Distribution Agreement

Essays on Entropy-based Robust Inference with Applications in Finance and Economics

By

Ke Wu
Doctor of Philosophy

Economics

Esfandiar Maasoumi, Ph.D.
Advisor

David Jacho-Chavez, Ph.D.
Committee Member

Elena Pesavento, Ph.D.
Committee Member

Guofu Zhou, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D. Dean of the James T. Laney School of Graduate Studies

---

Date

Essays on Entropy-based Robust Inference with Applications in Finance and Economics

By

Ke Wu
M.A., Emory University, 2014
M.A., Indiana University, 2008
B.A., University of International Business and Economics, 2006

Advisor: Esfandiar Maasoumi, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Economics

2015

Abstract

Essays on Entropy-based Robust Inference with Applications in Finance and Economics

By Ke Wu

The dissertation consists of three essays that center around entropy-based robust inference and its applications in the fields of asset pricing and labor economics.

In the first essay, I propose to use a metric entropy to measure asymmetric dependence in asset returns, i.e. the tendency that stocks co-move with the market more strongly during the market downturn than during the upside market. Using the entropy measure, I construct a model-free test for asymmetric dependence in stock returns, which is shown to have greater power than the existing correlation-based test proposed by Hong, Tu, and Zhou (2007). In stock portfolios sorted by size, book-to-market ratio and momentum, based on this new test I find statistically significant asymmetric dependence is much more pervasive than previously thought.

The second essay is an empirical extension to my first chapter, which examines how asymmetric dependence between stock return and the market return is priced in the cross-section of expected stock returns. Motivated by Ang, Chen, and Xing (2006), I construct proxies for the dependence with downside and upside market separately based on non-parametric kernel estimated joint return distributions. Empirically, I find a risk premium (discount) for stocks with high downside (upside) dependence. Moreover, downside dependence premium is almost twice as large as downside beta premium. Asymmetric dependence leaning toward the downside also earns a premium. The findings suggest that investors' aversion to downside losses are stronger than their attraction to the upside gains.

The third essay examines distributional wage gap between incumbents and newly hired workers in the US labor market from 1996 to 2012 based on metric entropy distances. We decompose the wage gap to structural and composition effects by identifying several counterfactual distributions using propensity score reweighting method as discussed in Firpo (2007). We consider weak uniform ranking of these counterfactual wage outcomes based on statistical tests for stochastic dominance as proposed in Linton, Maasoumi, and Whang (2005). Empirically, we find incumbent workers enjoy a better wage distribution, but the attribution of the gap to structural wage inequality and human capital characteristics varies among quantiles of the distribution.

Essays on Entropy-based Robust Inference with Applications in Finance and Economics

By

Ke Wu
M.A., Emory University, 2014
M.A., Indiana University, 2008
B.A., University of International Business and Economics, 2006

Advisor: Esfandiar Maasoumi, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Economics

2015

Acknowledgements

I would like to express my sincere gratitude to my dissertation committee members: Dr. David Jacho-Chavez, Dr. Elena Pesavento, Dr. Guofu Zhou; and, in particular, my advisor, Dr. Esfandiar Maasoumi for their constant guidance, support and invaluable feedback during the development of my dissertation. My interests in entropy-based inference arise when I was taking Dr. Maasoumi's advanced econometrics class. Inspired by him, entropy finally became the theme of my dissertation.

I am also indebted to faculty members in the Department of Economics at Emory University, especially Dr. Kaiji Chen, Dr. Andrew Francis, Dr. Junghoon Lee, Dr. Daniel Levy, Dr. Hugo Mialon, Dr. Sue Mialon, Dr. Vivian Yue and Dr. Tao Zha for their dedicated teaching and advising throughout my Ph.D. study and for their help during my job search. My thanks also go to our Directors of Graduate Studies, Dr. Maria Arbatskaya and Dr. Sara Markowitz for their continuous help and support since the very beginning when I got matriculated to the Ph.D. program. My appreciation also goes to Stephanie Gray, who helped me tremendously in almost every aspect during my study and job search. I am truly impressed by her professionality and dedication to graduate students.

My gratitude also goes to Dr. Melinda Pitts and Dr. Julie Hotchkiss for offering me a great opportunity to work for the Federal Reserve Bank of Atlanta as a research intern. I also benefited a lot through discussions with Dr. Jeffrey Racine and Dr. Le Wang. Their comments and suggestions helped enlarge my research tools and improve my research skills, for which I am really grateful. I would also like to thank our outstanding Ph.D. alumnus, Dr. Lei Jiang, as well as my fellow Ph.D. students Jiening Pan, Yifeng Zhu, and Tong Xu for numerous insightful discussions.

Finally, I owe my special thanks to my parents, especially my beloved father Jianchang Wu, who passed away in a tragic truck accident during his visit to Atlanta just after I finished my third-year study, for their dedication and constant support throughout my life that made this work possible. My deepest gratitude is due to my beloved wife, Jianqiu Wang. During those grief days, it was her endless love and the lovely smiles of our newborn baby girl, Emma Wu, that give me enough courage and strength to continue finishing my Ph.D. Seeing me become a doctor was always my father's wish. As a final tribute, I would dedicate this dissertation to him.

Contents

List of Figures

List of Tables

Preface

The dissertation centers around theoretical development and empirical applications of entropy-based statistical inference methods. Three essays in the dissertation can be categorized into two parts, each of which demonstrates the usefulness and advantages of entropy-based inference in the fields of asset pricing and labor economics, respectively.

Entropy, originated from physics and information theory as a measure of uncertainty, has gained increasing popularity in finance and economics research during recent years. Some recent notable papers with applications of entropy include Sims (2003); Backus, Chernov, and Martin (2011); Hansen (2012); Cabrales, Gossner, and Serrano (2013); Backus, Chernov, and Zin (2014) among others. In those studies, the usage of entropy has shown improvements upon conventional moment-based methods. For example, Backus, Chernov, and Zin (2014) use Kullback-Leibler relative entropy (Kullback, Leibler et al., 1951) to measure the differences between physical and risk-neutral probabilities and derive appropriate bounds for stochastic discount factors that can be used to compare the performance of popular theoretical asset pricing models. The entropy bounds show superior robustness compared to the traditional moment-based Hansen-Jagannathan bound (Hansen and Jagannathan, 1991), as it can easily extend to different time intervals and it is also robust to departures from log-normality. In my dissertation, I choose to employ a metric entropy measure proposed by Granger, Maasoumi, and Racine (2004) due to several desirable properties of the entropy. It belongs to a general K-class entropy family and mathematically it is the only metric entropy within its class, because it satisfies the triangular inequality. It is normalized to take values in-between 0 and 1. Moreover, the measure is invariant under continuous and strictly increasing transformations, such as the commonly used logarithm transformation.

The first part, comprised of two essays, answers an important question in empirical asset pricing, whether there exist statistically significant asymmetric dependence in asset returns, i.e. whether stocks co-move with the market more strongly during the market downturn than during the upside market. Moreover, if such asymmetric co-movements are statistically significant, then how do such asymmetric characteristics affect asset prices cross-sectionally? This question is very important because in asset management business, effective hedging relies on the dependence between assets hedged and the financial instruments used. If the dependence structure is varying with the state of the market, i.e. the dependence is stronger during the market downturn, the portfolio diversification may provide very little protection precisely when it is most needed, since there are very few places to hide when markets collapse.

From an asset allocation perspective, a simple example given in Ang and Chen (2002) has shown that when the underlying return generating process has asymmetric correlation but an investor has a belief that the joint return distribution (between individual stock and the market) follows bivariate normal, then she holds much more equity as a proportion of her investment (overestimate the benefits of diversification) than the optimal weight in the bad market state, while the investor holds too little equity (underestimate the benefits of diversification) in the good market state. They has also shown that the utility loss due to this mis-allocation is economically significant.

The first essay provides a metric entropy measure of asymmetric dependence in asset returns and use this measure to develop a model-free test for asymmetric dependence between stock and the market returns. The paper contributes to the literature in several aspects. First of all, in terms of methodology, the new test extends a robust entropy-based test of asymmetry for univariate process suggested by Racine and Maasoumi (2007) to bivariate case that is of interest in the field of finance.

Secondly, the metric entropy measures directly the distance between the probability density functions of the original joint distribution and the rotated distribution, so it could capture any asymmetry existed in all the moments. In contrast, in the finance literature, traditional tests of asymmetry only focus on testing for asymmetric correlation in the joint distribution, i.e. the asymmetry existed in the second moment. For example, Ang and Chen (2002) seems to be the first to propose a statistical test of asymmetric correlation in asset returns. Their test compares the sample exceedance (conditional) correlations with those implied by a statistical model. If there is a large difference, then the observed asymmetric correlations cannot be explained by the model. However, Ang and Chen (2002) test answers only the question whether the asymmetry can be explained by a given model. Under their joint normality assumption, a rejection of their test cannot rule out the possibility that the data features unexplained by a normal model may be explained by another model. Hong, Tu, and Zhou (2007) propose the first and the only model-free test of asymmetry to date. Their test compares sample conditional correlations at the downside and upside of the joint distribution. However, despite the novelty, their test detects only asymmetric correlations, and does not address asymmetry beyond the second moment. Moreover, its power seems low in empirical applications.

As is well known, the correlation coefficient is only a measure of linear dependence and thus has limitations in measuring general dependence. For example, except for the joint normal case, in general zero correlation does not imply independence, while several papers documented that realized stock returns are non-normally distributed (see, e.g., Embrechts, McNeil, and Straumann, 2002; Ang and Chen, 2002). Moreover, for heavy-tailed distributions without finite second moments, the correlation coefficient is not even defined, while Cont (2001) documented that distributions of many financial time series indeed have heavy tails and display nonexistence of higher order moments. Hence, conceptually the newly proposed entropy-based test is better.

Thirdly, using Monte Carlo simulations, I find that the newly proposed entropy-based test has correct empirical size and better finite sample power than the existing model-free test proposed in Hong, Tu, and Zhou (2007). The superior finite sample performance of the test is due to more information used, as the entropy measure summarizes all the information in the joint density function that uniquely defines the distribution while conditional correlation only uses the information in the second moment. Empirically, in commonly used decile stock portfolios sorted by size, book-to-market ratio and momentum, based on this new test I find statistically significant asymmetric dependence is much more pervasive than previously thought. Specifically, of the ten decile portfolios sorted by book-to-market ratio, I find asymmetry in 2 portfolios at the 5% significance level, and 7 portfolios at the 10% significance level, while Hong, Tu, and Zhou (2007) test fails to detect any asymmetry. My findings are consistent with empirical findings documented in a strand of prior research, like Ball and Kothari (1989); Bekaert and Wu (2000); Ang, Chen, and Xing (2006) among others.

