V}_{k+1}$ and $\bar{V}^*_{k+1,i}$ are defined, we can compare the value of $\bar{V}_{k+1}$ to the percentiles of $\bar{V}^*_{k+1,i}$, i=1, ..., N, to get the p-value to the test of the hypothesis that the best performing calendar rule among the first k+1 rules doesn't outperform.

6. To obtain the p-value from the comparison of $\bar{V}_k$ to the quartiles of $\bar{V}^*_{k,i}$, i=1, ..., N, we first sort the $\bar{V}^*_{k,i}$'s in ascending order to get

$$\bar{V}^*_{k,(1)} \leq \bar{V}^*_{k,(2)} \leq ... \leq \bar{V}_{k,(N)}$$

   Where $\bar{V}^*_{k,(i)}$'s is an ordered sequence of $\bar{V}^*_{k,i}$'s.

   Then we find the n such that

   $\bar{V}^*_{k,(n)} \leq \bar{V}_k < \bar{V}^*_{k,(n+1)}$

   And the White's Reality Check p-value is

$$P_{RC} = 1 - n/N$$

White (1997) also suggest a refinement of the p-value using interpolation.

Some critiques of the paper

We have the following critiques of the paper:

1. **the number of calendar rules tested is probably too large**

   The full set of calendar effects in the paper (including 9,452 different rules) are much more than what the researchers could have tested in reality, and many of these rules are effectively redundant. Although increasing the number of rules will not affect the final result asymptotically as the sample size goes to infinity, with finite sample it might affect the final result to certain extent. In particular, it could make some real anomalies appear as if they are from extensive data mining. It seems to us that the reduced set of calendar rules that they proposed (with 244 different rules) presents a more reasonable set for testing the effect of data snooping in calendar effects, therefore we should focus on this reduced set of rules.

2. **the reasonable measure of return of a trading strategy**

   The paper's definition of trading strategy return produces some bias. For example, when the calendar strategy under discussion is to short Monday and neutral on other days of the week, the return from this strategy (with return of $-r_{market}$ on Monday, and zero on other days of the week) is compared with the return from the market strategy (with return of r_{market} on every days of the week). Since the average return on market is positive, such a comparison considerably favors the null hypothesis that the strategy under consideration does not outperform the market strategy, and thus is biased against detection of any calendar anomaly.

   Such a problem can be ameliorated to a certain extent by appropriately defining the return on the strategy under discussion. The most "natural" choice in finance is the following: We assume that the investor has a certain amount to invest, and can choose to invest in either stock market or the riskless bond market. Therefore, the investor puts in $S_k(t)$ times his initial wealth in the stock market, and any residual amount in the riskless bond market, earning a riskless return (for our replication we assume a 5 percent per annum riskless rate). For example, if the strategy is to short the market on Monday and be neutral on the other days of the week, the return is $-r_{market} + 2r_f$ on Monday, and r_f on other days of the week, where r_{market} is the return on market, and r_f is the return on riskless bond (here assumed to be 5 percent per annum).

3. **log return versus simple return.**

   Fama (1998) argues that for long horizon studies of anomalies, the use of average abnormal return (AAR) should be preferred to buy-and-hold abnormal return (BHAR), because the former is less susceptible to bad model problem. Therefore, it would be nice if the paper also includes an analysis based on the simple return, in addition to the log return they use (which is a one-by-one and onto transformation of the buy-and-hold abnormal return), and see if there is a big difference in the results.

4. **the reliance on asymptotic distribution.**

   White's test statistics rely on the sample size T being large. How large should T be in order to be *sufficiently* large?[5]. We would want to do some simulation to test the real world performance of their test, under a finite sample.

5. **benchmark**

   We think that properly assigning the benchmark is very important and researchers have no consensus on **the** benchmark.

   Using market strategy as the benchmark biases the test towards not rejecting the null. The return on the market strategy, while uncertain, has a positive mean that is larger than the riskless rate due to a positive market risk premium. The return on *any* strategy other than the market strategy, however, requires sometimes investing in a riskless asset. Therefore, for any particular calendar strategy, the test statistic, which is the difference between the sample mean return of the calendar strategy and the sample mean return of the market strategy, tends to be negative (due to the market risk premium), even if there is no abnormal return on this particular strategy.

[5] Also, if T is really large, there might be structural breaks/regime switching occurring (how confident are we that the stock market today is the same as it was 100 years ago, when their sample starts?)

On the other hand, using the riskless rate as benchmark can significantly bias towards rejecting the null hypothesis of no abnormal calendar rule performance, because every calendar rule in the reduce set will long the market on some days and be neutral on the other days.

We would want to experiment what impact this will have on the likelihood of detection of calendar rule anomaly.

Our modification of the White's test

We modified White's reality check algorithm based on the reduced set of calendar effects in the paper. As discussed above, we feel that the reduced set is about the right universe of what in reality any researcher would choose from, in order to do a specification test of the calendar effects. It does not exaggerate the data-snooping problem nor discounting it.

Sullivan et al discussed both a mean criterion and a sharp ratio criterion. To save time and space, we only focus on the replication of the mean return criterion in their paper.

We use the daily CRSP value-weighted index, and to save computation time, most of our tests are performed on a reduced dataset of 800 observations, from 19620705 to 19650903. There is no reason for us to pick any particular time period, and we do intend to check whether the results we got can equally be obtained by using other time periods.

We made the following modifications to the White's test of calendar effects as discussed in the Sullivan et al paper:

1. **redefining the calendar rule return**

   Following one of the critiques about their paper, we redefined the return of a calendar strategy by assuming that the investor invests in the riskless bond at 5 percent per annum return whenever he/she is not in the stock market. With the reduced universe of strategy under analysis by their paper, for any particular calendar strategy, on any day, the strategy could dictate a long, neutral or short position in the market. If the strategy dictates a long position in the market during day t, then the strategy's return earned during the close of day t-1 to the close of day t is $r_{market,t}$[6][7], which

[6]for simplicity, as in many other researches, we take the close of day t-1 to be equal to the open of day t, alternatively, we could do without such assumption by assuming that a long position during day t entails buying at close of day t-1 at the day t-1 closing price, and selling at close of day t at the day t closing price.

[7]In this study, there is no "measurability problem" of the information set. Our test is based on the

is defined as $r_{market,t} = (p_t - p_{t-1})/p_{t-1} - 1$, where p_t is the closing price of day t; if the strategy dictates a neutral position, then the strategy's return during day t is r_f, which is the riskless interest rate (assumed to be 5 percent per annum); if the strategy dictates a short position in the market, then the strategy's return during day t is $-r_{market,t} + 2r_f$ (that is, investor is shorting the stock market by an amount equal to his initial investment, and invest that amount together with his wealth in the bond market).

2. **redefining the benchmark**

   We experiment the effect of different choices of benchmark on the likelihood of detecting anomaly. We test three different benchmarks:

   - the stock market as benchmark
   - the riskless bond market as benchmark
   - a 50-50 strategy, where 50% of the initial wealth every period is invested in the stock market, and the other 50% invested in the bond market.

   Our sense is that the market strategy might be biased against any rejection of the null, and the riskless strategy might be biased for a rejection of the null, while the 50-50 strategy might somehow maintain the balance.

3. **Do both Average Abnormal Return and Buy-and-Hold Abnormal Return**

   To see whether choosing Average Abnormal Return or Buy-and-Hold Abnormal Return matters to the result, we do a test on *both* the simple return and the log return. We want to see whether the results will differ significantly.[8]

Our tests on the White's method

After making the above modification, we used the revised test to do the following test to see how well they are at detecting anomalies.

marginal distribution, not conditional distribution, of returns.

[8] It turns out that, for all of our test results, the difference is really negligible. The reason is that for the test of *daily* returns as we do here, $r_{market,t}$ is very close to zero, so $log(1 + r_{market,t})$ is roughly equal to r_{market}.

1. **We did a "reality-check" on the reality check method**

   We artificially added an anomalous return to each Tuesday[9], and see how large a magnitude should this artificial anomaly be, in order for it to be detected by White's data-snooper. We also record the Sharpe-ratio of these artificial anomalies, and see whether it requires an absurdly high Sharpe ratio for the anomaly to be detected by the White's data-snooper.

   We did each of the following four specification:

   - add a positive return of magnitude $0.0001 * 4^i$ to Tuesday returns;
   - add a negative return of magnitude $0.0001 * 4^i$ to Tuesday returns;
   - multiply Tuesday returns by $(1 + 0.0001 * 4^i)$;
   - multiply Tuesday returns by $(1 - 0.0001 * 4^i)$;

   We change the magnitude of i to control the magnitude of the artificial anomaly and see at what level of magnitude the anomaly becomes detectable.

   We tried two types of artificial anomalies: adding an abnormal return and multiplying an abnormal return. The reason we do both is because these two methods will have very different effects on the Sharpe ratio. While adding an abnormal return does change the Sharpe ratio, multiplying the old return by a constant (larger than 1 or smaller than 1) generally has very little effect on the Sharpe ratio [10].

   We test both a positive abnormal return and a negative abnormal return, because we want to see whether the test can be done *symmetrically*: if there is sufficiently large negative abnormal returns for some particular calendar dates, then among the calendar rules under our considerations (altogether 244 rules), *some* rules should be able to pick it up, just like the case when a sufficiently large positive abnormal return on some particular calendar day pattern should be picked up by some calendar rules under our

[9] We chose Tuesday to do this experiment because while there has been documented abnormal return on Monday and Friday, there seems to be less evidence that Tuesday has any abnormal return

[10] A simple back of the envelope calculation is as followed: suppose the original Sharpe ratio of a strategy is $S_0 = (Er_0 - r_f)/\sqrt{Var(r_0)}$, now define another strategy by $r_1 = kr_0$, and the Sharpe ratio of the new strategy is $S_1 = (Er_1 - r_f)/\sqrt{Var(r_1)} = (kEr_0 - r_f)/k\sqrt{Var(r_0)}$, which is roughly equal to S_0 when r_f is close to zero.

consideration. We hope that the revised test statistics will have good performance in picking up both types of the artificial abnormal returns.

We report the results of our "reality check" on the White method at the end of the paper (see Figure 4.1 to Figure 4.4). Due to the limit of space, we only present the graphs for the case of a 50-50 benchmark. The riskless benchmark and the market benchmark cases will be available from us upon request.

We also tabulate the results for all of our tests (riskless, market and 50-50 benchmarks) at the end of the paper (see Table 4.2 to Table 4.4).

As shown in the tables, (also shown in some of the figures not reported here due to space limit), the "improved" White's test[11] can pick up only one of four types of the artificial anomalies that we put in: the case of a negative return being added to the return on Tuesday, and there only if the magnitude of the abnormal return is large enough[12].

We found that positive and negative artificial abnormal returns do **not** give symmetric results for the improved White's test when we use the market return as benchmark. When adding a negative abnormal return, the test statistic (which is the difference between the best performing strategy and the market strategy returns) is significantly positive; when adding a positive artificial abnormal return, the test statistic is very close to zero, especially for large values of the artificial abnormal return. Consequently, the best performing rule abnormal return and the corresponding p-value are very different for the cases of positive and negative abnormal returns.

Our explanation is as followed: we are always comparing a calendar strategy return to the benchmark, which is defined to be the market strategy returns here. When there is a large negative market return on Tuesdays, the market strategy return is pretty low. Some calendar strategies, for example the strategy that shorts the market every Tuesday and neutral on other days of the week, will significantly outperform the market

[11]By that we mean the White's test after the revision mentioned in Subsection 4.3.1, namely, after the calendar strategy return has been modified to be consistent with the real return of a trading strategy, but still using the market as the benchmark

[12]for our test, with a negative artificial abnormal daily return of -0.0064 or higher, the White test can pick up the anomaly. For the data we used, the daily return has a mean of 0.0007, and a variance 0.0052. The Sharpe ratio for the detectable artificial anomaly is -1.331.

strategy, thus resulting in a high value of the test statistic; when the market return on Tuesday is large and positive, the market strategy is pretty high. In addition, since the average return on other days of the week is positive, the calendar strategy to long Tuesday and neutral other days of the week *under-performs* the market. Any other strategy, because they sometimes involve shorting the market or being neutral on Tuesdays, will *significantly under-perform* the market. Thus, for significantly large positive abnormal returns on Tuesdays, the best calendar strategy (including the strategy dictating longing Tuesday and neutral other days of the week) will be under-performing the market, and the test of an outperforming calendar strategy is likely to give a high p-value[13].

We also found (not surprisingly) that the test using riskless benchmarks did just the opposite to the tests using the market benchmarks, namely, the best performing strategy is outperforming the benchmark too often. These problems can be ameliorated with the test using the 50-50 benchmark strategy. But we don't claim that 50-50 is the best benchmark to use, it is just approximately correct and easy to use for most tests of such purposes[14].

2. **We expand the reality check of White**

White only considers the maximum of a set of trading strategies, we extend that by also considering the kth largest rule, and see whether the top kth rule has significantly outperformed after considering the possibility of data-snooping. The theoretical justification that the White's methodology can be expanded is easily derived using an analogue of White's derivation of the asymptotic behavior of the max and min of K strategies, on p.7 of the paper "A Reality Check for Data Snooping" (proposition 2.2).

[13]indeed, the p-value for adding a positive abnormal return on Tuesday is very close to 1, even (especially) with large abnormal return and large Sharpe ratios

[14]To be more precise, we think the best benchmark in this case is a strategy that "compare oranges with oranges", namely, it should allow us to compare a calendar strategy with a benchmark that have the *same* exposure on the market risk. We could implement such a benchmark a as followed: for each trading strategy, we calculate the proportion of days in the market (for example, January effect will have roughly one twelfth of the days in the market), and then, we could design the benchmark to be one that has the same *proportion* in the market every day, and the rest in the riskless bond. To save space we did not report results based on this minor improvement. It suffice to notice that, since the exposure to market risk changes over all the calendar rules being considered, the market and riskless benchmarks which take extreme positions in the market can be biased, while the 50-50 strategy performs reasonably well.

The intuition is that the asymptotic normality assumption and the continuity theorem (e.g., Billingsley, 1968, theorem 2.2), as assumed by White. The test performed in their paper can be considered a special case of our expanded test (the case where k=1).

The detailed procedure of the expansion test is as followed:

(a) For each of the 244 calendar rules in the universe of simple strategies considered by their paper, we do a bootstrap to generate N iterations of the bootstrapped values. The result is a N by L matrix, where each column represents the N iterations of a particular calendar rule, and each row represents the performances of the L calendar rules for a particular iteration of the bootstrap.

(b) We record the realized mean performance of all L calendar rules, and we rank them from highest to lowest;

(c) We subtract the realized mean performance of each calendar rule from its bootstrapped values to get the de-meaned bootstrap values.

(d) For each iteration of the de-meaned bootstrap (one row of the N by L matrix), we rank the bootstrapped values of all calendar rules, from highest to lowest, thus we have N ranked sequences;

(e) To test whether the kth highest performing rule is outperforming, we pick the kth elements from each of the N iterations of the de-meaned bootstrap to get the *bootstrapped distribution of the kth largest rule*, and then compare the realized performance of the kth highest rule (this is the kth element of the ranked sequence of realized performance) with this distribution to get the empirical p-value of the hypothesis that the top kth calendar rule significantly outperformed.

Since from the previous test we know that the log return and simple return cases perform relatively similarly, for this test we will only do the log return case.

The result is as shown in Figure 4.5. We can see that the p-values associated with different rules are not very stable across different values of k. To make sure that our experiment is not biased by our choice of the bootstrap size, we chosen different size of bootstrap at N=200 and N=500 (the later is the size that Sullivan et al used in their paper), and find similar patterns.

We also did a "reality check" by using the data sequence with artificially generated anomaly (adding abnormal return of 0.1024 to dow=2, as we did before) with the size of bootstrap equal to 200 (see Figure 4.6). This significantly changed the pattern of the p-value distribution across different rules.

3. **We study the correlation among the L calendar strategies by using the bootstrapped values**

   As before, we record the performance of all the calendar rules in each of the iterations. Then, we construct the bootstrapped covariance among these trading rules as:

   $cov(f(K), f(J)) = \frac{1}{N}\sum_{i=1}^{N}(\hat{f}_{K,i} - \bar{f}_K)(\hat{f}_{J,i} - \bar{f}_J), \forall K, J,$

   where $\bar{f}_K = \frac{1}{N}\sum_{i=1}^{N}\hat{f}_{K,i}, \forall K,$

   is the average return for calendar rule K, and $\hat{f}_{K,i}$ is the return of calendar rule K, under the ith bootstrap iteration.

   The correlation among calendar rules can similarly be calculated.

   We report a histogram of the correlation coefficients among all calendar rules(see Figure 4.7). We also have a plot recording the correlation coefficients in a 3 dimension fashion, which we did not enclose, in order to limit the length of this paper. The plots seem to indicate that there is a pattern in correlation among the calendar rules, in particular, these calendar rules are definitely *not* iid distributed.

4.3.2 Order Statistics

One important reason why the reality check methods as in the White's papers is used is that, it is hard to get the closed-form distribution of the extreme value of a vector of correlated normals for the general case. It is interesting to see if we could derive some *approximate close form distribution* of these extreme values and make statistical inferences based on those. In this subsection, we try to do that.

The reason explicit close form distribution of the maximum return calendar rules is hard to derive is, these calendar rules are correlated. However, if we make a simplifying assumption that the calendar rule returns are equi-correlated, namely any two different calendar rule will have a common correlation coefficient ρ, then, we can derive some closed form representation of the distribution of the extreme value of this vector of correlated normals. Based on that, we plot the probability distribution functions for different values of the correlation coefficient ρ and different values of the number of rules considered, N. Then, we tabulate the moments and the critical values of the distribution of the extreme value, for different values of ρ and N.

As a direct application of our approximate close form characterization, we use the distributions thus derived to test the hypothesis that the best performing calendar rule significantly outperforms, and compare our results with those of the Sullivan et al paper.

closed form representation of order statistic in the case of equicorrelatation

Consider n *dependent* normal random variables $Y_1, Y2, ..., Y_n$, where $corr(Y_i, Y_j) = \rho$ if $i \neq j$. Without loss of generality, we could focus on the case where the $Y_i's$ are standard normal variables.

Since

$$0 \leq Var(\sum_{i=1}^{n} Y_i) = nVarY_i + n(n-1)cov(Y_i, Y_j), i \neq j,$$

it follows that the common correlation coefficient ρ must satisfy $\rho \geq -1/(n-1)$.

It is also easy to verify that when $\rho \geq 0$, the Y_i may be thought of being generated from *independent* normal random variables X_i, i=0,1, ... n, as follows:

$$Y_i = \rho^{\frac{1}{2}} X_0 + (1-\rho)^{\frac{1}{2}} X_i,$$

i=1, 2, ...n.

Following the discussion in David (1981)[15], for $\rho \geq 0$,

$Pr\{Y \leq y\} = Pr\{\sum_{i=1}^{n} a_i X_{i:n} \leq -(\frac{\rho}{1-\rho})^{\frac{1}{2}} (\sum a_i) X_0 + \frac{y}{(1-\rho)^{\frac{1}{2}}}\}$

where $Y = \sum_{i=1}^{n} a_i Y_{i:n}$.

Now let

[15] page 108 of David (1981)

$a_i = 0$ for $i < n$;

and $a_n = 1$.

We have,

$Pr\{Y_{n:n} \le y\} = Pr\{X_{n:n} \le -(\frac{\rho}{1-\rho})^{\frac{1}{2}} X_0 + \frac{y}{(1-\rho)^{\frac{1}{2}}}\}$

$= \int_{-\infty}^{\infty} Pr\{X_{n:n} \le -(\frac{\rho}{1-\rho})^{\frac{1}{2}} x + \frac{y}{(1-\rho)^{\frac{1}{2}}}\} d\Phi(x)$

$= \int_{-\infty}^{\infty} \Phi^n[-(\frac{\rho}{1-\rho})^{\frac{1}{2}} x + \frac{y}{(1-\rho)^{\frac{1}{2}}}] d\Phi(x)$

This can be computed for given value of ρ and y.

The probability density function can be computed as:

$f(y) = \frac{d\Pr\{Y_{n:n} \le y\}}{dy}$

$= \int_{-\infty}^{\infty} n\Phi^{n-1}[-(\frac{\rho}{1-\rho})^{\frac{1}{2}} x + \frac{y}{(1-\rho)^{\frac{1}{2}}}]\phi[-(\frac{\rho}{1-\rho})^{\frac{1}{2}} x + \frac{y}{(1-\rho)^{\frac{1}{2}}}]\frac{1}{(1-\rho)^{\frac{1}{2}}} d\Phi(x)$

In Figure 4.8 to Figure 4.17, we plot the graphs of the probability density functions for different values of ρ and N. We also present tables showing the different percentiles of the cumulative density functions for different parameters of ρ and N.

insert tables at end (Andy, could you please decide what to put in the table, and include the table?).

application of the order statistics rule, and comparison with the White test

In this subsection, we apply the result to the testing of the hypothesis that the best performing calendar rule does not outperform. To get easy comparison between our results and those in Sullivan et. al. (1998)), we do the test in the fashion that is similar to that in their paper.

In the application that we do, the distribution of $Y_i's$ are not standard normal, therefore we first do a transformation to get the standard normal by considering

$$Y' = \frac{Y_i - mean(Y_i)}{std(Y_i)}$$

.

In testing the performance of the calendar rules, we do the following:

1. We first get different realizations of the return of each calendar rule. We could do it in two ways: a stationary bootstrap following Politis and Romano, or an approximation

using the time series return of a particular calendar rule. We will discuss the two approaches in more detail later.

2. Based on the realizations of returns of each calendar rule, we estimate the correlation matrix of the universe of calendar rules (more on this later).

3. We take the *average* of the correlation between different calendar rules to be the estimate of ρ. So $\hat{\rho}$ is given by:

   $\hat{\rho} = \frac{1}{243*242/2} \sum_{i,j \neq i} \rho_{i,j}$,

   where $\rho_{i,j}$ is the i-jth *off-diagonal* element of the correlation matrix calculated in the first step[16]

4. As we discussed above, we need to convert the trading rule returns to standard normal to be able to use our closed form result. We do that by subtracting the mean and then dividing by standard deviation. The null hypothesis has that the mean of these trading strategies are zero, and we estimate the *average* standard deviation σ by

   $\hat{\sigma} = \frac{1}{L} \sum_{l=1}^{L} \sigma_l$,

   where σ_l is the standard deviation of the lth trading strategy.

5. We calculate the cumulative distribution function for $Y_{n:n}$: for $y = -5\hat{\sigma}$ to $5\hat{\sigma}$, with increment of $0.1\hat{\sigma}$, we calculate the probability that $Y_{n:n}$ is less than y, and plot $Pr\{Y_{n:n} \leq y\}$ as a function of y.

6. we test the hypothesis that "of all the calendar rules considered, the best performing rule does not outperform" by comparing the best performing rule's return with the 0.95 percentile (or 0.99 percentile) of the critical value that we obtained from last step. We could also directly compute the p-vale of the hypothesis by finding out the value of $1 - Pr\{Y_{n:n} \leq y^*\}$, where y^* is the best performing rule return over the sample period.

There are two ways to get the (random) realizations of a particular calendar rule return: bootstrap and approximation using time series.

[16]The diagonal elements of the correlation matrix are all 1, therefore including them will bias the estimated "average" correlation upwards.

approach one: using the PR bootstrap

In this approach, we first do a PR bootstrap that is similar to the one used in the Sullivan et al approach. After we get the correlation matrix from bootstrapping, we follow the step outlined above to get the test statistics.

Our test results are shown graphically in Figure 4.18 to Figure 4.20. On each plot, we first plot the cumulative distribution function of the maximum of trading rules, *under the assumption that these rules have the same correlation*, for the given parameters of ρ and σ that we estimated from the data using the PR bootstrap. We also draw as a dotted line the 95% confidence interval line for the one sided test of the hypothesis that the best performing rule does not outperform. Anything above the dotted line will be in the 5% tail of the distribution, which corresponds to the value of the best model return that will be occurring less than 5% of the time. We indicate on the plot by a "*" the point that corresponds to our best realized return. If the "*" is beneath the dotted line corresponding to the 95% confidence level, then our test does not reject the null hypothesis that the best rule is *not* over-performing. Otherwise, we reject the null.

To easily compare our results with those of Sullivan et al, we did the test in a way that is consistent to the revised tests of Sullivan et al, based on the market benchmark and log return[17]. We performed our tests under three scenario: using the real stock price (Figure 4.18), adding an abnormal return of magnitude 0.0016 on Tuesday (Figure 4.19), and adding an abnormal return of magnitude 0.0064 on Tuesday (Figure 4.20)[18].

The test results obtained are pretty close to those in Table 4.3. The test on the raw data does not reject the null, same as the test on the data with a "small" artificially added anomaly (of magnitude 0.0016 on Tuesday, or a Sharpe ratio of -0.443), while for a sufficiently large artificial anomaly (of magnitude 0.0064 on Tuesday, or a Sharpe ratio of -1.331), both our test and the revised White test reject the null.

In conclusion, the closed form distribution under equi-correlation assumption does not seem to radically change the result, at least not for this application on the calendar rules.

approach two: using the time series estimate

[17]We did revise the trading strategy return as we discussed in the last section, so as to make it more sensible in finance and easily comparable with results in Table 4.3.

[18]We choose these magnitude to be consistent with the "reality check" tests on the White method, as done in the last section

One reason why we would want to use the equi-correlation approximation is to avoid the Politis-Romano bootstrap which is computation intensive. The above approach, while interesting, will still need the PR bootstrap step to get the estimation of ρ and σ, thus it is useful to see how we could do without the PR bootstrap.

One implicit assumption about the calendar rule tests is that the return series is stationary. If we add the assumption that under the null hypothesis, the abnormal returns of calendar rules are independent with zero mean, then we could estimate the ρ and σ using the time series data.

Suppose the ith calendar rule abnormal return on the tth period is x_{it} with $x_{it} \sim NID(0, \sigma_i^2)$, and $cor(x_{it}, x_{jt}) = cor(x_{is}, x_{js}) = \rho_{ij}, \forall t, s$, but $cor(x_{it}, x_{js}) = 0, \forall t \neq s$ (otherwise there will be return predictability from the past performance of one calendar rule).

Since return of a calendar rule is iid, the *average* return of a calendar rule $\bar{x}_i$ (which is the basis of our test) will satisfy:

$$\bar{x}_i := \frac{1}{T}\sum_{t=1}^{T} x_{it} \sim N(0, \frac{1}{T}\sigma_i^2)$$

$cov(\bar{x}_i, \bar{x}_j) = cov(\frac{1}{T}\sum_{t=1}^{T} x_{it}, \frac{1}{T}\sum_{t=1}^{T} x_{it})$
$= \frac{1}{T}\rho_{ij}\sigma_i\sigma_j,$
and,
$cor(\bar{x}_i, \bar{x}_j) = \frac{cov(\bar{x}_i, \bar{x}_j)}{var(\bar{x}_i)var(\bar{x}_j)}$
$= \rho_{ij}.$

Therefore, we could estimate the correlation between calendar rules by the time series correlation of these calendar rules.

This will make the computation much easier, and our results show that there is generally no loss of power in detecting the anomalies.

We present at the end of the paper a table (Table 4.5) which has the comparison of the test using the bootstrapped estimation of ρ and σ and the test using the time series estimation of ρ and σ. There we noted that the bootstrapped estimation and the time series gives pretty close estimations of both the ρ and the σ: the time series estimation for these variables are slightly larger than the estimations obtained through the Politis Romano bootstrap, but the difference is negligible. As a consequence, the two methods have almost the same power in detecting artificial anomalies. When the artificial anomaly is around 0.0025 for Tuesdays, both methods (for the 800 observations case) will be able to reject the null at the 5% significance level, but both will not reject the null at the 1% significance level. For larger values of the artificial anomaly, both tests will be able to reject the null hypothesis at the 1% significance level, while for smaller magnitude of the artificial anomaly both test cannot reject the null at the 5% significance level.

We also present the result of the time series estimation using a much longer data set (8000 days instead of 800 days, this corresponds to roughly the data from 1962 to 1996), and we can see that with more data, the standard deviation of each trading rule returns actually increases[19], while the correlation between calendar rules does not change much (they increase slightly, but again the increase is almost negligible). With 1% significance level, the test using this longer data set can reject the null hypothesis under almost all the magnitude of the artificial anomaly.

These results show that the test using time series estimation will perform reasonably well (compared to the Politis Romano bootstrapping) when there is a sufficient long time series. This gives an alternative to the White's methodology which relies on the computation intensive bootstrapping method.

4.3.3 Bayesian Analysis

In empirical tests against data-mining, we need to consider the trade-off between the size and the power of the test, i.e., we want to balance the type I and type II error. In this section, we want to derive a way to gauge the tradeoff between these two types of errors and optimize over the two types of errors.

Theoretical derivation

In this subsection we develop a systematic approach in Bayesian analysis, to explicitly consider the trade-off between the type I error (rejecting true hypothesis) and the type II error

[19] We don't know for sure what might have caused this result, one conjecture is that these calendar rule returns shows some non-stationarity over time, thus the longer the time periods, the larger the variation of a particular calendar rule return

(not rejecting the false hypothesis). We first gives out theoretical results, and then use these in the empirical test about calendar rule anomalies. We consider several cases:

case I: a point null and a mixed posterior distribution

Suppose we have a point null, $H_0 : X = x_0$. If our posterior distribution is continuous, it will have a zero mass on $X = x_0$, thus the possibility of H_0 being true will be zero, and therefore we don't need to consider any Type I loss at all. To avoid such degenerate case, we assume that the posterior distribution of the parameter that we are interested in testing has a mixed posterior distribution:

Assume the posterior distribution P of the parameter is:

$Pr(H_0 : X = x_0) = p$,

$Pr(H_1) = 1 - p$,

where H_1: not H_0

And $\tilde{X}|H_1 \sim N(x_1, \sigma^2)$.

A *loss function* in Bayesian analysis is defined as the negative utility resulting from making a wrong decision. Here, in this context, the total decision space D consists of two possible decisions:

d_1: not reject H_0;

d_2: reject H_0.

Correspondingly, two types of errors can be made: type I and type II. We assume the following loss functions[20]:

L($d_1|H_1$ is true, the real parameter value is x) $= (x - x_0)^2$;

L($d_2|H_0$ is true) = K;

and

L($d_1|H_0$ is true) = L($d_2|H_1$ is true) = 0.

Here, K is a parameter that allows us to control for the relative importance of the two types of errors[21].

Let P be the posterior probability distribution of the parameter W. For any decision

[20]There is no necessity in assuming a quadratic loss, although it is probably the most widely used, because of its tractability.

[21]$K = \sigma^2 + (x_1 - x_0)^2$ corresponds to the case that we assign equal importance to the two types of loss. This can be shown as followed: the *average* loss incurred when H_1 is true and decision d_1 is made is: E$[(X - x_0)^2|H_1$ is true$]$=$E[(x - x_1)^2 + 2(x - x_1)(x_1 - x_0) + (x_1 - x_0)^2|H_1] = \sigma^2 + (x_1 - x_0)^2$, thus by setting $K = \sigma^2 + (x_1 - x_0)^2$ we are giving equal importance to the two types of losses.

d, the *risk* for making such decision is the expected loss, given the posterior distribution. Therefore, the risk for making the two types of decisions are:

$\rho(P,d_1) = (1-p)\int_{-\infty}^{\infty}(\sigma t + x_1 - x_0)^2\phi(t)dt + 0$

$= (1-p)(\sigma^2 + (x_1 - x_0)^2)$,

and,

$\rho(P,d_2) = pK + 0$

The *Bayes risk*, $\rho^*(P)$, is defined to be the greatest lower bound for the risk $\rho(R,d)$ for all the feasible decisions in D, i.e., we have

$$\rho^*(P) = \inf_{d\in D}\rho(P,d)$$

.

Correspondingly, the decision d^* whose risk is equal to the Bayes risk is called a *Bayes Decision against the distribution* P[22].

In our question, the Bayes risk is:

$$\rho^*(P) = min((1-p)(\sigma^2 + (x_1 - x_0)^2, pK),$$

and the decision corresponding to the Bayes risk is the Bayes decision.

case II: a *generalized point null* and a continuous posterior distribution

Given that most of the time the posterior probability distribution is continuous rather than mixed, we want to generalize the above method to deal with such a case. Instead of assuming a point null, we assume a small region around the point null. We illustrate our idea using the following simple example.

Posterior probability distribution:

$X \sim N(x_0, \sigma^2)$;

$H_0 : |X - x_0| < B\sigma$,

H_1: otherwise

[22]For some problem there could be multiple Bayesian decisions; for some others, there could be no Bayesian decision.

Note that for simplicity of treatment we have made the mean of the distribution the same as the center point of the region that the null hypothesis would be true (corresponding to the case that $x_1 = x_0$ in the first case). This could be easily generalized.

B is a small and positive number. As B approaches zero, we are back to the original point null.

Two decisions can be made:

d_1: not reject H_0;

d_2: reject H_0.

The loss function can similarly be defined as:

L($d_1|H_1$ is true, the real parameter value is $x > x_0 + B\sigma$) $= K_1(x - x_0)^2$,

L($d_1|H_1$ is true, the real parameter value is $x < x_0 - B\sigma$) $= K_2(x - x_0)^2$,

$L(d_2|H_0$ is true$) = 1$,

otherwise L = 0.

Here $\{K_1, K_2\}$ control for the relative importance of the errors, and K_1, K_2 can potentially be different. Intuitively, $K_1 > K_2$ means the researcher is more concerned with the loss incurred when the real parameter is too large (thus he assigns a bigger loss to it).

The risks associated with the two decisions are:

$\rho(P, d_1) = \int_{-\infty}^{-B\sigma} K_2(x - x_0)^2 \phi \frac{x-x_0}{\sigma} d\frac{x-x_0}{\sigma} + \int_{B\sigma}^{\infty} K_1(x - x_0)^2 \phi \frac{x-x_0}{\sigma} d\frac{x-x_0}{\sigma}$, and,

$\rho(P, d_2) = \int_{-B\sigma}^{B\sigma} \phi \frac{x-x_0}{\sigma} d\frac{x-x_0}{\sigma} + 0$

In this context, the Bayes risk is:

$$\rho^*(P) = min(\rho(P, d_1), \rho(P, d_2))$$

,

and the decision corresponding to the Bayes risk is the Bayes decision.

case III: a one sided test and a continuous posterior distribution

Consider the following example:

The posterior distribution is $X \sim N(x_0, \sigma^2)$,

The two hypothesis are:

$H_0 : X < x_1$,

$H_1 : X \geq x_1$.

We define the decision as:

d_1: not reject H_0;

d_2: reject H_0.

The loss function can be defined as:

L($d_1|H_1$ is true, the real parameter value is x) $= (x - x_1)^2$,

L($d_2|H_0$ is true, the real parameter value is x) $= K(x - x_1)^2$,

otherwise L $= 0$.