In the second essay, I further examine how the asymmetric dependence between individual stock return and market return is priced in the cross-section of expected stock returns. Motivated by Ang, Chen, and Xing (2006), I construct proxies for the dependence with downside and upside market separately based on non-parametric kernel estimated joint cumulative return distributions. Asymmetric dependence is measured using the entropy test statistic from the first essay, modified to reflect to which side the dependence is stronger. All else being equal, stocks with stronger downside dependence than upside dependence with the market is more risky, as those stocks face large downside risk while have limited upside potential. Risk averse investors should require positive risk premium for holding such stocks. Empirically, using monthly returns to U.S. common stocks traded on the NYSE/AMEX/NASDAQ from January 1962 to December 2013, I indeed find a significant risk premium (discount) for stocks with high downside (upside) dependence. Asymmetric dependence leaning toward the downside also earns a risk premium. The positive risk premium associated with the downside dependence is higher than the discount due to upside dependence.

The findings suggest that investors' aversion to downside losses is stronger than their attraction to the upside gains, which can be implied from a theoretical optimal asset allocation example, where a representative agent with disappointment aversion utility (Gul, 1991) maximizes her utility by allocating wealth among one risk-free and two risky assets, as described in Ang, Chen, and Xing (2006). Fama and MacBeth (1973) regressions show that the contemporaneous impacts of the dependence measures cannot be explained by traditional risk factors, like the market beta (Sharpe, 1964; Lintner, 1965), downside or upside betas (Ang, Chen, and Xing, 2006), coskewness (Harvey and Siddique, 2000), and cokurtosis (Dittmar, 2002). They are also different from the effects of firm level characteristics, like size (Banz, 1981) and book-to-market ratio (Fama and French, 1992), illiquidity (Amihud, 2002), and the momentum (Jegadeesh and Titman, 1993).

The estimated cross-sectional excess return premium for bearing downside dependence risk is approximately 11.6% per annum, almost twice as large as the downside beta premium (6% as reported in Ang, Chen, and Xing (2006)). The downside asymmetric dependence is not persistent over time and shows limited predictability (R-squared is very low). However, a zero investment trading strategy that forms portfolios based on past asymmetric dependence can still earn an average equal-weighted annualized return of 4.5%, significant both economically and statistically. A similar trading strategy based on downside beta fails to yield an economically meaningful return spread. Such comparisons with traditional linear dependence measures (the betas) used in finance literature suggest that there exist gain when going beyond linear risk framework. Entropy-based nonlinear dependence measures may better capture the market risk than their linear counterparts.

The first part of the dissertation shows that entropy has the capacity to incorporate more information in a distribution than traditional moment-based correlations and linear regression coefficients (betas). Such capacity leads to statistical tests with higher power and also empirically more significant risk premium. The second part, comprised of one essay, demonstrates that the metric entropy can serve as an effective measure of distributional wage gap, and can tell us whether two wage distributions are significantly different from each other, while traditional methods focusing on the mean, median, or certain particular quantile appear to place too much weight on a part of the population, or too equal a weight everywhere.

Some recent papers have examined the wage differentials at the entire distribution level, e.g. Maasoumi and Wang (2013) employs the same metric entropy measure to examine the gender wage gap based on the metric distance between wage distributions of female and male workers. Using similar methodologies, the third essay (a joint work with Esfandiar Maasoumi and Melinda Pitts) examines distributional wage gap between incumbent and newly hired workers in the US labor market. We explore weak uniform rankings between wage distributions based on the concept of stochastic dominance that allow assessments over entire classes of welfare functions. Furthermore, we decompose observed gaps to those differentials associated with discrimination in the wage structure, or to human capital composition effect.

The classic Oaxaca (1973) and Blinder (1973) decomposition is a regression based method focusing only on linear conditional mean decomposition. One major limitation of the Oaxaca-Blinder procedure as discussed by Barsky et al. (2002) is that the decomposition provides consistent estimates of the structure and composition effects only under the assumption that the conditional expectation is linear. However, such assumption is not quite likely to hold in many empirical applications. As advocated in DiNardo, Fortin, and Lemieux (1996), we use an alternative non-parametric decomposition based on propensity score reweighting methods. A key advantage of this reweighting approach is that it identifies the entire counterfactual distribution under much less restrictive assumptions, and hence can easily be applied to more general distributional statistics besides the simple mean and quantiles, such as the metric entropy.

The empirical analysis focuses on employees who work at least 35 hours per week using monthly Current Population Survey (CPS) data from 1996 to 2012. Among others, we find incumbent workers generally enjoy a better distribution of wages, but the attribution of the gap to wage inequality and human capital characteristics varies between quantiles. For instance, highly paid new workers are mainly due to human capital components, and in some years, even better wage structure.

References

Amihud, Y. 2002. Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets* 5:31–56.

Ang, A., and J. Chen. 2002. Asymmetric correlations of equity portfolios. *Journal of*

Financial Economics 63:443–94.

Ang, A., J. Chen, and Y. Xing. 2006. Downside risk. *Review of Financial Studies* 19:1191–239.

Backus, D., M. Chernov, and I. Martin. 2011. Disasters implied by equity index options. *The Journal of Finance* 66:1969–2012.

Backus, D., M. Chernov, and S. Zin. 2014. Sources of entropy in representative agent models. *The Journal of Finance* 69:51–99.

Ball, R., and S. Kothari. 1989. Nonstationary expected returns: Implications for tests of market efficiency and serial correlation in returns. *Journal of Financial Economics* 25:51–74.

Banz, R. W. 1981. The relationship between return and market value of common stocks. *Journal of Financial Economics* 9:3–18.

Barsky, R., J. Bound, K. K. Charles, and J. P. Lupton. 2002. Accounting for the black–white wealth gap: a nonparametric approach. *Journal of the American Statistical Association* 97:663–73.

Bekaert, G., and G. Wu. 2000. Asymmetric volatility and risk in equity markets. *Review of Financial Studies* 13:1–42.

Blinder, A. S. 1973. Wage discrimination: reduced form and structural estimates. *Journal of Human Resources* 8:436–55.

Cabrales, A., O. Gossner, and R. Serrano. 2013. Entropy and the value of information for investors. *The American Economic Review* 103:360–77.

Cont, R. 2001. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance* 1:223–36.

DiNardo, J., N. M. Fortin, and T. Lemieux. 1996. Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica* 64:1001–44.

Dittmar, R. F. 2002. Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns. *The Journal of Finance* 57:369–403.

Embrechts, P., A. McNeil, and D. Straumann. 2002. Correlation and dependence in risk management: properties and pitfalls. *Risk management: value at risk and beyond* 176–223.

Fama, E. F., and K. R. French. 1992. The cross-section of expected stock returns. *The Journal of Finance* 47:427–65.

Fama, E. F., and J. D. MacBeth. 1973. Risk, return, and equilibrium: Empirical tests. *The Journal of Political Economy* 81:607–36.

Granger, C., E. Maasoumi, and J. Racine. 2004. A dependence metric for possibly nonlinear processes. *Journal of Time Series Analysis* 25:649–69.

Gul, F. 1991. A theory of disappointment aversion. *Econometrica* 59:667–86.

Hansen, L. P. 2012. Dynamic valuation decomposition within stochastic economies. *Econometrica* 80:911–67.

Hansen, L. P., and R. Jagannathan. 1991. Implications of security market data for models of dynamic economies. *The Journal of Political Economy* 99:225–62.

Harvey, C. R., and A. Siddique. 2000. Conditional skewness in asset pricing tests. *The Journal of Finance* 55:1263–95.

Hong, Y., J. Tu, and G. Zhou. 2007. Asymmetries in stock returns: Statistical tests and economic evaluation. *Review of Financial Studies* 20:1547–81.

Jegadeesh, N., and S. Titman. 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance* 48:65–91.

Kullback, S., R. Leibler, et al. 1951. On information and sufficiency. *The Annals of Mathematical Statistics* 22:79–86.

Lintner, J. 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics* 47:13–37.

Maasoumi, E., and L. Wang. 2013. The gender earnings gap: Measurement and analysis. *Emory University, Working Paper* .

Oaxaca, R. 1973. Male-female wage differentials in urban labor markets. *International Economic Review* 14:693–709.

Racine, J. S., and E. Maasoumi. 2007. A versatile and robust metric entropy test of time-reversibility, and other hypotheses. *Journal of Econometrics* 138:547–67.

Sharpe, W. F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance* 19:425–42.

Sims, C. A. 2003. Implications of rational inattention. *Journal of Monetary Economics* 50:665–90.

Chapter 1

Asymmetric Dependence in Stock Returns: a Robust Entropy-Based Test

Abstract

In this paper, I propose to use a metric entropy to measure asymmetric dependence in asset returns, i.e. the tendency that stocks co-move with the market more strongly during the market downturn than during the upside market. Using the entropy measure, I construct a model-free test for asymmetric dependence in stock returns, which is shown to have greater power than the existing correlation-based model-free test proposed by Hong, Tu, and Zhou (2007). The new test extends a robust entropy-based test of asymmetry for univariate process proposed by Racine and Maasoumi (2007) to bivariate case. In stock portfolios sorted by size, book-to-market ratio and momentum, based on this new test I find statistically significant asymmetric dependence is much more pervasive than previously thought.

Keywords: Asymmetric dependence, metric entropy, copulas, GARCH, simulation.

JEL Classification: C12, C15, C32, G12

1.1 Introduction

Prospect theory and disappointment aversion preferences indicate that investors treat gains and losses unequally and hence behave differently in different states of the market (Kahneman and Tversky, 1979; Kahneman, Knetsch, and Thaler, 1990; Gul, 1991). It may explain why risk premia required by investors are different for assets that exhibit asymmetric co-movements to the upside and downside changes of the market returns (see, e.g., Harvey and Siddique, 2000; Ang, Chen, and Xing, 2006). Empirically, asymmetric characteristics of asset returns, i.e. stocks tend to co-move more strongly during downside market than during upside market, have been found by a number of studies. Ball and Kothari (1989); Schwert (1989); Conrad, Gultekin, and Kaul (1991); Cho and Engle (1999), Bekaert and Wu (2000); Ang and Chen (2002); Bae, Karolyi, and Stulz (2003); Ang, Chen, and Xing (2006), among others, document asymmetries in covariances, correlations, volatilities, and betas of stock returns.

Studying the phenomenon of asymmetric co-movements is important because effective hedging relies on the dependence between assets hedged and financial instruments used. If the dependence structure is varying with the state of the market, portfolio diversification may provide little protection precisely when it is most needed. Furthermore, if all stocks tend to fall with the market during bad times, the value of diversification may be exaggerated for portfolio managers who do not take into account the increasing downside dependence.

Despite the importance of this topic, relatively few studies have proposed statistical tests of asymmetric dependence. Furthermore, almost all prior studies are model dependent and focus on the return asymmetry up to the second moment, i.e. asymmetric covariances, correlations and betas. For example, Ang and Chen (2002) propose a model dependent test and find correlation asymmetries among various portfolios under normality assumption, which allows for the possibility that the asymmetry unexplained by the joint normal model may be explain by some other symmetric models. Hong, Tu, and Zhou (2007) propose the first and the only model-free test of asymmetry to date. Despite the novelty, Hong, Tu, and Zhou (2007) test, however, has two major weaknesses. First, it detects only asymmetric correlations, and does not address asymmetry beyond the second moment. Second, the power of the test seems to be low in empirical applications. For example, the test cannot detect any asymmetry in portfolios sorted by book-to-market ratio and finds only one portfolio among ten decile momentum sorted portfolios with statistically significant asymmetry.

In this paper, I propose an entropy-based model-free test for asymmetric dependence between individual stock/portfolio return and the market return. The test statistic is a normalized metric entropy measure proposed by Granger, Maasoumi, and Racine (2004) that have been widely applied in many previous studies in econometrics (see, e.g., Maasoumi and Racine, 2002, 2008). The test is a bivariate extension of a robust entropy-based test of asymmetry for univariate processes proposed by Racine and Maasoumi (2007). The new test improves upon Hong, Tu, and Zhou (2007) test in the following two aspects. Firstly, Hong, Tu, and Zhou (2007) use asymmetric correlation to proxy for asymmetric dependence, which has several issues. The correlation coefficient is only a measure of linear dependence and thus has some well-known limitations in measuring dependence. For example, except for the joint normal case, in general zero correlation does not imply independence, while several papers documented that realized stock returns are non-normally distributed (see, e.g., Embrechts, McNeil, and Straumann, 2002; Ang and Chen, 2002). Moreover, for heavy-tailed distributions without finite second moments, the correlation coefficient is not even defined, while Cont (2001) documented that distributions of many financial time series have heavy tails and display nonexistence of higher order moments. In contrast, the entropy measure summarizes all the information of a given distribution, and hence can capture asymmetry that exist in all the moments. So conceptually, the newly proposed entropy-based test is better. Secondly, with Monte Carlo simulations, I find that the entropy-based test has correct empirical sizes and higher power than Hong, Tu, and Zhou (2007) test in finite sample cases. Therefore, in terms of finite sample performance the entropy test is also better.

Empirically, sorting portfolios based on size, book-to-market ratio and momentum, the entropy test detects statistically significant asymmetry in all three groups of the portfolios.

For example, in contrast to the Hong, Tu, and Zhou (2007) test, I find asymmetry in 2 portfolios at the 5% significance level, and 7 portfolios at the 10% significance level, out of the ten decile portfolios sorted by the book-to-market ratio.

The rest of the paper is organized as follows. Section 1.2 reviews the literature and introduces the entropy-based test for asymmetric dependence. Section 1.3 examines the test size and power using Monte Carlo Simulations. Section 1.4 applies the entropy test to investigate asymmetry in returns of commonly used stock portfolios. Section 1.5 concludes.

1.2 Tests of Asymmetry

For the ease of understanding, in this section I first review the standard asymmetric correlation tests, then extend the concept to general asymmetry, and finally provide the entropy-based test.

1.2.1 Asymmetric Correlation

In the finance literature, Ang and Chen (2002) and Hong, Tu, and Zhou (2007) provide tests of asymmetry in bivariate return series, but they test only asymmetric correlation instead of general asymmetric dependence.

To see why, let us consider two standardized strictly stationary return series denoted by $\tilde{x}_{i,t}$ and $\tilde{y}_t$.[1]. Both of the tests rely on exceedance correlations defined as

$$\rho^+(c) = corr(\tilde{x}, \tilde{y} | \tilde{x} > c,\ \tilde{y} > c), \tag{1.1}$$

$$\rho^-(c) = corr(\tilde{x}, \tilde{y} | \tilde{x} < -c,\ \tilde{y} < -c), \qquad \forall c \geq 0. \tag{1.2}$$

Clearly, both $\rho^+(c)$ and $\rho^-(c)$ measure conditional correlations between two return series conditioning on both series are above or below a certain exceedance level c. The null hypothesis of interest is

[1] In practice, they may stand for stock return and the market return series respectively

$$H_0: \qquad \rho^+(c) = \rho^-(c), \quad \text{for all}\, c \geq 0.$$

Ang and Chen (2002) seems to be the first to propose a formal statistical test for the asymmetric correlation hypothesis, whose test statistic H is defined as

$$H = \left[\sum_{i=1}^{m} w(c_i)(\rho(c_i, \phi) - \hat{\rho}(c_i))^2 \right]^{1/2} \tag{1.3}$$

where $c_1, \ldots, c_m$ are m pre-selected exceedance levels, $w(c_1), ..., w(c_m)$ are weights, $\hat{\rho}(c_i)$ stands for sample realization of $\rho^+(c_i)$ or $\rho^-(c_i)$ and $\rho(c_i, \phi)$ is the population exceedance correlation implied by a given model with parameters ϕ. Their test addresses the interesting question whether the asymmetric correlations observed in the data can be explained by the given model. Therefore, the test is model dependent and the testing results rely on the choice of the pre-specified model. One weakness associated with the model-dependent test is that the data may still have symmetric correlations, even if a given symmetric model, like the normality model they used, cannot explain it.

To overcome the weakness, Hong, Tu, and Zhou (2007) propose a model-free test. Their test statistic is defined as

$$J_\rho = T(\hat{\rho}^+ - \hat{\rho}^-)'\hat{\Omega}^{-1}(\hat{\rho}^+ - \hat{\rho}^-) \tag{1.4}$$

where T is the sample size, $\hat{\rho}^+$ and $\hat{\rho}^-$ are $m \times 1$ vectors of sample exceedance correlations, and $\hat{\Omega}$ is a consistent estimator of the covariance matrix of $\sqrt{T}(\hat{\rho}^+ - \hat{\rho}^-)$. Under the null of symmetric correlations and certain regularity conditions, the test has a simple asymptotic chi-square distribution, $J_\rho \xrightarrow{d} \chi^2_m$. The test answers the question whether there exists asymmetric correlations at all in the data. In other words, if the test rejects the null, it implies that no distributions with symmetric correlations can fit the data well.

1.2.2 Asymmetric Dependence

Both existing tests try to test for asymmetric dependence in bivariate stock return data by testing for asymmetric correlations. Yet, it is well known that linear correlation coefficient is only a measure of linear dependence. In general, zero correlation does not imply independence except for the joint normal case. Hence, testing for linear correlation ignores possible higher order dependence entirely. On the other hand, it is also documented that financial time series usually display heavy tails and have non-standard higher order moments (see, e.g., Embrechts, McNeil, and Straumann, 2002; Cont, 2001). Therefore, it is of interest to have a test for general asymmetric dependence that involves all the higher order moments.

Since the joint density function uniquely defines a joint distribution, directly testing for asymmetry in the joint probability density function certainly involves all higher order moments. Motivated by a univariate test of asymmetry proposed by Racine and Maasoumi (2007), I focus on testing the symmetry of the joint density function.

Analogous to the exceedance correlations, I define exceedance densities by

$$f^{+}(c) = f(\tilde{x},\tilde{y}|\tilde{x} > c,\ \tilde{y} > c \ or\ \tilde{x} < -c,\ \tilde{y} < -c), \tag{1.5}$$

$$f^{-}(c) = f(-\tilde{x},-\tilde{y}|\tilde{x} > c,\ \tilde{y} > c \ or\ \tilde{x} < -c,\ \tilde{y} < -c), \tag{1.6}$$

where $f(\tilde{x},\tilde{y})$ is the joint probability density function of return series $\tilde{x}_{i,t}$ and $\tilde{y}_t$ in ranges of $\tilde{x} > c,\ \tilde{y} > c$ or $\tilde{x} < -c,\ \tilde{y} < -c$. Restriction to these ranges where both returns are above or below certain exceedance level basically follows Ang and Chen (2002) and Hong, Tu, and Zhou (2007), since we want to capture the co-movements of both returns. Since $\tilde{x}_{i,t}$ and $\tilde{y}_t$ are standardized to have zero means, $f(-\tilde{x},-\tilde{y})$ denotes the joint probability density function of the rotated return series around the mean.[2] If the joint distribution is truly symmetric, then the two densities should be the same almost everywhere. Intuitively,

[2] Note that when the data series are not standardized, the rotation can also be easily done by pre-multiplying a rotation matrix P to the original series to get the new data pair $(-x_{i,t} + 2\hat{\mu}_X, -y_t + 2\hat{\mu}_Y)$. Specifically,

the distance between the two density functions reflects the degree of asymmetry of the joint return distribution. Hence, the null hypothesis for testing asymmetric dependence is

$$H_0: \qquad f^+(c) = f^-(c), \quad \text{for all } c \geq 0. \tag{1.7}$$

If this hypothesis is rejected, then the data must possess asymmetry as their density functions must be different at least in one of the two symmetric regions.

[Insert Figure 1.1 about here]

As an example, Figure 1.1 illustrates a case of symmetric dependence and compares it with a case of asymmetric dependence. Subfigure (a) shows a scatter plot of 2000 data points that are generated by a Clayton copula model, which is known to have stronger left tail dependence than right tail dependence. Subfigure (b) is a similar plot but is generated using a bivariate normal copula model that has symmetric dependence at both tails. When I consider symmetric/asymmetric dependence, I examine the dependence structure over the shaded areas that represent the first and third quadrants, i.e. I take the exceedance level $c = 0$ here. In subfigure (a), it is clear that the data are more concentrated in the third quadrant than in the first quadrant, indicating stronger dependence (greater mass of the joint densities) in the lower tail. Subfigure (b) has roughly equal joint densities in both tails, indicating symmetric dependence. The lines in both figures are fitted linear regression lines that indicate overall linear dependence. With visual inspection, we can clearly see that the linear dependence line does not differ very much in the two cases, but the actual dependence

---

$$P \cdot \begin{bmatrix} x_{i,1} & y_1 \\ \vdots & \vdots \\ x_{i,T} & y_T \end{bmatrix} = \begin{bmatrix} -x_{i,1} + 2\hat{\mu}_X & -y_1 + 2\hat{\mu}_Y \\ \vdots & \vdots \\ -x_{i,T} + 2\hat{\mu}_X & -y_T + 2\hat{\mu}_Y \end{bmatrix}$$

where the rotation matrix P takes the following form

$$P = \begin{bmatrix} -1 + \frac{2}{T} & \frac{2}{T} & \frac{2}{T} & \cdots & \frac{2}{T} \\ \frac{2}{T} & -1 + \frac{2}{T} & \frac{2}{T} & \cdots & \frac{2}{T} \\ \frac{2}{T} & \frac{2}{T} & -1 + \frac{2}{T} & \cdots & \frac{2}{T} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{2}{T} & \frac{2}{T} & \frac{2}{T} & \cdots & -1 + \frac{2}{T} \end{bmatrix}$$

structures are quite different. Therefore, this example also highlights the danger of failing to discover asymmetry when we focus only on linear dependence.