The risks associated with the two decisions are:

$\rho(P, d_1) = \int_{x_1}^{\infty}(x - x_1)^2\phi\frac{x-x_0}{\sigma}d\frac{x-x_0}{\sigma} + 0,$

and,

$\rho(P, d_2) = \int_{-\infty}^{x_1} K(x - x_1)^2\phi\frac{x-x_0}{\sigma}d\frac{x-x_0}{\sigma} + 0$

In this context, the Bayes risk is:

$$\rho^*(P) = min(\rho(P, d_1), \rho(P, d_2)),$$

and the decision corresponding to the Bayes risk is the Bayes decision.

Empirical Test

In order to apply the results we have just derived to the calendar rules study, and to really see the trade-off between Type I and Type II errors, we need to carefully define the null and alternative hypothesis, so as to determine the appropriate size and power of the test.

One natural null hypothesis is that all the calendar rules abnormal returns will have a distribution with mean zero. The simplest case is that they all have a normal distribution with mean zero and same variance. Furthermore, we could allow these trading rule abnormal returns to be equi-correlated[23]. We propose an alternative hypothesis that seems natural, too: that, comparing with the null, *one* strategy might have a different and positive mean,

[23]Granted this is still a restrictive assumption in our applications. There are, however, some natural applications, where the assumption that all trading strategies returns are correlated to some extent is not very far from the truth. For example, in the mutual fund studies, one might assume that there are some systematic components of the mutual fund returns, so that the returns are naturally correlated.

the rest of L-1 of the calendar rules will have the same return distribution as in the null. More explicitly, we make the following assumption to the structure of the hypothesis testing:

Assumptions:

We assume that all L calendar rules have the same standard deviation. In addition, out of the L calendar rules that we are interested in, L-1 of them will have mean returns that all follow normal distribution with mean zero. These L-1 rules are equi-correlated with a common correlation coefficient ρ (a special case could be that $\rho = 0$). However, *one* of the L rules, rule L^*, might have a mean return that could be non-zero. We don't know which one of these L rules is L^*, and we don't know what the true mean of L^* is, but we know a prior distribution of the mean of L^*, which we denote as $\xi(\mu_{L^*})$. (Later we will assume an uninformative prior). We further assume that rule L^* is independent with the other L-1 rules[24].

With the above assumptions about the structure, we define the two hypothesis that we are testing as:

H_0: $\mu_{L^*} \leq 0$

and

H_1: $\mu_{L^*} > 0$

In what follows we do three things:

1. We first characterize the conditional distribution of the data for a given value of the μ_{L^*}.

2. Using the prior distribution of μ_{L^*} and the realization of μ_{L^*} in the data, which we call μ^*, we characterize the posterior distribution of μ_{L^*}.

3. We do the test based on the methods discussed in the theoretical part.

conditional distribution of the data for given μ^*

We first characterize the conditional distribution of the maximum abnormal returns among L rules, conditional on the value of μ_{L^*}.

[24]we assume this for tractability: with this assumption we don't need to worry about the correlation between the maximum of the first L-1 rules return and the return from L^*, and this allows us to characterize the test statistics in closed form. If we don't care about getting close form result, we could drop such assumption and use Monte Carlo methods to simulate the distribution when the rule L^* is correlated with the other rules.

Denote by Y_1 the best abnormal return among the L-1 rules that all have mean zero, and by Y_2 the abnormal return of rule L^*. Let

$Y := max(Y_1, Y_2)$.

Then we can derive the cumulative density function:

$Pr(Y \leq y) = Pr(Y_1 \leq y)Pr(Y_2 \leq y)$

(by independency)

$= Pr(\frac{Y_1}{\sigma} \leq \frac{y}{\sigma})Pr(\frac{Y_2}{\sigma} \leq \frac{y}{\sigma})$

(normalization to standard deviation of 1)

$= \int_{-\infty}^{\frac{y}{\sigma}} f(x)dx \int_{-\infty}^{\frac{y}{\sigma}} \phi(x - \mu_{L^*})dx$

where, f(x) is the probability distribution function of the maximum of (L-1) equi-correlated rules, which we have derived earlier in the order statistics section.

Given the cumulative density function, we could derive the conditional probability density function of the data:

$f(y|\mu_{L^*}) = \frac{dPr(Y \leq y)}{dy}$

$= \frac{1}{\sigma} \int_{-\infty}^{\frac{y}{\sigma}} f(x)dx\phi(\frac{y}{\sigma} - \mu_{L^*}) + \frac{1}{\sigma} f(\frac{y}{\sigma}) \int_{-\infty}^{\frac{y}{\sigma}} \phi(x - \mu_{L^*})dx$

In what follows, to save space, we define $z = \frac{y}{\sigma}$, and sometimes call it "data".

the posterior distribution of μ_{L^*}

With the conditional distribution of data and the prior distribution of μ_{L^*}, we can do the Bayesian updating to get the posterior:

$\Xi(\mu_{L^*}|data) = \frac{f(y|\mu_{L^*})\xi(\mu_{L^*})}{\int_{-\infty}^{\infty} f(y|\mu)\xi(\mu)d\mu}$

We assume an uninformative prior: $\xi(x) = c, \forall x$, thus we have:

$\Xi(\mu_{L^*}|data) = \frac{f(y|\mu_{L^*})}{\int_{-\infty}^{\infty} f(y|\mu)d\mu}$

$= \frac{\int_{-\infty}^{z} f(x)dx\phi(z-\mu_{L^*})+f(z)\int_{-\infty}^{z} \phi(x-\mu_{L^*})dx}{\int_{-\infty}^{\infty}[\int_{-\infty}^{z} f(x)dx\phi(z-\mu)+f(z)\int_{-\infty}^{z} \phi(x-\mu)dx]d\mu}$

This is the posterior distribution of the μ_{L^*}, given data.

applying the theoretical result

Now, we could apply the derivation in the theoretical part, case III, to derive the Bayes risk and the Bayes decision.

We use a quadratic loss function:

L($d_1|H_1$ is true and the real value of μ is μ_{L^*}) $= \mu_{L^*}^2$;

and,

L($d_2|H_0$ is true and the real value of μ is μ_{L^*}) $= K\mu_{L^*}^2$.

where

d_1: not reject H_0

and

d_2: reject H_0.

As in the theoretical derivation, K measures the relative importance of the two types of error. A high value of K means that in the researcher's mind the Type I error (rejecting a true null) is more costly than the Type II error (accepting a false null).

Then,

$\rho(P, d_2) = K \int_{-\infty}^{0} \mu^2 \Xi(\mu|data) d\mu + 0,$

and

$\rho(P, d_1) = \int_{0}^{+\infty} \mu^2 \Xi(\mu|data) d\mu + 0.$

In what follows, we characterize the decision over rejecting and not rejecting the null by a simple metric, which we would call the "indifference K". We define this to be the value of the K that will make the researcher indifferent between the two decisions d_1 and d_2, i.e., the K that makes $\rho(P, d_2) = \rho(P, d_1)$. The researcher's decision over rejecting and not rejecting the null can be conveniently characterized by this metric K, because as K becomes higher and higher, it will require the researcher to have a higher and higher fear about the loss on the type I error in order to refrain from rejecting the null. In other words, for a fixed concern on the type I error, as K increases, the optimal choice of the researcher will be more and more likely to be d_2 (choose to reject the null).

Since K will satisfy

$\rho(P, d_2) = \rho(P, d_1),$

we have,

$K \int_{-\infty}^{0} \mu^2 \Xi(\mu|data) d\mu = \int_{0}^{+\infty} \mu^2 \Xi(\mu|data) d\mu,$

re-organize we have,

$$K = \frac{\int_{0}^{+\infty} \mu^2 \Xi(\mu|data) d\mu}{\int_{-\infty}^{0} \mu^2 \Xi(\mu|data) d\mu}$$

$$= \frac{\int_{0}^{+\infty} \mu^2 \phi(z-\mu) d\mu \int_{-\infty}^{z} f(x)dx + f(z) \int_{0}^{+\infty} \int_{-\infty}^{z} \mu^2 \phi(x-\mu) dx d\mu}{\int_{-\infty}^{0} \mu^2 \phi(z-\mu) d\mu \int_{-\infty}^{z} f(x)dx + f(z) \int_{-\infty}^{0} \int_{-\infty}^{z} \mu^2 \phi(x-\mu) dx d\mu},$$

To calculate the indifference K value for our calendar rule test, we follow three steps:

1. either through time series or through bootstrap, get the estimation of the best calendar

rule abnormal return, y, as well as the estimation of average standard deviation among the calendar rule abnormal returns, σ and their average correlation coefficient, ρ.

2. using the parameter estimates of {y, σ,ρ}, and a flat prior, derive the posterior distribution of μ_{L^*}.

3. plug in the above formulae to get the value of K in the test. treatment.

We implemented the methods that we discussed above. Furthermore, we did some experiment with an artificially generated anomaly, and see how the indifference K value changes with the magnitude of the artificial anomaly. We report the result of the experiment in Table 4.6. There we could see that, as the magnitude of the artificial anomaly increases, the indifference K value increases very quickly, indicating more and more confidence in reporting an anomaly.

4.3.4 Empirical Assessment of the Goodness-of-Fit of Conditional Predictive Models

There are basically two families of the predictive asset pricing models: ones that rely on the marginal distribution, and ones that rely on the conditional distribution, given a set of prediction variables. The calendar effect models belong to the first family of models, and models belonging to the second family include, among others, accounting information based models and macroeconomic variable based models.

Ou and Penman (1989) performed a financial statement analysis that combined a large set of financial statement items into one summary measure and used it to form arbitrage portfolio in stocks. They found that the two-year holding-period return to the portfolio is about 7%, even after adjusting for size. The effect cannot be attributed to other firm risk characteristics. Recent researches by Lev and Thiagarajan (1993), Abarbanell and Bushee (1997, 1998) and Kraft (1999) all tried to explain the effect of accounting variables in predicting the stock return.

Papers incorporating the macroeconomic variables into stock return estimation include Chen, Roll and Ross (1986) and Chen (1991).

One potential problem with the conditional predictive models, as many researchers have realized, is the data-snooping issue. Researchers frequently justify their choice of regressors as "guided by economic theory *and empirical evidence*". So, to certain extent their choices of regressors are affected by previous researches and there is a potential for data mining. Recently, Foster et al (1997) documented the potential problem of "choosing the best k out of m regressors". They studied the setting that the researcher can chooses k out of m candidate regressors to perform the regression. Since there are $\binom{m}{k}$ ways to do the regression, if the researcher reports the best measure of the goodness of fit among the $\binom{m}{k}$ measures, we should not do a test *as if* he only had one fixed set of the k parameter at hand and just did his experiment once. Indeed, such "naive" test statistics will severely underestimate the possibility of data-mining, and overestimate the likelihood of an anomaly. In view of this effect, Foster et al provided simple procedure to adjust the critical R^2 values to account for the data-snooping effect. They went on to use their methodology to re-examine several studies by estimating "the 95 percent confidence limit for m^*", which is "the minimum number of regressors required to achieve an R-squared at least as high as that reported in a number of recent studies. If the researcher had access to more than the listed m^* regressors (either directly, or through the reading of other researcher's papers), then their reported R-squared would occur at least five percent of the time through random chance, according to the test used[25]". They found that, after considering the data-snooping effect, while some effects seem to remain significant, some others are less anomalous.

This subsection discusses the implementation of the Foster et al methodology into real data. We assume the following setting for our basic test: the researcher has at hand some M candidate regressors which he hope could be used to explain the variation in a dependent variable y. Granted many if not all of the candidate regressors might have zero prediction power. We want to appraise, in such a extent, what is the effect of data-mining if the researcher reports his "best fit" in a k regressor model out of the $\binom{m}{k}$ alternatives.

We discuss the detail of the implementation with an example below. In implementing the test, we extend the Foster et al methodology in two ways:

1. First, we expand the analysis to include other measures of goodness of fit.

   We have included the following measures:

   (a) maximum absolute t value among all non-constant regressors, maxt;

[25] Quoted from the heading of Table IV in their paper

(b) the sum of squares of t values for all non-constant regressors, ssqrt;

(c) the number and proportion of t values that are significant at 95% level[26], nsgt and propt, respectively;

(d) the R^2.

What we reported are some simple goodness-of-fit measures. This can be generalized to a large extent by including functions on these goodness-of-fit measures, in the form of $f(maxt, ssqrt, nsgt, R^2)$. Or, more generally, we could create metrics that evaluate the welfare impact of these apparent anomalies on an economic agent. For example, the Sharpe ratio of the strategy exploiting this anomaly, or, as in a recent paper by Brennan and Xia (1999), the welfare effect on the agent in a dynamic setting with learning.

2. Second, we adopt a methodology that combines the usage of real world data and simulation. In Foster et al, a *simple* Monte Carlo simulation is discussed, where, to gauge the effect of data-mining, a randomly generated dependent variable is regressed on five out of fifty independently generated random variables. They use this simulation to show the effect that, under the null hypothesis of no dependence, what the empirical distribution of the maximum R^2 would look like. Our design allows us to do more than that. In particular, to generate an empirical distribution of the test statistics *under the Null Hypothesis that all regressors are irrelevant*, we propose the following:

   (a) randomly generate a vector of dependent variable that is independent with the fixed set of regressors.

   (b) use the randomly generated dependent variable *and the real data on the regressors*, form all possible combination of k out of m regressors to do the regression, and record the maximum test statistics among all these combinations.

   (c) do N iterations of the steps 1 and 2 above to generate the empirical distribution of the best performing measure of the goodness of fit, under the null hypothesis that there is no explanatory power from the regressors used.

[26]For a fixed number k, these two measure have a one-on-one relationship

The approach that we adopt is an improvement over the Foster et al method, because by design we allow for structure in regressors, in particular, dependence among the regressors and non-normality among the regressors, to name just a few. The real data almost surely have some structure that differs from the set of randomly generated normal regressors that Foster et al proposed, and this is likely to affect the test statistics[27].

example

As an example showing the implementation of Foster et al method in the real data, we consider the prediction power of a set of (mostly) economic and accounting variables on the future stock returns. As typical in any exercise of predictability, the information set that we could condition on can be categorized into two classes:

1. stock specific information

   such as lagged stock return and volume information, and accounting variables.

2. general, market level information

   such as the market return (proxied by S&P 500 return or CRSP value weighted index return), the general economic indicators.

In designing our example, we choose both types of information. In doing this, no attempt is made to make our list exhaustive, for two reasons: first, the limit of computing power; second, rather than to identify all the possible anomalies in conditional prediction models, the purpose of this subsection is to discuss the real world implementation of methodologies designed to gauge the effect of data-snooping on the result through examples. Any interested researchers will be able to follow these examples and choose their own set of possible information to condition on in a regression, and thus it is not our job to do that.

[27]Our expansion should not be viewed as a criticism to the Foster et al approach. In fact they are discussing the methods in a more abstract way, and they did not have any particular data set that has a universe of all possible variables to be used in the data-mining exercise. Therefore they do not have real data on which to discuss the fine structure while we do have in a real world application. The theoretical discussion has its own merit: by not being involved into the fine structure of data, they could talk generally about the effect of increasing the number of regressors, m, on the result (as they discussed in their evaluation of previous studies, in table IV). Our method will not be able to directly address that question, since we do not know the new data structure when additional regressors are added in. We could, however, *approximate* the new data structure. We discuss that as the end of this subsection.

Choice of variables

We choose our accounting variables mostly following Lev and Thiagarajan (1993) and Kraft (1999). We add size and book-to-market following many studies in finance documenting the significance of these variables. In addition, we add lagged turnover, lagged return, and return seven periods ago (to capture the momentum effect) as other stock specific variables.

The stock specific variables included are[28]:

1. Inventory = change in sales (2)[29] - change in inventory (38)

2. Accounts receivable = change in sales - change in accounts receivable

3. Gross margin = change in sales - change in cost of goods (30)

4. Change in earnings = [earnings per share (t) - earnings per share (t-1)] / price (t) (14)

5. market cap

6. BE/ME = Stockholder's Equity (60) / Market Cap

7. turnover

8. return

9. return six periods before

We choose our economic variables mostly following Chen, Roll and Ross (1986) and Chen (1991), We also added some more.

The market-wide variables included are:

1. M1_chg: change in M1

2. retail_chg: change in total retail sales

3. housing_chg: change in new private housing units started

4. wage_chg: change in weekly wages for workers in private non-farming economy in US

[28]All variables are one period lagged variables comparing to the dependent variable, as in any predictive model

[29]Numbers in parenthesis are identification codes in Quarterly COMPUSTAT

5. CPI_chg: change in CPI
6. indp_chg: change in US industrial production
7. inddur_chg: change in US industrial production in durable goods
8. defauprm: default premium (the difference in average returns between Baa and Aaa grade corporate bond)
9. termprm: term premium (the difference between the average US 10 year treasury bond return and th average US 3 month treasury bill return)
10. GDP_chg: change in GDP
11. CRDINT_chg: change in the total of US credit market instruments

Most of the accounting variables came from COMPUSTAT, several stock specific and market wide indicators came from CRSP, and most of the economic variables came from DATASTREAM. The common available data period from all sources goes from the second quarter of 1989 to the last month of 1998[30].

Sometimes we need to convert quarterly information into monthly. In doing so, we have taken special care regarding the "adaptive ness" of data, namely, we are not using information only available in the future to "predict" current returns. For example, we assume that the

[30]We have decided to use monthly data over other frequencies.
There are two reasons why we don't choose a higher frequency of data:

1. One reason is data availability. While CRSP data are available at weekly or even daily frequencies, the COMPUSTAT and many of the macro variables are only available at a quarterly or annual frequency. We strike a compromise by choosing the monthly.
2. Another reason, higher frequency data will involve issues related with market microstructure, such as bid-ask bounce, and the exact dates of the information announcement might be after the end of that period, so there could be a problem of measurability by using information only available at a later date to predict current stock returns. We think by using monthly frequencies, we could substantially reduce these problems.

The major reason why we don't use a lower frequency is because of concern with the number of observations we can get. We have taken the intersection of the available data periods from CRSP and COMPUSTAT. COMPUSTAT data goes from 1989 to 1999, CRSP data goes from 1962 to 1998 for most stocks. That leaves about 10 years of common date periods. If we use quarterly or even annually data that will give us very few observations. Even with monthly, we have only 113 data points, barely enough to make some statistically sound inferences. Cramer (1987), for example, argued that "R^2 should not be quoted for samples of less than fifty observations".

name	tick symbol	permno	cusip
IBM	IBM	12490	45920010
Ford	F	25785	34537010
Merck	MRK	22752	58933110
Intel	INTC	59328	45814010

Table 4.1: Stocks Used In Our Example

second quarter GDP will be available only after June[31], therefore it will be used for months June, July and August as currently available information about GDP. April and May, while also in the second quarter, will only have first quarter GDP information available to condition on when people is trying to make prediction about future stock returns.

Stocks that we choose

We have randomly picked four stocks in the demonstration. These are as listed in Table 4.1:

We checked that for all four stocks, we have data for the periods that we are interested in[32].

null hypothesis

There is *no* linear relationship between the regressors and the dependent variable, but the dependent variables could be themselves correlated[33].

Standardization

To improve accuracy in regression, we normalized the regressors so that they have a mean of zero and standard deviation of 1. All of our goodness-of-fit measures will not be affected by this standardization.

[31]For simplicity we assume it is available *immediately* after June. Such an assumption is not too far away from truth: we checked DataStream, which is our source for economic data, and on October 15th, we have all the quarterly data for the third quarter (ending September) available.

[32]the turnover measure for IBM in December 1998 is termed "inf" in our data set, but since we use the predictive model specification, the last month we include for regressors are November 1998, so this is not a problem

[33]Note that we are **not** assuming that there is a perfect correlation among the regressors, otherwise the estimation of regression parameters will be impossible.

In the end, we have as dependent variable a vector of y_{t+1}, which is the net return for each month, for 113 periods, and we also have as independent variables a matrix of 20 variables for 113 periods, X_t. The predictive nature of the regression requires that X_t's are one period lagged of y_{t+1}'s.

the basic test

The procedure of the test

We show as an example the practice of pick the best performing regression from three out of the twenty regressors. We have $\binom{20}{5}$ regressions to choose from to form the five variable regression.

We do the following experiments:

1. Randomly generate a dependent variable, $\tilde{y}$, (so there is no explanation power of the regressors). For any possible combination of five out of the twenty regressors, do the regression using the above described random dependent variables.

   For each of the four stocks and each of the four metrics, we record the highest value of these metrics among all $\binom{20}{5}$ combinations. This is one realization of the maximum goodness-of-fit measures.

2. Generate N (N=200 in our study) replications of the above goodness-of-fit measures to estimate an empirical distribution of these "best" statistics of goodness-of-fit measures under the null hypothesis of no explanatory power.

3. Then, with the real dependent variable y replacing the randomly generated $\tilde{y}$, do the regression analysis in the first step.

   For each of the four stocks and each of the four metrics, we record the highest value of these metrics, as well as the coefficient values (which is not directly meaningful, since we have standardized the independent variables), the t value associated with each coefficients, the identity of these independent variables (a number from 1 to 20, identifying the numbers among the 20 variables, that are used in the regression) and all four measures of the goodness-of-fit for this regression specification.

4. compare the statistics we get from step 3 to the distribution of the statistics under the null, to compute the empirical p-value.

We report the test statistics in a table.

In the table, we see that when these measures of goodness-of-fit are reported individually, many of them look very significant: the maximum t statistics range from 2.374 to 3.556 for all the stocks, and for two of the four stocks (IBM and FORD), in the "best" specification all the coefficients are significant at 95% level. The R^2 is in general also pretty high for this type of test: ranging from 7.618

However, when data-mining is explicitly considered, most of these effects go away. When judged against the empirical distribution of test statistics under the null hypothesis of no predictive power, the lowest empirical p-value we get is $4.5(R^2)$ of IBM, which is barely significant at a 5% level. All other "patterns" seem to be easily attributable to the extensive specification searching.

extension of the basic test

The above approach is useful when the researcher is designing a study of data-mining. In particular, he can, through thorough designing, generate an *exhaustive* list of m potential explanatory variables, which includes all possible regressors that could have been used by any researcher in this type of study[34]. In that case, he can perform the above test with the (possibly very large) universe of the regressors, to get the empirical distribution of the best metric of goodness-of-fit under the null hypothesis, and then analyze whether the most significant specification of a k variable regression is likely the result of data-mining.

Sometime, however, the researcher might not be able to construct an exhaustive list of all possible explanatory variables, or he might be only interested in a simpler question: what is the likelihood that the specification he(she) is interested in testing is a result of data-mining. In those scenarios, a more relevant question, as asked by Foster et al in their paper, is: given the level of goodness-of-fit of a particular specification that a researcher is interested in, how large should the potential size of the universe of regressors be, i.e., how many regressors (m^*) will he have access to, in order for that level of significance to be generated by chance

[34]The universe of calendar rules discussed in Sullivan et al (1998) is an example of exhaustive listing, in a different context with the marginal prediction models.

at least 5% of the time, or any other significance level, under the null hypothesis. Then he will be able to evaluate whether that number of regressors is reasonable. If, for example, it turns out that for the statistical result be attributable to data-mining he would have to have access to thousands of regressors, then he could be relatively more confident than, say, in the case when the results he has can be attributable to data-mining with a dozen or so regressors at hand.

As is clear from the above discussion, in such a case the researcher does not have an exhaustive list of the potential regressors. Thus our basic model will not work. To answer that question, we need to extend our basic test, and we do it in two steps.

expanded test 1: random X's with a fixed correlation structure

As a preparation for the later discussion, let us show a method that works in a relatively simpler setting: that the researcher actually has knowledge about the full universe of the regressors, *and* he(she) knows the correlation structure among these variables, V[35]. The only thing that he *does not* know is the full set of observations for some of these regressors.

Such incidences, although not very common, could happen if the collection of large amount of observations for some of the regressors poses a problem, but the researcher has access to a small amount of observations of each of the regressors in the universe, from which he could estimate the correlation structure among these regressors.

In that case, we propose the following approach:

Let us assume that there are T observations used in the specification that the researcher is testing, and the universe of all possible regressors contains m regressors. We first randomly generate a set of T by m random variables which has the same correlation structure as the real regressors. Then, we can compute the empirical distribution of the "best measures" by regressing a randomly generated dependent variable on k out of the m randomized regressors that we just generated. To be more specific, we take the following steps:

1. For the given correlation structure $V_{m\times m}$ among the m potential regressors, we compute the "square root matrix" of V, $\Lambda_{m\times m}$, such that $\Lambda'\Lambda = V$;

2. we generate a matrix of the standard normal random variable, $L_{T\times m}$, where each

[35]We have been using correlation and covariance interchangeably, since we assume that the regressors have been standardized so that their variance are all equal to one

column of L is a standard normal distribution;

3. we define $\tilde{X} = L\Lambda$, so that $Var(\tilde{X}) = \Lambda'\Lambda = V$.

4. we generate a T by 1 vector of a standard normal random variable, y.

5. for every possible combination of the k out of m regressors, we perform the regression of y on the subset of regressors, and record the best measures of the goodness-of-fit out of all $\binom{m}{k}$ such regressions. This gives one realization of the empirical distribution of the maximum measure of the goodness-of-fit.

6. we repeat steps 2 to 4 for N times, to get the empirical distribution of the maximum measure of the goodness-of-fit. We then compare the best measure of goodness-of-fit from the real data to the distribution we just obtained, to get the empirical p-value.

As an illustration of the method, we estimate the correlation structure for all regressors for IBM stock used in our previous example, and then we use that structure to do the test of data-mining following the steps we outlined just now. We report the results in a table.

When we compare this table with the previous table, where real regressors are used, we noticed that they are pretty close. In particular, none of our statistical inference about these best performing measures have changed.

expanded test 2: random X's with a hypothetical correlation structure

In some cases, the research has no knowledge about the set of candidate regressors, but still want to have a rough idea about the likelihood that the significance of the specification he (she) is interested comes out of data-mining.

In that case, we propose an approximate method using a hypothetical correlation structure. In particular, we assume that there might be other regressors out there, and the correlation coefficients among all these unknown regressors are random, but with the same mean and variance conditions as the existing correlation structure among the k regressors that the researcher knows. That is to say, if for the existing k regressors, the cross-correlations, $\rho_{i,j}$ for $i > j$[36], has sample mean $\hat{\mu}$ and sample standard deviation $\hat{\sigma}^2$, then we assume that all other

[36]We only need to count $i > j$ case since the correlation matrix is symmetric

unknown regressors will have correlations with each other, where the correlation coefficients will be random variables with the same $\{\hat{\mu}, \hat{\sigma}^2\}$ as their *population* moments. Therefore, we could first generate a correlation structure among any m hypothetical regressors using these mean and variance condition, and then use this correlation structure to generate random variables to act as regressors. From there, we could do the test of data-mining effect as we did before.

However, we need to be a little bit careful here: the correlation coefficient is in the range of -1 to 1, and we cannot directly assume a normal distribution on it. There are many ways to fix that. We choose to do a transformation through a tangent function: we first transform the correlation coefficients through

$$\theta_{ij} = tan(\rho_{ij}\frac{\pi}{2})$$

which transforms ρ_{ij} to θ_{ij} on the real line, then, after we estimate the sample mean and sample variance, we generate random normal variables with that mean and variance, and convert these random normal variables back to actual correlation coefficients through

$$\rho'_{ij} = ctan(\theta'_{ij})\frac{2}{\pi}$$

The hypothetical correlation approximation can be performed in the following way:

1. We estimate the real correlation matrix among the k regressors used in the regression specification that the researcher is interested in.
2. we apply the tangent transformation to the $\frac{k\times(k-1)}{2}$ upper-triangular elements of the correlation matrix.
3. we estimate the sample mean and sample variance of the transformed variables, $\{\hat{\mu}, \hat{\sigma}^2\}$.
4. For a fixed m, we generate $\frac{m\times(m-1)}{2}$ random variables from a normal distribution with mean and variance $\{\hat{\mu}, \hat{\sigma}^2\}$.

5. We apply an inverse tangent transformation to get back "correlations" estimates, and create a correlation matrix accordingly.

6. we use the correlation matrix thus created to randomly generate regressors that have this correlation structure, as discussed in the case of "fixed correlation structure".

7. Now we can gauge the effect of data-mining for the case of m regressors, even if we don't know the m regressors.

50-50		anomaly	sharp ratio	best model abnormal return	p-value
log					
	adding	-0.0004	-0.221	0.0124	0.105
		-0.0016	-0.443	0.0158	0.035
		-0.0064	-1.331	0.0518	0.000
		-0.0256	-4.88	0.2135	0.000
		-0.1024	-19.08	0.8465	0.000
		0.0004	-0.073	0.0105	0.155
		0.0016	0.149	0.0133	0.075
		0.0064	1.036	0.0255	0.000
		0.0256	4.586	0.0778	0.000
		0.1024	18.785	0.2761	0.000
	multiplying	(1-0.0004)	-0.147	0.0113	0.145
		(1-0.0016)	-0.147	0.0113	0.145
		(1-0.0064)	-0.148	0.0113	0.145
		(1-0.0256)	-0.148	0.0112	0.145
		(1-0.1024)	-0.151	0.0110	0.145
		(1+0.0004)	-0.147	0.0113	0.145
		(1+0.0016)	-0.147	0.0113	0.145
		(1+0.0064)	-0.147	0.0113	0.145
		(1+0.0256)	-0.146	0.0113	0.145
		(1+0.1024)	-0.144	0.0114	0.140
simple					
	adding	-0.0004	-0.221	0.0126	0.095
		-0.0016	-0.443	0.0160	0.035
		-0.0064	-1.331	0.0519	0.000
		-0.0256	-4.88	0.2149	0.000
		-0.1024	-19.08	0.8670	0.000
		0.0004	-0.073	0.0107	0.155
		0.0016	0.149	0.0135	0.070
		0.0064	1.036	0.0258	0.000
		0.0256	4.586	0.0794	0.000
		0.1024	18.785	0.2968	0.000
	multiplying	(1-0.0004)	-0.147	0.0101	0.100
		(1-0.0016)	-0.147	0.0114	0.140
		(1-0.0064)	-0.148	0.0114	0.140
		(1-0.0256)	-0.148	0.0114	0.140
		(1-0.1024)	-0.151	0.0113	0.145
		(1+0.0004)	-0.147	0.0115	0.140
		(1+0.0016)	-0.147	0.0115	0.140
		(1+0.0064)	-0.147	0.0115	0.140
		(1+0.0256)	-0.146	0.0115	0.140
		(1+0.1024)	-0.144	0.0116	0.135

Table 4.2: The 50-50 Benchmark

market		anomaly	sharp ratio	best model abnormal return	p-value
log					
	adding	-0.0004	-0.221	0.0069	0.360
		-0.0016	-0.443	0.0136	0.145
		-0.0064	-1.331	0.0634	0.000
		-0.0256	-4.88	0.2808	0.000
		-0.1024	-19.08	1.1541	0.000
		0.0004	-0.073	0.0027	0.780
		0.0016	0.149	0.0021	0.885
		0.0064	1.036	0.0008	0.995
		0.0256	4.586	0.0000	0.995
		0.1024	18.785	0.0000	0.995
	multiplying	(1-0.0004)	-0.147	0.0046	0.490
		(1-0.0016)	-0.147	0.0046	0.490
		(1-0.0064)	-0.148	0.0046	0.490
		(1-0.0256)	-0.148	0.0045	0.495
		(1-0.1024)	-0.151	0.0042	0.540
		(1+0.0004)	-0.147	0.0046	0.490
		(1+0.0016)	-0.147	0.0046	0.490
		(1+0.0064)	-0.147	0.0046	0.490
		(1+0.0256)	-0.146	0.0047	0.485
		(1+0.1024)	-0.144	0.0050	0.475
simple					
	adding	-0.0004	-0.221	0.0068	0.365
		-0.0016	-0.443	0.0136	0.150
		-0.0064	-1.331	0.0631	0.000
		-0.0256	-4.88	0.2804	0.000
		-0.1024	-19.08	1.1499	0.000
		0.0004	-0.073	0.0026	0.795
		0.0016	0.149	0.0020	0.890
		0.0064	1.036	0.0008	0.995
		0.0256	4.586	0.0000	0.995
		0.1024	18.785	0.0000	0.995
	multiplying	(1-0.0004)	-0.147	0.0045	0.495
		(1-0.0016)	-0.147	0.0045	0.495
		(1-0.0064)	-0.148	0.0045	0.495
		(1-0.0256)	-0.148	0.0044	0.510
		(1-0.1024)	-0.151	0.0042	0.545
		(1+0.0004)	-0.147	0.0045	0.495
		(1+0.0016)	-0.147	0.0045	0.495
		(1+0.0064)	-0.147	0.0045	0.495
		(1+0.0256)	-0.146	0.0046	0.490
		(1+0.1024)	-0.144	0.0049	0.475

Table 4.3: The Market Benchmark

riskless		anomaly	sharp ratio	best model abnormal return	p-value
log					
	adding	-0.0004	-0.221	0.018	0
		-0.0016	-0.443	0.018	0
		-0.0064	-1.331	0.041	0
		-0.0256	-4.88	0.147	0
		-0.1024	-19.08	0.556	0
		0.0004	-0.073	0.018	0
		0.0016	0.149	0.025	0
		0.0064	1.036	0.05	0
		0.0256	4.586	0.157	0
		0.1024	18.785	0.566	0
	multiplying	(1-0.0004)	-0.147	0.018	0
		(1-0.0016)	-0.147	0.018	0
		(1-0.0064)	-0.148	0.018	0
		(1-0.0256)	-0.148	0.018	0
		(1-0.1024)	-0.151	0.018	0
		(1+0.0004)	-0.147	0.018	0
		(1+0.0016)	-0.147	0.018	0
		(1+0.0064)	-0.147	0.018	0
		(1+0.0256)	-0.146	0.018	0
		(1+0.1024)	-0.144	0.018	0
simple					
	adding	-0.0004	-0.221	0.018	0
		-0.0016	-0.443	0.018	0
		-0.0064	-1.331	0.041	0
		-0.0256	-4.88	0.149	0
		-0.1024	-19.08	0.584	0
		0.0004	-0.073	0.019	0
		0.0016	0.149	0.025	0
		0.0064	1.036	0.051	0
		0.0256	4.586	0.159	0
		0.1024	18.785	0.594	0
	multiplying	(1-0.0004)	-0.147	0.018	0
		(1-0.0016)	-0.147	0.018	0
		(1-0.0064)	-0.148	0.018	0
		(1-0.0256)	-0.148	0.018	0
		(1-0.1024)	-0.151	0.018	0
		(1+0.0004)	-0.147	0.018	0
		(1+0.0016)	-0.147	0.018	0
		(1+0.0064)	-0.147	0.018	0
		(1+0.0256)	-0.146	0.018	0
		(1+0.1024)	-0.144	0.018	0