1.2.3 An Entropy Measure

The important question is how to test the above null hypothesis given in the null hypothesis 1.7. Intuitively, the joint distribution is symmetric if the distance between $f^+(c)$ and $f^-(c)$ is zero almost everywhere and is otherwise asymmetric if the distance is not zero on a set with positive measure. To do so, I have to rely on certain measures of distance between two probability density functions.

In statistics and information theory, entropy has a long history of being used as a measure of divergence between distributions. It was first introduced by Shannon (1948), and later extended by Kullback, Leibler et al. (1951). Ullah (1996) and Maasoumi (1993) provide excellent surveys of various entropy measures and their applications in econometrics. More recently, entropy has drawn great attention from financial economists and has been more and more applied in finance research, e.g. Backus, Chernov, and Zin (2014) use Kullback-Leibler relative entropy (Kullback, Leibler et al., 1951) to measure the differences between physical and risk-neutral probabilities and derive appropriate bounds for stochastic discount factors that can be used to compare the performance of popular theoretical asset pricing models.

The entropy measure I use belongs to the same K-class entropy as the Kullback-Leibler divergence measure. First proposed by Granger, Maasoumi, and Racine (2004), the measure is a special case of K-class entropy with $K = 1/2$, which is a normalization of the Hellinger distance measure and is the only metric entropy within its class. Besides being a metric, as shown by Granger, Maasoumi, and Racine (2004), this measure has been proved to have many other desirable properties as a measure of distance between distributions.

Consider, for simplicity, first the case where we have only one exceedance level c. The entropy measure of asymmetry is defined as

$$S_\rho(c) \;=\; \frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(f^+(c)^{\frac{1}{2}} - f^-(c)^{\frac{1}{2}})^2\, d\tilde{x}\, d\tilde{y}, \tag{1.8}$$

which is clearly a function of the exceedance level c. In practice, c is chosen according to empirical interests. For example, when $c = 0$, $S_\rho(0)$ measures asymmetric dependence in the first quadrant (where both standardized returns are positive) and the third quadrant (where both standardized returns are negative). If the asymmetric dependence in the tails of the distribution are of interests, $S_\rho(c)$ can be measured at other exceedance levels, e.g. 0.5 or 1 standard deviations away from the mean.

The entropy measure $S_\rho(c)$ is well defined for both continuous and discrete data. It takes values in-between 0 and 1, and equals 0 if and only if the densities are equal, which indicates symmetric dependence. Mathematically, it is a true measure of "distance" because it satisfies the triangular inequality. Moreover, the measure is invariant under continuous and strictly increasing transformations, such as the commonly used logarithm transformation.

Consider now the case where we have multiple exceedance levels, $c_1, \ldots, c_m$, which we want to test whether there exists symmetric dependence at each exceedance level jointly. For example, while the singleton set of $c = \{0\}$ is usually of interest in the literature, the set of the levels $c = \{0; 0.5; 1; 1.5\}$ is also commonly used by previous studies, such as in Ang and Chen (2002) and Hong, Tu, and Zhou (2007). For the multiple level case, I can also apply the statistic in equation 1.8 for each of the individual levels, and then aggregate the estimates using some function. Since $S_\rho(c)$ is a metric and always non-negative, we may simply take arithmetic average,

$$S_\rho = \frac{S_\rho(c_1) + \cdots + S_\rho(c_m)}{m}, \tag{1.9}$$

where $S_\rho(c_j)$ is computed from equation 1.8 for $j = 1, \ldots, m$. Thus, the entropy test statistic is well defined for either the singleton test case with one exceedance level or the joint test case with multiple exceedance levels.

To carry out the entropy test in practice, we need to estimate first the joint density functions from the data, and then compute the integral in equation 1.8 to obtain the statistic $\hat{S}_\rho(c)$. Finally, we need to have a procedure to determine the distribution of the test statistic under the null hypothesis and hence the P-values of the test statistic. The task is unfortunately much more complex than that of the asymmetric correlation tests. These issues are addressed in the following two subsections.

1.2.4 Non-parametric Estimation

Consider now how to estimate the density functions in equation 1.8 given the data. Following Maasoumi and Racine (2002); Racine and Maasoumi (2007) among others, I use non-parametric kernel smoothing method to consistently estimate the unknown joint densities. Specifically, the popular "Parzen-Rosenblatt" kernel density estimator (see Rosenblatt, 1956; Parzen, 1962) is used. For the univariate case, the "Parzen-Rosenblatt" kernel density estimator is defined as

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} k\left(\frac{X_i - x}{h}\right), \tag{1.10}$$

where n is sample size of the data $\{X_i\}$, h is a smoothing parameter (commonly referred to as the bandwidth) and $k(\cdot)$ is a nonnegative bounded kernel function. In this paper, we have to deal with bivariate density functions, so in the kernel estimation we need to employ a "product kernel function", which is constructed as the product of univariate kernel functions. That is, our candidate kernel density estimator of the data is given by

$$\hat{f}(x, y) = \frac{1}{nh_1h_2} \sum_{i=1}^{n} k\left(\frac{x_i - x}{h_1}\right) \times k\left(\frac{y_i - y}{h_2}\right) \tag{1.11}$$

where n is sample size, $k(\cdot)$ is a suitable univariate kernel function, h_1 and h_2 are bandwidths for each of the two variables, and $\{(x_i, y_i)\}$ are the observed data pairs. It should be noted that n is equal to T, the length of the return series in the empirical applications of this paper.

Econometrically, the accuracy of the nonparametric kernel density estimator clearly relies on the selection of both the kernel function and the bandwidth. It turns out that the choice of kernel function plays a much less important role than the selection of bandwidth. The return data are continuous, so I use standard Gaussian kernel, $k(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$, as the univariate kernel function in the estimation. On selecting the bandwidth, I use the well-known Kullback-Leibler cross-validation (the likelihood cross-validation method) (see Li and Racine, 2007, for details). This cross-validation procedure minimizes the Kullback-Leibler divergence measure between the actual density and the estimated one. Numerically, it solves the following maximization problem of the log-likelihood function,

$$\max_{h_1,h_2} \mathcal{L} = \sum_{i=1}^{n} \ln \left[\hat{f}_{-i}(x_i, y_i) \right], \tag{1.12}$$

where

$$\hat{f}_{-i}(x_i, y_i) = \frac{1}{nh_1h_2} \sum_{j \neq i}^{n} k\left(\frac{x_j - x}{h_1} \right) \times k\left(\frac{y_j - y}{h_2} \right), \tag{1.13}$$

which is equal to $\hat{f}(x, y)$ without the i-th realization. Based on the efficient market hypothesis (Fama, 1970), stock returns can be seen as i.i.d. or stationary and weakly dependent series, and under such assumptions the estimated density converges to the actual density. (see, e.g., Li and Racine, 2007, for technical details).

With the method described above, the density functions in equation 1.8 can be consistently estimated. Then the test statistic $\hat{S}_\rho(c)$ can be obtained by computing the integral using a standard numerical procedure in the case of a single exceedance level. In the presence of multiple exceedance levels, the test statistic is computed from equation 1.9.

1.2.5 Distribution of the Test Statistic

To conduct statistical inference based on the entropy test of asymmetry, we need to know the sampling distribution of the test statistic under the null hypothesis of symmetry. There are in general two ways to derive the sampling distribution. One is to rely on asymptotic theory and the other way is to use bootstrap resampling method.

The asymptotic theory for the class of entropy measures with similar functional forms has been developed by Skaug and Tjøstheim (1993); Tjøstheim (1996) and by Hong and White (2005). Under certain regularity conditions, $\hat{S}_\rho(c)$ follows asymptotically normal distribution and the distribution derived under the null hypothesis does not depend on the choice of the bandwidths. This is partially because the bandwidth is a quantity that vanishes in the limit. However, for a given finite sample size, the computed value of the test statistic depends critically on the bandwidth selection (Maasoumi and Racine, 2008, see). It really raises concerns on using the simple asymptotic distribution to conduct statistically inference in empirical applications, since the results of such asymptotic-based tests tend to be highly sensitive to the bandwidth and there are many competing approaches for bandwidth selection. Therefore, following the suggestions of Hong and White (2005) and many others, rather than relying on asymptotic distribution for inference, I use a bootstrap resampling approach to determine the empirical null distribution of $\hat{S}_\rho(c)$ (see Efron, 1982; Hall, 1992; Horowitz, 2001, for more discussions on bootstrap resampling approach).

To construct a sample under the null hypothesis in the bootstrap resampling, let

$$Z = \{(x_{i,1}, y_1), (x_{i,2}, y_2), ..., (x_{i,T}, y_T); (-x_{i,1}, -y_1), (-x_{i,2}, -y_2), ..., (-x_{i,T}, -y_T)\},$$

which is a vector obtained by stacking together the original data pairs $(x_i,\ y_i)$ with the rotated data pairs $(-x_i,\ -y_i)$. Through bootstrapping samples from Z, we construct the empirical distribution of $\hat{S}_\rho(c)$. I repeat the bootstrapping draws B times from Z, and then can obtain B resamples of $\hat{S}_\rho(c)$.

There are many different kinds of bootstrap resampling procedures, e.g. the simple bootstrap, wild bootstrap, block bootstrap, and so on. The choice of which bootstrap resampling procedure to use depends on the nature of the data. As stock return are known to be stationary and weakly dependent, the block bootstrap that takes such dependence structure into account seems to be the natural choice (see Künsch, 1989).