Table 4.4: The Riskless Benchmark

		subtracting	$\hat{\rho}$	$\hat{\sigma}$	best rule abnormal return	p value
PR, 800 days						
	1	0.001	0.64079855	0.00547007	0.01026016	0.40841329
	2	0.0015	0.64097546	0.00549935	0.01309569	0.20164552
	3	0.002	0.64093695	0.00553126	0.01593264	0.07637345
	4	0.0025	0.6407029	0.00556576	0.01921501	0.01753883
	5	0.003	0.64029651	0.0056028	0.02487453	0.00054137
	6	0.0035	0.63974079	0.00564231	0.03053408	0.00000555
	7	0.004	0.63905617	0.00568425	0.03619365	0.00000002
	8	0.0045	0.63825938	0.00572852	0.04185325	0
	9	0.005	0.63736336	0.00577508	0.04751287	0
Time series 800 days						
	1	0.001	0.64300185	0.00554347	0.01026016	0.41777972
	2	0.0015	0.64326878	0.00559088	0.01309569	0.21344371
	3	0.002	0.6434861	0.00564634	0.01593264	0.08596857
	4	0.0025	0.64365328	0.00570959	0.01921501	0.02224925
	5	0.003	0.64377089	0.00578038	0.02487453	0.00093024
	6	0.0035	0.64384045	0.00585844	0.03053408	0.00001545
	7	0.004	0.6438643	0.00594346	0.03619365	0.00000012
	8	0.0045	0.64384538	0.00603515	0.04185325	0
	9	0.005	0.64378707	0.0061332	0.04751287	0
Time series 8000 days						
	1	0.001	0.64464312	0.00902857	0.03242855	0.0113445
	2	0.0015	0.64467263	0.0090551	0.04720392	0.00001534
	3	0.002	0.64468266	0.00908684	0.06543859	0
	4	0.0025	0.64467348	0.00912375	0.08367332	0
	5	0.003	0.64464553	0.00916575	0.10190810	0
	6	0.0035	0.64459932	0.0092128	0.12014295	0
	7	0.004	0.6445355	0.0092648	0.13837787	0
	8	0.0045	0.64445481	0.00932167	0.15661289	0
	9	0.005	0.64435805	0.00938334	0.17484799	0

Table 4.5: The comparison between tests using PR bootstrap and Time series estimation of ρ and σ.

experiment number	magnitude of anomaly	indifference k
1	-0.0001	0.02883554
2	-0.00015	0.0676278
3	-0.0002	0.17790587
4	-0.00025	0.69169196
5	-0.0003	17.92682708
6	-0.00035	1213.718624
7	-0.0004	180672.274
8	-0.00045	50545263.9
9	-0.0005	22750571080

Table 4.6: Indifference K Value Associated With Different Magnitude of Artificial Anomaly

Figure 4.1: SUBTRACTING AN ARTIFICIAL ANOMALY

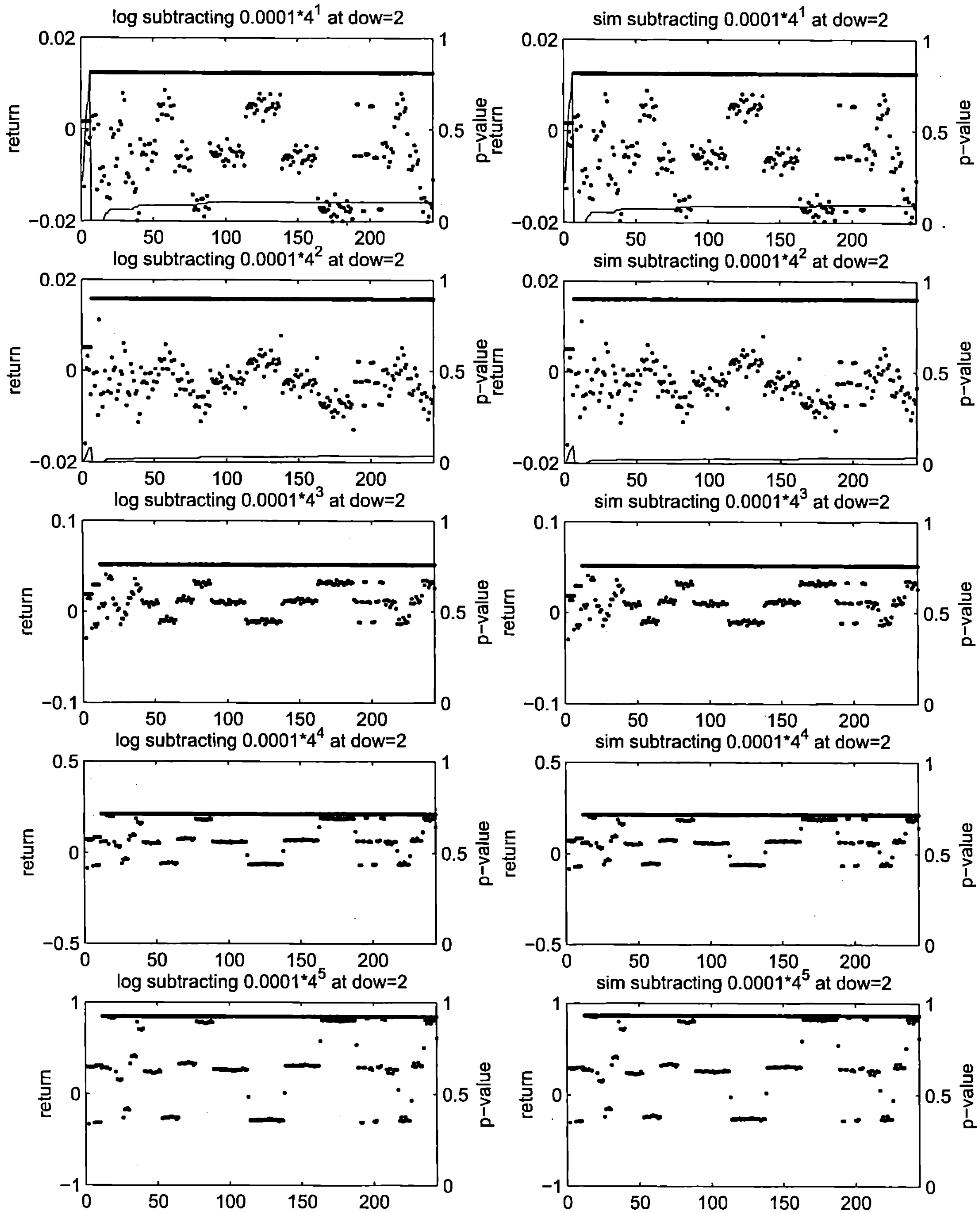


Figure 4.2: ADDING AN ARTIFICIAL ANOMALY

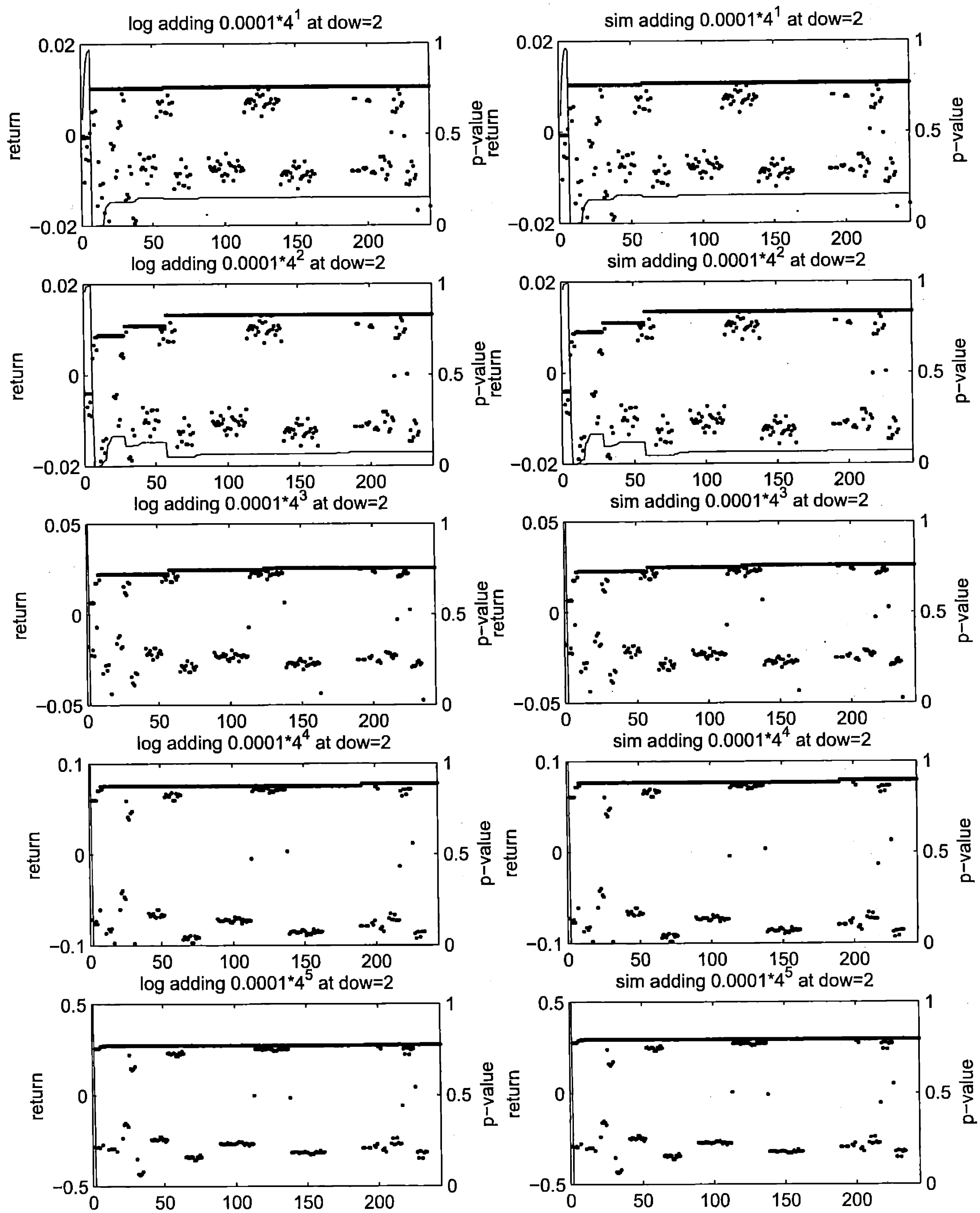


Figure 4.3: MULTIPLYING AN ARTIFICIAL ANOMALY (I)

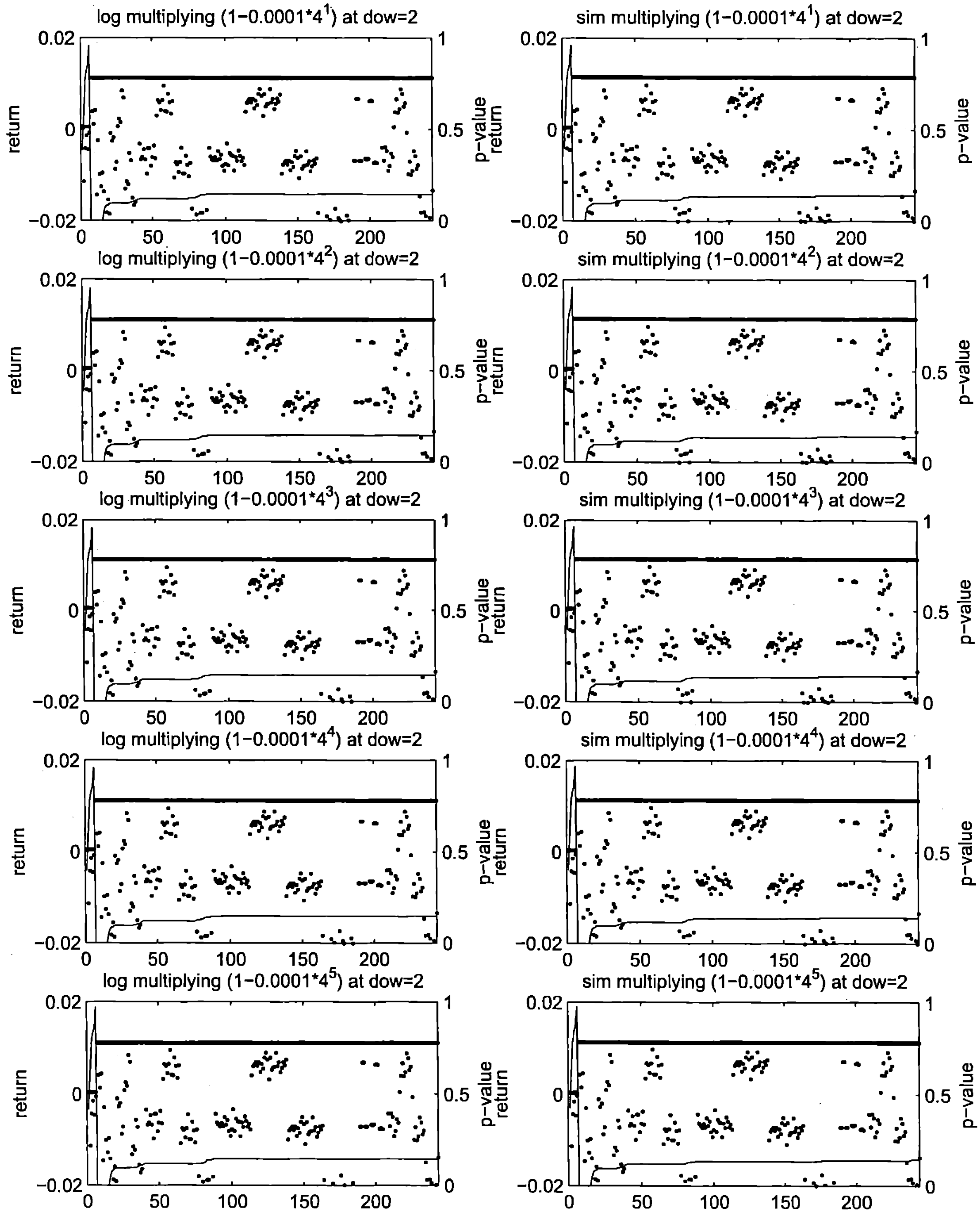


Figure 4.4: MULTIPLYING AN ARTIFICIAL ANOMALY (II)

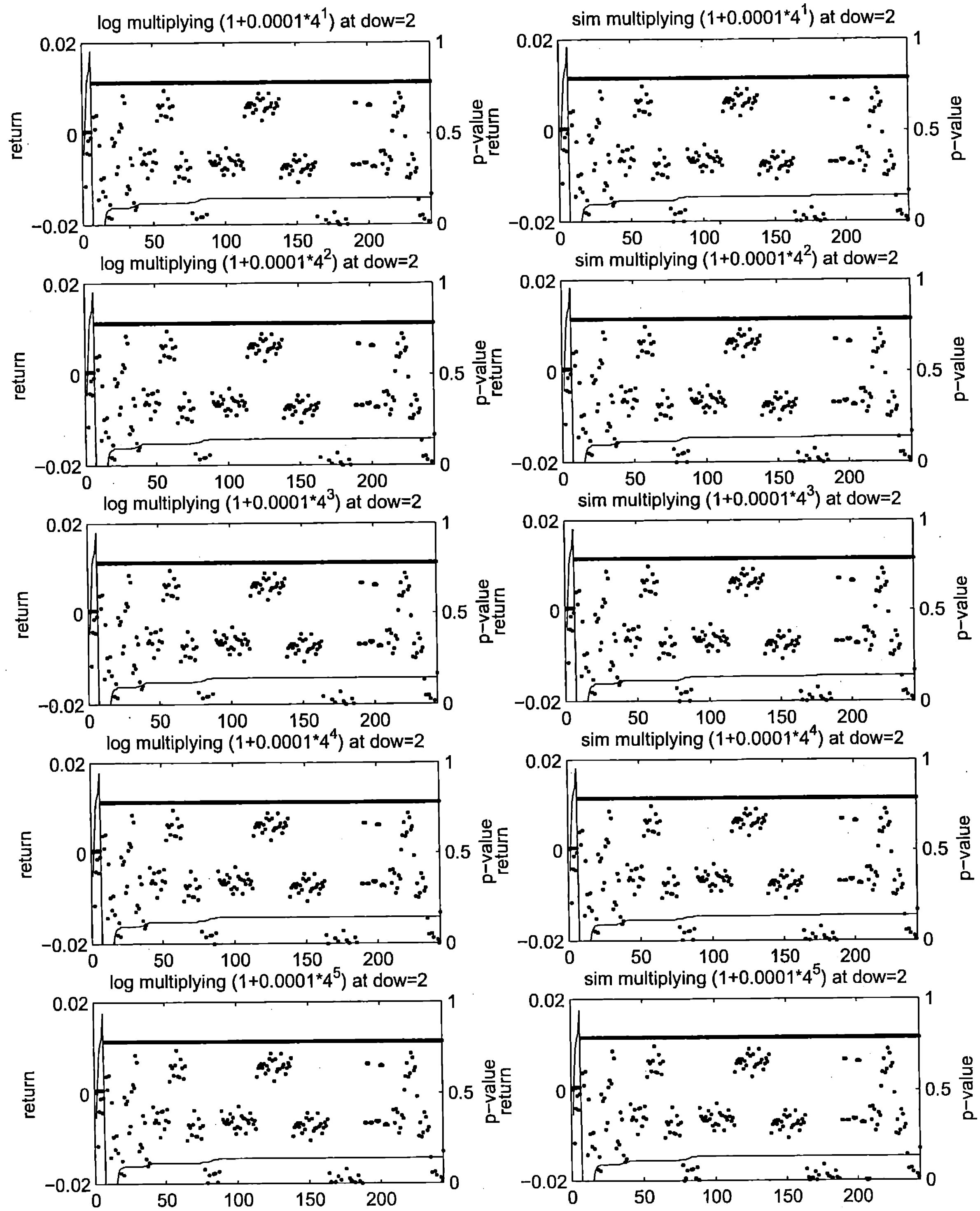


Figure 4.5: EXPANSION TEST RESULT: NO ANOMALY

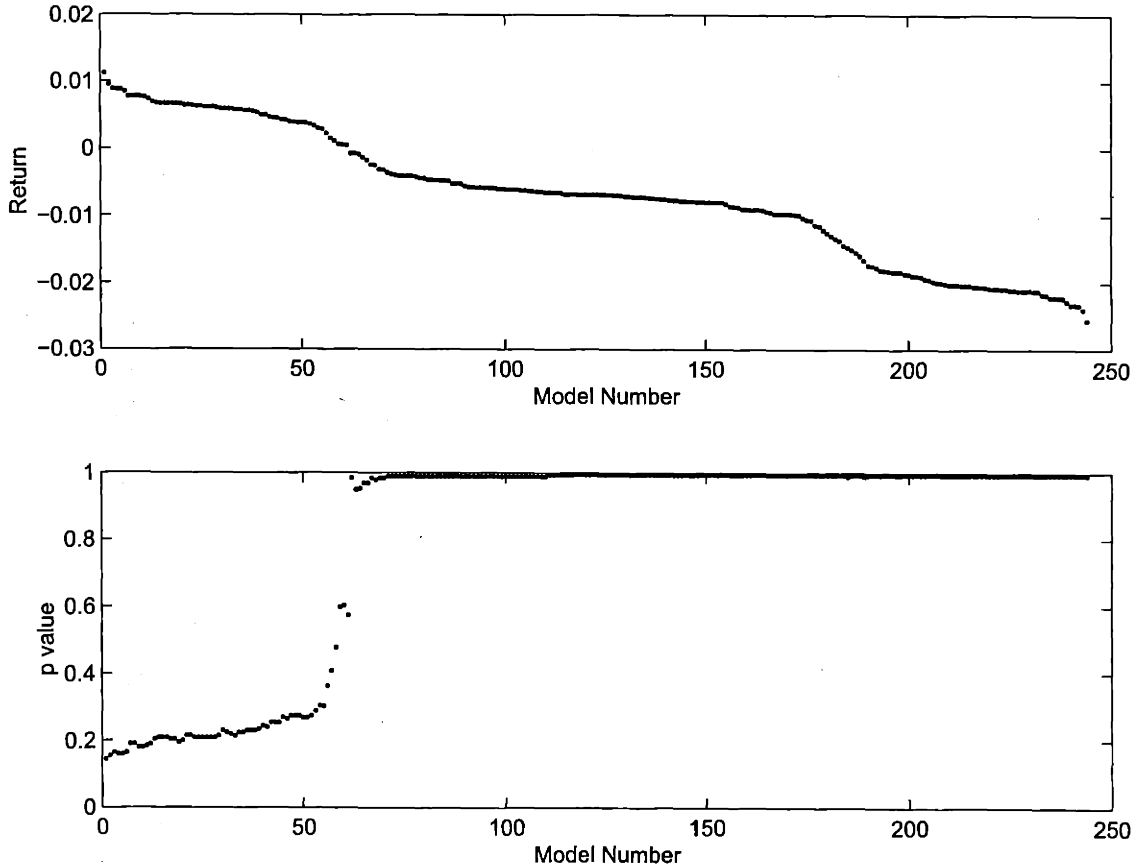
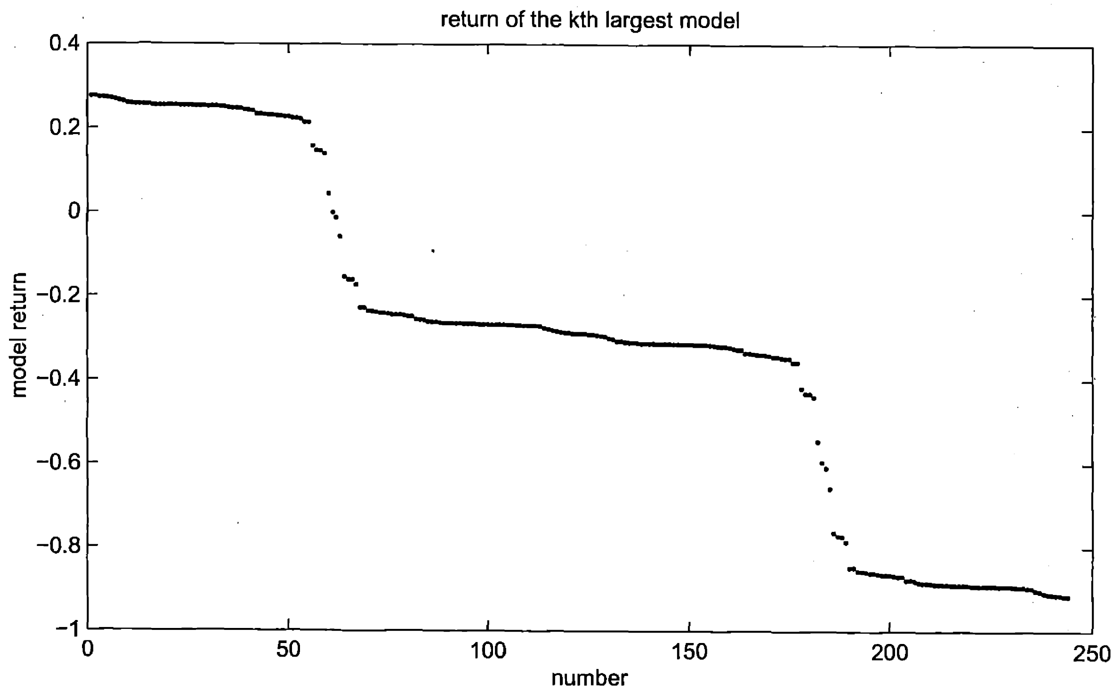