Politis and Romano (1994) shows that using overlapping blocks with lengths that are sampled randomly from a geometric distribution, with the mean equal to the pre-selected block length l, yields stationary bootstrapped data samples, while overlapping or non-overlapping blocks with fixed lengths may not ensure the stationarity. This procedure proposed by Politis and Romano (1994) is called stationary bootstrap, which is a special kind of block bootstrap. Due to the merit of stationary bootstrap, I choose to employ the procedure in this paper.

How to select the average block length l used in the stationary bootstrap is another important issue. I apply the data-driven and automatic method suggested by Politis and White (2004); Patton, Politis, and White (2009) to select the optimal block length. Econometrically, their method is beneficial as it minimizes the mean squared error of the estimated long-run variance of the time series.

In terms of selecting B, the number of bootstrap samples, it is obviously true that the greater the B, the more accurate the bootstrapped distribution is. However, unlike the commonly used bootstrap procedures used in linear regressions, kernel estimation can be enormously time-consuming. In some similar problems, Davidson and MacKinnon (2000) suggest the use of $B = 399$ for simulations that compute the P-value of a test at the 5% nominal significance level. In this paper, although I find that a value of $B = 199$ yields similar results, following the suggestion of Davidson and MacKinnon (2000), I choose to report the empirical testing results and all the simulation results based on stationary bootstrap with $B = 399$.

After having computed B replications of $\hat{S}_\rho(c)^*$, the sampling distribution of $\hat{S}_\rho(c)$ can be easily obtained. To find out the critical values for rejection at different confidence levels, I can reorder the bootstrapped estimates from smallest to largest and denote the list by $\hat{S}_{\rho,1}(c)^*$, $\hat{S}_{\rho,2}(c)^*$, ... , $\hat{S}_{\rho,B}(c)^*$, and then determining those percentiles from these ordered statistics. For example, to conduct the symmetry test at the 5% level, the null hypothesis H_0 in 1.7 will be rejected if $\hat{S}_\rho(c) > \hat{S}_{\rho,379}(c)^*$, where $\hat{S}_{\rho,379}(c)^*$ is the 95th percentile of the ordered bootstrapped estimates. Empirical p-values may also be obtained by counting the proportion of the ordered bootstrapped statistics that exceeds $\hat{S}_\rho(c)$, the test statistic estimated from the original sample.

1.3 Monte Carlo Simulations

In this section, using copula-GARCH based Monte Carlo simulation, I examine the size and power of the entropy-based test and show that the entropy test has reliable sizes, and has higher power in finite samples than the Hong, Tu, and Zhou (2007) test.

1.3.1 Modeling Dependence with Copulas

Since we are testing the joint distribution of two random variables, the simulation procedures involve in generating random samples from some joint distribution with certain dependence structure. Copulas are probably the most commonly used method to model the complete dependence structure between random variables (see Patton, 2004; Rodriguez, 2007; Okimoto, 2008, for some applications of copulas in finance). Sklar (1959) proves that all bivariate distribution functions $F(x_1, x_2)$ can be completely described by the univariate marginal distributions $F_1(x_1)$ and $F_2(x_2)$ and a copula function $C : [0,1]^2 \mapsto [0,1]$. Copula, a word chosen by Sklar, is a multivariate probability distribution function that describes such dependence structure between the two (or more) marginal distributions (see Nelsen, 1999, for a more detailed introduction to copulas).

Many copulas with different dependence structures have been developed and commonly applied in the literature. Some of those parametric copulas, such as Gaussian, Student's t and Frank copulas, are known to have symmetric tail dependence structure. Some copulas are constructed to have asymmetric tail dependence. For example, Clayton copula is known for strong left tail dependence, whereas Gumbel copula shows strong right tail dependence.

As stock returns usually show stronger left tail dependence than right tail dependence with the market return (see Ang and Chen, 2002), Clayton copula seems to be a natural choice. However, it is not wise to completely rule out those copulas with symmetric dependence. Figure 1.2 gives the scatter plots of random samples generated by Gaussian, Clayton and mixed Gaussian-Clayton copulas, as well as the actual data plots of the smallest decile size portfolio returns. It is clear that Clayton copula generated data with strong left tail dependence, as the plots are highly concentrated at the left tail, but the dependence seems to be much stronger than that is actually reflected in the scatter plot of the smallest decile size portfolio. Comparing to the smallest size portfolio, which has shown to have the strongest asymmetric dependence in the following section and in Hong, Tu, and Zhou (2007), the generated data plots do not look much like the actual data plots. As shown in subfigure (C), the scatter plots generated by equal-weighted mixed Gaussian-Clayton copula look more similar to the actual data plots in subfigure (D). Therefore, I choose to use those mixed copulas as the data generating process in simulations. Similar mixture copula models are also used in Hong, Tu, and Zhou (2007).

[Insert Figure 1.2 about here]

A bivariate Gaussian copula is given by

$$C_{nor}(u,\, v;\, \rho) = \Phi_\rho(\Phi^{-1}(u),\, \Phi^{-1}(v)) \tag{1.14}$$

where $\rho \in (-1, 1)$ is correlation coefficient between the marginal distributions. Φ^{-1} is standard normal CDF inverse, and Φ_ρ is the standard bivariate normal distribution function with correlation ρ.

A bivariate Clayton copula is defined as

$$C_{clay}(u,\, v;\, \tau) = (u^{-\tau} + v^{-\tau} - 1)^{-\frac{1}{\tau}} \tag{1.15}$$

where $\tau > 0$ governs the dependence between the marginals. Higher τ indicates stronger dependence. Tawn (1988) proves that every convex combination of existing copula functions is still a copula.[3] I can construct a mixture Gaussian-Clayton copula with pre-chosen weights by a convex combination of the two copulas.

The mixture Gaussian-Clayton copula used in this paper takes the following specifica-

[3]The formal statement of the theorem is given in the appendix.

tion.

$$C_{mix}(u,\, v; \rho, \tau, \kappa) = kappaC_{nor}(u,\, v;\, \rho) + (1-\kappa)C_{clay}(u,\, v;\, \tau) \tag{1.16}$$

where κ indicates the weight put on the bivariate normal (Gaussian) copula. The mixture copula shown in equation (1.16) nests both Gaussian and Clayton copulas as special cases. When $\kappa = 1$, the mixture copula reduces to Gaussian copula. When $\kappa = 0$, the mixture copula reduces to Clayton copula. In the following simulation, I take different κ values of 0, 0.25, 0.375, 0.5 and 1 to generate random samples with different levels of asymmetric dependence from the highest to lowest.

1.3.2 Simulation with Copula-GARCH Model

The joint distribution of generated random samples is governed by the mixture Gaussian-Clayton copula model. I still need to specify some model to best mimic the marginal distributions of asset returns. GARCH(1,1) process is a well-known parsimonious model for stock returns. Following Hong, Tu, and Zhou (2007), I model the marginal distributions of the return series with a GARCH(1,1) with no ARMA components. Basically, the return series is modeled to be equal to an expected return component plus a random error term that follows a GARCH(1,1) process. I first fit the copula-GARCH model to the data to estimate the related parameters using Maximum Likelihood (ML) approach. I then plug the ML estimates back in the model and use it as the data generating process (DGP) in simulation. To be conservative, instead of using portfolios that show clear asymmetric dependence, such as the smallest stock portfolio or momentum portfolios, etc., I use the 5^{th} smallest value-weighted size portfolio and the market return to estimate copula and GARCH parameters. Empirically, I do not find any evidence for asymmetric dependence for size 5 portfolio, hence using it to calibrate the parameters impose a harder challenge for the tests. It is interesting to see whether the tests have reasonable power under such parameter settings. Hong, Tu, and Zhou (2007) has done a similar practice in their simulation exercises.

[Insert Table 1.1 about here]

Table 1.1 gives the ML estimates from fitting the data to the GARCH(1,1) process using the full sample period. Panel A lists the parameter estimates for return series of value-weighted size 5 portfolio and Panel B lists the estimates for the market return series. All the estimated parameters are statistically significant at 5% level.

Taking those estimates as the population parameters, I are able to simulate the data with the copula-GARCH model using the following detailed steps.

1. For a given κ, draw a bivariate uniform random sample of size T from the mixture Gaussian-Clayton copula model;

2. Apply inverse standard normal CDF transformation to get a bivariate standard normal random sample with pre-specified dependence structure;

3. Feed each series of the joint normal random sample into the univariate GARCH(1,1) process as the innovation terms to generate simulated joint return series;

4. The simulated data vectors will each follow a GARCH(1,1) process and the perceived dependence structure governed by the mixture copula model.

5. Repeat step 1 to 4 for 1,000 times to get 1,000 simulated random samples.

6. Repeat step 1 to 5 with different sample sizes T. Specifically, I consider T=240, 420 and 600.

[Insert Table 1.2 about here]

The sample sizes follow common choices used in the literature. $T = 240$ stands for 20 years of monthly frequency data. $T = 420$ is the length of the subsample data period as used in Hong, Tu, and Zhou (2007). $T = 600$ stands for 50 years of monthly frequency data and is close to the full sample data length ($T = 588$) used in this paper. In simulation, I use one fixed bandwidth for each 1,000 random samples generated from the same DGP. The fixed bandwidth is set to be equal to the average of the 1,000 bandwidths computed for each of the 1,000 random samples via likelihood cross-validation. Similar practice is conducted for the optimal block length selection. The expected block length for each 1,000 random samples generated from the same DGP is fixed to be the average of the 1,000 optimal block lengths computed using Patton, Politis, and White (2009) algorithm. Averaging bandwidth and block length across random samples drawn from the same DGP could potentially reduce some sampling randomness and make the simulation results more stable.

Table 1.2 reports the empirical size and power for both tests when the nominal size is set at 5% based on 1,000 simulations. Powers are reported with different DGPs of different degrees of asymmetric dependence levels (from $\kappa = 1$ to $\kappa = 0$) and at various sample sizes. I report size and power of Hong, Tu, and Zhou (2007) test (HTZ test hereafter) computed based on both asymptotic distribution and stationary bootstrap with 399 replications.

Based on the standard paired bootstrap procedure described in Cameron and Trivedi (2005) and in Horowitz (2001), I construct a pivotal (standardized) statistic when bootstrapping HTZ test statistic to achieve asymptotic refinement. I obtain the variance estimates of HTZ test statistic via sub-bootstrap, i.e. within each bootstrap replication, I bootstrap the replicated sample again to estimate the standard error based on a series of sub-bootstrapped statistics. Since I am estimating the variance rather than tail quantiles or critical values, a fairly small number of resamples is sufficient for consistent estimates. Following Racine (1997), I set the number of sub-bootstrap replications at one tenth of the original number bootstrap replications, i.e. $B_{sub} = 20$. But the bootstrap results of HTZ test does not yield better power than their asymptotic counterparts. However, the empirical sizes are much closer to the nominal values than those based on asymptotic theory.

The last column reports the power increase when inference of both tests is based on stationary bootstrap and the exceedance level is set at 0. Since the inference method is the same, we attribute this power increase to better information summarized by the entropy measure. The average power increase is computed as mean of differences among all the simulation scenarios considered in this paper.

I find a pattern that the average power increase is getting more significant as the nominal test size decreases, i.e. the entropy test gives better inference results when I want to report the testing results in a more accurate manner. At nominal size of 10%, the average power difference is only 0.03 or 4%. While at nominal size of 5%, the average power increase is 0.103 or 17.3% and when the nominal size is set at 1%, the the average power increase is 0.245 or a huge 84.6% increase. I can see that the role of information is very significant in making better statistical inference.

It indicates that the entropy test on average has higher power than HTZ test for different DGP that reflects various degree of asymmetric dependence. The difference in power varies with the dependence structure of the DGP. When the simulated data have very strong asymmetric dependence, the performance of both tests are close to each other. If the DGP is a bivariate Clayton copula ($\kappa = 0$), the difference in power is quite small (about 0.14 for $T = 240$) and the difference vanishes as sample size increase to $T = 600$. The power difference is most pronounced when the degree of asymmetric dependence is not very strong. When the DGP is a 37.5% mixed Gaussian-Clayton copula, the power of the entropy test is about 4 times higher than the power of HTZ test for smaller sample sizes ($T = 240$ or $T = 420$). The difference shrinks as the sample size increases to 600, but the power of the entropy test is still twice as large as the power of HTZ test.

The improvement of power for both tests with larger sample size is expected, especially for HTZ test based on asymptotic distribution. I tried to make inferences of HTZ test using stationary bootstrap, but the results do not show improvement upon inference based on asymptotic distribution, so I report the size and power of their test based on asymptotic theory. When the underlying DGP is of symmetric dependence, i.e. the bivariate Gaussian case with $\kappa = 1$, the probability of rejection is the empirical size of the tests. The sizes of both tests are reported in the top left panel in Table 1.2.

[Insert Table 1.3 about here]

Table 1.3 and Table 1.4 report the empirical size and power for both tests when the nominal size is set at 1% and 10% respectively based on 1,000 simulations. The results reaffirm the conclusions drawn from Table 1.2. The entropy test shows higher power for almost all different DGPs.

[Insert Table 1.4 about here]

1.4 Is Asymmetry Rare?

In this section, I apply the entropy measure to test whether there exists statistically significant asymmetry in common portfolios sorted by size (market capitalization), book-to-market ratio and momentum (past return).

1.4.1 Data

Following existing studies on testing for asymmetric correlations, I consider portfolios of stocks sorted by popular characteristics, i.e. size, book-to-market ratio, and momentum. As Ang and Chen (2002), I use value-weighted returns of for both size and book-to-market decile portfolios, and use equal-weighted returns for decile momentum portfolios which are formed based on prior 2 to 12 month cumulative return. Return on CRSP (Center for Research in Security Prices) value-weighted market index based on stocks listed in NYSE/AMEX/NASDAQ is used as a proxy for the market return. All returns are at the monthly frequency and are in excess of the risk-free rate which is taken as the one-month T-bill rate. The entire data are available from Kenneth French's site.[4] The sample period is from January 1965 to December 2013 (588 observations in total).

1.4.2 Empirical Testing Results

Table 1.5 provides the testing results on the size portfolios. At the usual 5% level, the entropy test rejects symmetry for all size portfolios from the 1st to 6th smallest size port-

[4] I are grateful to Kenneth French for making the data available at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

folios. In contrast, the existing model-free test of Hong, Tu, and Zhou (2007) can only reject symmetry for the smallest size portfolio based on either the singleton exceedance level $c = \{0\}$ or the multiple exceedance levels $c = \{0, 0.5, 1, 1.5\}$. It is interesting to observe that the entropy test statistics decrease monotonically as the firm size increases with the only exception of the 9th decile portfolio. Similar patterns also hold for the asymmetric correlation test statistic. Intuitively, this should be true too. The larger the firm, the more it resembles the market, and hence the asymmetry relative to the market reduces.

[Insert Table 1.5 about here]

Note that while the P-values of the entropy test are computed based on 399 stationary bootstraps, the P-values of the asymmetric correlation test are computed from the asymptotic Chi-square distribution with degree of freedom 1 in the singleton and multiple exceedance levels, respectively. For the entropy test, because consistent nonparametric kernel estimation requires a fairly large sample size, I use only the single exceedance level. This is also consistent with earlier Monte Carlo simulations that the entropy test based on a singleton exceedance level yields uniformly better power and size properties than the multiple exceedance level. In contrast, estimating correlations do not require as many samples as in the density estimation case. Hence, it is not surprising that the asymmetric correlation test yields similar results with the singleton or the multiple exceedance levels.

The empirical testing results from both tests for value-weighted book-to-market ratio portfolios are reported in Table 1.6. In the shorter time period from 1965 to 1999 (the same period as used in Hong, Tu, and Zhou (2007)), neither test detects any significant asymmetry in value-weighted book-to-market ratio portfolios, nor the test statistics show any clear pattern (See Table 1.B.2 in the appendix).

In the full sample, the singleton entropy test finds that the 9th highest book-to-market ratio portfolio shows significant asymmetric dependence with the market return at 1% level. The joint entropy test detects more significance in book-to-market ratio portfolios. The S_ρ measure shows a roughly increasing pattern when I go from low to high book-to-market ratio portfolios. In Ang and Chen (2002), they also find an increasing pattern of their H statistic when moving from growth (low book-to-market ratio) stocks to value (high book-to-market ratio) stocks.

[Insert Table 1.6 about here]

Table 1.7 gives the empirical testing results for equal-weighted momentum portfolios. Both tests find significant asymmetry for the return of the highest momentum portfolio (the highest past winner portfolio). The entropy test, in addition, finds statistically significant asymmetric dependence in the return of the lowest momentum portfolio (the biggest past loser portfolio). This finding is consistent with Ang and Chen (2002), which shows that bivariate normal model is rejected when fitting to the past loser portfolio returns, i.e. the returns exhibit asymmetric correlations. However, HTZ test fails to detect such asymmetric correlation in the past loser portfolio. I also find that all of the equal-weighted decile momentum portfolios shows significant asymmetric dependence at conventional significance levels. The test statistic S_ρ increases when I go to either lower or higher ends and is the lowest in the middle deciles. The pattern is consistent with that of the J statistic in HTZ test, but again due to lower power in finite samples, their test fails to attain statistical significance.

[Insert Table 1.7 about here]

1.5 Conclusion

Asymmetric dependence in stock returns is important for both portfolio management and risk hedging. However, existing tests focus only on asymmetric correlations, a special case of asymmetric dependence because correlation coefficient is only a measure of linear dependence that ignores higher order dependence. In this paper, I propose to use a metric entropy to measure and construct a model-free test for asymmetric dependence in bivariate return data. Econometrically, the test extends the univariate test of asymmetry proposed by Racine and Maasoumi (2007) to the bivariate case that is of interest in finance.

This paper makes several original contributions to the literature. First of all, the model-free entropy test of asymmetric dependence extends the univariate test of asymmetry proposed by Racine and Maasoumi (2007) to the bivariate case that is of interest in finance.

Secondly, with Monte Carlo simulations, I find that the entropy test has correct size, and has greater power in finite samples than the existing model-free test of asymmetric correlation proposed by Hong, Tu, and Zhou (2007).

Thirdly, I have find that based on the entropy test, statistically significant asymmetries are detected for most common portfolios, such as those sorted on size, book to market ratio and momentum. In contrast, Hong, Tu, and Zhou (2007) only identify a few. Specifically, I find patterns that are more consistent with findings documented in prior studies. For example, I find that smaller size portfolios show stronger asymmetric dependence, which is consistent with the findings in Ang and Chen (2002) and Kroner and Ng (1998). I also find that asymmetric dependence increases with the book-to-market ratio, which is consistent with an empirical fact that "Value stocks are more asymmetric than growth stocks" as described in Ang and Chen (2002).

Finally, the proposed entropy test is very flexible. It works well for both continuous and discrete data types. It is also applicable to either i.i.d or stationary time series data. Therefore, the test has great potentials to be applied to other studies. While the paper applies the test of asymmetric dependence to the US stocks, it will be of interest to apply the new method to the international markets to assess cross country asymmetric dependence of stock returns. It will also be of interest to apply the methodology of this paper to bonds, currencies and other asset classes. These will be potential topics of future research.

1.A Appendix: Tawn (1988) Theorem

Theorem 1.A.1. Tawn (1988) *If* $C_1(u_1, u_2), C_2(u_1, u_2), \ldots, C_n(u_1, u_2)$ *are bivariate copula functions, then*

$$C(u_1, u_2) = w_1 \cdot C_1(u_1, u_2) + w_2 \cdot C_2(u_1, u_2) + \cdots + w_n \cdot C_n(u_1, u_2)$$

is again a copula for $w_i \geq 0$ *and* $\sum_{i=1}^{n} w_i = 1$.

1.B Appendix: Additional Tables

Table 1.B.1: Subsample Test Results of Asymmetric Dependence: Size Portfolios

	Entropy-based Test C={0}		HTZ Test C={0}		HTZ Test C={0, 0.5, 1,1.5}	
Portfolio	$S_\rho \times 100$	P-value	Test-stat	P-value	Test-stat	P-value
Size 1	1.820	0.105	2.458	0.117	9.728	0.045
Size 2	1.591	0.083	0.790	0.374	0.942	0.918
Size 3	1.473	0.175	0.549	0.459	0.856	0.931
Size 4	1.280	0.221	0.339	0.560	0.584	0.965
Size 5	1.385	0.165	0.252	0.616	4.878	0.300
Size 6	1.237	0.286	0.120	0.729	3.924	0.416
Size 7	0.971	0.561	0.016	0.898	0.706	0.951
Size 8	1.015	0.454	0.023	0.878	0.401	0.982
Size 9	0.881	0.526	0.001	0.972	0.008	1.000
Size 10	0.954	0.544	0.001	0.980	0.111	0.999

The sample period is from January 1965 to December 1999. P-values of entropy test are computed based on 399 stationary bootstraps. P-values of HTZ (2007) test are computed based on asymptotic Chi-square(4) distribution.