Figure 4.6: EXPANSION TEST RESULT: ADDING AN ARTIFICIAL ANOMALY OF MAGNITUDE 0.1024



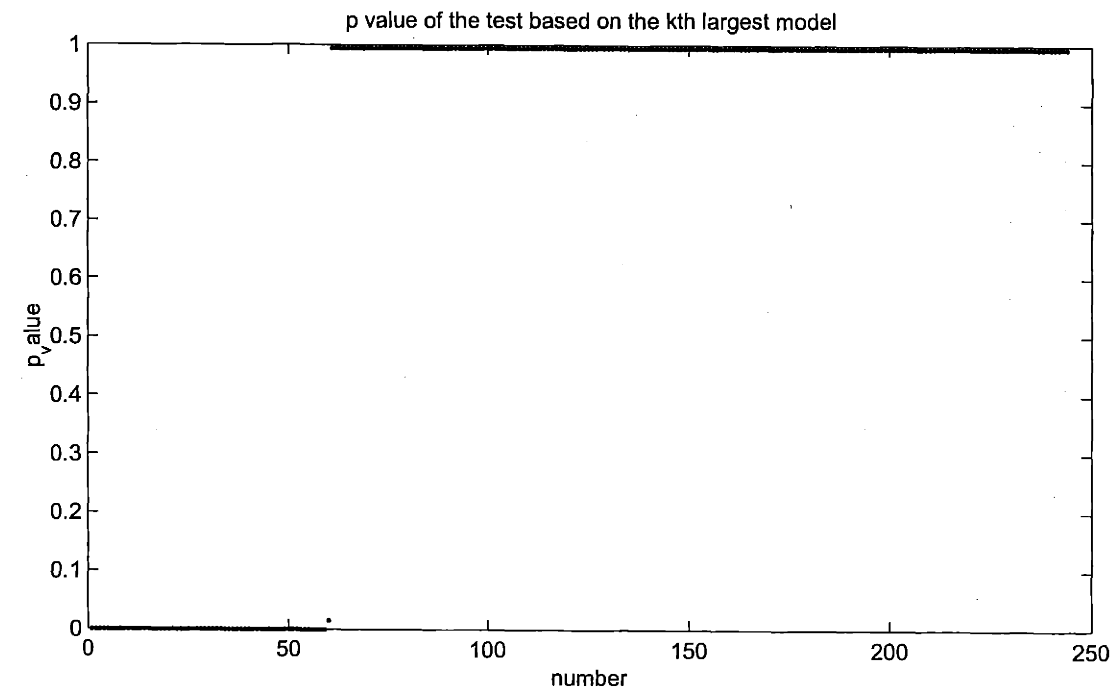


Figure 4.7: HISTOGRAM OF CORRELATION AMONG CALENDAR RULES

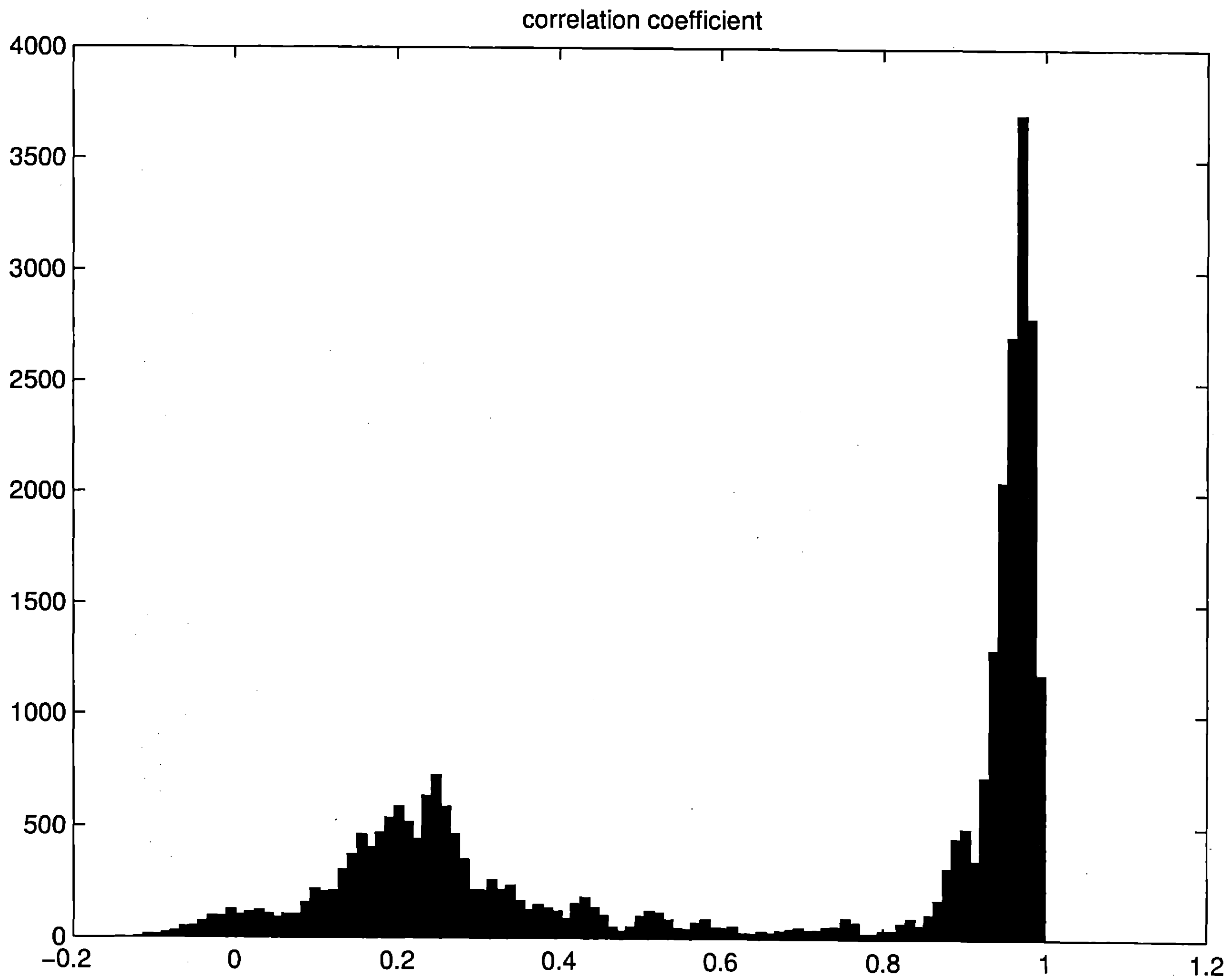


Figure 4.8: CUMULATIVE DISTRIBUTION OF THE MAXIMUM OF INDEPENDENT NORMAL RANDOM VARIABLES

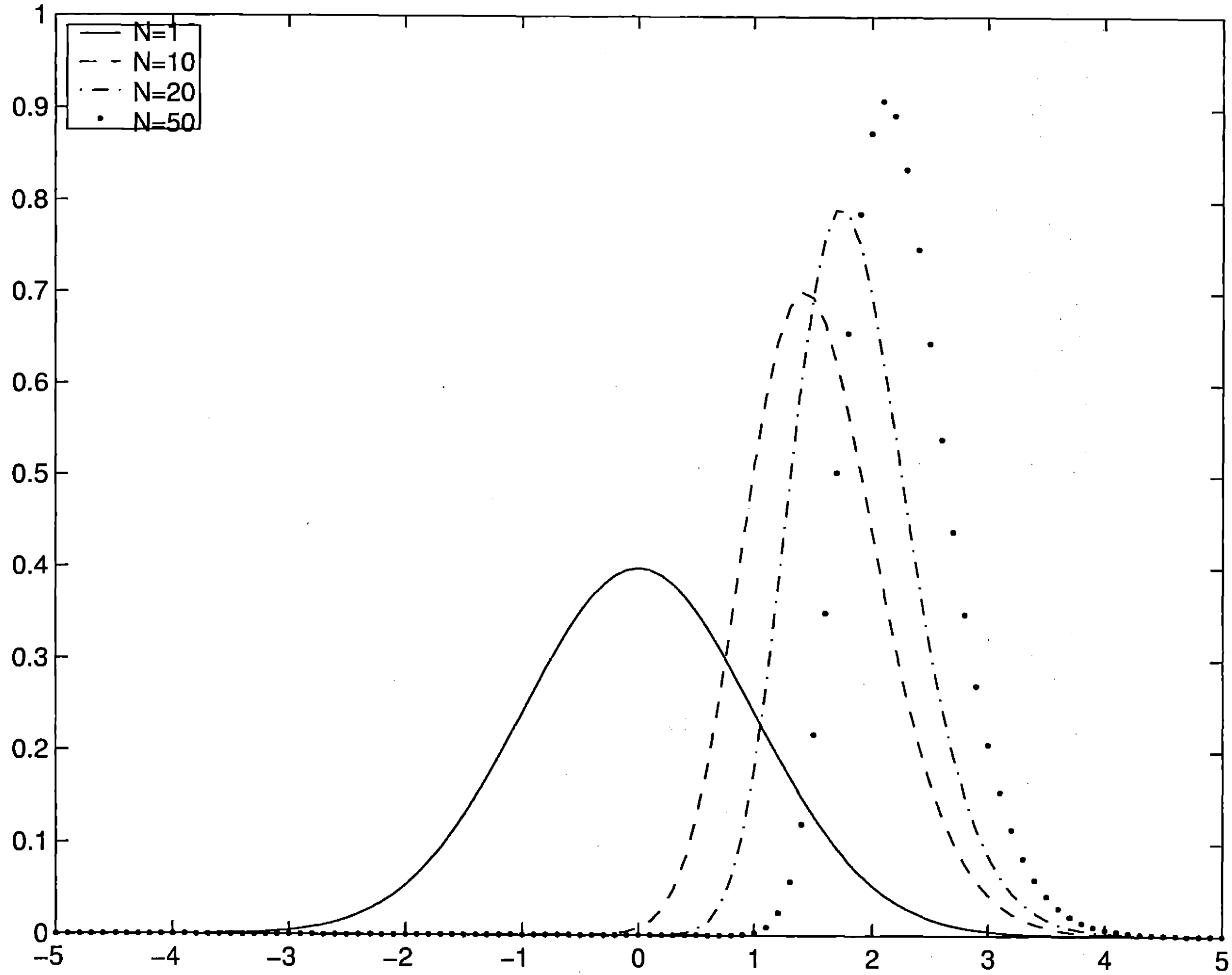


Figure 4.9: CUMULATIVE DISTRIBUTION OF THE MAXIMUM OF EQUI0-CORRELATED NORMAL RANDOM VARIABLES: ρ=0.1

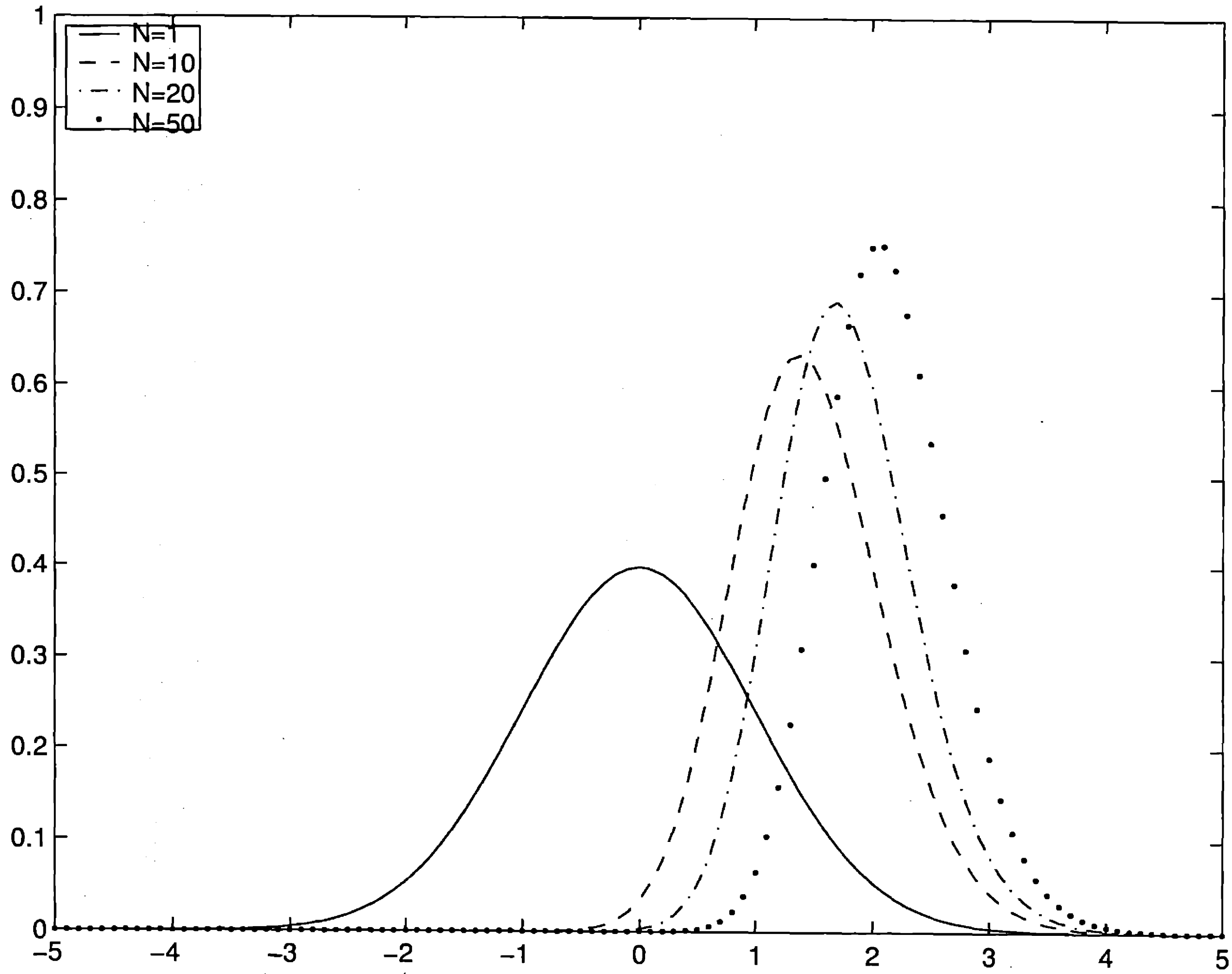


Figure 4.10: CUMULATIVE DISTRIBUTION OF THE MAXIMUM OF EQUI0-CORRELATED NORMAL RANDOM VARIABLES: ρ=0.2

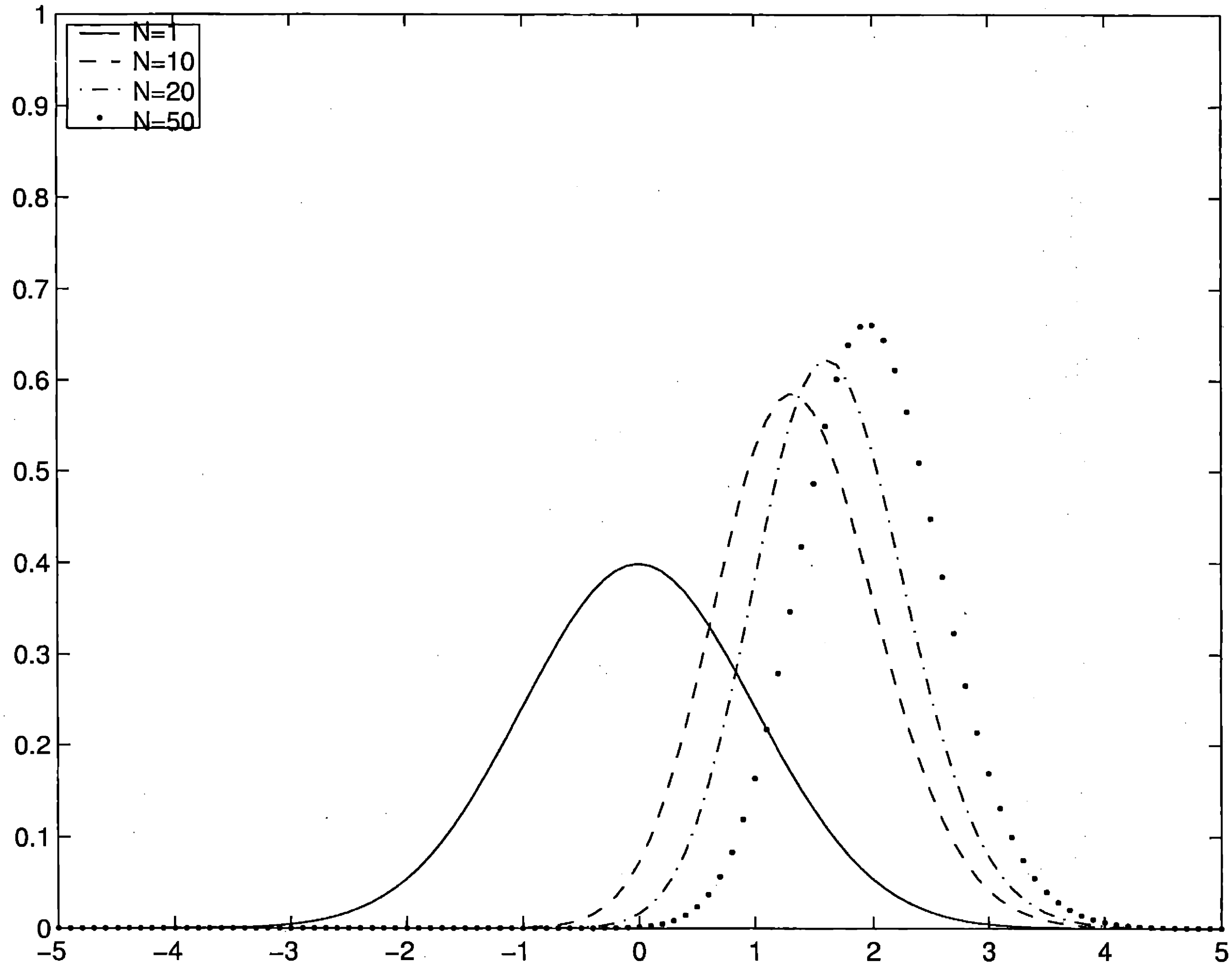


Figure 4.11: CUMULATIVE DISTRIBUTION OF THE MAXIMUM OF EQUI0-CORRELATED NORMAL RANDOM VARIABLES: ρ=0.3

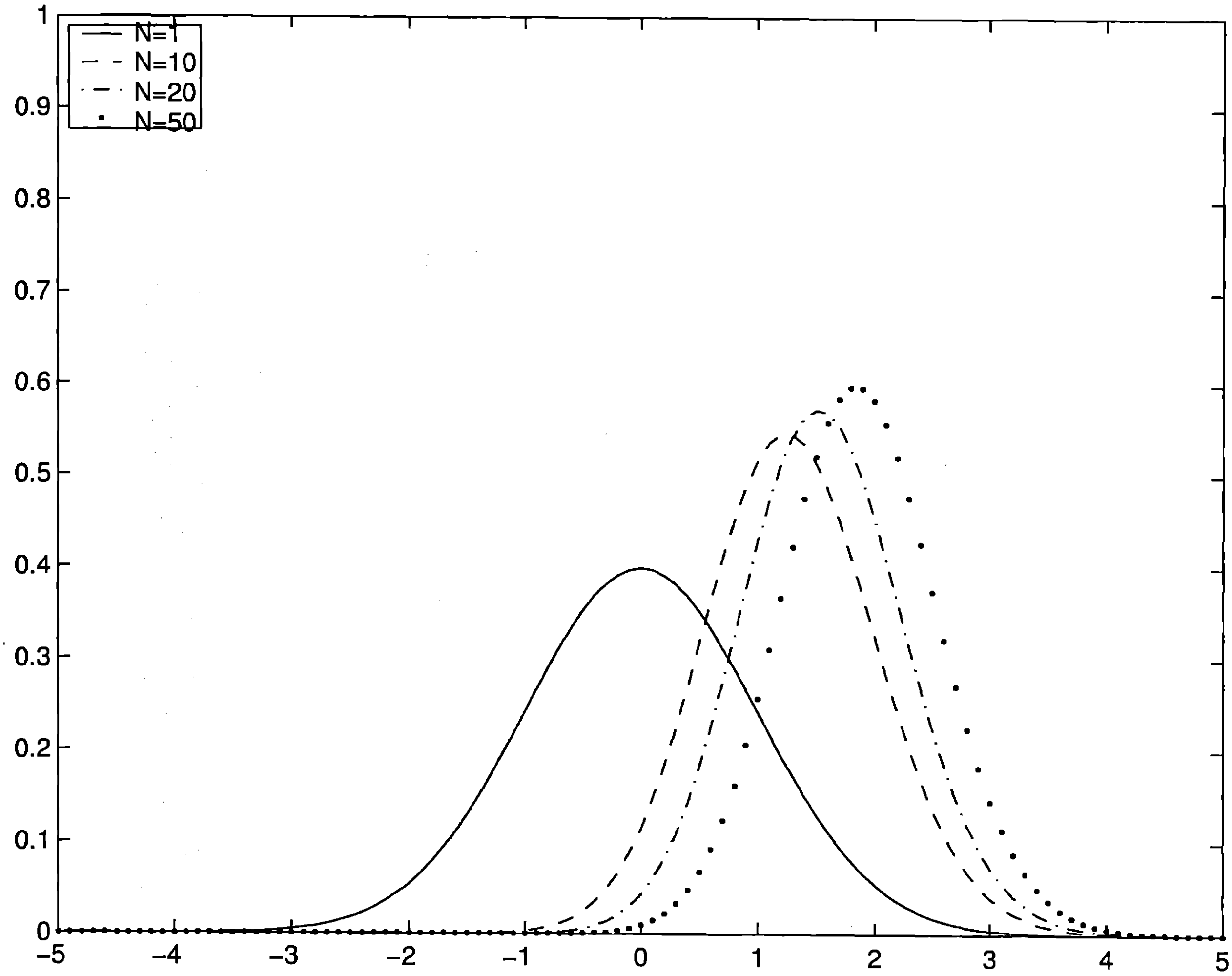


Figure 4.12: CUMULATIVE DISTRIBUTION OF THE MAXIMUM OF EQUI0-CORRELATED NORMAL RANDOM VARIABLES: ρ=0.4

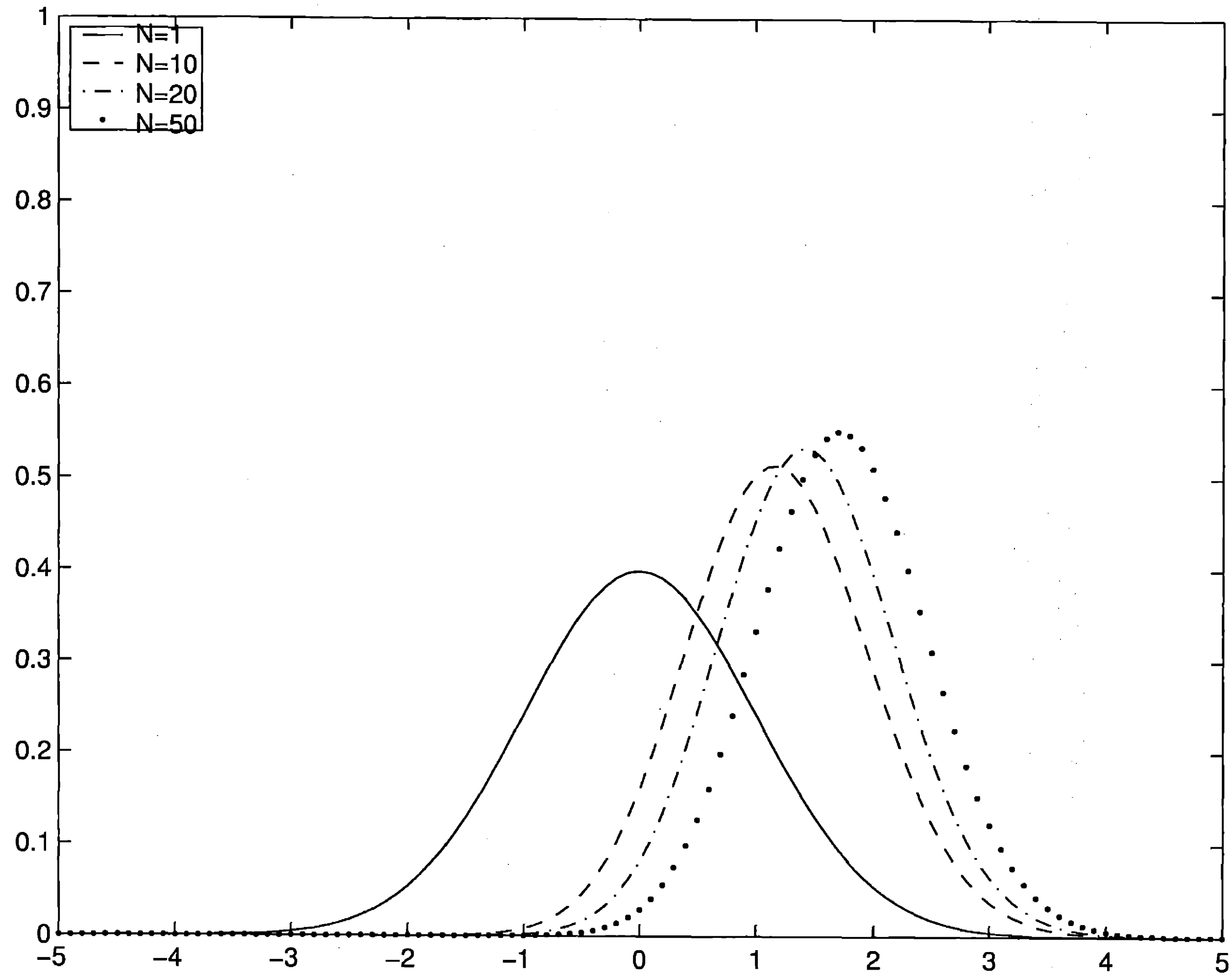


Figure 4.13: CUMULATIVE DISTRIBUTION OF THE MAXIMUM OF EQUI0-CORRELATED NORMAL RANDOM VARIABLES: ρ=0.5

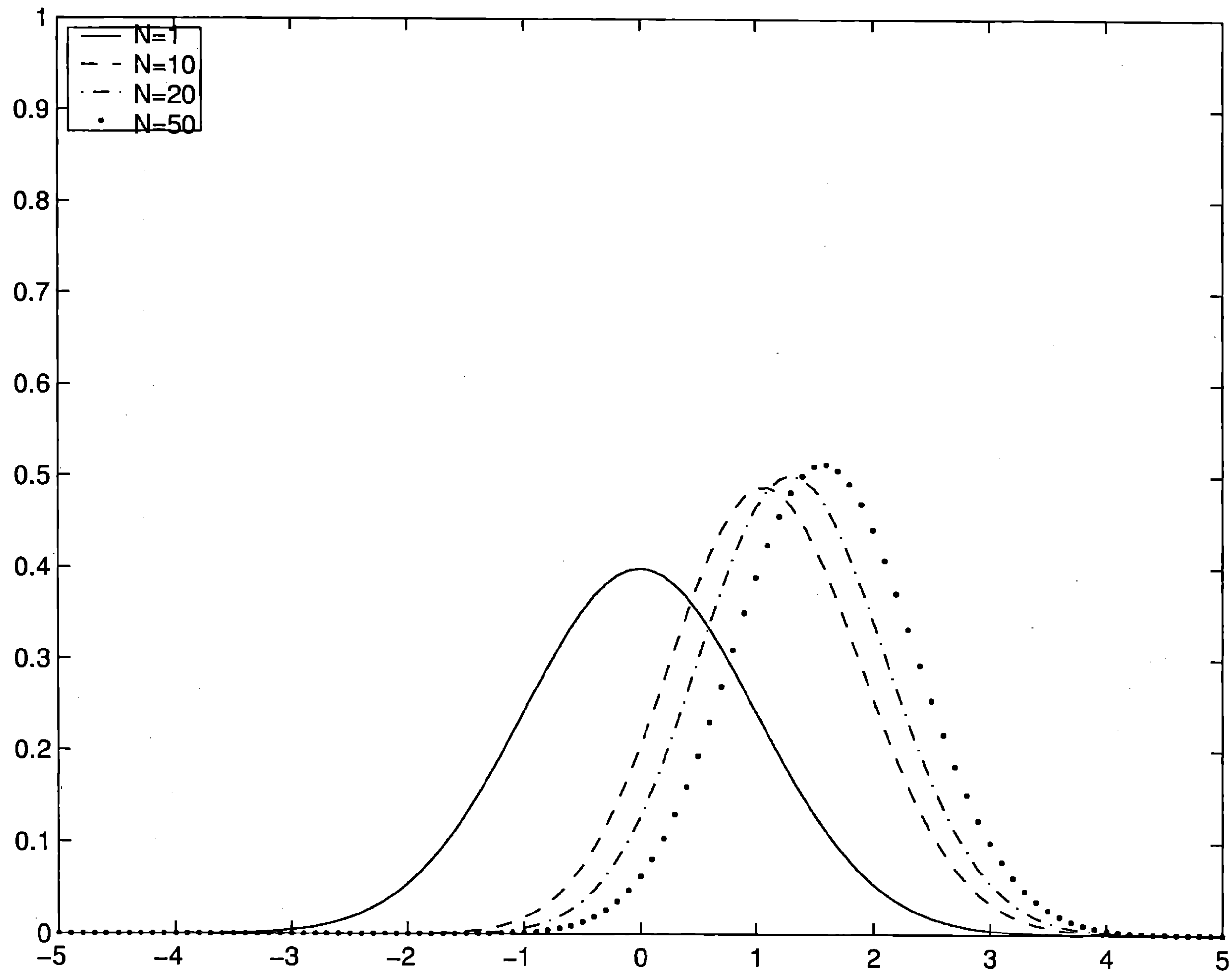


Figure 4.14: CUMULATIVE DISTRIBUTION OF THE MAXIMUM OF EQUI0-CORRELATED NORMAL RANDOM VARIABLES: ρ=0.6

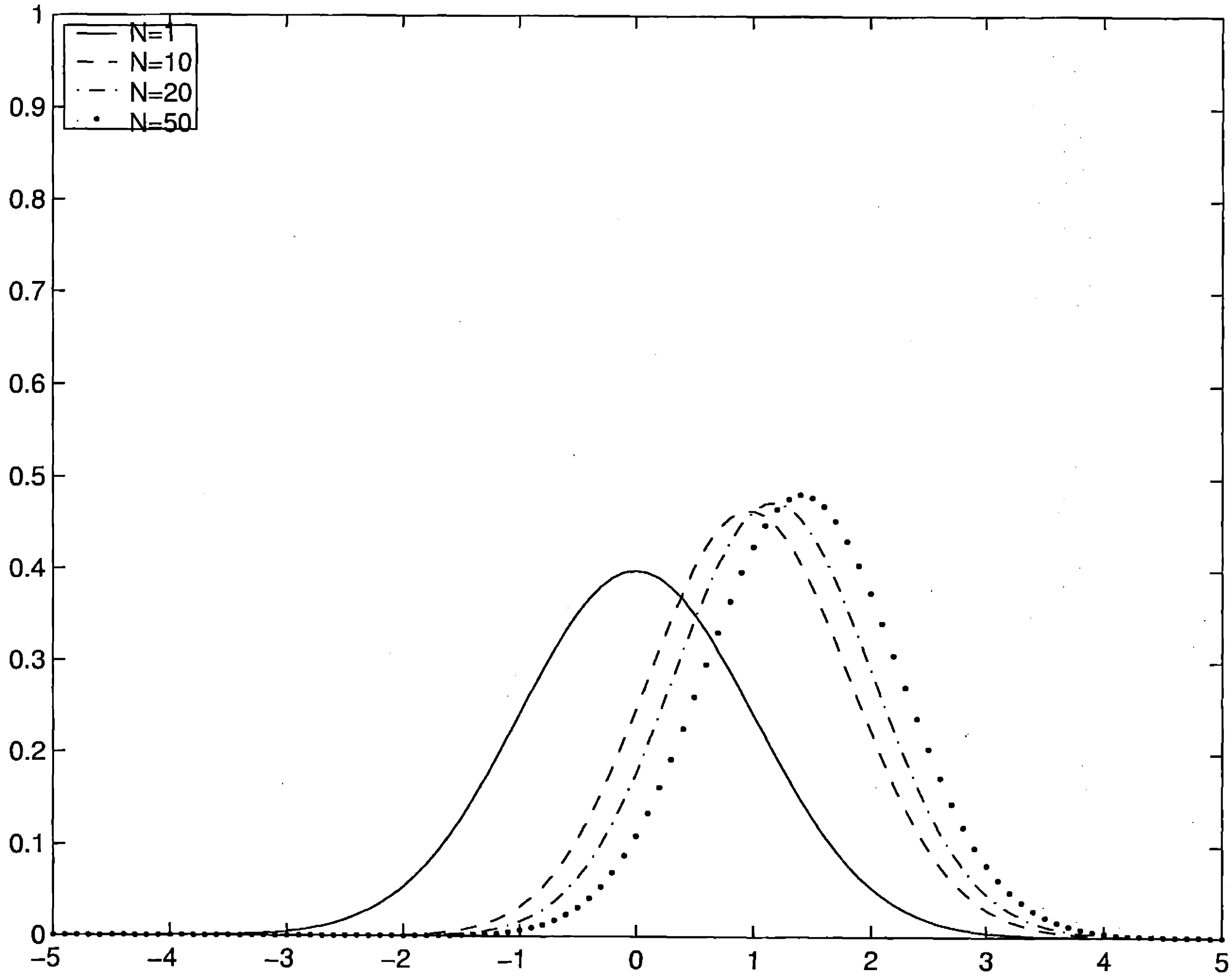


Figure 4.15: CUMULATIVE DISTRIBUTION OF THE MAXIMUM OF EQUI0-CORRELATED NORMAL RANDOM VARIABLES: ρ=0.7

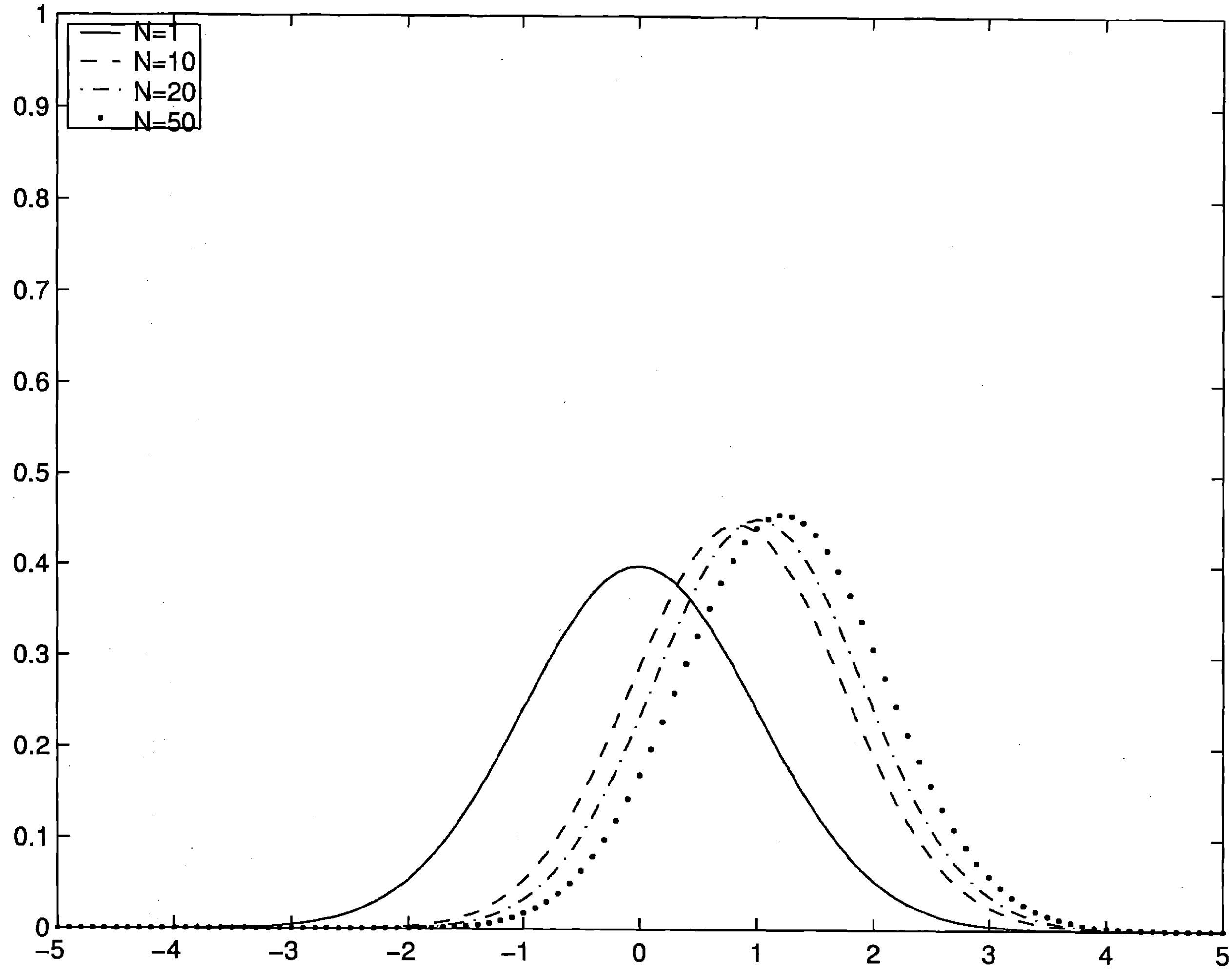


Figure 4.16: CUMULATIVE DISTRIBUTION OF THE MAXIMUM OF EQUI0-CORRELATED NORMAL RANDOM VARIABLES: ρ=0.8

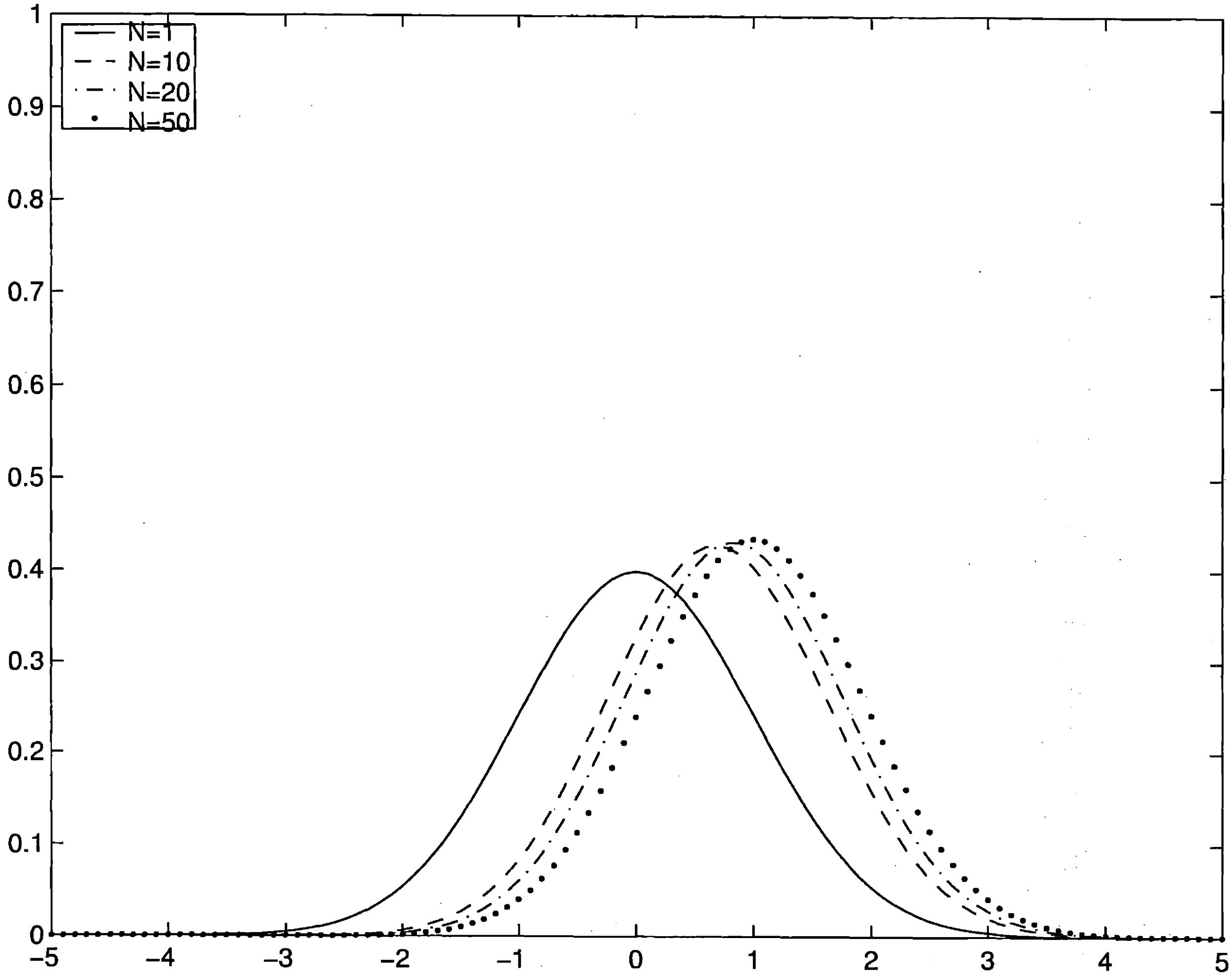


Figure 4.17: CUMULATIVE DISTRIBUTION OF THE MAXIMUM OF EQUI0-CORRELATED NORMAL RANDOM VARIABLES: ρ=0.9

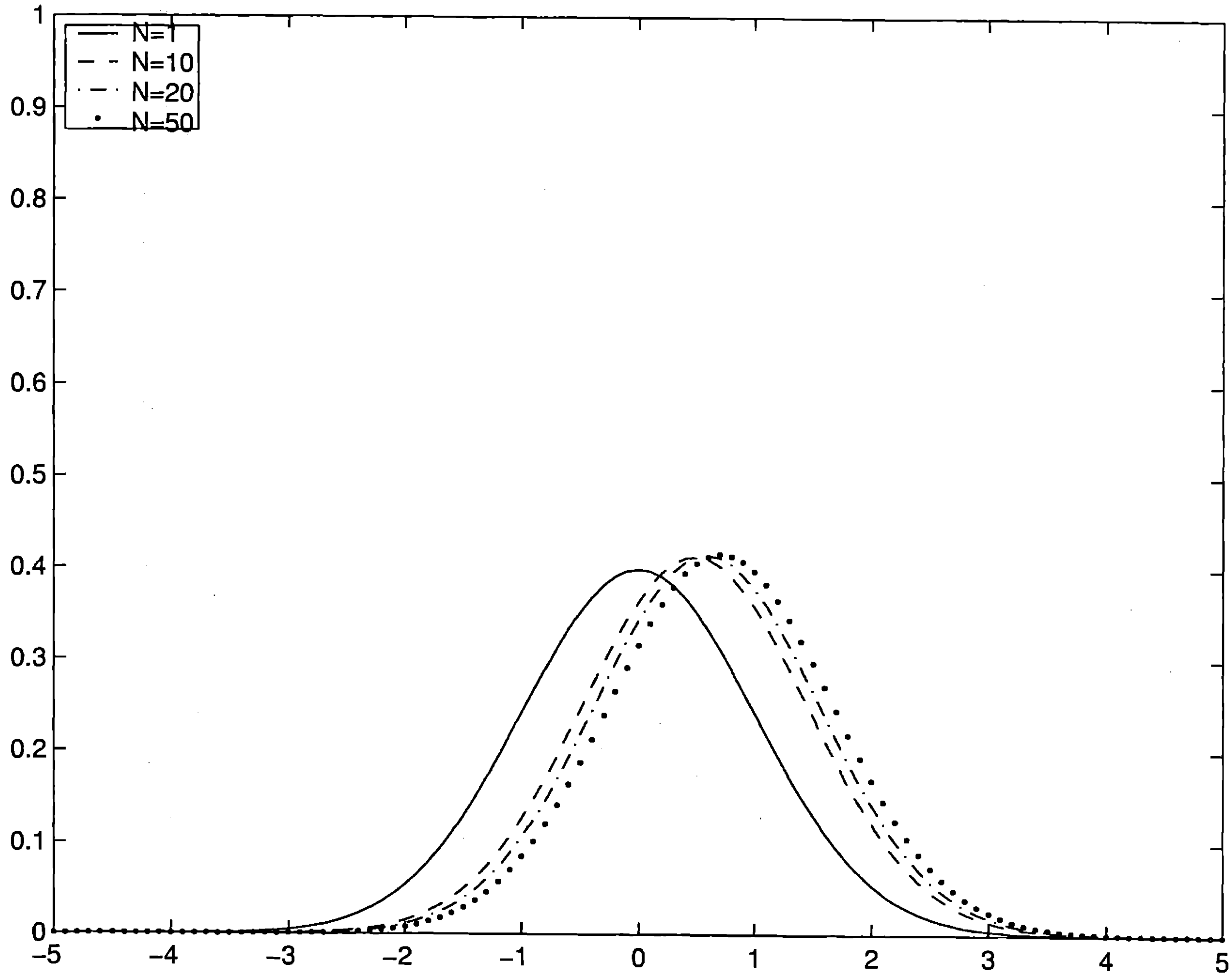


Figure 4.18: TEST BASED ON EQUI-CORRELATED NORMAL ASSUMPTION: CASE WHEN THERE IS NO ANOMALY

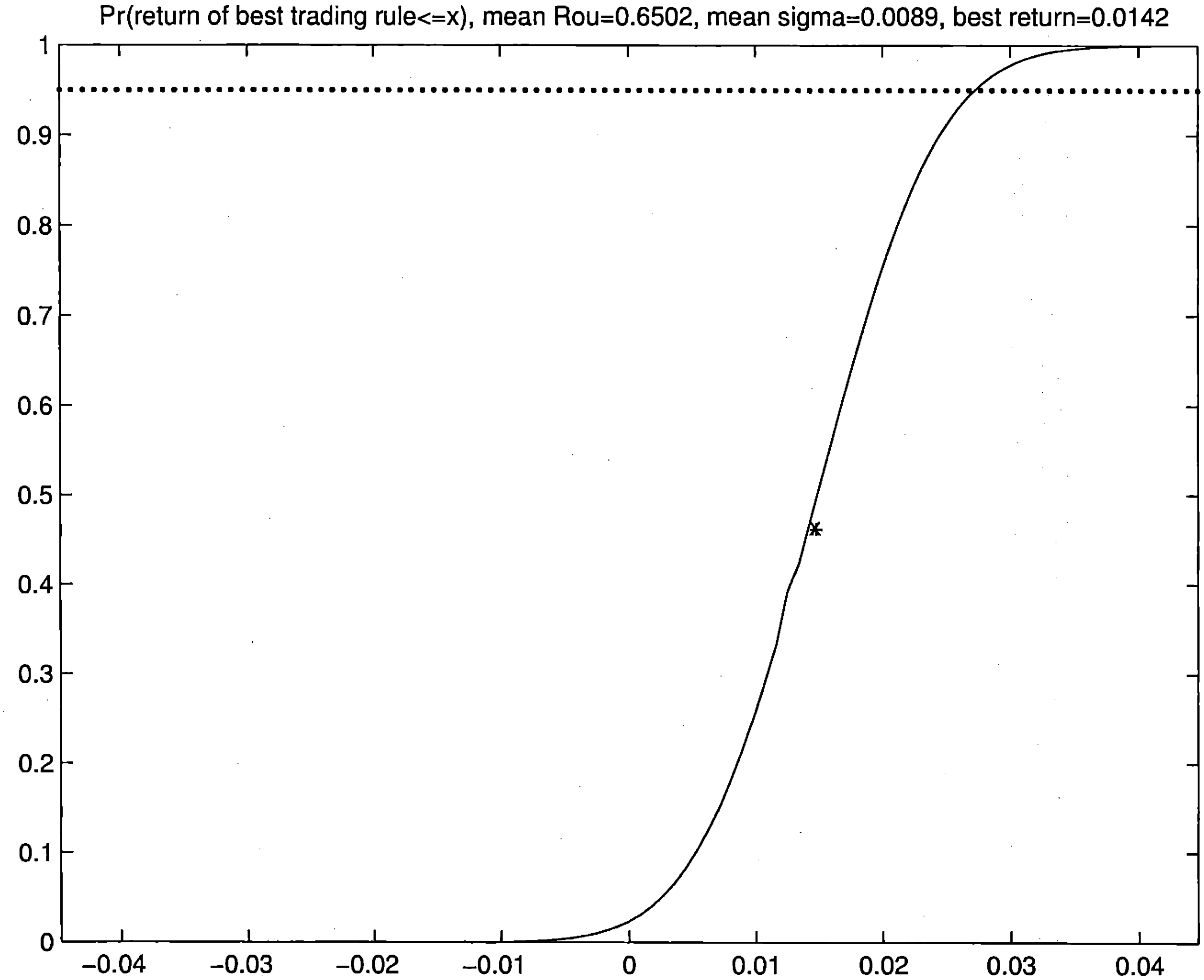


Figure 4.19: TEST BASED ON EQUI-CORRELATED NORMAL ASSUMPTION: CASE WHEN AN ANOMALY OF MAGNITUDE 0.0016 IS ADDED

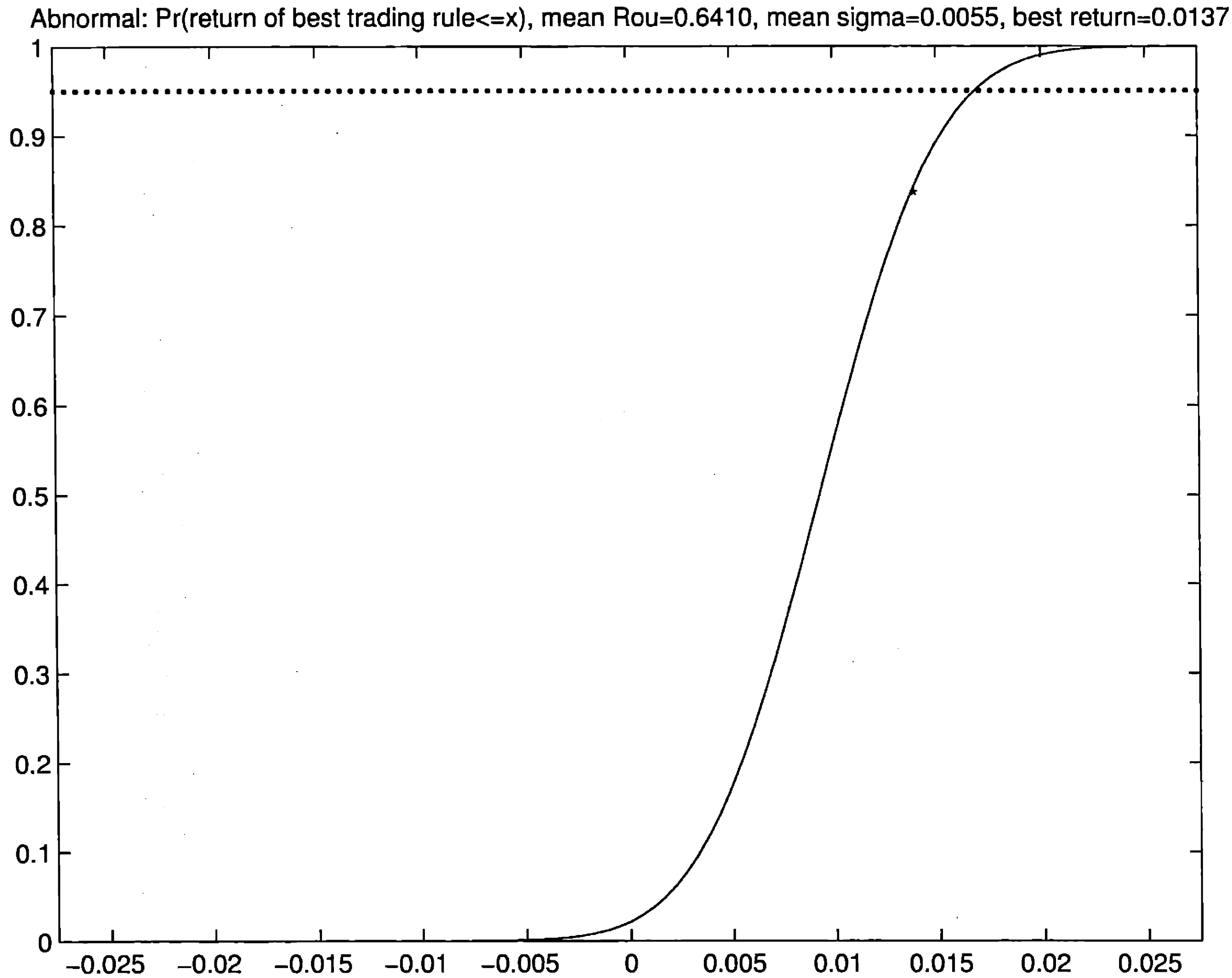
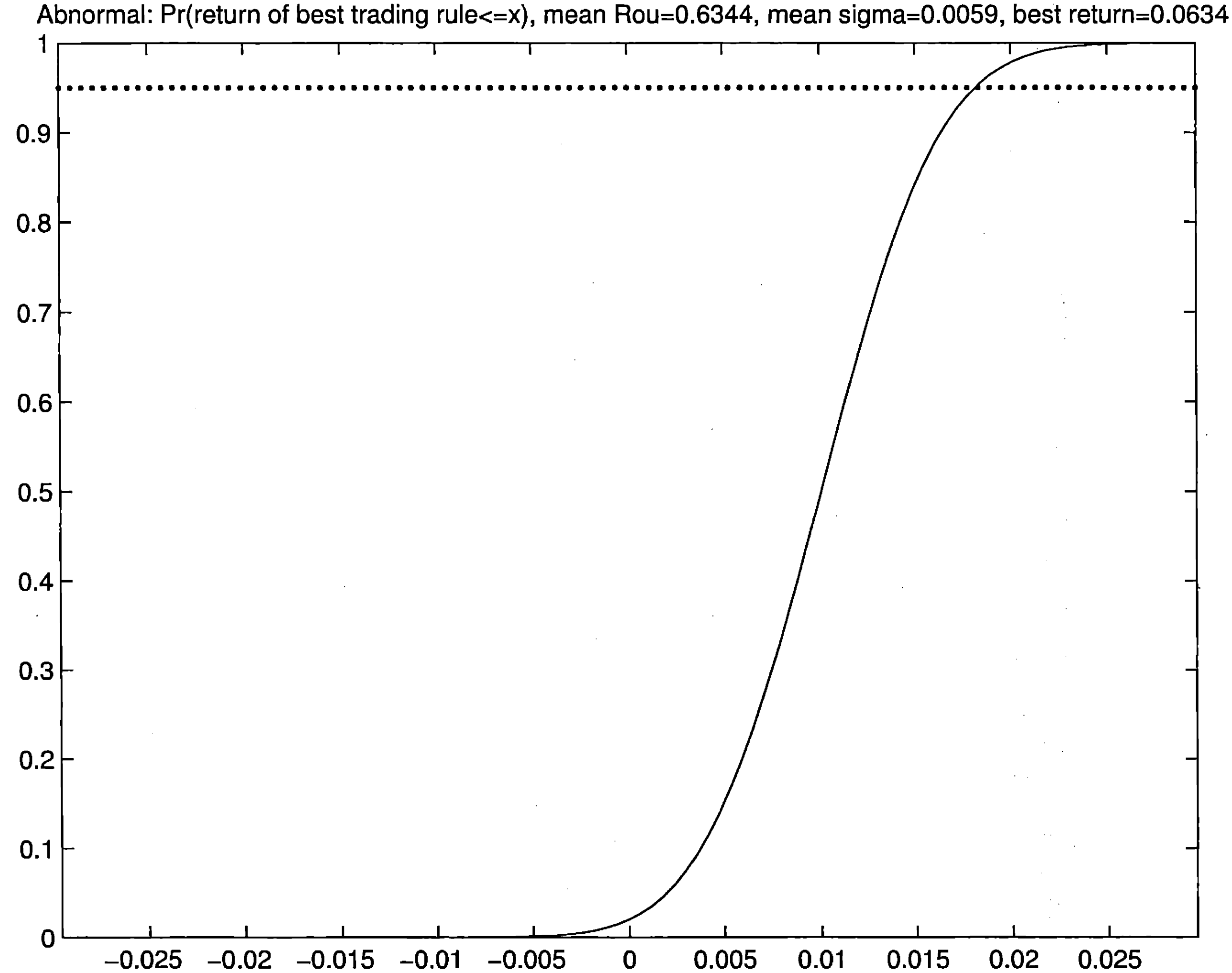


Figure 4.20: TEST BASED ON EQUI-CORRELATED NORMAL ASSUMPTION: CASE WHEN AN ANOMALY OF MAGNITUDE 0.0064 IS ADDED



Appendix A

appendices

A.1 Appendices of Chapter 1 on the theory part

A.1.1 The Prendergast (2000) critique

Prendergast (2000) argues that although theoretically the principal agent framework should produce a negative relation between risk and incentive, in practice we often do not detect it, because risk is correlated with other variables. Prendergast differentiates between two types of contracts: monitoring the input (effort level) versus monitoring the output. In input monitoring, the principal chooses the *ex ante* most productive action, and pays the agent his cost of effort if the ex ante most productive action is taken (otherwise pays the agent nothing); in output monitoring, the principal lets the agent choose an action to take, and use pay-for-performance to align incentive.

An options interpretation

In Prendergast's model, output based contract allows agent discretion. Since agent is assumed to observe the realization of the productivity of actions before he chooses the action, he can choose the ex post most productive action. This is potentially better than the ex ante optimal action that the principal picks, and the value of agent discretion can be thought of as a real option. Intuitively, the higher the uncertainty level, the more valuable this option of agent discretion, thus the contract is more likely to take the form of output monitoring and pay-on-output. This implies that we could observe a positive relation between risk and pay-for-performance. For a detailed analysis, see Prendergast (2000).

Several aspects of the Prendergast model make it hard to directly compare his results with my results. First, Prendergast assumes risk neutral agent, which assumes away the risk sharing-incentive tradeoff in my paper, and as a result, he always gets first best incentive and effort level[1]. Second, we have different interpretation of effort and the cost of effort. While in my model cost of effort can potentially include all opportunity costs for the action that the agent takes, in the Prendergast model, cost of effort is more narrowly defined. For example, the private benefit to agent in his model would be a cost of effort in my model if the agent does not choose his own favorite action (which allows him to realize the private benefit). Third, the setting in Prendergast (2000) is discontinuous, so that he only gets corner solutions, where the principal uses either input monitoring or output monitoring, but never both at the same time, and when output monitoring is adopted, agent always get 100% of incentive.[2] Therefore, a naive application of Prendergast would predict that either the CEO is paid a fixed pay, or he is paid 100% of the performance. Certainly this is too strong to be what Prendergast intends to show.

I think that in theory, the way to incorporate the Prendergast critique into our study consists of two directions. First, Prendergast points out that the agent's discretion is valuable to the firm, and the value of agent discretion is higher when risk is higher. This link could be picked up in my model by the productivity of CEO effort on firm value, because in my model, agent's discretion is part of his effort. Prendergast (2000) shows that it is reasonable that higher risk increases the value of discretion, thus, k in my model will be positively linked to risk: $k = k(\sigma_\delta^2)$. Then, from (1.2.7) and (1.2.18), now we see that the effect of firm-specific risk σ_δ^2 could be ambiguous, as predicted by Prendergast. Second, Prendergast points out that the higher the cost of monitoring effort level, or, for that matter, the lower the quality-to-cost ratio for the signal on effort, the lower the level of input monitoring and the higher the level of incentive pay. That is, there should be a substitution between monitoring and incentive, so that when monitoring is not efficient, incentive pay will be needed[3]. In an appendix, this feature is studied in more detail as an extension of my theory model. There I allow the agent to be risk-averse to explicitly address the risk sharing-incentive tradeoff.

[1]In fact, when output monitoring is used and a performance based pay is given, Prendergast's model assumes that the agent knows exactly the output, thus there is no risk at all to the agent. Thus it does not even matter whether or not the agent is risk averse in that case.

[2]Even in section 4.2 where Prendergast discusses partial delegation, the result is a corner solution of either zero or full-scale output monitoring.

[3]As argued by Prendergast: "When monitoring of inputs is costly, firms will respond by offering output-based contracts with delegation of tasks.", P24.

Empirically, Prendergast (2000) offers a very important perspective, that is, there could be important missing variables in the empirical work trying to estimate the relation between risk and incentive. Omitting these variables in the regression analysis could potentially eliminate or even reverse the observed relation between risk and incentive. As Prendergast himself puts it:

In standard econometric parlance, a difficulty with the existing empirical work is omitted variable bias. This model argues that uncertainty affects the responsibilities offered to workers, which in turn affects incentive. But responsibilities and discretion are rarely observed by the econometrician, so that omitted variable bias arises.[4]

The omitted variables problem can never be fully addressed. Still, I try to work in the direction pointed out by Prendergast, by studying how adding the proxy for the productivity of CEO effort might mitigate the omitted variable problem.

A.1.2 Proof of proposition 1

For the problem represented by equations (1.2.4) to (1.2.6), we can write the Lagrangine:

$$L = (1-a)(Z_0+ke)-b-\frac{1}{2}(1-a)^2\gamma_P\eta^2\sigma_m^2+\lambda[a(Z_0+ke)+b-f(e)-\frac{1}{2}a^2\gamma_A(\eta^2\sigma_m^2+\sigma_\delta^2)-\overline{CE_A}]+\mu[ak-f'(e)] \tag{A.1.1}$$

The first order conditions are (with respect to a, b, e, λ and μ):

$$-(Z_0+ke)+(1-a)\gamma_P\eta^2\sigma_m^2+\lambda(Z_0+ke)-a\lambda\gamma_A(\eta^2\sigma_m^2+\sigma_\delta^2)+uk=0 \tag{A.1.2}$$

$$-1+\lambda=0, or\lambda=1 \tag{A.1.3}$$

$$(1-a)k+\lambda ak-\lambda f'(e)-\mu f"(e)=0 \tag{A.1.4}$$

$$[a(Z_0+ke)+b-f(e)]-\frac{1}{2}a^2\gamma_A(\eta^2\sigma_m^2+\sigma_\delta^2)=\overline{CE_A} \tag{A.1.5}$$

[4]Prendergast (2000), P4.

$$f'(e) = ak \tag{A.1.6}$$

By plugging (A.1.6) into (A.1.4), we simplify it and get

$$(1-a)k - \mu f''(e) = 0 \tag{A.1.7}$$

which gives μ as a function of other parameters:

$$\mu = \frac{(1-a)k}{f''(e)} \tag{A.1.8}$$

This has clean interpretation. The equation (A.1.4) gives the tradeoff in increasing the effort level: on the positive side, a higher effort will bring *extra* benefit to the principal, which the agent ignores when designing on her own effort level tradeoff; on the negative side, it induces the agent to work harder, and, given there is an increasing cost of effort, this will result in ever higher cost to her, as captured by the curvature of the f(e) function.

The equation (A.1.8) gives the *shadow price* of the incentive compatibility constraint, which is the ratio of the extra benefit to principal over the measure of the increase in cost to the agent.

(A.1.3) and (A.1.2) give:

$$(1-a)\gamma_P \eta^2 \sigma_m^2 - a\gamma_A(\eta^2\sigma_m^2 + \sigma_\delta^2) + uk = 0 \tag{A.1.9}$$

Interpretation:

Here higher incentive (an increase in a) will:

- reduce principal's welfare cost of bearing risk by $(1-a)\gamma_P\eta^2\sigma_m^2$ on the margin, which is good to the principal;

- increase agent's welfare cost of bearing risk by $a\gamma_A(\eta^2\sigma_m^2 + \sigma_\delta^2)$ on the margin, which is bad to the principal, since he has to increase the b term to make up the reservation utility for the agent;

- improved incentive for agent to work harder, and loosen the incentive compatibility constraint. The benefit of this is captured by the shadow price of IC constraint, μ, and the productivity of effort on firm performance, k.

(A.1.9), combined with (A.1.8), gives

$$(1-a)\gamma_P\eta^2\sigma_m^2 - a\gamma_A(\eta^2\sigma_m^2+\sigma_\delta^2) + \frac{(1-a)k}{f"(e)}k = 0 \qquad \text{(A.1.10)}$$

which upon reorganization gives the result for a^*. The equation for e^* and b^* are both implicitly determined once a^* is determined.

∎.

A.1.3 Proof of proposition 2

We can write the Lagrangine:

$$L = [(1-a)(Z_0+ke)-b] + \frac{1}{2}\frac{(\overline{r_m}-r_f)^2}{\gamma_P\sigma_m^2} - (1-a)\eta(\overline{r_m}-r_f) + \lambda IR + \mu[ak - f'(e)], \quad \text{(A.1.11)}$$

where $IR = a[Z_0 + ke - \eta(\overline{r_m}-r_f)] + b - f(e) + \frac{1}{2}\frac{(\overline{r_m}-r_f)^2}{\gamma_A\sigma_m^2} - \frac{1}{2}\gamma_A a^2\sigma_\delta^2 - \overline{CE_A}$.

The first order conditions are (with respect to a, b, e, λ and μ):

$$-(Z_0+ke) + \eta(\overline{r_m}-r_f) + \lambda[Z_0 + ke - \eta(\overline{r_m}-r_f)] - a\lambda\gamma_A\sigma_\delta^2 + \mu k = 0 \qquad \text{(A.1.12)}$$

$$-1+\lambda = 0, or \lambda = 1 \qquad \text{(A.1.13)}$$

$$(1-a)k + \lambda ak - \lambda f'(e) - \mu f"(e) = 0 \qquad \text{(A.1.14)}$$

$$a[\overline{Z} - \eta(\overline{r_m}-r_f)] + b - f(e) + \frac{1}{2}\frac{(\overline{r_m}-r_f)^2}{\gamma_A\sigma_m^2} - \frac{1}{2}\gamma_A a^2\sigma_\delta^2 = \overline{CE_A} \qquad \text{(A.1.15)}$$

$$f'(e) = ak \qquad \text{(A.1.16)}$$

By plugging (A.1.16) into (A.1.14), we simplify it and get

$$(1-a)k - \mu f"(e) = 0 \tag{A.1.17}$$

which gives μ as a function of other parameters:

$$\mu = \frac{(1-a)k}{f"(e)} \tag{A.1.18}$$

(A.1.13) and (A.1.12) give:

$$a\gamma_A \sigma_\delta^2 = \mu k \tag{A.1.19}$$

Combining the last two equations, we get the result for a^*.The equation for e^* and b^* are both implicitly determined once a^* is determined.

We could also study the effect of an increase in the firm level firm-specific risk σ_δ^2 on 1) the level of firm-specific risk that the CEO receives, $a^{*2}\sigma_\delta^2$, and 2) the intercept of linear compensation contract, b^*. Since the market risk can be hedged out by trading the market portfolio, *and* market risk is orthogonal to firm-specific risk by definition, we could study the effect of firm-specific risk on these parameters without worrying about the market risk. For this reason, we for simplicity assume that the market risk is zero in the following analysis.

- The relation between σ_δ^2 and $a^{*2}\sigma_\delta^2$;

  It is easier to study the relation between σ_δ^2 and $\sqrt{a^{*2}\sigma_\delta^2}$.

$$\sqrt{a^{*2}\sigma_\delta^2} = \frac{k^2}{\gamma_A \sigma_\delta f"(e) + \frac{k^2}{\sigma_\delta}} \tag{A.1.20}$$

  It is easy to see that the above expression decreases with σ_δ if and only if $\gamma_A f"(e)\sigma_\delta^2 \geq k^2$, which in turn is equivalent to $\gamma_A\sigma_\delta^2 - \mu k \geq 0$.

  The intuition:

  The term $\gamma_A\sigma_\delta^2$ is the marginal cost for agent to bear more firm-specific risk;

  The term μk is the marginal benefit of loosening up the incentive compatibility constraint, when a increases;

Therefore, whether total firm-specific risk of agent ($a^{*2}\sigma_\delta^2$) will increase with the firm level firm-specific risk (σ_δ) will depend on whether the benefit outweighs the cost of incentive at the margin.

- The relation between σ_δ^2 and b^*.

  For two agents with same reservation compensation level (in dollar terms) and same preference, do we see that the intercept of linear compensation contract, b, increase with the level of idiosyncratic risk? Theoretically, it could again be ambiguous.

  Instead of looking at b directly, we look at a transformation of b, b', which is defined by $\pi = a[Z - E(\tilde{Z})] + b'$. b and b' are related by: $b = b' - aE(\tilde{Z})$.

  The mean dollar value of compensation $E(\pi)$ will be equal to b'. The *certainty equivalent* payoff, $CE(\pi)$, however, will also depend on the total firm-specific risk that the agent bears. For a fixed b' and same preference for agents, the higher the total firm-specific risk that the agent bears, the lower the certainty equivalent payment will be. Thus, given that the two agents need to have the same certainly equivalent payment to satisfy their reservation compensation level, the agent bearing more of the firm-specific risk should have a higher b' also. Since there is an ambiguity in the relation between σ_δ and $a^{*2}\sigma_\delta^2$, the level of b' is also ambiguously related to σ_δ.

  To get the exact relation between σ_δ and b, we notice that $b = b' - aE(\tilde{Z})$. Theoretically there could be an ambiguity between σ_δ and b, but when $E(\tilde{Z})$ is sufficiently large, the term $aE(\tilde{Z})$ will dominate in the expression of b, thus empirically b is typically increasing in σ_δ because a is decreasing in σ_δ. This empirical pattern is found by Garen (1994), and confirmed by my own analysis (not reported to save space).

∎.

A.1.4 When principal holds a diversified portfolio of stocks

Strictly speaking, the model in Section 2.1 is not complete. On the one hand, we assume the principal has sufficient holdings in other stocks to be well-diversified. On the other hand, in the model, the only risky asset (therefore market risk exposure) that we model is his holdings of the firm. Presumably, holding other firms' stocks – required to diversify away the un-systematic risk – should also be reflected by more systematic risk that the principal holds. Thus, by only analyzing the principal's holdings in the firm, we miss a large amount of systematic risk that he holds.

In what follows, we add the missing piece of the remaining systematic risk, by assuming that the principal already holds a large number of other stocks. As a result, the idiosyncratic risks are all diversified away, but there is a total of $S systematic risk that the principal already holds. The holdings of the firm under discussion is a *marginal* investment, and we calculate the principal's wealth change as a result of his holdings in this new firm. We will see that in the end, it is the firm's covariance with the existing systematic risk, rather than the variance in firm risk, that matters. But we still can conclude that an increase in the firm's systematic risk does not necessarily decrease the equilibrium level pay-performance sensitivity, but an increase in the firm's non-systematic risk does.

Additional Assumption:

The principal already holds $S of the market portfolio. The variance of his market portfolio is σ_S^2.

Now, the principal's incremental certainty equivalent payoff from holding (1-a) of the firm and paying a $b fixed pay is:

$$(1-a)\bar{Z} - b - \frac{1}{2}\gamma_P[(1-a)^2\eta^2\sigma_m^2 + 2(1-a)\eta\sigma_m\sigma_S] \qquad \text{(A.1.21)}$$

We implicitly assume that $\sigma_S \gg \eta\sigma_m$, that is, to the principal, the market risk from the existing portfolio is much larger than the market risk in the new firm. To a first order approximation, in the above equation, we could ignore the variance of firm's market risk, and only retain the co-variance term between the firm and the market portfolio. Thus we can write the incremental certainty equivalent payoff of the principal as $(1-a)[\bar{Z} - \eta\gamma_P\sigma_m\sigma_S] - b$, where the term $\gamma_P\sigma_m\sigma_S$ can be thought of as a "per dollar risk premium" that the principal imposes, and the whole term is the risk-adjusted payoff that the principal gets from the firm.

The principal's problem can be written as:

$$max_{a,b,e}(1-a)[\overline{Z}-\eta\gamma_P\sigma_m\sigma_S]-b \tag{A.1.22}$$

$$st.\ [a\overline{Z}+b-f(e)]-\frac{1}{2}a^2\gamma_A(\eta^2\sigma_m^2+\sigma_\delta^2)\geq\overline{CE_A} \tag{A.1.23}$$

and

$$f'(e)=ak \tag{A.1.24}$$

The Lagrangine can be written as

$$(1-a)[\overline{Z}-\eta(\gamma_P\sigma_m\sigma_S)]-b+\lambda[a(\overline{Z}+b-f(e)-\frac{1}{2}a^2\gamma_A(\eta^2\sigma_m^2+\sigma_\delta^2)-\overline{CE_A}]+\mu[ak-f'(e)] \tag{A.1.25}$$

The first order conditions are (with respect to a, b, e, λ and μ):

$$-[(Z_0+ke-\eta(\gamma_P\sigma_m\sigma_S)]+\lambda(Z_0+ke)-a\lambda\gamma_A(\eta^2\sigma_m^2+\sigma_\delta^2)+\mu k=0 \tag{A.1.26}$$

$$-1+\lambda=0, or \lambda=1 \tag{A.1.27}$$

$$(1-a)k+\lambda ak-\lambda f'(e)-\mu f"(e)=0 \tag{A.1.28}$$

$$[a(Z_0+ke)+b-f(e)]-\frac{1}{2}a^2\gamma_A(\eta^2\sigma_m^2+\sigma_\delta^2)=\overline{CE_A} \tag{A.1.29}$$

$$f'(e)=ak \tag{A.1.30}$$

Solving it, we get

$$a = \frac{\eta(\gamma_P \sigma_m \sigma_S) + \frac{k^2}{f''(e)}}{\gamma_A(\eta^2\sigma_m^2 + \sigma_\delta^2) + \frac{k^2}{f''(e)}} \tag{A.1.31}$$

When the firm has higher market risk, that means η is higher. It has two effects:

1. It increases the cost of the agent to bear the market risk in the firm, represented by the term $\gamma_A(\eta^2\sigma_m^2 + \sigma_\delta^2)$;

2. It also increases the cost of the market to bear the market risk in the firm, represented by the term $\eta(\gamma_P \sigma_m \sigma_S)$

Depending on which effect dominates, a higher market risk loading in the firm could either increase or decrease the pay-performance sensitivity, a.[5]. Therefore the conclusion of the model in Section 1.2.1 still holds.

One key insight that we gain is, once we realize that the principal holds a large amount of other assets to diversify away the non-systematic risk, that should imply he has an existing exposure to the market risk. In that case, it is the *covariance* of the new firm with the existing portfolio that affects the pay-performance sensitivity, but not the *variance* of the new firm's payoff that matters.[6] This insight carries over to the extension below, when we truly "open up" the principal's trading opportunity in the market.

extension

Here we make an extension to the above extension. The key to the model in section 1.2.1 is that the agent cannot trade the market portfolio. However, there is no reason to restrict the principal's trading opportunity. In what follows, we assume that the principal is able to trade the market portfolio. Compared to the result above, assuming the principal can trade the market portfolio allows him to *optimize* on his loading in the systematic risk. Thus the S parameter in the above model will be *endogenous.*

[5]Note that the term Z_0 does not appear in the expression, thus it does not matter whether or not the increase in market risk loading is simultaneously compensated by an increase in the mean value of the firm, to get a zero price impact on the firm.

[6]A similar point has been discussed in Garen (1994).

We assume that the principal holds a large portfolio of different stocks to diversify, and then adjusts his loadings in the systematic risk by trading the market portfolio. As a result, the idiosyncratic risk in the firm as well as in any other stocks in his portfolio are diversified away, *and* principal holds an optimal amount of the market risk.

Similar to the discussion in Section 1.2.2, the principal's new problem is:

$$max_{a,b,e}[(1-a)\overline{Z}-b]+\frac{1}{2}\frac{(\overline{r_m}-r_f)^2}{\gamma_P\sigma_m^2}-(1-a)\eta(\overline{r_m}-r_f) \tag{A.1.32}$$

$$st.\ [a\overline{Z}+b-f(e)]-\frac{1}{2}a^2\gamma_A(\eta^2\sigma_m^2+\sigma_\delta^2)\geq\overline{CE_A} \tag{A.1.33}$$

and

$$f'(e)=ak \tag{A.1.34}$$

Note that the certainty equivalent of the principal can be written as:
$\{(1-a)[\bar{Z}-\eta(\bar{r}_m-r_f)]-b\}+\frac{1}{2}\frac{(\overline{r_m}-r_f)^2}{\gamma_P\sigma_m^2}$,
and the term $[\bar{Z}-\eta(\bar{r}_m-r_f)]$ can be thought of as a market risk-adjusted mean firm value to the principal.

The Lagrangine can be written as

$$(1-a)[\overline{Z}-\eta(\bar{r}_m-r_f)]-b+\frac{1}{2}\frac{(\overline{r_m}-r_f)^2}{\gamma_P\sigma_m^2}+\lambda[a\overline{Z}+b-f(e)-\frac{1}{2}a^2\gamma_A(\eta^2\sigma_m^2+\sigma_\delta^2)-\overline{CE_A}]+\mu[ak-f'(e)] \tag{A.1.35}$$

The first order conditions are (with respect to a, b, e, λ and μ):

$$-[(Z_0+ke-\eta(\bar{r}_m-r_f)]+\lambda(Z_0+ke)-a\lambda\gamma_A(\eta^2\sigma_m^2+\sigma_\delta^2)+\mu k=0 \tag{A.1.36}$$

$$-1+\lambda=0, or\lambda=1 \tag{A.1.37}$$

$$(1-a)k + \lambda ak - \lambda f'(e) - \mu f"(e) = 0 \tag{A.1.38}$$

$$[a(Z_0 + ke) + b - f(e)] - \frac{1}{2}a^2\gamma_A(\eta^2\sigma_m^2 + \sigma_\delta^2) = \overline{CE_A} \tag{A.1.39}$$

$$f'(e) = ak \tag{A.1.40}$$

Solving it, we get

$$a = \frac{\eta(\bar{r}_m - r_f) + \frac{k^2}{f''(e)}}{\gamma_A(\eta^2\sigma_m^2 + \sigma_\delta^2) + \frac{k^2}{f''(e)}} \tag{A.1.41}$$

When the firm has higher market risk, that means η is higher. It has two effects:

1. It increases the cost of the agent to bear the market risk in the firm, represented by the term $\gamma_A(\eta^2\sigma_m^2 + \sigma_\delta^2)$;

2. It also increases the cost of the market to bear the market risk in the firm, represented by the term $\eta(\bar{r}_m - r_f)$

Depending on which effect dominates, a higher market risk loading in the firm could either increase or decrease the pay-performance sensitivity, a.

The conclusion is, once we "open up" the principal's trading opportunity on the market to allow him to adjust his loading of the market risk, we get essentially the same intuition as in Section 1.2.1. The only difference is that now the principal's personal cost of bearing the market risk is replaced by the market's cost of bearing the same risk at the margin.

A.1.5 Linear contract on both the market and firm-specific movements

In this appendix we relax the assumption of Model I and assume that the linear contract between the principal and the agent could have separate sensitivity on the market and idiosyncratic movements. We show that this is functionally equivalent to Model II, where we allow the agent to freely trade the market portfolio.

Note that the contract we study here does not necessarily equal to relative performance evaluation, which proposes a zero sensitivity of pay to market movement. In fact, as become clear later, the sensitivity of pay to market movement is always positive here, and in some cases could even be higher than the sensitivity of pay to the idiosyncratic movement.