Table 1.B.2: Subsample Test Results of Asymmetric Dependence: Book-to-Market Portfolios

	Entropy-based Test C={0}		HTZ Test C={0}		HTZ Test C={0, 0.5, 1,1.5}	
Portfolio	$S_\rho \times 100$	P-value	Test-stat	P-value	Test-stat	P-value
BE/ME 1	0.820	0.516	0.022	0.883	0.341	0.987
BE/ME 2	0.928	0.391	0.020	0.887	0.208	0.995
BE/ME 3	0.704	0.739	0.042	0.837	0.251	0.993
BE/ME 4	1.054	0.411	0.117	0.733	1.716	0.788
BE/ME 5	1.164	0.451	0.167	0.683	2.638	0.620
BE/ME 6	0.866	0.714	0.102	0.749	1.500	0.827
BE/ME 7	1.410	0.356	0.121	0.728	1.008	0.909
BE/ME 8	1.523	0.185	0.278	0.598	2.570	0.632
BE/ME 9	1.623	0.183	0.504	0.478	1.180	0.881
BE/ME 10	1.420	0.308	0.588	0.443	2.896	0.575

The sample period is from January 1965 to December 1999. P-values of entropy test are computed based on 399 stationary bootstraps. P-values of HTZ (2007) test are computed based on asymptotic Chi-square(4) distribution.

Table 1.B.3: Subsample Test Results of Asymmetric Dependence: Momentum Portfolios

	Entropy-based Test C=0		HTZ Test C=0		HTZ Test C=0, 0.5, 1,1.5	
Portfolio	$S_\rho \times 100$	P-value	Test-stat	P-value	Test-stat	P-value
L	2.526	0.000	2.162	0.141	4.449	0.349
2	1.503	0.080	1.231	0.267	3.009	0.556
3	1.402	0.123	0.946	0.331	4.572	0.334
4	1.434	0.133	0.758	0.384	4.412	0.353
5	1.779	0.033	0.694	0.405	4.088	0.394
6	1.563	0.063	0.722	0.396	0.794	0.939
7	1.505	0.108	0.585	0.444	3.445	0.486
8	1.431	0.123	0.670	0.413	0.911	0.923
9	1.528	0.110	1.088	0.297	1.636	0.802
W	1.750	0.100	1.648	0.199	10.266	0.036

The sample period is from January 1965 to December 1999. P-values of entropy test are computed based on 399 stationary bootstraps. P-values of HTZ (2007) test are computed based on asymptotic Chi-square distribution.

References

Ang, A., and J. Chen. 2002. Asymmetric correlations of equity portfolios. *Journal of Financial Economics* 63:443–94.

Ang, A., J. Chen, and Y. Xing. 2006. Downside risk. *Review of Financial Studies* 19:1191–239.

Backus, D., M. Chernov, and S. Zin. 2014. Sources of entropy in representative agent models. *The Journal of Finance* 69:51–99.

Bae, K.-H., G. A. Karolyi, and R. M. Stulz. 2003. A new approach to measuring financial contagion. *Review of Financial Studies* 16:717–63.

Ball, R., and S. Kothari. 1989. Nonstationary expected returns: Implications for tests of market efficiency and serial correlation in returns. *Journal of Financial Economics* 25:51–74.

Bekaert, G., and G. Wu. 2000. Asymmetric volatility and risk in equity markets. *Review of Financial Studies* 13:1–42.

Cameron, A. C., and P. K. Trivedi. 2005. *Microeconometrics: methods and applications*. Cambridge university press.

Cho, Y.-H., and R. F. Engle. 1999. Time-varying betas and asymmetric effect of news: Empirical analysis of blue chip stocks. Working Paper, National Bureau of Economic Research.

Conrad, J., M. N. Gultekin, and G. Kaul. 1991. Asymmetric predictability of conditional variances. *Review of Financial Studies* 4:597–622.

Cont, R. 2001. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance* 1:223–36.

Davidson, R., and J. G. MacKinnon. 2000. Bootstrap tests: How many bootstraps? *Econometric Reviews* 19:55–68.

Efron, B. 1982. *The jackknife, the bootstrap, and other resampling plans*, vol. 38. Society for industrial and applied mathematics.

Embrechts, P., A. McNeil, and D. Straumann. 2002. Correlation and dependence in risk management: properties and pitfalls. *Risk management: value at risk and beyond* 176–223.

Fama, E. F. 1970. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance* 25:383–417.

Granger, C., E. Maasoumi, and J. Racine. 2004. A dependence metric for possibly nonlinear processes. *Journal of Time Series Analysis* 25:649–69.

Gul, F. 1991. A theory of disappointment aversion. *Econometrica* 59:667–86.

Hall, P. 1992. The bootstrap and edgeworth expansion. *Springer series in statistics* .

Harvey, C. R., and A. Siddique. 2000. Conditional skewness in asset pricing tests. *The Journal of Finance* 55:1263–95.

Hong, Y., J. Tu, and G. Zhou. 2007. Asymmetries in stock returns: Statistical tests and economic evaluation. *Review of Financial Studies* 20:1547–81.

Hong, Y., and H. White. 2005. Asymptotic distribution theory for nonparametric entropy measures of serial dependence. *Econometrica* 73:837–901.

Horowitz, J. L. 2001. The bootstrap. *Handbook of econometrics* 5:3159–228.

Kahneman, D., J. L. Knetsch, and R. H. Thaler. 1990. Experimental tests of the endowment effect and the coase theorem. *Journal of Political Economy* 98:1325–48.

Kahneman, D., and A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 263–91.

Kroner, K., and V. K. Ng. 1998. Modeling asymmetric comovements of asset returns. *Review of Financial Studies* 11:817–44.

Kullback, S., R. Leibler, et al. 1951. On information and sufficiency. *The Annals of Mathematical Statistics* 22:79–86.

Künsch, H. R. 1989. The jackknife and the bootstrap for general stationary observations. *Annals of statistics* 17:1217–41.

Li, Q., and J. S. Racine. 2007. *Nonparametric econometrics: Theory and practice*. Princeton University Press.

Maasoumi, E. 1993. A compendium to information theory in economics and econometrics. *Econometric Reviews* 12:137–81.

Maasoumi, E., and J. Racine. 2002. Entropy and predictability of stock market returns. *Journal of Econometrics* 107:291–312.

Maasoumi, E., and J. S. Racine. 2008. A robust entropy-based test of asymmetry for discrete and continuous processes. *Econometric Reviews* 28:246–61.

Nelsen, R. B. 1999. *An introduction to copulas*. Springer.

Okimoto, T. 2008. New evidence of asymmetric dependence structures in international equity markets. *Journal of Financial and Quantitative Analysis* 43:787–815.

Parzen, E. 1962. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics* 33:1065–76.

Patton, A., D. N. Politis, and H. White. 2009. Correction to "automatic block-length selection for the dependent bootstrap" by d. politis and h. white. *Econometric Reviews* 28:372–5.

Patton, A. J. 2004. On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. *Journal of Financial Econometrics* 2:130–68.

Politis, D. N., and J. P. Romano. 1994. The stationary bootstrap. *Journal of the American Statistical Association* 89:1303–13.

Politis, D. N., and H. White. 2004. Automatic block-length selection for the dependent bootstrap. *Econometric Reviews* 23:53–70.

Racine, J. 1997. Consistent significance testing for nonparametric regression. *Journal of Business & Economic Statistics* 15:369–78.

Racine, J. S., and E. Maasoumi. 2007. A versatile and robust metric entropy test of time-reversibility, and other hypotheses. *Journal of Econometrics* 138:547–67.

Rodriguez, J. C. 2007. Measuring financial contagion: A copula approach. *Journal of Empirical Finance* 14:401–23.

Rosenblatt, M. 1956. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics* 27:832–7.

Schwert, G. W. 1989. Why does stock market volatility change over time? *The Journal of Finance* 44:1115–53.

Shannon, C. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27:379–423.

Skaug, H. J., and D. Tjøstheim. 1993. A nonparametric test of serial independence based on the empirical distribution function. *Biometrika* 80:591–602.

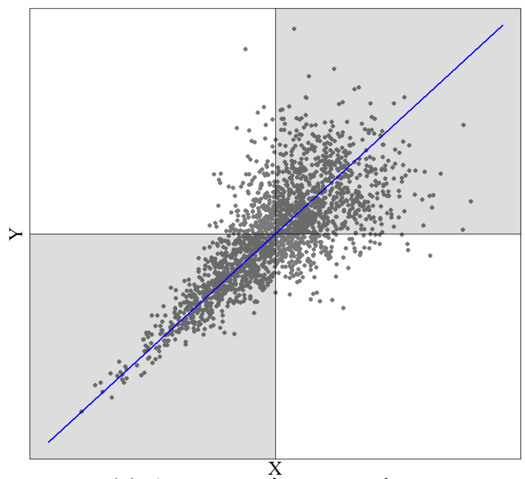
Sklar, M. 1959. *Fonctions de répartition à n dimensions et leurs marges.* Université Paris 8.

Tawn, J. A. 1988. Bivariate extreme value theory: models and estimation. *Biometrika* 75:397–415.

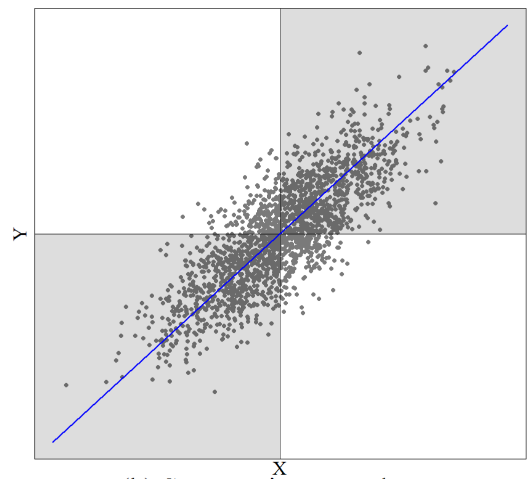
Tjøstheim, D. 1996. Measures of dependence and tests of independence. *Statistics: A Journal of Theoretical and Applied Statistics* 28:249–84.

Ullah, A. 1996. Entropy, divergence and distance measures with econometric applications. *Journal of Statistical Planning and Inference* 49:137–62.