We assume that the setup of the model is exactly the same as in Section 1.2.1, except that the compensation contract rewards the market performance and the idiosyncratic performance differently:

$$\pi_A = a_1\eta(\tilde{r}_m - \bar{r}_m) + a_2(\bar{Z} + \tilde{\delta}) + b \tag{A.1.42}$$

The new certainty equivalents for the agent and the principal are, respectively,

$$CE_A = a_2\tilde{Z} + b - f(e) - \frac{1}{2}\gamma_A a_1^2\eta^2\sigma_m^2 - \frac{1}{2}\gamma_A a_2^2\sigma_\delta^2, \tag{A.1.43}$$

and

$$CE_P = (1 - a_2)\tilde{Z} - b - \frac{1}{2}\gamma_P(1 - a_1)^2\eta^2\sigma_m^2. \tag{A.1.44}$$

The principal's problem now becomes,

$$max_{a_1,a_2,b,e}[(1 - a_2)\overline{Z} - b] - \frac{1}{2}(1 - a_1)^2\gamma_P\eta^2\sigma_m^2 \tag{A.1.45}$$

$$st.\ [a_2\overline{Z} + b - f(e)] - \frac{1}{2}\gamma_A(a_1^2\eta^2\sigma_m^2 + a_2^2\sigma_\delta^2) \geq \overline{CE_A} \tag{A.1.46}$$

and

$$a_2k = f'(e) \tag{A.1.47}$$

We can write out the Lagrangine:

$$[(1-a_2)\bar{Z}-b]-\frac{1}{2}\gamma_P(1-a_1)^2\eta^2\sigma_m^2+\lambda IR+\mu[a_2k-f'(e)], \tag{A.1.48}$$

where

$$IR=a_2\tilde{Z}+b-f(e)-\frac{1}{2}\gamma_A(a_1^2\eta^2\sigma_m^2+a_2^2\sigma_\delta^2)-\overline{CE_A}. \tag{A.1.49}$$

The first order conditions are (with respect to a_1, a_2, b, e, λ and μ):

$$\gamma_P(1-a_1)\eta^2\sigma_m^2=\lambda\gamma_A a_1\eta^2\sigma_m^2 \tag{A.1.50}$$

$$-\bar{Z}+\lambda Z-\lambda\gamma_A a_2\sigma_\delta^2+\mu k=0 \tag{A.1.51}$$

$$-1+\lambda=0 \tag{A.1.52}$$

$$(1-a_2)k+\lambda a_2k-\lambda f'(e)-\mu f''(e)=0 \tag{A.1.53}$$

$$a_2\bar{Z}+b-f(e)-\frac{1}{2}\gamma_A[a_1^2\eta^2\sigma_m^2+a_2^2\sigma_\delta^2]-\overline{CE_A}=0 \tag{A.1.54}$$

$$a_2k=f'(e) \tag{A.1.55}$$

By plugging (A.1.55) into (A.1.53), we get

$$\mu=\frac{(1-a_2)k}{f''(e)} \tag{A.1.56}$$

Plugging (A.1.53) into (A.1.50),

$$\gamma_P(1-a_1)\eta^2\sigma_m^2 - \gamma_A a_1 \eta^2 \sigma_m^2 = 0 \tag{A.1.57}$$

or

$$a_1 = \frac{\gamma_P}{\gamma_A + \gamma_P} \tag{A.1.58}$$

That is, the principal and the agent divide the market risk according to their respective risk aversion, to equalize their marginal costs of bearing the market risk.

Plugging (A.1.53) into (A.1.51), we can solve for a_2

$$a_2 = \frac{\mu k}{\gamma_A \sigma_\delta^2} \tag{A.1.59}$$

which by (A.1.56) becomes

$$a_2 = \frac{(1-a_2)k^2}{\gamma_A f''(e)\sigma_\delta^2} \tag{A.1.60}$$

Thus,

$$a_2 = \frac{k^2}{\gamma_A f''(e)\sigma_\delta^2 + k^2} \tag{A.1.61}$$

This is the same result on the sharing of the firm-specific risk as in Model II. Thus, when CEO cannot trade the market portfolio, but the firm can design contracts on both the market movement and the idiosyncratic movement, the resulting contract will equalize the marginal cost of bearing market risk for both the principal and the agent. Then, the sharing of the firm-specific risk will be independent on the level of market risk.

A.1.6 The effect of monitoring

This appendix extends the basic model into a case when monitoring is available. We prove that under certain circumstances monitoring and incentive are substitutes: higher incentive is needed when monitoring is inefficient. The empirical works of Lambert and Larcker (1987), Yermack (1995) and Bryan, Hwang and Lilien (2000) provide some support of this.

We assume that at zero cost, the principal can have a noisy signal $\widetilde{S}$ on the true effort level e:

$$\widetilde{S} = ke + \widetilde{\epsilon} \tag{A.1.62}$$

where $\widetilde{\epsilon}$ is a white noise with $\widetilde{\epsilon} \sim N(0, \sigma_\epsilon^2)$.

Again, we assume linear contracting, and recognize that now the principal can contract on both the output $\widetilde{Z}$ and the signal $\widetilde{S}$. Therefore the general form of the contract will be:

$$\pi_A = a_1\widetilde{Z} + a_2\widetilde{S} + b \tag{A.1.63}$$

and

$$\pi_P = (1 - a_1)\widetilde{Z} - a_2\widetilde{S} - b \tag{A.1.64}$$

We assume that the term ϵ is idiosyncratic, thus the principal will not care about the risk induced by it. However, agent will care about the ϵ risk since he will not be able to diversify it. Now, the certainty equivalent payoff of the principal is:

$$CE_P = [(1 - a_1)\overline{Z} - a_2ke - b] - \frac{1}{2}(1 - a)^2\gamma_P\eta^2\sigma_m^2, \tag{A.1.65}$$

whereas the certainty equivalent payoff of the agent changes to:

$$CE_A = [a_1\overline{Z} + a_2ke + b - f(e)] - \frac{1}{2}\gamma_A[a_1^2(\eta^2\sigma_m^2 + \sigma_\delta^2) + a_2^2\sigma_\epsilon^2]. \tag{A.1.66}$$

The maximization problem now becomes:

$$max_{a_1,a_2,b,e}[(1 - a_1)\overline{Z} - a_2ke - b] - \frac{1}{2}(1 - a)^2\gamma_P\eta^2\sigma_m^2 \tag{A.1.67}$$

$$st.\ [a_1\overline{Z} + a_2ke + b - f(e)] - \frac{1}{2}\gamma_A[a_1^2(\eta^2\sigma_m^2 + \sigma_\delta^2) + a_2^2\sigma_\epsilon^2] \geq \overline{CE_A} \tag{A.1.68}$$

and

$$G(e) \geq G(e'), \forall e', \tag{A.1.69}$$

where $G(e) = [a_1\overline{Z}(e) + a_2ke + b - f(e)] - \frac{1}{2}\gamma_A[a_1^2(\eta^2\sigma_m^2 + \sigma_\delta^2) + a_2^2\sigma_\epsilon^2]$. When effort is a continuous choice, we could again replace (A.1.69) with:

$$a_1k + a_2k = f'(e) \tag{A.1.70}$$

Proposition 3 *The above problem has a solution:*

$$a_1^* = \frac{[\gamma_P\eta^2\sigma_m^2 + \frac{k^2}{f"(e)}][\gamma_A\sigma_\epsilon^2 + \frac{k^2}{f"(e)}] - [\frac{k^2}{f"(e)}][\frac{k^2}{f"(e)}]}{[\gamma_P\eta^2\sigma_m^2 + \gamma_A(\eta^2\sigma_m^2 + \sigma_\delta^2) + \frac{k^2}{f"(e)}][\gamma_A\sigma_\epsilon^2 + \frac{k^2}{f"(e)}] - [\frac{k^2}{f"(e)}][\frac{k^2}{f"(e)}]} \tag{A.1.71}$$

and

$$a_2^* = \frac{\gamma_A(\eta^2\sigma_m^2 + \sigma_\delta^2)\frac{k^2}{f"(e)}}{[\gamma_P\eta^2\sigma_m^2 + \gamma_A(\eta^2\sigma_m^2 + \sigma_\delta^2) + \frac{k^2}{f"(e)}][\gamma_A\sigma_\epsilon^2 + \frac{k^2}{f"(e)}] - [\frac{k^2}{f"(e)}][\frac{k^2}{f"(e)}]} \tag{A.1.72}$$

$$e^* = f'^{(-1)}((a_1^* + a_2^*)k) \tag{A.1.73}$$

and

$$b = \overline{CE_A} - [a_1^*\overline{Z} - a_2^*ke - f(e^*)] - \frac{1}{2}\gamma_A[a_1^{*2}(\eta^2\sigma_m^2 + \sigma_\delta^2) + a_2^*\sigma_\epsilon^2] \tag{A.1.74}$$

Proof:

We can write the Lagrangine:

$L = (1 - a_1)(Z_0 + ke) - a_2 - b - \frac{1}{2}\gamma_P(1 - a_1)^2\eta^2\sigma_m^2 + \lambda[a_1(Z_0 + ke) + a_2ke + b - f(e)$
$-\frac{1}{2}\gamma_A[a_1^2(\eta^2\sigma_m^2 + \sigma_\delta^2) + a_2^2\sigma_\epsilon^2] - \overline{CE_A}] + \mu[(a_1 + a_2)k - f'(e)]$

The first order conditions are (with respect to a_1, a_2, b, e, λ and μ):

$$-(Z_0 + ke) + (1 - a_1)\gamma_P\eta^2\sigma_m^2 + \lambda(Z_0 + ke) - a_1\lambda\gamma_A(\eta^2\sigma_m^2 + \sigma_\delta^2) + uk = 0 \tag{A.1.75}$$

$$-ke + \lambda ke - \gamma_A \lambda a_2 \sigma_\epsilon^2 + uk = 0 \tag{A.1.76}$$

$$-1 + \lambda = 0, or \lambda = 1 \tag{A.1.77}$$

$$(1 - a_1)k - a_2 k + \lambda(a_1 k + a_2 k - f'(e)) - \mu f''(e) = 0 \tag{A.1.78}$$

$$[a_1 \overline{Z} + a_2 ke + b - f(e)] - \frac{1}{2}\gamma_A [a_1^2(\eta^2 \sigma_m^2 + \sigma_\delta^2) + a_2^2 \sigma_\epsilon^2] = \overline{CE_A} \tag{A.1.79}$$

$$a_1 k + a_2 k - f'(e) = 0 \tag{A.1.80}$$

Solving them we get the solution in proposition 3.

∎.

We give some interpretation of this result. When a signal exists, it can be used to refine the estimated effort level and thus improve the contract. However, such signal also introduces another *noise* and would add some randomness to the contract. The usefulness of signal therefore depends on its variance. When the signal gets very noisy (represented by a large σ_ϵ), it becomes less useful. In the extreme, when the signal has an infinite variance, we are back to the case of no monitoring. This can be seen by letting σ_ϵ in equation (A.1.71) approach infinity, and verify that the expression does go to equation (1.2.7). However, as long as the variance of the noise term is not infinity, there will be *some* usefulness of the signal[7].

In the intermediate cases when σ_ϵ^2 is neither zero nor infinity, it can be proved that the optimal pay performance sensitivity a_1^* is increasing in σ_ϵ^2. On the other hand, a_2 is decreasing in σ_ϵ^2. The intuition is: when a more precise signal is available, we could afford

[7]In his classical paper, Holmstrom (1979) points out that "... one might conjecture that in some situations a sufficiently noisy, yet informative signal would add too much randomness to the contract to be acceptable by risk averse parties. But ... since both parties are on the margin risk-neutral towards randomness in [the signal] y, given [the output level] x, the new contract can be designed so that marginally it does not increase risk, but still improves incentive for action." (Page 87) This argument, although made in a slightly different context, accords well with the intuition we have here.

to load CEOs with less incentive, but then penalize the CEOs relatively heavily when the signal indicates that they are not exerting the right effort. Thus, there is a substitution effect between monitoring on effort and incentive pay.

A.2 Appendices of chapter 1 on the empirical part

A.2.1 Measuring performance

Dollar versus percentage

The executive compensation literature has examined performance as defined both in terms of dollars and percentages. As posited by Aggarwal and Samwick (1999a) and Murphy (1999), the percentage performance makes less economic sense because it ignores the difference in firm size and can treat a large dollar return in a large firm and a smaller dollar return in a small firm equally. As seen in the theory model, the optimal contract should equalize the marginal cost of effort to some measure of the performance on the margin. It is very awkward to equalize the cost of effort to the percentage rate performance without involving the market size. In short, in our model, the percentage rate version of the model is mis-specified.[8] Another advantage for focusing on the dollar performance is that it allows us to ignore the issue of firm leverage on total risk. When the firm is levered, the percentage risk of the firm's equity is not the same as the percentage risk of the firm asset, which is what we should measure. But as long as we assume that debt is not risky, the total dollar risk of the firm equity will be equal to the total dollar risk of the firm asset. Therefore, in what follows, I focus on the dollar performance. Performance is defined as the market value of the firm at the beginning of the period multiplied by some benchmark-adjusted rate of return during that period.

Performance benchmark

What is the benchmark against which performance can be calculated? Some examples of benchmarks are:

- Zero

  The performance is defined as nominal dollar return.

[8]Gibbons and Murphy (1992) give an empirical defense for using the percentage performance. They argue that the regressions using percentage performance or elasticity measure empirically fit the data better.

- Inflation rate

  The performance is defined as real dollar return.

- Market return

  Performance is defined as market value*$(r_i - r_m)$, which is the dollar return on the stock in excess of a stock of comparable market value, but with the market return.

- Industry average return

  Performance is defined as market value*$(r_i - r_{IND})$, which is the dollar return on the stock in excess of a stock with comparable size, but with the industry average return.

The first two do not assume relative performance evaluation, whereas the next two are relative performance evaluation benchmarks.

As discussed in the first section, although the ideal arrangement is to link CEO pay to relative performance, empirically there is no strong evidence that this is what firms actually do. Furthermore, I have argued that it might be justified to reward CEOs for systematic risk also. Empirical studies have used both the total performance and the relative performance benchmarks. For example, Coughlan and Schmidt (1985) use the abnormal return from CAPM. Murphy (1985) reports that it does not seem to matter whether the raw return or the relative performance return is used, as long as the two are not used simultaneously.

A.2.2 Measuring pay and pay-performance sensitivity

This section closely follows the discussions in Jensen and Murphy (1990a) and Hall and Liebman (1998). Two components of CEO wealth are influenced by firm performance — human capital and financial capital. The changes in these two components can be measured by total direct compensation, and re-evaluation of stock and stock options holdings, respectively.

1. Total direct compensation

   This includes the salary and bonus, plus the value of any grants in stock, options, long-term incentive compensations and others, that CEOs receive in a given period.

2. Re-evaluation of stock and stock options holdings

CEOs often have large holdings of stock and stock options from the firm, the value of which changes with firm performance.

Total direct compensation change

As Jensen and Murphy (1990a) point out, two different assumptions can be made about how the firm performance affects the total direct compensation. The conservative assumption is that the firm's performance only affects the current period's total direct compensation and has no influence on future levels of total direct compensation. If this is so, we can add the total direct compensation from the current period to the change in CEO stock and options value due to re-evaluation, to calculate the total change in CEO wealth.

The second assumption is that there is a permanent effect of performance on total direct compensation. A good firm performance in the current period not only increases current pay, but also increases the future pay for the CEO as well. Under this assumption, firm performance has a lingering effect on CEO pay, and as Jensen and Murphy (1990a) point out, we should compute the present value of the changes in future compensation cash flows, as a result of this year's performance. In order to do so, Jensen and Murphy (1990a) first use a regression to estimate the effect of firm performance on the current and next period pay change. Then, using these estimates, they assume a real annual interest rate of 3 percent, and further assume that all the changes in salary and bonus are permanent after the second year. They also assume that the CEO will retire at age 70, and calculate the present value of the direct compensation cash flow for the remaining tenure of the CEO.

The assumption that firm performance only influences current period direct compensation may be too conservative. On the other hand, the assumption that there is a permanent effect of firm performance on CEO pay could be too strong. Jensen and Murphy (1990a) notice that such treatment may exaggerate the effect of performance on pay. For one thing, the bonus component of direct compensation is likely to be transitory. I expect the truth to lie somewhere in between. Thus, I calculate the two as lower bound and upper bound, respectively, and test both in my regression analysis later on. I also adhere to all the assumptions that Jensen and Murphy (1990a) made when calculating the upper bound of human capital change due to performance.

Re-evaluation of stock and stock options

CEOs typically hold a fair amount of stock and stock options in the firm. Re-evaluation of these financial assets can cause major change in CEO wealth due to firm performance. Hall and Liebman (1998) conclude that the re-evaluation effect dwarfs the change in total direct compensation in terms of providing incentive.

To calculate the magnitude of CEO wealth change due to re-evaluation of stock and stock options, we need to know the amount of stock holdings, the amount, strike price and maturity of options, and other parameters that affect the option value, such as volatility of stock, the dividend yield, the continuously compounded riskless interest rate, and the current stock price. We have full information about stock holdings in our data set, but some information about option holdings is missing, and therefore must be approximated.

As with most of the other data set used in the compensation analysis, the ExecuComp data set has complete information only for the current year options grant. For the previous year options grants, the data set divides those options into exercisable options and un-exercisable options, and reports the amount of options and the intrinsic value of in-the-money options in each category, where the intrinsic value of an in-the-money option is defined as the payoff one can get by exercising the option immediately. To get the full information of holdings of options granted in previous years, some assumptions need to be made and the options holdings must be estimated. For example, Murphy (1999) develops a method that is adopted by Aggarwal and Samwick (2000), and Hall and Liebman (1998) offer another method. I use the method developed by Core and Guay (1999), which is similar to the Hall and Liebman approach. Core and Guay (1999) demonstrate that their method yields estimates of option portfolio sensitivities that are effectively unbiased and 99% correlated with the measures that would be obtained if the parameters of a CEO's option portfolio were known.

Two measures of CEO stock and stock option value change are of interest here: the realized change and the hypothetical change for a $1,000 increase in firm performance.

One caveat in calculating the realized CEO stock and option value change is that this measure cannot be directly calculated as the difference between year t and t-1 of the end-of-year value of CEO stock and options portfolio. The reason is that not all of the changes in the year-to-year financial wealth can be attributed to firm performance. CEOs might sell off some stock or exercise some options or they might be awarded some more stock or stock options during a year. This will affect the year-to-year change in CEO financial wealth, but it has nothing to do with firm performance.

I calculate the realized change in idiosyncratic financial wealth as follows. I first determine the amount of stock and stock options that a CEO holds at the beginning of year. Then, for the stock, I calculate the beginning-of-year market value and the realized return over the year. The product of the two will be my measure of the change in CEO stock portfolio value for that year. For options, I calculate the Black-Scholes value for each option at the beginning and end of a year, and the difference between the two. I then multiply the difference with the amount of options that the CEO holds at the beginning of the year to determine the value change in CEO option holdings.

I also calculate the hypothetical increase in CEO financial wealth due to a \$1,000 increase in shareholder value. To do so, I first calculate the change in CEO financial wealth when the firm return increases from the median return of the market to 1% above the median return. Then I divided this number by the change in total shareholder value due to the 1% increase in firm return, and multiply the result by 1,000.

One area where I differ from existing literature involves calculation of the sensitivity of an option's value to the stock price. Many investigators use the option's delta δ (hedge ratio) to measure the pay-performance sensitivity. While this is roughly correct, it fails to consider the time value of options. For any options that a CEO holds at the beginning of the year, if the stock price over the year increases by \$1, then the option's value does not increase by $\$\delta$, because there is shortening by one year of the option's maturity. In fact, if the stock price increase is small, the option value can actually decrease. My measure is typically lower than that calculated from directly using the option hedge ratio, because of the time value of options.

The CEO wealth change due to options is added to the CEO wealth change due to stock, to arrive at the total CEO wealth change due to the re-evaluation of stock and options. This value is then added to CEO human capital value change to obtain total CEO wealth change due to firm performance. Following Jensen and Murphy (1990a), I measure the Pay-Performance Sensitivity (PPS) as the change in CEO total wealth for a \$1000 increase in firm value. In the regression analysis, all the measures of pay-performance sensitivity are calculated using 1994 constant dollars.

Truncating the outliers

The raw measures of pay-performance sensitivity are highly skewed and fat-tailed as a consequence of outliers. In subsequent analysis, I employ a truncated measure where I replace the top and bottom 1% of the data with the 99th percentile and 1st percentile values, respectively.

A.2.3 Constructing risk measures through the market model regression

To construct the industry average firm-specific risk, I first define the industry as a two-digit SIC code. Then, I estimate the idiosyncratic risk using a market model. I use all the CRSP firms (rather than only firms in the executive compensation data) to construct the measure of industry average risk.

I notice that the equal weighted firm-specific risk measures are rather high: mean of 0.026433 and median of 0.026763 for monthly firm-specific risk, and commensurate measures for daily and weekly firm-specific risk. This implies an annual variance in the region of 0.25, or 50% standard deviation annually. The average of firm-specific risk measure is also of the same magnitude.

However, by restricting to the executive compensation database only, we get an annualized firm-specific risk of 0.0322, almost an order of magnitude lower. We thus suspect that there are some difference in the characteristics in the riskness of data in CRSP in general, and the data in executive compensation data set, and we propose the value weighted average firm-specific risk measure instead, for the following argument.

Campbell et al. (2000) documented that smaller firms have higher firm-specific risk variance. The CRSP data that I use to construct the industry average risk measures include many small firms, whereas the executive compensation data set contains some of the largest firms. Thus it is likely that the equal weighted industry average firm-specific risk variance will pick up more of the effect of the small stocks. This caused me to use the value weighted average variance to measure the industry level of firm-specific risk. This measure is analogous to those that Campbell et al. (2000) proposed.

We measure firm-specific risk, market risk and total risk. The later two types are included for a consistency check with the literature on the pay-performance sensitivity studies, and to demonstrate the validity of using firm-specific risk.

We do the estimation with the CRSP monthly data file, crsp.msf, available in SAS data format from WARD research database maintained by Wharton School.

To estimate the several risk measures, we estimate the CAPM model in Equation (??). To be consistent with the convention, unlike in the compensation calculation, every rate here is nominal. Changing everything into real terms will not affect the resulting level of firm-specific risk, although it will slightly affect the alpha for each firm.

As a consistency check, we compare the calculated risk measures with those calculated by the PERC Research Network at the Tuck School, which estimates the various risk measures using CRSP data. PERC Research Database is a database that uses the CRSP data to calculate many summary statistics. It is previously maintained by Terence Lim, formerly assistant professor at Dartmouth and currently with Goldman Sachs Asset Management.

For each year and each stock we want to study, we estimate the market model for the 60 months that end in the last month of the preceding year.

$R_{i,t} - r_{f,t} = \alpha_i + \beta_i(r_{m,t} - r_{f,t}) + \epsilon_{i,t}$

I require that at least 48 monthly observations exist for the last 60 months before we include that firm in the regression. The requirement of sufficient data points ensures that the estimated regression coefficients are less affected by measurement errors.

We record the firm-specific risk $VAR(\epsilon_i)$, the market risk $\beta_i^2 Var(r_m)$ and the total risk σ_i^2, and perform a correlation analysis with the reported result from the PERC database.

The resulting measures are very highly correlated with the measure for beta and firm-specific risk of PERC. The correlation coefficient is always above 95% and usually above 99%, indicating a good match with existing measures, which is reassuring.

For a robustness check, I also calculated the residual variance of the following specification.

$r_{i,t} = \alpha_i + \beta_i r_{m,t} + \epsilon_{i,t}$.

The resulting risk measures are almost the same. For simplicity, I will only do regression with the market model specification.

The default risk measures are calculated using a five-year monthly regression. As a robustness check, I also included risk measures for several shorter horizon: 1 year, weekly data, and 3 year, weekly data. In the weekly regressions, to guard against non-synchronous trading, I adopt an approach as in Fama and French (1992) and included 4 lagged terms.

A.2.4 Campbell et al. (2000) method

Some detailed procedures to get the Campbell et al. (2000) measure of firm-specific risk.

This is the refinement that they described in Section II.C, individual industries, on p19 of their paper.

1. To get the industry level risk, we do a CAPM regression and find the residual return for each industry, after the market component, $\beta_{industry} R_{market}$ is removed. We then define the industry specific risk as the variance of the residual return over the sample period, which in our basic model is 5 years. This amounts to doing a CAPM regression and using the mean squared error as the industry specific risk measure.

2. to get the average firm-specific risk within an industry, we follow the discussion of equations (26) and (27) in Campbell et al. (2000). We further note that from (26) and (27), we can write as follows.

$R_{ijt} = R_{it} + \eta^*_{ijt}$, thus $\eta^*_{ijt} = R_{ijt} - R_{it}$

Thus, we can estimate the η^*_{ijt}, and then take the variance over the sample period, which will be the measure of variance of an individual stock's firm-specific risk, $Var(\eta^*_{ijt})$. Then, we take the weighted average over all the stocks in the industry. The weight is proportional to the lagged market cap.

For general case, the non-industry specific version is similarly defined. First, the FIRM risk is the same as before, just that at above the industry average level of firm-specific risk, there is now a layer of value weighted industry average.

The industry risk IND, is even simpler to calculate: it is

$\epsilon it = R_{it} - R_{mt}$,

Then we estimate the variance of ϵ_{it} over the sample period, and take the value weighted average.

The only thing that needs to be computed is the industry total (lagged) market value.

A.2.5 Proxies for the productivity of CEO effort

One goal of the present research is to empirically test for the other side of the diversification-incentive tradeoff, and determine how productivity of CEO effort is related to the pay-performance sensitivity. Several proxies for CEO effort productivity are proposed.

- Investment

  Productivity of CEO effort could be positively related to the amount of investment the firm is making. If firms are undertaking a great deal of investments, the CEOs will likely have major influence on the firms' value through making important investment decisions. Alternatively, if the CEOs just sit on top of existing assets, then their effort might not contribute too much to firm value.

  This prompts the use of investment to proxy for the productivity of CEO effort on firm value. Following Aggarwal and Samwick (2000), I measure the investment as capital expenditures for property, plant and equipment (COMPUSTAT annual file, item Data30) divided by the stock of net property, plant and equipment (COMPUSTAT Annual File, item Data8). Again, here the main reason to use this measure of investment is for consistency with existing literature. But because it does not take into account the effect of intangible assets, it is likely to create some bias in industries with a large percentage of investment in intangibles. Also, because divestment could be equally important to firm value as investment, it is desirable for a measure of the absolute level of investment plus divestment to be constructed.

- Firm age

  It has been argued that younger firms might have more growth potential, and thus their CEOs are more important in developing the firm through selecting the right projects and exerting effort. I compute the firm age using the starting year that the firm is listed on the CRSP database.

  Again, there are some drawbacks in the measurement of firm age. First, a new (old) firm could have old (new) divisions. Second, the year a firm lists on CRSP is typically the year it goes public, but a firm could stay private for many years.

- Tobin's Q.

  Tobin's Q can be related to future growth opportunity. A firm with a higher Tobin's Q will likely be affected more by CEO effort. We measure Tobin's Q as the total market value of firm (equity plus debt), divided by the book value of firm (equity plus debt). Data limitations forces us to use the book value of debt in places where we would have to use the market value of debt. As Bowman (1980) pointed out, the cross-sectional correlation between the book value and market value of debt is pretty large, thus we hope this measurement error will not affect our result severely. As a robustness check, we also tried to focus on the ratio of market value of equity over book value of equity to isolate the most accurately measured portion of the Tobin's Q, and use that measure in the regression analysis. That does not change the result substantially. In COMPUSTAT annual file, our measure of the Tobin's Q is $\frac{(DATA199*DATA25+DATA9+DATA34)}{(DATA60+DATA9+DATA34)}$.

A.2.6 Robustness of the results

This subsection contains some of the robustness checks that I conducted.

CEO's own valuation of stock and options package

The approach to value CEO's stock and stock options package, although conventional, is likely to exaggerate the real value of the financial wealth to the CEOs. Due to inefficient risk bearing, CEOs find it costly to bear the firm firm-specific risk embedded in the stocks and stock options of their own firms. They therefore value these less than the market valuation. Meulbroek (2000a) directly addresses this issue. As a robust check, I also estimate the CEOs' own valuation of their stock and stock options portfolio according to the Meulbroek (2000a) method, and re-estimate the pay-performance sensitivity measures accordingly. I then perform a regression analysis to assess the relation between this revised measure of pay-performance sensitivity and the measures of firm-specific risk. I obtain similar results as before.

Another issue that is specifically related to options is the question of whether the Black-Scholes formula is appropriate to characterize the CEO's valuation in this context. To use the Black-Scholes formula, we implicitly assume that CEOs can dynamically replicate the options through trading on a portfolio of the firm stock and bond. This assumption is unrealistic in the setting of CEO compensation. As a robustness check to that, I further reduce the valuation of all the options by 30%, and found that such a reduction does not change the main results of the paper.

Accounting measure

By using the market model to estimate CEOs' risk, I implicitly assume that the CEO is just like a typical outside investor, i.e., the firm-specific risk is neither controllable nor observable to the CEO, just as it is to the outsider. What if the CEO know and/or have some control over the firm return process? If so, some components of the risk to outside investor might not be risky at all to the CEO, and thus will not represent a cost to him. Since CEO compensation trades off the cost of bearing firm-specific risk with incentive, we can refine our measure of the real risk to CEOs by filtering out those components that CEOs might know and/or can control. One strong assumption is that the CEO can observe the accounting return before investors do. Thus the firm-specific risk of real concern to the CEO is the component that is unpredicted by the accounting returns. By filtering out what the CEO might know, I could refine the measure of the risk that the CEO neither control nor know beforehand.[9]

I first run the market model regression. Then, I regress the residual terms on the accounting returns. The mean squared error of this second regression will measure how much of the market model regression residual cannot be explained by the accounting return, and I use this adjusted measure of firm-specific risk to measure the firm firm-specific risk that is not predicted by accounting performance. I test whether using this measure to replace the previous definition of firm-specific risk will influence the result. It gives very similar results as before.

Sub-sample of firms with full options data

I employ the ExecuComp data to determine the estimation of CEO compensation. One limit of the data set is the lack of exact exercise price information for the out-of-money options. The method that I use to approximate the option holding information is developed by Core and Guay (1999). Although it is one of the best methods available, it has the potential drawback of underestimating the true exercise prices of the out-of-money options. Thus the pay-performance sensitivity we calculate could be biased. I want to guard against the miscalculation of options exercise price and, therefore, the total incentive. Since typically

[9] A more general framework would be to endogenize the risk level of the firm, and address both the incentive level and risk level of the firm simultaneously. There have been theoretical studies addressing this issue, for example, Core and Qian (2000). However, their model generates a non-monotonic (first decreasing, then increasing) and non-linear pay schedule, which is not observed often in the real compensation contract. Since the purpose of this paper is mainly to empirically address the effect of risk on pay, I choose instead to use the simplified approach above to address the endogeneity of risk.

CEO stock options are granted for 10 years and they are granted at the money, if the current stock price is higher than the stock prices in any of the previous 10 years, then the options are all in-the-money and there is no miscalculation of the options exercise price.

For that reason, I constructed a sub-sample where the current stock price is higher than the stock prices in any of the previous 10 years (after adjusting for stock split, but not adjusting for dividend because typically executive stock options are not protected for dividend). I find that for this sub-sample, we have a similar result: there is a negative relation between non-systematic risk and incentive, but no statistically significant relation between systematic risk and incentive. The effect is similar to the full sample result.

Sub sample of firms with positive dividends

One hypothesis is that firms with severe cash constraints might refrain from using cash compensation such as salary and bonus, and resort to restricted stock and options instead. If so, we might observe a higher pay-performance sensitivity just because of cash constraint. Also, to some extent, it might be argued that firm age might be associated with cash constraint. If so, one of the proxies for CEO effort productivity could be picking up the effect of cash constraint.

One way in the literature to address the cash constraint is to assume that cash constrained firms do not pay dividend. Out of 9714 firm-year observations, 6261 firm-year, or roughly two thirds of the observations, have positive dividend payment. I therefore constructed a sum-sample with positive dividend and assume that at least for this sub-sample the cash constraint is not severe. The main results of the paper is not changed when we focus on this sub-sample.

Hardwired relation between risk and incentive in options

One question to address is: is the observed relation between risk and incentive a result of the relation between total risk and the delta of the options, which is hard-wired in my structure? If the sensitivity of option value to stock price decreases with total volatility, then, for a fixed options portfolio that the CEO holds, it will *look like* that his pay-performance sensitivity is decreasing in firm-specific risk.

We argue that my result is unlikely all due to that. In fact, for many of the options, the hard-wired relation between risk and option's sensitivity to stock will probably go the other direction than as predicted by the risk sharing-incentive tradeoff model.

The sensitivity of option value to the stock price is approximately captured by the delta of the options: $\Delta = \frac{\partial C}{\partial S}$. We need to know when a higher total risk leads to a higher delta.

It can be shown that (Cf Hull (2000), P316):

$$\Delta = e^{-dT} N(Z) \tag{A.2.81}$$

Where d is the expected dividend rate over the life of the option, and $Z = \frac{ln(\frac{S}{X})+T(r-d+\frac{\sigma^2}{2})}{\sigma\sqrt{T}}$, r is the continuously compounded riskfree rate, σ is the standard deviation of continuously compounded return, S is current stock price, X is the exercise price of options, T is the maturity of options.

Note that typically $S = X$, thus $ln(\frac{S}{X})$ drops out. The partial derivative of Z to σ is:

$$\frac{\partial Z}{\partial \sigma} = \sqrt{T}(-\frac{r-d}{\sigma^2} + \frac{1}{2}), \tag{A.2.82}$$

which is positive if $\sigma^2 > 2(r-d)$.

Our assumption of r is 0.06. We calculate the stocks for which the above condition holds, using the dividend yield and the standard deviation of continuously compounded return, both reported in the ExecuComp database. We find that 5703 out of 9491 CEO/year observations with full option information satisfy this condition. For these stocks, a higher total risk will increase the delta of options. So, roughly speaking, the predicted pay-performance sensitivity for the options of these stocks should *increase* with firm-specific risk. This is contrary to what we observed. Thus the hardwired relation between risk and option delta cannot explain the observed negative relationship between firm-specific risk and pay-performance sensitivity in the data.

Impact of early exercise of CEO stock options

To calculate the CEOs' exposure to firm risk through financial asset holdings, I implicitly assume that the executive stock options are European and thus the CEOs hold their stock options to the expiration date. In practice, CEOs can, and often do, exercise their options before expiration. This might affect the accuracy of our results.

As a robustness check, I also calculate the pay-performance sensitivity, assuming that CEOs actually exercise their options three years before it matures. Very similar results were obtained.

Controlling for size effect

The cross sectional level of CEO's pay-performance sensitivity changes predictably with firm size: larger firm's CEO has lower pay-performance sensitivity. To make sure that the main result I document is not just driven by size, in the main regression analysis, I have added the market capitalization and the square of market capitalization as control variables. I have also tried to use the log of market capitalization.

Also, we could argue that for large firms, the job of controlling the whole firm is too complicated to be handled by any single manager. Therefore, to some extent, the top management team, rather than a single CEO, is acting the role of CEO, while at small firms, the CEO alone is making all the important decisions. Thus, in a large firm, although the CEO himself might be taking a smaller responsibility and thus a smaller share in the firm performance, the whole management team should still be taking a comparable responsibility as a smaller counterpart. The ExecuComp database reports the compensation of the five most highly compensated executives in the firm in each year, which can be used to construct the total pay-performance sensitivity of whole the management team. Compared to the pay-performance sensitivity measure of a single firm, this measure is likely to be more robust to size related heterogeneity in pay-performance sensitivities. I used this pay-performance sensitivity for the whole team to re-do the empirical analysis, and the results are qualitatively similar.

As argued by Berk (1995), size might act as a catch-all variable for any mis-specified risk factors. Thus, by using size as a direct control variable, we might actually mix some of the priced risk with this control variable. For that reason, book value of assets and the sales are also tried to replace the market capitalization as control variables. The results are qualitatively similar.

Convertibility of securities and dilution effect

The main analysis implicitly assumes that the company has fixed amount of equity outstanding. Thus, the pay-performance sensitivities can be interpreted as the proportion of the total outstanding equity that the CEO holds. In reality, things can be complicated due to existence of convertible securities. When the firm does well, many convertible securities might be converted into equity. This implies that the CEO's pay performance sensitivity is lower on the upside and higher on the downside.

Since COMPUSTAT does not give the exact terms of conversion, we could only approximately address the issue of convertibility. One measure is the book value of convertible securities as a percentage of the sum of book value of equity plus book value of convertible securities. If that measure is large, then the convertibility of securities might have a large impact on pay-performance sensitivities. In our dataset, less than 10% of the firm-year has more than 10% dilution. If we define the dilution as the ratio of total book value of convertible securities over the sum of *market value* of equity plus book value of convertible securities, less than 5% of the firm-year would have more than 10% dilution. So generally speaking, the effect of dilution is not large.

As a robustness check, we also constructed a modified measure of CEO pay-performance sensitivity, where we treat the convertible securities as if they were common stocks. When we use this modified measure of CEO pay-performance in our regression analysis, the results are qualitatively the same.

The impact of stock options resetting on pay-performance sensitivity

In practice, firms sometimes reset the options after a severe underperformance. We want to know how that affects our analysis.

Intuitively, option resetting will decreases the ex-ante incentive, because CEOs will build in the expectation that when firms performs poorly, the CEO is less punished, since they are given a "windfall" of wealth through option resetting. However, the effect is small for practical concerns, as shown by a recent paper by Brenner et al. (2000).

Resetting typically occurs when the firm has a severe underperformance. The probability of resetting increases with the degree of underperformance. Take 1995 as example, 75% of firms have return of 2% or higher. According to Brenner et al.(2000), at that performance level, the chance of a resetting is only about 1%. More than 90% firms perform at -20% or better, and the resetting probability at that point is about 12%. In the event of a resetting, the value of CEO's option increases for about 40%. Thus the ex-ante value change due to re-setting is probably less than 1%.

Sub-sample of firms that has been listed for at least 10 years

One of the proxies, firm age, might be related to the natural life cycle of firms. One possibility is, CEOs of recently listed firms hold a lot of stocks in their firms, only because they have not been able to sell off these stocks. Thus, their pay-performance sensitivities might be high for reasons other than the optimal tradeoff between risk sharing and incentive. To make sure that my results are not due to that, I do the analysis for the sub-sample of firms that has been listed for at least 10 years. The results are qualitatively the same.

The impact of past return on current period pay-performance sensitivity

One concern that people could have is: could there by a relation between past stock return and current period pay-performance sensitivity? For example, could it be that firms suffering adverse shocks in previous years have lower pay-performance sensitivity? If so, and if past return is negatively related to current level of firm-specific risk, then, what I am picking up might be the fact that compensation plans do not have much sensitivity on the down side.

Two issues are relevant here. First, CEO's compensation plan is convex in stock performance, due to the convexity of CEO options holdings. Thus, pay-performance sensitivity on the upside is larger than pay-performance sensitivity on the downside. In an unreported test, I estimate the empirical difference between upside pay-performance sensitivity and downside pay-performance sensitivity, and find that the difference is small for reasonable magnitudes of return fluctuation. Therefore, to a first order approximation, we could assume away the issue of option convexity. Second, when we focus on the point estimate of the slope of compensation contract, do we see that firms that have performed badly in the past have lower slope and thus lower pay-performance sensitivity, to protect their CEOs from further negative shocks in stock price?

In the data, the correlation between the past return and current period pay-performance sensitivity is slightly negative (-0.04), which means bad-performing firms have slightly *higher* pay-performance sensitivity. Furthermore, when we limit our sample to only the firms that have negative returns over the previous year, the correlation between past return and current period pay-performance sensitivity is more negative, indicating that the very bad firm's pay-performance sensitivity is much higher. A direct comparison of the pay-performance sensitivity levels for the whole sample and the decile of firms with the lowest return last year (which could be proxy for distressed firm) reveals that the mean (median) pay-performance sensitivity for the whole sample is $29.0/$1000 ($12.7/$1000), while for the distressed firms they are $50.6/$1000 ($26.3/$1000), thus distressed firms seem to have much higher pay-performance sensitivity than the rest of firms. This is reasonable if the other side of the tradeoff is also considered: firms that are in distress will be more dependent on their CEO's effort, thus the productivity of CEO effort could be higher, which implies the pay-performance sensitivity should be higher.

other control variables

The empirical literature has documented other factors that seem to be related to the pay-performance sensitivity. For example, Lambert and Larcker (1987) and Yermack (1995) has documented that when monitoring is less efficient, a stronger incentive is needed to align the interest between shareholders and CEO. Also, firm leverage has been linked to the severity of agency cost of debt. Since higher pay-performance sensitivity on CEOs increases their incentive to maximize equity value, even at the cost of debt holders, as leverage increases, equity holders may find it optimal to lower the pay-performance sensitivity of managers to reduce the expected agency cost of debt. This is modeled by John and John (1993), and empirically tested by Yermack (1995) and Bryan et al. (2000)[10]. In addition, factors like CEO's age (horizon problem), CEO's tenure (entrenchment), and total fixed pay of CEO might affect the pay-performance sensitivity. I have re-done the regression analysis, controlling for these variables. The results are qualitatively similar as those reported in the paper.

[10]Another argument, advocated by Jensen (1986), suggested that leverage itself is an incentive scheme, because the fear of losing their jobs if their firms go bankrupt will be enough to keep the CEO work hard. If there is a substitution between leverage and pay-performance sensitivity as incentives, then qualitatively we would also expect pay-performance sensitivity to decrease when leverage increases.

A.3 Appendices of chapter 2

A.3.1 Proof of Lemma 1

The maximum of N independent normal random variables, X_i, i=1,...,N, where $X_i \sim NID(\mu, \sigma^2)$, has mean

$E(X_{N:N}) = \sigma f_1(N) + \mu$, where $f_1(N)$ is increasing in N,

and variance

$Var(x_{N:N}) = \sigma^2 f_2(N)$, where $f_2(N)$ is decreasing in N.

Proof:

first, make the following transformation:

define

$X_i' = \frac{x_i - \mu}{\sigma}$,

then, $X_i' \sim NID(0, 1)$

define $X_{N:N}' := \max\limits_i \tilde{X}_i'$

then,

$Pr(X_{N:N}' < y) = Pr(Max(X_1', ...X_N') < y)$

$= Pr(X_1' < y, ..., X_N' < y)$

$= Pr(X_1' < y)...Pr(X_N' < y)$

$Pr(X_{N:N}' < y) = \Phi^n(y)$

$f^*(y) = n\Phi^{n-1}(y)\phi(y)$

Therefore,

$E(X_{N:N}') = \int y f^*(y)dy := f_1(N)$

$Var(X_{N:N}') = \int y^2 f^*(y)dy - E(X_{N:N}')^2 := f_2(N)$

Then, we have

$E(X_{N:N}) = \sigma E(X_{N:N}') + \mu = \sigma f_1(N) + \mu$, and

$Var(X_{N:N}) = \sigma^2 Var(X_{N:N}') = \sigma^2 f_2(N)$.

It is also easy to verify that $f_1(N)$ is increasing in N, while $f_2(N)$ is not monotonic in N. I have plotted values of $f_1(N)$ and $f_2(N)$ for N from 1 to 1000, and put the plot at the end of this paper.

QED.

A.3.2 Proof of Lemma 2

Proof: agents choose today's consumption level c_1, and the amount of goods (in units of current consumption) to be invested in the risky asset ξ. Since the risky asset has a unit price of C, the amount of risky asset bought is $\frac{\xi}{C}$. The agent's optimization problem is:

$$\max_{c_1,\xi} U(c_1) + \beta E\{U[(W - c_1 - \xi)A + \tfrac{\xi}{C}\tilde{X}]\}$$

$$\Leftrightarrow$$

$$\max_{c_1,\xi} -e^{-\alpha c_1} - \beta e^{-\alpha[(W-c_1)A+\xi(\frac{\bar{X}}{C}-A)-\frac{1}{2}\alpha(\frac{\xi}{C})^2\sigma_x^2]}$$

Since the first term does not involve ξ, we could focus on the second term in determining the maximum value of ξ.

Then we have:

$$\max_{\xi} -\beta e^{-\alpha[(W-c_1)A+\xi(\frac{\bar{X}}{C}-A)-\frac{1}{2}\alpha(\frac{\xi}{C})^2\sigma_x^2]}$$

$$\Leftrightarrow$$

$$\max_{\xi} \alpha[(W - c_1)A + \xi(\tfrac{\bar{X}}{C} - A) - \tfrac{1}{2}\alpha(\tfrac{\xi}{C})^2\sigma_x^2]$$

The first order condition is:

$$-\alpha(\tfrac{\bar{X}}{C} - A) + \alpha^2\tfrac{\xi}{C^2}\sigma_x^2 = 0$$

solve it, we will get:

$$\xi = \frac{C(\bar{X}-AC)}{\alpha\sigma_x^2}$$

Therefore ξ does not depend on wealth level;

Plugging the optimal ξ back into the original problem, we have:

$$\max_{c_1} -e^{-\alpha c_1} - \beta e^{-\alpha[(W-c_1)A+\xi^*(\frac{\bar{X}}{C}-A)-\frac{1}{2}\alpha(\frac{\xi^*}{C})^2\sigma_x^2]}$$

$$\Leftrightarrow$$

$$\max_{c_1} -e^{-\alpha c_1} - \beta e^{-\alpha[(W-c_1)A+\frac{1}{2}\frac{(\bar{X}-AC)^2}{\sigma_x^2}]}$$

If we write

$$K = \beta e^{-\frac{1}{2}\frac{(\bar{X}-AC)^2}{\sigma_x^2}},$$

then we have,

$$\max_{c_1} -e^{-\alpha c_1} - Ke^{-\alpha[(W-c_1)A]}$$

Note that this is the same problem as in Theorem 2.1, therefore the solution is similarly:

$$V(W) = -e^{-\frac{\alpha AW - lnAK}{1+A}} - Ke^{-\frac{\alpha AW + AlnAK}{1+A}}.$$

QED.

A.3.3 Proof of Theorem 1

We note that the cash flow of the labor intermediation firm is as followed:

during the set-up period, a cost C is paid to set up the labor intermediation firm;

Then, one period later, three cash flows:

- get the MRP_L of all its members from the production firms
- pay each member the mean of his MRP_L,
- charge each member a fixed membership fee

Therefore the total cash flow one period later will be:

$\tilde{X} = \frac{N}{K_1K_2}\int_{\phi_{12}^0}^{K_1} d\phi_{12}\int_0^{\phi_2^*(\phi_{12})} d\phi_2[\tilde{W}_{LIF}(\phi_{12},\phi_2) - E(\tilde{w}_{LIF}(\phi_{12},\phi_2)) + c]$.

Note that the conditions for Central Limit Theorem for sum of independent random variables (Liapounov) are satisfied (c.f., for example, DeGroot 1989). Therefore, we know that

$\tilde{X} \sim N(E(\tilde{X}), Var(\tilde{X}))$,

where,

$E(\tilde{X}) = E(\frac{N}{K_1K_2}\int_{\phi_{12}^0}^{K_1} d\phi_{12}\int_0^{\phi_2^*(\phi_{12})} d\phi_2[\tilde{W}_{LIF}(\phi_{12},\phi_2) - E(\tilde{w}_{LIF}(\phi_{12},\phi_2)) + c])$

$= \frac{N}{K_1K_2}\int_{\phi_{12}^0}^{K_1} d\phi_{12}\int_0^{\phi_2^*(\phi_{12})} d\phi_2 c$ by Law of iterative expectations,

$= \frac{N}{K_1K_2}\int_{\phi_{12}^0}^{K_1} \phi_2^*(\phi_{12})d\phi_{12}c$

$= \frac{N}{K_1K_2}\int_{\phi_{12}^0}^{K_1}[\frac{1}{4}\alpha\sigma_{B1}^2\phi_{11}^2\phi_{12}^2 + \sigma_{B1}f_1(N)\phi_{11}\phi_{12} - c]d\phi_{12}c;$

$= \frac{N}{K_1K_2}[\frac{1}{12}\alpha\sigma_{B1}^2\phi_{11}^2\phi_{12}^3|_{\phi_{12}^0}^{K_1} + \frac{1}{2}\sigma_{B1}f_1(N)\phi_{11}\phi_{12}^2|_{\phi_{12}^0}^{K_1} - c(K_1 - \phi_{12}^0)]c.$

And,

$Var(\tilde{X}) = Var(\frac{N}{K_1K_2}\int_{\phi_{12}^0}^{K_1} d\phi_{12}\int_0^{\phi_2^*(\phi_{12})} d\phi_2[\tilde{W}_{LIF}(\phi_{12},\phi_2) - E(\tilde{w}_{LIF}(\phi_{12},\phi_2)) + c])$

$= \frac{N}{K_1K_2}\phi_{11}^2\sigma_{B1}^2f_2(N)\int_{\phi_{12}^0}^{K_1} d\phi_{12}\int_0^{\phi_2^*(\phi_{12})} d\phi_2\phi_{11}^2\phi_{12}^2\sigma_{B1}^2f_2(N)$

$= \frac{N}{K_1K_2}\phi_{11}^2\sigma_{B1}^2f_2(N)\int_{\phi_{12}^0}^{K_1}\phi_{12}^2\phi_2^*(\phi_{12})d\phi_{12}$

$= \frac{N}{K_1K_2}\phi_{11}^2\sigma_{B1}^2f_2(N)[\frac{1}{20}\alpha\sigma_{B1}^2\phi_{11}^2\phi_{12}^5|_{\phi_{12}^0}^{K_1} + \frac{1}{4}\sigma_{B1}f_1(N)\phi_{11}\phi_{12}^4|_{\phi_{12}^0}^{K_1} - \frac{1}{3}c\phi_{12}^3|_{\phi_{12}^0}^{K_1}]$

QED.

A.3.4 Proof of Theorem 2

We prove a more general case, and then show that our theorem is a special case of the general case.

To begin with, we assume that firms design a linear sharing contract by choosing parameters $0 \leq a \leq 1$, $0 \leq b \leq 1$, and c, such that for period t,

Per capita labor income:

$w_t = (1-a)AK_t + b\tilde{B}_t + c;$

Per capita capital income:

$q_t = aAK_t + (1-b)\tilde{B}_t - c$

Where $\tilde{B}_t$ is a random variable, and $\tilde{B}_t \sim N(B, \sigma^2_{B2})$

This is a labor risk sharing contract, whereby labor takes on b of the labor risk, and investor takes on the rest $(1-b)$. The capital productivity can potentially be shared, too. The term c represents a constant transfer between laborers and investors.

In particular, since capital product is non-random, there is no gain in introducing risk sharing to capital by setting a less than 1: this only distorts the incentive of old generation to save. Therefore, the optimal risk sharing contract should have $a = 1$ [11][12].

The individual's maximization problem is:

$$\max_{s_t} \underset{\tilde{B}_{t+1}}{E} \; U(c_t, c_{t+1})$$

s.t.

$$c_t + s_t = (1-a)AK_t + bB_t + c$$

and

$$c_{t+1} = aAK_{t+1} + (1-b)\tilde{B}_{t+1} - c$$

and

$$s_t = K_{t+1}$$

Lemma 3 (risk sharing at firm level) *In this setting, savings is determined by*

$$s_t^* = \frac{(1-a)AK_t + bB_t + 2C + \frac{\ln \beta a A}{\alpha} - (1-b)\bar{B} + \frac{1}{2}\alpha(1-b)^2\sigma^2_{B2}}{(1+aA)}.$$

In the particular case when $a = 1$, the solution can be further simplified. In particular, when a=1 we have:

$$s_t^* = \frac{bB_t + 2C + \frac{\ln \beta A}{\alpha} - (1-b)\bar{B} + \frac{1}{2}\alpha(1-b)^2\sigma^2_{B2}}{(1+A)}.$$

[11] we assume that labor supply is fixed at 1, therefore there is no distortion on the incentive to supply labor by introducing this risk sharing on labor income. However, labor risk sharing cannot be too excessive even in our model, otherwise it might violate the incentive constraint to solve the moral hazard problem. We suspect that in real world, labor risk sharing is not carried out to the extreme, precisely because of moral hazard problems. Our model, however, does not take on such a complication.

[12] Also, in my simple model, when there is no capital productivity risk, the only effect of $a < 1$ is to "recycle" the labor risk back to laborer. The author thank Peter Diamond for pointing out this

and, the utility of an investor, conditioning on observing $\tilde{B}_t$, is,

$U^*(B_t) = -\exp(-\alpha\frac{\lambda_1 - \frac{\ln A\beta}{\alpha}}{1+A}) - \beta\exp(-\alpha\frac{\lambda_1 + \frac{\ln A\beta}{\alpha}A}{1+A}),$

where, $\lambda_1 = AbB_t + (A-1)c + (1-b)\bar{B} - \frac{1}{2}\alpha(1-b)^2\sigma_{B2}^2$.

The ex-ante utility of individual, before observing $\tilde{B}_t$, is,

$\underset{\tilde{B}_t}{E}\, U = -\exp(-\alpha\frac{\lambda_2 - \frac{\ln A\beta}{\alpha}}{1+A}) - \beta\exp(-\alpha\frac{\lambda_2 + \frac{\ln A\beta}{\alpha}A}{1+A}),$

where, $\lambda_2 = (Ab + (1-b))\bar{B} + (A-1)c - \frac{1}{2}\alpha[(1-b)^2 + \frac{A^2}{1+A}b^2]\sigma_{B2}^2$.

Proof:

By the time that the choice of s_t is made, the worker has already observed B_t.

The maximization problem of a worker can be re-written as:

$\max_{s_t} \underset{\tilde{B}_{t+1}}{E} \{-e^{-\alpha[(1-a)AK_t + bB_t + c - s_t]} - \beta e^{-\alpha[aAs_t + (1-b)\tilde{B}_{t+1} - c]}\}$

The first order condition to the maximization problem, after re-organization, is:

$\underset{\tilde{B}_{t+1}}{E} \{-e^{-\alpha[(1-a)AK_t + bB_t + c - s_t]} + \beta aAe^{-\alpha[aAs_t + (1-b)\tilde{B}_{t+1} - c]}\} = 0$

taking the expectation, noticing that $B_{t+1} \sim N(\bar{B}, \sigma_{B2}^2)$, we have,

$e^{-\alpha[(1-a)AK_t + bB_t + c - s_t]} = \beta aAe^{-\alpha[aAs_t + (1-b)\bar{B} - \frac{1}{2}\alpha(1-b)^2\sigma_{B2}^2 - c]};$

taking log on both sides,

$-\alpha[(1-a)AK_t + b\bar{B}_t + c - s_t] = ln\beta aA - \alpha[aAs_t + (1-b)\bar{B} - \frac{1}{2}\alpha(1-b)^2\sigma_{B2}^2 - c];$

divide by $-\alpha$ on both sides,

$(1-a)AK_t + bB_t + c - s_t = \frac{-ln\beta aA}{\alpha} + aAs_t + (1-b)\bar{B} - \frac{1}{2}\alpha(1-b)^2\sigma_{B2}^2 - c;$

reorganize to get,

$(1+aA)s_t = (1-a)AK_t + bB_t + 2c + \frac{ln\beta aA}{\alpha} - (1-b)B + \frac{1}{2}\alpha(1-b)^2\sigma_{B2}^2 + c;$

therefore,

$S_t = \frac{(1-a)AK_t + bB_t + 2c + \frac{ln\beta aA}{\alpha} - (1-b)\bar{B} + \frac{1}{2}\alpha(1-b)^2\sigma_{B2}^2}{1+aA}.$

The other results of Theorem 2.2 can be obtained by plugging the optimal s_t back into the formulae.

QED.

Therefore, we have

$\underset{\tilde{B}_t}{E}\, U = -\exp(-\alpha\frac{\lambda_2 - \frac{\ln A\beta}{\alpha}}{1+A}) - \beta\exp(-\alpha\frac{\lambda_2 + \frac{\ln A\beta}{\alpha}A}{1+A})$

$= -\theta\exp(-\alpha\frac{\lambda_2}{1+A})$

where, $\lambda_2 = (Ab + (1-b))\bar{B} + (A-1)c - \frac{1}{2}\alpha[(1-b)^2 + \frac{A^2}{1+A}b^2]\sigma_B^2$,

and

$\theta = [e^{\frac{lnA\beta}{1+A}} + e^{-\frac{AlnA\beta}{1+A}}] > 0.$

The problem of maximizing $\underset{\bar{B}_t}{E} U$ is equivalent to the problem of maximizing λ_2, which in turn is equivalent to:

$\max_{\phi} [Ab + (1-b)](B_0 + \phi \bar{B}_1) + (A-1)c - \frac{1}{2}\alpha[(1-b)^2 + \frac{A^2}{1+A}b^2]\phi^2\sigma_L^2$

The first order condition is:

$[Ab + (1-b)]\bar{B}_1 = \alpha[(1-b)^2 + \frac{A^2}{1+A}b^2]\sigma_L^2\phi$

So $\phi^* = \frac{[Ab+(1-b)]\bar{B}_1}{\alpha[(1-b)^2 + \frac{A^2}{1+A}b^2]\sigma_L^2}$.

The maximum of ex-ante utility is obtained by plugging in the optimal level of ϕ.

QED.

A.4 Appendices of Chapter 3

A.4.1 Discussion

We assume that the factor f and firm specific θ follows a stationary AR(1) process. We also considered two other candidates which does not work well: 1) a random walk (or more generally martingale) process, 2) an IID process.

Case I: a martingale process

The martingale process does not work well for the following reasons:

1. The martingale process does not give a smooth price process: there is always a good chance of sudden default;

In the case that insider has hidden info on the θ_t, he has to absorb the change in K_t all by himself. There is a chance that, following a series of bad luck (θ_t all being negative), the insider might find the project sufficiently unattractive, and give it up!

Outsider should definitely take this into consideration when forming his expected future cash flow: because at the date when the insider gives up the firm, he will realize a LARGE drop of price!!!

This is when all the hidden information about θ_t's are suddenly coming out.

The deep issue here is: these hidden information about θ_t's, when things are sufficiently good, will not come out (that case, insider absorbs all the shocks to the price). But they are permanent shocks and they accumulate over time: this is a property of the random walk, or the weaker assumption of martingale of the cash flow process.

When these information accumulated enough, the market price of asset will SUDDENLY adjust: insider defaults on the dividend payment because it has cumulatively become VERY unattractive to him, and the outsider will have to re-evaluate the asset value, and assign a much lower price to it. Those hidden information suddenly enters price at this time.

There is a "wedge" between insider and outsider's evaluation of the asset value: θ_t. If θ_t is cumulative, then each period it does not go away after being realized (so that the game starts all over again). Instead, all the small increments of θ_t will accumulate, and then, eventually, this wedge can be big enough, and if it is to the insider's disadvantage, a re-assessment of the firm value will be forced: insider will suddenly let out all his accumulated shocks on θ_t, and re-match the level of information between insider and outsider.

2. There is time-nonstationarity in the solution and the incentive mechanism.

If both f and θ are martingales, then $Et(C_{t+1}|C_t) = C_t$;

Insider and outsider's valuation of the firm value will deviate each period by the θ term. As the market price never incorporates this price deviation, the deviation will not be corrected to let the game start afresh each period. Instead, it will stay there and over time will grow larger. The value deviation itself will follow a martingale (or random walk), and the variance of it will increase over time.

If there is no sudden large correction of the price, then this deviation will grow over time. This can have two possibilities:

1) if the realization of this cumulative deviation $\sum(\theta_t) \sim N(0, T\sigma^2)$ is very negative, then it might break the insider's incentive to continue with the firm, and the firm will go bust; 2) On the contrary, if the insider does not default, that gives the outsider some signal about the likelihood of a good realization of the $\sum(\theta_t)$. Conditioning on the insider not defaulting, the outsider will charge a higher value of continuation valuation. The good thing about this is, the higher valuation (conditional on not defaulting) is only a function of time and thus deterministic, thus it does not complicate the story about regression R^2.

Given this pattern, the outsider will rationally increase the dividend requirement every period, since he has a higher expected value of the firm given that it has not default yet. Thus the dividend process gradually increases over time and depends on how long the firm has already been in business. Both counter-factual and also harder to do empirical work on.

The upshot:

With θ_t being a martingale (or random walk), there can be large jumps in the return process: although typically the idiosyncratic information is not revealed, it can suddenly come out and default the firm; also, the equilibrium will be non-stationary over time.

Case II: if θ_t is IID.

Then, this period's realization of θ_t does not tell ANYTHING about next period's θ_t, then bad: no need to incorporate θ_t into the market value: it affects the current period cash flow, but NOT the future cash flow streams, and thus NOT the value of asset. Then θ_t is irrelevant to anything in our model!

A.4.2 Proofs

Proof of Equation (3.2.31)

$$\begin{aligned}
Y_t &= \alpha\tfrac{1+r-\varphi}{\varphi}\{E(K_t|f_t,\theta_{1,t}) - [\tfrac{1}{r}\tfrac{1}{m}C_0r_0]\} - \alpha[a(1-\varphi)+bC_0r_0] \\
&= \alpha\tfrac{1+r-\varphi}{\varphi}\{a + bE(C_t|f_t,\theta_{1,t}) - [\tfrac{1}{r}\tfrac{1}{m}C_0r_0]\} - \alpha[a(1-\varphi)+bC_0r_0] \\
&= \alpha\tfrac{1+r-\varphi}{\varphi}bE(C_t|f_t,\theta_{1,t}) + \alpha(1+r-\varphi)a - \alpha\tfrac{1+r-\varphi}{\varphi}bC_0r_0 - \alpha a(1-\varphi) - \alpha bC_0r_0 \\
&= \tfrac{\alpha}{m}E(C_t|f_t,\theta_{1,t}) + \alpha ar - \tfrac{\alpha}{m}\tfrac{1+r}{1+r-\varphi}C_0r_0
\end{aligned}$$

The second and third terms are:

$$\begin{aligned}
&\alpha ar - \tfrac{\alpha}{m}\tfrac{1+r}{1+r-\varphi}C_0r_0 \\
&= \alpha r\tfrac{1}{m}\tfrac{C_0r_0}{1-\varphi}(\tfrac{1}{r} - \tfrac{\varphi}{1+r-\varphi}) - \tfrac{\alpha}{m}\tfrac{1+r}{1+r-\varphi}C_0r_0 \\
&= \tfrac{\alpha}{m}C_0r_0(\tfrac{1}{1-\varphi} - \tfrac{\varphi r}{(1+r-\varphi)(1-\varphi)} - \tfrac{1+r}{1+r-\varphi}) \\
&= 0
\end{aligned}$$

∎.

Proof of Equation (3.2.35)

using (3.2.29) and (3.2.31),

$$r_{total,t+1} = \frac{E(\alpha K_{t+1}|f_{t+1},\theta_{1,t+1}) + \frac{\alpha}{m}E(C_{t+1}|f_{t+1},\theta_{1,t+1})}{E(\alpha K_t|f_t,\theta_{1,t})} - 1 \qquad \text{(A.4.83)}$$

using (??) and (3.2.25),

$$r_{total,t+1} = \frac{\frac{1}{r}\frac{C_0 r_0}{1-\varphi} - \frac{\varphi}{1+r-\varphi}\frac{C_0 r_0}{1-\varphi} + \frac{1+r}{1+r-\varphi}C_0(f_{t+1} + \theta_{1,t+1})}{\frac{1}{r}\frac{C_0 r_0}{1-\varphi} - \frac{\varphi}{1+r-\varphi}\frac{C_0 r_0}{1-\varphi} + \frac{\varphi}{1+r-\varphi}C_0(f_t + \theta_{1,t})} - 1 \tag{A.4.84}$$

The denominator can be simplified to:

$\frac{C_0 r_0}{1-\phi}\frac{(1+r)(1-\phi)}{r(1+r-\phi)} + \frac{\phi}{1+r-\phi}C_0(f_t + \theta_{1,t})$

After some re-organization, we can write:

$$r_{total,t+1} = r + \frac{-r\frac{C_0 r_0}{1-\phi}\frac{(1+r)(1-\phi)}{r(1+r-\phi)} + C_0\frac{1+r}{1+r-\phi}[f_{t+1} + \theta_{1,t+1} - \phi f_t - \phi\theta_{1,t}]}{\frac{C_0 r_0}{1-\phi}\frac{(1+r)(1-\phi)}{r(1+r-\phi)} + \frac{\phi}{1+r-\phi}C_0(f_t + \theta_{1,t})} \tag{A.4.85}$$

using (3.2.6) and (3.2.7), the numerator becomes: $\frac{C_0(1+r)}{1+r-\phi}(\epsilon_{t+1} + \xi_{t+1})$.

re-organizing we can get,

$$r_{total,t+1} = r + \frac{(1+r)(\epsilon_{t+1} + \xi_{t+1})}{\frac{r_0(1+r)}{r} + \phi(f_t + \theta_{1,t})} \tag{A.4.86}$$

∎.

A.4.3 Discussion about idiosyncratic risk versus market risk

Morck et al. (1999) proposes using the noise trading as in De Long et al (1989,1990) to explain their finding. They argue that

"If risk arbitrage is relatively unattractive in countries that protect private property rights poorly, informed trading may be relatively thin. De Long et al. (1989, 1990) show how stock markets bereft of informed traders can be characterized by large fluctuations due to noise trading. If these fluctuations are primarily market-wide, as they propose, economies with governments that do not respect property rights should exhibit intensified market-wide stock price variation and highly synchronous stock price".

On first look, De Long et al (1989, 1990) indeed look like they need to assume that noise trader risks are all market risk. In fact, they even write that:

"It is important for the analysis below that noise trader risk be market-wide rather than idiosyncratic. If noise traders' misperceptions of the returns to individual assets are uncorrelated and if each asset is small relative to the market, arbitrageurs would eliminate any possible mis-pricing for the same reasons that idiosyncratic risk is not priced in the standard capital asset pricing model".

But De Long et al. in fact does not study multiply risky asset case. Their model has one riskless asset and one risky asset (the aggregate equity, or the "market" portfolio). Intuitively, their concern is that different noise trader takes on different beliefs about the return on the same risky asset (here the market portfolio), so that they bet against each other, and in the end the price on that asset would (by the assumption that all idiosyncratic "beliefs" cancel out) average out, and the sophisticated investor (risk arbitrager) will face the same price on the market portfolio as if noise trader does not have any impact. In that sense, the idiosyncratic risk in the mind of De Long et al. should be different opinions about the pricing of market portfolio in the minds of different noise traders, rather than the "unanimous opinion" among different noise traders of the prices of different assets.

To really see the impact of idiosyncratic risk (that is, the unanimous opinion among noise traders about a certain asset, which in aggregate still shows up in the pricing of that individual asset as a non-market movement in price), we would have to extend the original model in De Long et al. to a multiple risky asset setting. A simple such setting is tried, where we have two risky asset and one riskless asset. The dividend process are as assumed in De Long (1990), but now noise traders have two separate mis-pricing signals: ρ_1 and ρ_2. We further assume that the two mis-pricing signals are independent, so that we allow the noise trader's mis-pricing to be idiosyncratic rather than systematic. Given the model of De Long et al., we could solve for the individual demand curves and pricing dynamics for each asset, and get that these two risky assets will have totally separate risks:

similar to equation (12) in De Long et al. (1990), we have,

$$p_{1t} = 1 + \frac{\mu(\rho_{1t} - \rho^*)}{1+r} + \frac{\mu\rho^*}{r} - \frac{(2\gamma)\mu^2\sigma_\rho^2}{r(1+r)^2} \tag{A.4.87}$$

and

$$p_{2t} = 1 + \frac{\mu(\rho_{2t} - \rho^*)}{1+r} + \frac{\mu\rho^*}{r} - \frac{(2\gamma)\mu^2\sigma_\rho^2}{r(1+r)^2} \tag{A.4.88}$$

where ρ_{1t} and ρ_{2t} are independent pricing signals that noise traders have on each risky assets. Thus, the critical assumption in De Long et al. (1990, 1991) is indeed that all the noise traders have the same belief on any specific asset, rather than that all the individual risky assets have the same mis-pricing.

Our results appears contradictory to the intuition that "if all noise are asset specific, then they all cancel out, thus the risk arbitragers can simply ignore can such noise". The reason behind it is the fact that the noise trader's information, although specific to individual asset, do not cancel out. This is easily illustrated by the simple example we have just seen. There, although the noise on the two risky assets are *independent*, they *do not* cancel out. If the market is composed of one unit of each risky asset, then the total market risk is $2\sigma_\rho^2$, rather than 0. On the contrary, if these two noise have to cancel out, then they would be *perfectly dependent*: one would have to be the opposite of the other. In the same way, if all the idiosyncratic risk among multiple (more than 2) assets all cancel out, then these risks are perfectly linearly correlated, rather than independent!

The fact that "independency" and "cancel out" are not equal, combined with the fact that the Constant Absolute Risk Aversion utility functions are "wealth independent" (meaning that the initial wealth does not affect the risk appetite for additional risk) and "risk independent" (meaning that the addition of one risk does not affect the risk appetite of another *independent* risk), gives rise to the result earlier that the market prices for the two risky assets will be as if there is no other risky asset existing in the market.

Therefore, there is no reason to believe that once risk-arbitragers are driven out, noise trading will inject a lot of non-fundamental "market risk", thus necessarily increase the market risk level. In fact, it is still likely that each stock has its unique misperception by the noise traders, and thus they fluctuate non-synchronously, as long as these noise trader's beliefs about noise are unanimous. That implies, Morck et al. (2000)'s resort to the noise trader model in De Long et al. (1990, 1991) is not a convincing reason to explain the observed pattern.

A.4.4 Calibration studies

We calibrate the economies for different levels of "transparency", while fixing the ratio of idiosyncratic risk to market risk. For the US stock market, we assume that the transparency level, η, is 1, and we assume the ratio of idiosyncratic risk to systematic risk κ is 9, since the average regression R-squared in the market model is about 10 percent in the US. For other economies, we assume there is no fundamental difference in the ratio of idiosyncratic risk to market risk, so we fix the level of κ to be 9. We decrease the parameter of transparency from 1 to 0.1. We also assume different parameters of the autoregression, ϕ from 0.1 to 0.9, with increment of 0.1. We assume there are 100 stocks in the market, each with an independent source of idiosyncratic risk, and we estimate the R-squared for each of the stock, using 100 periods of returns from both the market portfolio and the specific stocks. We then average these 100 stocks to get the average level of R-squared. We repeat this experiment for 100 times, to get the average of these average R-squared.

Other parameters in the calibration are:

r=15%, $E(r_t) = 1$ (this parameter can be arbitrarily scaled, just re-adjust the parameter on C_0), $\sigma_\epsilon^2 = 1$ (this parameter does not matter. Only the ratio of the market over idiosyncratic risk, $\kappa = \frac{\sigma_\xi^2}{\sigma_\xi^2}$).

The results are in Table 1.

We can see that when ϕ is not very large, the R-squared is roughly equal to $\frac{1}{\kappa\eta+1}$, which is close to 0.1 when η is 1, close to 0.5 when η is 0.1, and decrease with η for fixed κ. As ϕ increases, the level of auto-regression increases, and from the expressions of $r_{total,t}$ and $r_{m,t}$, we see that the market model is increasingly mis-specified, and the R-squared decreases for the same η.

Table 1: The calibrated R-squared for different parameter values of transparency and autoregression

		coefficient of transparency η									
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
coefficient of autoregression φ	0.1	0.527751	0.360804	0.269666	0.223847	0.18727	0.159759	0.142691	0.130497	0.117329	0.106421
	0.2	0.527803	0.359497	0.27509	0.224596	0.185511	0.160672	0.145076	0.12872	0.115889	0.107062
	0.3	0.526726	0.358081	0.274228	0.216404	0.182916	0.161776	0.141053	0.12568	0.115589	0.104069
	0.4	0.514578	0.349045	0.263968	0.215562	0.182616	0.156186	0.132655	0.119068	0.1034	0.095026
	0.5	0.511369	0.338924	0.255371	0.193496	0.153119	0.119582	0.09194	0.076275	0.063314	0.051813
	0.6	0.479673	0.268535	0.158187	0.096663	0.064801	0.047486	0.03575	0.029462	0.024632	0.022428
	0.7	0.226983	0.086697	0.048079	0.033366	0.025208	0.021801	0.019903	0.018138	0.016436	0.015761
	0.8	0.05021	0.028271	0.020675	0.01844	0.016688	0.016787	0.015028	0.014621	0.013773	0.013142
	0.9	0.02993	0.017931	0.01533	0.015374	0.012227	0.013694	0.012803	0.010784	0.012462	0.011756

References

Abarbanell, J. and B. Bushee, 1997, "Fundamental Analysis, Future Earnings, and Stock Prices", *Journal of Accounting Research*, 35, 1-24.

Abarbanell, J. and B. Bushee, 1998, "Abnormal Returns to a Fundamental Analysis Strategy", *The Accounting Review*, 73, 19-45.

Abowd, J. and , Kaplan DS, 1999, Executive compensation: Six questions that need answering, Journal of Economic Perspectives 13 (4) 145-168.

Abraham, K and S. Taylor, 1996, "Firms' Use of Outside Contractors: Theory and Evidence", *Journal of Labor Economics*, 14, 394-424.

Aggarwal, R. and A. Samwick, 1999a, "The Other Side of the Trade-off: The Impact of Risk on Executive Compensation", *Journal of Political Economy* 107 (1) 65-105.

Aggarwal, R, and A. Samwick, 1999b, "Executive compensation, strategic competition, and relative performance evaluation: Theory and evidence", *Journal of Finance* 54 (6) 1999-2043

Aggarwal, R, and A. Samwick, 1999c "Empire-Builders and Shirkers: Investment, Firm Performance, and Managerial Incentives", Working Paper, NBER W7335.

Aharony, J. and I. Swary, 1980, "Quarterly Dividend and Earnings Announcements and Stockholders' Returns: An Empirical Analysis", *Journal of Finance*, 35 (1), 1-12.

Allen and Santomero, 1999 , Journal of Banking and Finance) for several reasons).

Ariel, Robert A., 1987, "A Monthly Effect in Stock Returns", *Journal of Financial Economics*. Vol. 18 (1). p 161-74.

Ariel, Robert A., 1988, "Evidence on Intra-month Seasonality in Stock Returns", 109-19, in *Stock market anomalies*, Dimson, Elroy, ed, Cambridge; New York and Melbourne: Cambridge University Press.

Ariel, Robert A., 1990, "High Stock Returns before Holidays: Existence and Evidence on Possible Causes", *Journal of Finance*, 45 (5). 1611-26.

Arrow, K, 1971, *Essays in the Theory of Risk-Bearing.* Chicago: Markham Publishing Company.

Asquith, P. and D. Mullins, 1983, "The Impact of Initiating Dividend Payments on Shareholders' Wealth", *Journal of Business*, 56 (1), 77-96.

Atkins, A. and E. Dyl, 1990, "Price Reversals, Bid-Ask Spreads, And Market-Efficiency", *Journal of Financial and Quantitative Analysis* 25, 535-547.

Baker, G. and B. Hall, 1999, CEO Incentives and Firm Size, Harvard Business School Working Paper 99-060.

Ball, R. and P. Brown, 1968, "An Empirical Evaluation of accounting income numbers", *Journal of Accouting Research*, 6, 159-178.

Ball, R., 1978, "Anomalies in Relationships between Securities' Yields and Yield-Surrogates", *Journal of Financial Economics* 6 (2/3) 103-26.

Banz, R. 1981, "The Relationship between Return and Market Value of Common Stocks", *Journal of Financial Economics* 9, 3–18

Barro, R. and R. Barro, 1990, Pay, Performance and Turnover of Bank CEOs, Journal of Labor Economics 8(3), 448-481.

Basu, S., "Investment Performance of Common Stocks in Relation to Their Price-Earnings Ratios: A Test of the Efficient Market Hypothesis", *Journal of Finance*, 32 (3) 663-82.

Basu, S. 1983, "The Relationship between Earnings' Yield, Market Value and Return for NYSE Common Stocks: Further Evidence", *Journal of Financial Economics*, 12 (1) 129-56.

Beatty, R. and E. Zajac, 1994, Managerial incentives, monitoring, and risk bearing. A study of executive compensation, ownership, and board structure in initial public offerings, Administrative Science Quarterly 39 (2) 313-335.

Berger, J., 1985, *Statistical Decision Theory and Bayesian Analysis*, Second Edition. New York: Springer-Verlag.

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

Bernard, V. and J. Thomas, 1989, "Post-earnings-announcement drift: Delayed price response or risk premium?", *Journal of Accounting Research*, 27, 1-47.

Bertrand, M. and S. Mullainathan, 1999a, "Corporate Governance and Executive Pay: Evidence from Takeover Legislation", Working Paper, Princeton University

Bertrand, M. and S. Mullainathan, 1999b, "Do CEOs Set Their Own Pay? The Ones Without Principals DO", Working Paper, Princeton University.

Bertrand, M. and S. Mullainathan, 2000, "Agents with and without Principals", Working Paper, Princeton University.

Black, F., 1973, "Yes, Virginia, there is hope: Tests of the Value Line ranking system", *Financial Analyst Journal*, 29, 10-14.

Bloom, M., G. Milkovich, 1998, Relationships among risk, incentive pay, and organizational performance, Academy of Management Journal 41 (3), 283-297.

Bodie, Z., 1975, Hedging Against Inflation, Ph.D. Dissertation, Massachusetts Institute of Technology.

Booth, L. and D. Johnston, 1984, "The Ex-Dividend Day Behavior of Canadian Stock Prices: Tax Changes and Clientele Effects", *Journal of Finance*, 39 (2), 457-76.

Bowman, J., 1980, The Importance of a Market Value Measurement of Debt in Assessing Leverage, Journal of Accounting Research 18(1) 242-254.

Box, G. and G. Tiao, 1973, *Bayesian Inference in Statistical Analysis.* New York: Addison-Wesley.

Boudoukh, J., M. Richardson and R. Whitelaw, 1994, "A Tale of Three Schools: Insights on Autocorrelations of Short-Horizon Stock Returns", *Review of Financial Studies* 7, 539-73.

Bremer, M. and R. Sweeney, 1991, "The Reversal of Large Stock-Price Decreases", *Journal of Finance* 46, 747-754.

Brennan, M. and T. Copeland, 1988, "Stock Splits, Stock-Prices, and Transaction Costs", *Journal of Financial Economics* 22, 83-101

Brock, W., J. Lakonishok and B. LeBaron, 1992, "Simple Technical Trading Rules and the Stochastic Properties of Stock Returns", *Journal of Finance* 47, 1731-64.

Brown K. and W. Harlow, 1988, "Market Overreaction - Magnitude And Intensity", *Journal of Portfolio Management* 14. 6-13.

Bryan, S., L. Hwang and S. Lilien, 2000, CEO Stock-Based Compensation: An Empirical Analysis of Incentive-Intensity, Relative Mix, and Economic Determinants, Journal of Business 73 (4) 661-693.

Campbell, J., Lo, A. and C. MacKinlay, 1996, *The Econometrics of Financial Markets.* Princeton, NJ: Princeton University Press.

Campbell, J., M. Lettau, B. Malkiel and Y. Xu, 2000, "Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk", *Journal of Finance*, forthcoming.

Carpenter, J. 1998, "The Exercise and Valuation of Executive Stock Options", *Journal of Financial Economics* 96 (3), 453-473.

Carpenter, J. 2000, Does Option Compensation Increase Managerial Risk Appetite?, Journal of Finance 55 (5), 2311-2331.

Chan, K., 1993, "Imperfect Information and Cross-Autocorrelation among Stock Prices", *Journal of Finance* 48, 1211-30.

Chan, L., Jegadeesh, N. and J. Lakonishok, 1995, "Evaluating the Performance of Value Versus Glamour Stocks: The Impact of Selection Bias", *Journal of Financial Economics* 38, 269–296.

Chen, N., Roll, R. and S. Ross, 1986, "Economic Forces and the Stock Market: Testing the APT and Alternative Asset Pricing Theories", *Journal of Business* 59, 383–403.

Chen, NF, 1991, "Financial Investment Opportunities and the Macroeconomy", *Journal of Finance*, 46, 529-554

Comment, R. and G. Jarrell, 1995, "Corporate Focus And Stock Returns", *Journal of Financial Economics* 37, 67-87.

Copeland, T. and D. Mayers, 1982, "The Value Line Enigma (1965-1978): A Case Study of Performance Evaluation Issues", *Journal of Financial Economics*, 10 (3), 289-321.

Core J, Guay W, 1999, "The use of equity grants to manage optimal equity incentive levels", *Journal of Accounting and Economics* 28 (2), 151-184.

Core J. and W. Guay, 2000a, Estimating the Value of Stock Option Portfolios and Their Sensitivities to Price and Volatility, Unpublished working paper, University of Pennsylvania.

Core J. and W. Guay, 2000b, The Increasing Relation Between Risk and CEO Incentives, Unpublished working paper, University of Pennsylvania.

Core J., W. Guay and R. Verrecchia, 2000, Are Performance Measures Other Than Price Important to CEO Incentives? Unpublished working paper, University of Pennsylvania.

Core J. and J. Qian, 2000, Option Like Contracts for Innovation and Production, Unpublished working paper, University of Pennsylvania.

Coughlan, T. and R. Schmidt, 1985, Executive Compensation, Management Turnover, and Firm Performance: an Empirical Investigation, Journal of Accounting and Economics 7(2), 43-66.

Cross, F., 1973, "The behavior of stock prices on Fridays and Mondays", *Financial Analysts Journal*, 29, 67-69.

Cusatis, P, J. Miles and J. Woolridge, 1993, "Restructuring Through Spinoffs-The Stock-Market Evidence", *Journal of Financial Economics* 33, 293-311.

Dalton, D., C. Daily, A. Ellstrand and J. Johnson, 1998, "Meta-Analytical Reviews of Board Composition, Leadership Structure, and Financial Performance", *Strategic Management Journal* 19, 269-290.

Dann, L., 1981, "Common Stock Repurchases: An Analysis of Returns to Bondholders and Stockholders", *Journal of Financial Economics*, 9 (2), 113-38.

David, H. A., *Order Statistics*, Second Edition. New York: John Wiley & Sons.

Daves, P. and Ehrhardt, M., "Liquidity, Reconstitution, and the Value of U.S. Treasury Strips", *Journal-of-Finance* 48(1), March 1993, pages 315-29.

Demsetz, H. and K. Lehn, 1985, The Structure of Corporate Ownership: Causes and Consequences, Journal of Political Economy 93, 1155-77.

Diaconis, P. and F. Mosteller, 1989, "Methods for Studying Coincidences", *Journal of the American Statistical Association* 84, 853–861.

DeAngelo, L., 1988, "Managerial Competition, Information Costs, and Corporate Governance: The Use of Accounting Performance Measures in Proxy Contests", *Journal of Accounting & Economics*, 10 (1), 3-36.

Dear, K. and C. Begg, 1992, "An Approach for Assessing Publication Bias Prior to Performing Meta-Analysis", *Statistical Science* 7, 237–245.

Debondt W. and R. Thaler, 1985, "Does The Stock Market Overreact", *Journal of Finance*, 40, 793-805.

Debondt W. and R. Thaler, 1987, "Further Evidence On Investor Overreaction And Stock-Market Seasonality", *Journal of Finance* 42, 557-581.

Dodd, P. and J. Warner, 1983, "On Corporate Governance: A Study of Proxy Contests", *Journal of Financial Economics*, 11 (1-4), 401-38.

Eades, K., 1982, "Empirical Evidence on Dividends as a Signal of Firm Value", *Journal of Financial & Quantitative Analysis*, 17 (4), 471-500.

Eades, K., P. Hess, and E. Kim, 1984, "On Interpreting Security Returns during the Ex-Dividend Period", *Journal of Financial Economics*, 13 (1), 3-34.

Ellert, J., 1976, "Mergers, Antitrust Law Enforcement and Stockholder Returns", *Journal of Finance*, 31 (2), 715-32.

Elton, E. and M. Gruber, 1970, "Marginal Stockholder Tax Rates and the Clientele Effect", *Review of Economics & Statistics*, 52 (1), 68-74.

Elton, E., M. Gruber and J. Rentzler, 1983, "A Simple Examination of the Empirical Relationship between Dividend Yields and Deviations from the CAPM", *Journal of Banking & Finance*, 7 (1) 135-46.

Fama, E., 1980, Agency Problems and the Theory of the Firm, Journal of Political Economy 88(2), 288-307.

Fama, E. and K. French, 1992, The Cross-Section of Expected Stock Returns, Journal of Finance 47(2), 427-465.

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

Fama, E., 1998, "Market efficiency, long-term returns, and behavioral finance", *Journal of Financial Economics* 49, 283-306.

Fay, J. and J. Medoff, 1985, "Labor and Output over the Business Cycle: Some Direct Evidence", *American Economic Review*, 75, 638-55.

Finkelstein, S. and B. Boyd, 1998, How much does the CEO matter? The role of managerial discretion in the setting of CEO compensation, Academy of Management Journal, 41 (2), 179-199.

Foster, D., Smith, T. and R. Whaley, 1997, "Assessing the Goodness-of-Fit of Asset Pricing Models: The Distribution of the Maximal R^2", *Journal of Finance* 52, 591–607.

Franke G., R. Stapleton and M. Subrahmanyam, 1998, Who buys and who sells options: The role of options in an economy with background risk, Journal of Economic Theory 82 (1) 89-109.

French, K. 1980 "Stock Returns and the Weekend Effect", *Journal of Financial Economics*, 8 (1), 55-69.

Galambos, J., 1987, *The Asymptotic Theory of Extreme Order Statistics*, Second Edition. Malabar: Krieger Publishing Company.

Garen, J., 1994, "Executive-Compensation and Principal-Agent Theory", *Journal of Political Economy* 102 (6) 1175-1199.

Garvey G., 1997, Marketable incentive contracts and capital structure relevance, Journal of Finance 52 (1) 353-378.

Gibbons, M. and P. Hess, 1981, "Day of the Week Effects and Asset Returns", *Journal of Business*, 54 (4) 579-96.

Gibbons R., 1998, Incentives in organizations, Journal of Economic Perspective 12 (4), 115-132.

Gibbons, R. and K. Murphy, 1990, "Relative Performance Evaluation for Chief Executive Officers", in *Industry and Labor Relations Review* 43, S30-S51.

Gibbons, R. and K. Murphy, 1992, "Optimal incentive contracts in the presence of career concerns: Theory and evidence", *Journal of Political Economy* 100 (3) 468-505.

Gilles, C. and S. LeRoy, 1991, "Econometric Aspects of the Variance-Bounds Tests: A Survey", *Review of Financial Studies*, 4 (4), 753-91.

Golec, J. and Tamarkin, M., "Bettors Love Skewness, Not Risk, at the Horse Track", *Journal of Political Economy*; 106(1), February 1998, pages 205-25.

Goldsmith, R, 1969, *Financial Structure and Development.* New aven, Conn: Yale Univ. Press.

Goulet, W., 1974, "Price Changes, Managerial Actions and Insider Trading at the Time of Listing", *Financial Management* 3, 30-36.

Greenwood, J and B., Smith, 1997, "Financial Markets in Development, and the Development of Financial Markets",*Journal of Economic Dynamics and Control* 21, 145–81

Greene, W., 1993, "Econometric Analysis", New York: Macmillan Publishing Company.

Grossman, S. and O. Hart, 1983, An analysis of the principal-agent problem, Econometrica, 51 (1) 7-45.

Gruber, M., 1996, "Another Puzzle: The Growth in Activity Managed Mutual Funds", *Journal-of-Finance*; 51(3), July 1996, pages 783-810.

Hall, J and J. Liebman, 1998, "Are CEOs really paid like bureaucrats?", *Quarterly Journal of Economics* 113 (3) 653-691.

Hall, J and K. Murphy, 2000, Optimal Exercise Prices for Executive Stock Options, American Economic Review 90 (2) 209-214.

Hall, J and K. Murphy, 2000, Stock Options For Undiversified Executives, NBER working paper 8052.

Hamilton, J. 1994, Time Series Analysis, Princeton University Press.

Hamilton, L., 1991, How Robust is Robust Regression? Stata Technical Bulletin, 2.

Harvey, C and G. Zhou, 1990, "Bayesian Inference In Asset Pricing Tests", *Hournal of Financial Economics* 26 (2) 221-254.

Haubrich, J., 1994, Risk aversion, performance pay, and the principal-agent problem, Journal of Political Economy, 102 (2) 258-276.

Hedges, L., 1992, "Modeling Publication Selection Effects in Meta-Analysis", *Statistical Science* 7, 246–255.

Hermalin B., and N. Wallace, 1997, Firm Performance and Executive Compensation in the Savings and Loan Industry, unpublished working paper, University of California at Berkeley.

Hertzel, M., 1991, "The Effects of Stock Repurchases on Rival Firms", *Journal of Finance* 46, 707-716.

Hess, P., 1982, "The Ex-Dividend Day Behavior of Stock Returns: Further Evidence on Tax Effects", *Journal of Finance*, 37 (2) 445-56.

Himmelberg, C. and G. Hubbard, 2000, Incentive Pay and the Market for CEOs: An Analysis of Pay-for-Performance Sensitivity, Unpublished working paper, Columbia University.

Himmelberg, C., G. Hubbard and D. Palia, 1999, Understanding the determinants of managerial ownership and the link between ownership and performance, Journal of Financial Economics 53: (3) 353-384.

Holloway, C., 1981, "A Note on Testing an Aggressive Investment Strategy Using Value Line Ranks", *Journal of Finance*, 36 (3), 711-19.

Holmstrom, B., 1979, "Moral Hazard and Observability", *Bell Journal of Economics* 10 (1) 74-91.

Holmstrom, B., 1982, Moral Hazard in Teams, Bell Journal of Economics 13 (2) 324-40.

Holmstrom, B., 1999, "Managerial Incentive Problems: A dynamic perspective", *The Review of Economic Studies* 66 (1) 169-82.

Holmstrom, B and P. Milgrom, 1987, "Aggregation and Linearity in the Provision of Intertemporal Incentives", *Econometrica* 55 (2) 303-328.

Holmstrom, B and P. Milgrom, 1987, Aggregation and Linearity in the Provision of Intertemporal Incentives, Econometrica 55 (2) 303-328.

Holmstrom, B and P. Milgrom, 1991, Multi-Task Principal Agent Analysis: Linear contracts, Asset Ownership and Job Design, Journal of Law, Economics, and Organization, 7 (Special Issue) 24-52.

Hubbard, R. and D. Palia, 1995, Executive pay and performance: evidence from the US banking industry, Journal of Financial Economics 39 (1) 105-130.

Hull, J., 2000, Options, Futures and Other Derivatives, Prentice Hall.

Ibbotson, R., 1975, "Price Performance of Common Stock New Issues", *Journal of Financial Economics*, 2 (3), 235-72.

Ingersoll, J. 2000, Incentive Stock Options, Unpublished Working Paper, Yale University.

Iyengar, S. and J. Greenhouse, 1988, "Selection Models and the File Drawer Problem", *Statistical Science* 3, 109–135.

Jaffe, J., 1974, "Special Information and Insider Trading", *Journal of Business* 47, 410-428.

Jaffe, J., D. Keim and R. Westerfield, 1989, "Earnings Yields, Market Values, and Stock Returns", *Journal of Finance*, 44 (1) 135-48.

Jagannathan, R. and E. McGrattan, 1995, "The CAPM Debate", *Federal Reserve Bank of Minneapolis Quarterly Review* 19, 2–17.

Jegadeesh, N. and S. Titman, 1993, "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency", 48, 65-91.

Jensen, M and K. Murphy, 1990a, "Performance Pay and Top Management Incentives", *Journal of Political Economy* 98 (2) 225-264.

Jensen, M and K. Murphy, 1990b, "CEO Incentives-It's Not How Much You Pay, but How", *Harvard Business Review* 68 (3) 138-149.

Jogernson? "the accumulation of human and nonhuman capital, 1948-84", in "The measurement of saving, investment and wealth", NBER studies in income and wealth, vol 52, 227-82)

John, T. and K. John, 1993, Top-management compensation and capital structure, Journal of Finance 48 (3) 949-974.

Jung, C., 1973, *Synchronicity: An Acausal Connecting Principle* (translated by R. Hull). Princeton, NJ: Princeton University Press.

Keim, D. 1983, "Size-Related Anomalies and Stock Return Seasonality: Further Empirical Evidence", *Journal of Financial Economics* 12, 13–32.

Keim, D and R. Stambaugh, 1984, "A Further Investigation of the Weekend Effect in Stock Returns", *Journal of Finance*, 39 (3), 819-35.

Kandel S, and R. Stambaugh, 1996, "On the predictability of stock returns: An asset-allocation perspective", *Journal of Finance* 51 (2), 385-424.

Kim, C. and J. Park, 1994, "Holiday Effects and Stock Returns: Further Evidence", *Journal of Financial & Quantitative Analysis*, 29 (1), 145-57.

Klein L. and D. Peterson, 1989, "Earnings Forecast Revisions Associated with Stock Split Announcements", *Journal of Financial Research* 12, 319-328.

Kohers, T. and Patel J., 1999, "A new time-of-the-month anomaly in stock index returns", *Applied Economics Letters*, 6, 115–120

Kothari, S., Shanken, J. and R. Sloan, 1995, "Another Look at the Cross-Section of Expected Stock Returns", *Journal of Finance* 50, 185–224.

Kole, Stacey R., 1997, The Complexity of Compensation Contracts, Journal of Financial Economics 43 (1) 79-104.

Koenker, R. and G. Bassett, Robust Tests for Heteroscedasticity Based on Regression Quantiles, Econometrica 50 (1) 43-61.

Kraft, A, 1999, "Accouting-Based and Market-Based Trading Rules", working paper, University of Rochester

Kramer, W. and Runde, R. , 1997, "Stocks and the Weather: An Exercise in Data Mining or Yet Another Capital Market Anomaly?", *Empirical-Economics*; 22(4), 1997, pages 637-641

Krueger, A, 1999, "Measuring Labor's Share, Alan Krueger", NBER Working Paper No. W7006.

Krueger, T. and Kennedy, W., 1990, "An Examination of the Super Bowl Stock Market Predictor", *Journal-of-Finance*; 45(2), June 1990, pages 691-97.

Lakonishok, J., Shleifer, A. and R. Vishny, 1994, "Contrarian Investment, Extrapolation, and Risk", *Journal of Finance* 49, 1541–1578.

Lakonishok, J. and S. Smidt, 1988, "Are Seasonal Anomalies Real? A Ninety-Year Perspective", *Review of Financial Studies* 1, 403–426.

Lambert, R. and D. Larcker, 1987, An analysis of the use of accounting and market measures of performance in executive compensation contracts, Journal of Accounting Research 25 Suppl. 85-125.

Lambert, R., D. Larcker and R. Verrecchia, 1991, Portfolio considerations in valuing executive-compensation, Journal of Accounting Research 29 (1) 129-149.

Langetieg, T., 1978, "An Application of a Three-Factor Performance Index to Measure Stockholder Gains from Merger", *Journal of Financial Economics*, 6 (4), 365-83.

Leamer, E., 1978, *Specification Searches.* New York: John Wiley and Sons.

Lehmann, B. and D. Modest, 1988, "The Empirical Foundations of the Arbitrage Pricing Theory", *Journal of Financial Economics* 21, 213–254.

LeRoy, S. and R. Porter, 1981, "The Present-Value Relation: Tests Based on Implied Variance Bounds", *Econometrica*, 49 (3), 555-74.

Lev, B and Thiagarahan, S, 1993, "Fundamental Information Analysis", *Journal of Accounting Research*, 31, 190-215.

Lintner, J., 1965, "The Valuation of Risky Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets", *Review of Economics and Statistics* 47, 13–37.

Lippert, R. and W. Moore, 1995, Monitoring versus Bonding - shareholder rights and management compensation, Financial Management, 24 (3) 54-62.

Litzenberger, R. and K. Ramaswamy, 1979, "The Effect of Personal Taxes and Dividends on Capital Asset Prices: Theory and Empirical Evidence", *Journal of Financial Economics* 7 (2), 163–95.

Lo, A. and C. MacKinlay, 1990a, "Data-Snooping Biases in Tests of Financial Asset Pricing Models", *Review of Financial Studies* 3, 431–469.

Lo, A. and C. MacKinlay, 1990b, "When Are Contrarian Profits Due to Stock-Market Overreaction", *Review of Financial Studies* 3, 175-205.

Lo, A. and C. MacKinlay, 1997, "Maximizing Predictability in the Stock and Bond Markets", *Macroeconomic Dynamics* 1, 102–134.

Loughran, T., 1997, "Book-to-Market Across Firm Size, Exchange, and Seasonality: Is There An Effect?", *Journal of Financial and Quantitative Analysis* 32, 249–268.

Loughran, T. and J. Ritter, 1995, "The New Issues Puzzle", *Journal of Finance*, 50 (1), 23-51.

MacKinlay, C., 1987, "On Multivariate Tests of the CAPM", *Journal of Financial Economics* 18, 341–371.

MacKinlay, C., 1994, "Multi-Factor Models Do Not Explain Deviations From the CAPM", *Journal of Financial Economics* 38, 3–28.

Mandelker, G., 1974, "Risk and Return: The Case of Merging Firms", *Journal of Financial Economics*, 1 (4), 303-35.

Madhavan A and S. Smidt, 1991, "A Bayesian Model of Intraday Specialist Pricing", *Journal of Financial Economics* 30 (1) 99-134.

Mankiw, G., 1986, "The Equity Premium and the Concentration of Aggregate Shocks", *Journal of Financial Economics*, 17 (1), 211-19.

Manne, H., 1965, "Mergers and the Market for Corporate Control", *Journal of Political Economy*, 73(2), 110-120.

Marcus, A., 1982, Risk Sharing and the Theory of Firm, Bell Journal of Economics 13 (2) 369-378.

Mcnichols, M. and A. Dravid, 1990, "Stock Dividends, Stock Splits, and Signaling" *Journal of Finance*, 857-879.

McQueen, G., Pinegar, M. and Thorley, S., 1996, "Delayed Reaction to Good News and the Cross-Autocorrelation of Portfolio Returns", *Journal of Finance* 51, 889-919.

McQueen, G. and Thorley, S., 1997, "Do Investors Learn? Evidence from a Gold Market Anomaly", *Financial-Review*; 32(3), August 1997, pages 501-25.

Mehra, R. and E. Prescott, 1985, "The Equity Premium: A Puzzle", *Journal of Monetary Economics*, 15, 145-161

Mehra, R. and E. Prescott, 1988, "The Equity Risk Premium: A Solution?", *Journal of Monetary Economics*, 22 (1), 133-36.

Merton, S., 1992, Options pricing in the real world of uncertainties: Educating towards a flexible labor force, Bachelor's thesis, Harvard University.

Meulbroek, L., 2000a, "The Efficiency of Equity-Linked Compensation: Understanding the Full Cost of Awarding Executive Stock Options", Harvard Business School Working Paper No. 00-056.

Meulbroek, L., 2000b, "Does Risk Matter? Corporate Insider Transactions in Internet-Based Firms", Harvard Business School Working Paper No. 00-062.

Morck, R., A. Shleifer and R. Vishny (1988), Management Ownership and Market Valuation, Journal of Financial Economics 20, 293-315.

Morck, R., B. Yeung and W. Yu, 1999, The Information Content of Stock Markets: Why Do Emerging Markets Have Synchronous Stock Price Movements? Unpublished working paper. University of Alberta.

Murphy, K., 1985, Corporate Performance and Managerial Remuneration, Journal of Accounting and Economics 7(2), 11-42.

Murphy, K., 1986, Incentives, Learning and Compensation: a Theoretical and Empirical Investigation of Managerial Labor Contracts, Rand Journal of Economics 17(2), 59-76.

Murphy, K., 1999, "Executive Compensation", in Orley Ashenfelter and David Card (eds.), *Handbook of Labor Economics*, Vol. 3, North Holland.

Myers, S., 2000, Outside equity, Journal of Finance 55 (3).

Ofek, E. and D. Yermack, 2000, Taking stock: Does equity-based compensation increase managers ownership, Journal of Finance, 55 (3) 1367-1384.

Officer, R., 1975, "Seasonality in Australian Capital Markets: Market Efficiency and Empirical Issues", *Journal of Financial Economics*, 2 (1), 29-51.

Pastor, L., 2000, "Portfolio selection and asset pricing models", *Journal of Finance* 55 (1), 179-223.

Penman, S.,, 1987, "The Distribution of Earnings News over Time and Seasonalities in Aggregate Stock Returns", *Journal of Financial Economics*, 18 (2). 199-228.

Pettengill, G., 1989, "Holiday Closings and Security Returns", *Journal of Financial Research*, 12 (1). 57-67.

Politis, D. and J. Romano, 1994, "The Stationary Bootstrap", *Journal of the American Statistical Association* 89, 1303-1313.

Prendergast, Canice, 2000, "The Tenuous Tradeoff of Risk and Incentives", working paper, University of Chicago.

Pruitt, S. and R. White, 1988, "The CRISMA Trading System: Who Says Technical Analysis Can't Beat the Market?", *Journal of Portfolio Management* 14, 55-58.

Reinganum, M., 1981, "Misspecification of Capital Asset Pricing: Empirical Anomalies Based on Earnings' Yields and Market Values", *Journal of Financial Economics* 9 (1) 19-46.

Richards, A., 1997, "Winner-Loser Reversals in National Stock Market Indices: Can They Be Explained?", *Journal of Finance* 52, 2129-44.

Rosen, S., 1992, Contracts and the Market for Executives, in Lars Werin and Hans Wijkander (eds.), Contract Economics, Blackwell.

Ross, S. 1973, "The Economic Theory of Agency: The Principal's Problem", *American Economic Review* 63 (1), 134-139.

Ross, S., 1987, "Regression to the Max", Working Paper, Yale School of Organization and Management.

Rozeff, M. and W. Kinney, 1976, "Capital Market Seasonality: The Case of Stock Returns",*Journal of Financial Economics*, 3 (4), 379-402.

Sanger, G. and J. McConnell, 1986, "Stock exchange listings, firm value, and security market efficiency: the impact of Nasdaq", *Journal of Financial and Quantitative Analysis* 21, 1-25.

Schizer, D., 2000, "Executives and hedging: The fragile legal foundation of incentive compatibility", *Columbia Law Review* 100 (2): 440-504.

Schwert, W. 1983, "Size and Stock Returns, and Other Empirical Regularities", *Journal of Financial Economics*, 12 3-12.

Shanken, J. 1987, "A Bayesian Approach to Testing Portfolio Efficiency", *Journal of Financial Economics* 19 (2), 195-215.

Shaw, E, 1973, *Financial Deepening in Economic Development.* New York: Oxford Univ. Press.

Shiller, R. , 1981, "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?", *American Economic Review*; v71 n3 June 1981, pp. 421-36.

Shiller, R, 1993, *Macro markets : creating institutions for managing society's largest economic risks.* Oxford University Press.

Shleifer, A. and R. Vishny, 1986, "Large Shareholders and Corporate Control", *Journal of Political Economy* 94 (3) 461-488.

Shleifer, A. and R. Vishny, 1997, "A Survey of Corporate Governance", *Journal of Finance*, 52(2), 737-783.

Sloan, R., 1993, Accounting Earnings and Top Executive Compensation, Journal of Accounting and Economics 16 (1-2-3), 55-100.

Smith, C., R. Watts, 1992, The Investment Opportunity Set and Corporate Financing, Dividend, and Compensation Policies, Journal of Financial Economics 32 (3) p 263-92.

Stein, J., 1989, "Overreactions In The Options Market", *Journal of Finance* 44, 1011-1023.

Stoll, H. and A. Curley, 1970, " Small Business and the New Issues Market for Equities", *Journal of Financial & Quantitative Analysis*, 5 (3), 309-22.

Sullivan, R., A. Timmermann and H. White, 1999, "Data-Snooping, Technical Trading Rule Performance, and the Bootstrap", *Journal of Finance, forthcoming.*

Sullivan, R., A. Timmermann and H. White, 1998, "Dangers of Data-Driven Inference: The Case of Calender Effects in Stock Returns", Working Paper, University of California, San Diego, Department of Economics.

Townsend, R, 1983, "Financial Structure and Economic Activity",*American Economic Review* 73, 895-911.

Ule, M., 1937, "Price Movements of Newly-listed Common Stocks", *Journal of Business* 10, 346-369.

Van Horne, J. 1970, "New Listings and Their Price Behavior", *Journal of Finance* 25, 783-794.

Vermaelen, T., 1981, "Common Stock Repurchases and Market Signalling: An Empirical Study", *Journal of Financial Economics*, 9 (2), 139-83.

von Hippel, E, 1986, *Cooperation between rivals : informal know-how trading* MIT Working Paper.

Weil, P., 1989, "The Equity Premium Puzzle and the Risk-Free Rate Puzzle", *Journal of Monetary Economics*, 24, 401-21.

West, K., 1996, "Asymptotic Inference About Predictive Ability", *Econometrica* 64, 31067-1084.

White, H., 1980, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, Econometrica 48 (4) 817-830.

White, H., 1997, "A Reality Check for Data Snooping", San Diego, NRDA Technical Report 97-10.

Yermack, D, 1995, "Do Corporations Award CEOs Stock-Options Effectively", *Journal of Financial Economics* 39 (2-3) 237-269.

Ying, L, W. Lewellen, G. Schlarbaum and R. Lease, 1977, "Stock Exchange Listings and Securities Returns", *Journal of Financial and Quantitative Analysis* 12, 415-432.

Zhou, X. 2000, "CEO pay, firm size, and corporate performance: evidence from Canada", *Canadian Journal of Economics* 33 (1) 213-251.