Figure 1.1: Illustration of Asymmetric Dependence



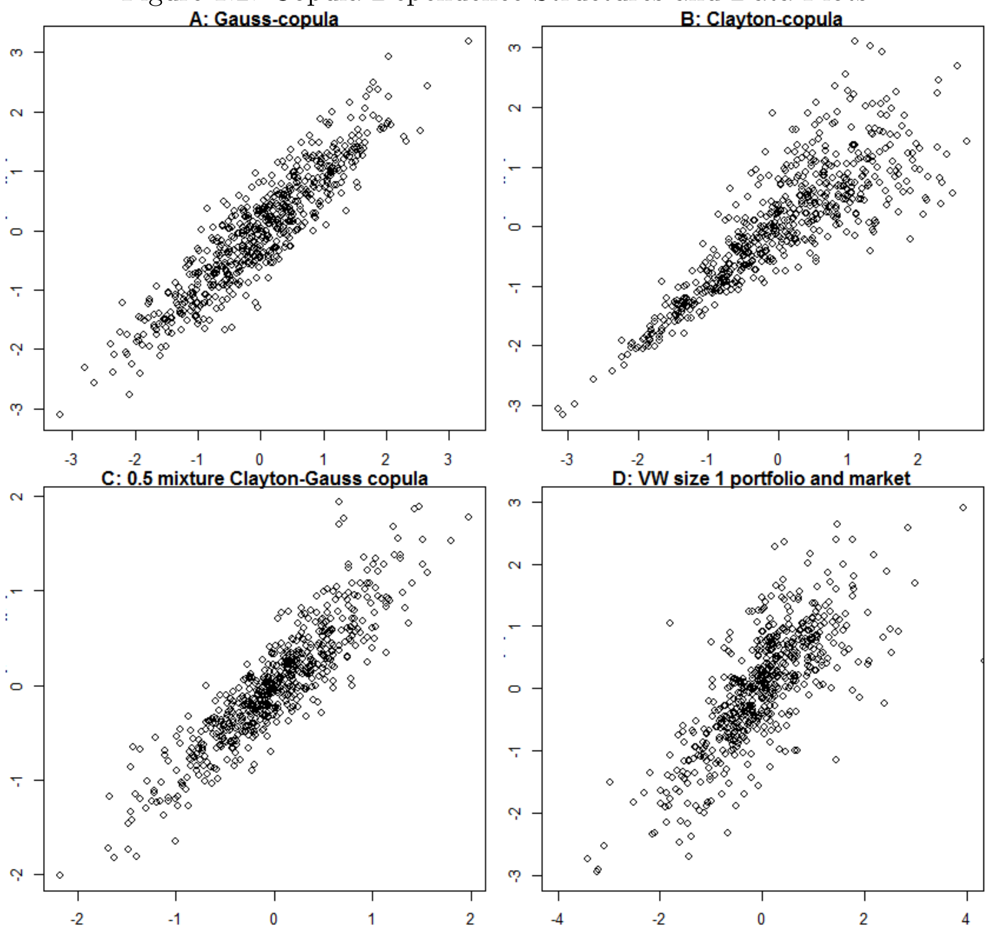
(a) Asymmetric Dependence



(b) Symmetric Dependence

This figure shows two scatter plots from two data-generating processes with different dependence structures. Subfigure (a) shows a scatter plot of of 2000 points generated from a Clayton copula that is known to have stronger left tail dependence than right tail dependence. Subfigure (b) is a similar plot where the DGP is a bivariate normal distribution that has symmetric dependence at both tails. The blue lines in both subfigures are linear regression lines fitted to the data.

Figure 1.2: Copula Dependence Structures and Data Plots



This figure shows scatter plots of random samples generated by Gaussian copula (A), Clayton copula (B) and mixed Gaussian-Clayton copula with mixing weights of 0.5 each (C), as well as the actual data plots of the value-weighted returns of the smallest size portfolio and the market returns.

Table 1.1: ML estimates for GARCH(1,1) processes

Panel A: Value-Weighted Size 5 Portfolio Return Series

	Estimate	S.E.	t-value	Pr(>|t|)
μ	0.94	0.23	4.02	0.00
ω	4.97	2.31	2.15	0.03
α	0.14	0.05	2.99	0.00
β	0.73	0.09	7.82	0.00

Panel B: Value-Weighted Market Return Series

	Estimate	S.E.	t-value	Pr(>|t|)
μ	0.56	0.17	3.29	0.00
ω	1.14	0.55	2.06	0.04
α	0.11	0.03	3.71	0.00
β	0.84	0.04	23.28	0.00

The table reports maximum likelihood estimates for parameters of GARCH(1,1) processes used to fit the 5th smallest size portfolio return (Panel A) and the market return (Panel B) data. The GARCH models are then used as the data-generating processes to simulate stock return data. The specification is a standard GARCH(1,1) process: $r_{i,t} = \mu_i + \varepsilon_{i,t}$ where $\varepsilon_{i,t}$ is normally distributed with a time-varying variance $\sigma^2_{i,t} = \omega_i + \alpha_i \varepsilon^2_{i,t-1} + \beta_i \sigma^2_{i,t-1}$. μ is the unconditional mean for the stock return process. α is the autoregressive parameter and β is the moving average parameter in the GARCH process. ω is the constant term in the time-varying variance process.

Table 1.2: Size and Power: Entropy-based test and HTZ test (5% nominal size)

	Entropy Test	HTZ Test					Power Increase
		Asymptotic Theory				Bootstrap	
	C={0}	C={0}	C={0, 0.5}	C={0, 0.5, 1}	C={0, 0.5, 1,1.5}	C={0}	C={0}
			Panel A: $\kappa = 100\%$ **(size)**				
T = 240	0.022	0.000	0.000	0.003	0.023	0.039	N/A
T = 420	0.033	0.000	0.000	0.001	0.008	0.045	N/A
T = 600	0.044	0.000	0.000	0.000	0.004	0.043	N/A
			Panel B: $\kappa = 50\%$				
T = 240	0.094	0.005	0.021	0.046	0.092	0.233	-0.139
T = 420	0.223	0.003	0.022	0.034	0.068	0.323	-0.100
T = 600	0.405	0.014	0.050	0.060	0.088	0.469	-0.064
			Panel C: $\kappa = 37.5\%$				
T = 240	0.312	0.086	0.086	0.104	0.167	0.423	-0.111
T = 420	0.729	0.215	0.142	0.140	0.176	0.582	0.147
T = 600	0.937	0.426	0.299	0.249	0.263	0.758	0.179
			Panel D: $\kappa = 25\%$				
T = 240	0.748	0.478	0.325	0.299	0.380	0.549	0.199
T = 420	0.991	0.791	0.618	0.504	0.510	0.725	0.266
T = 600	1.000	0.969	0.867	0.763	0.723	0.854	0.146
			Panel E: $\kappa = 0\%$				
T = 240	0.952	0.857	0.742	0.690	0.717	0.614	0.338
T = 420	1.000	0.983	0.937	0.895	0.880	0.766	0.234
T = 600	1.000	0.993	0.982	0.972	0.958	0.859	0.141
Avg. Power	0.699	0.485	0.424	0.396	0.419	0.596	0.103 (17.3%)

The nominal size of the tests is set at 5%. The table reports the probabilities of rejecting the null hypothesis of symmetric dependence based on 1,000 Monte Carlo simulations. Different values of $\hat{\mathrm{I}}^{o}$ governs the degree of left tail dependence of the underlying data generating process (DGP). When $\kappa = 100\%$, the DGP is a joint normal distribution and the respective rejecting probabilities are empirical sizes. In all other cases, the rejection probabilities are powers. The last column reports power increases when inferences of both tests are based on 399 stationary bootstraps and the exceedance level is set at 0. The average power increase is computed as mean of differences among all the simulation cases considered in this paper.

Table 1.3: Size and Power: Entropy-based test and HTZ test (1% nominal size)

	Entropy Test	HTZ Test					Power Increase
		Asymptotic Theory				Bootstrap	
	C={0}	C={0}	C={0, 0.5}	C={0, 0.5, 1}	C={0, 0.5, 1,1.5}	C={0}	C={0}
Panel A: $\kappa = 100\%$ **(size)**							
T = 240	0.006	0.000	0.000	0.001	0.009	0.006	N/A
T = 420	0.005	0.000	0.000	0.000	0.000	0.007	N/A
T = 600	0.008	0.000	0.000	0.000	0.000	0.007	N/A
Panel B: $\kappa = 50\%$							
T = 240	0.013	0.003	0.003	0.011	0.039	0.074	-0.061
T = 420	0.055	0.000	0.003	0.004	0.024	0.092	-0.037
T = 600	0.129	0.000	0.006	0.009	0.025	0.133	-0.004
Panel C: $\kappa = 37.5\%$							
T = 240	0.084	0.010	0.026	0.034	0.074	0.145	-0.061
T = 420	0.380	0.032	0.029	0.035	0.054	0.241	0.139
T = 600	0.733	0.084	0.060	0.069	0.103	0.367	0.366
Panel D: $\kappa = 25\%$							
T = 240	0.370	0.159	0.111	0.113	0.204	0.221	0.149
T = 420	0.915	0.415	0.262	0.218	0.259	0.325	0.590
T = 600	0.998	0.740	0.537	0.419	0.401	0.507	0.491
Panel E: $\kappa = 0\%$							
T = 240	0.745	0.559	0.466	0.446	0.506	0.273	0.472
T = 420	0.992	0.865	0.754	0.668	0.668	0.437	0.555
T = 600	1.000	0.960	0.922	0.877	0.857	0.659	0.341
Avg. Power	0.535	0.319	0.265	0.242	0.268	0.290	0.245 (84.6%)

The nominal size of the tests is set at 1%. The table reports the probabilities of rejecting the null hypothesis of symmetric dependence based on 1,000 Monte Carlo simulations. Different values of $\hat{\mathrm{I}}^{\mathrm{o}}$ governs the degree of left tail dependence of the underlying data generating process (DGP). When $\kappa = 100\%$, the DGP is a joint normal distribution and the respective rejecting probabilities are empirical sizes. In all other cases, the rejection probabilities are powers. The last column reports power increases when inferences of both tests are based on 399 stationary bootstraps and the exceedance level is set at 0. The average power increase is computed as mean of differences among all the simulation cases considered in this paper.

Table 1.4: Size and Power: Entropy-based test and HTZ test (10% nominal size)

	Entropy Test	HTZ Test					Power Increase
		Asymptotic Theory				Bootstrap	
	C={0}	C={0}	C={0, 0.5}	C={0, 0.5, 1}	C={0, 0.5, 1,1.5}	C={0}	C={0}
Panel A: $\kappa = 100\%$ **(size)**							
T = 240	0.058	0.000	0.003	0.005	0.035	0.085	N/A
T = 420	0.077	0.000	0.000	0.004	0.017	0.093	N/A
T = 600	0.107	0.000	0.000	0.000	0.008	0.100	N/A
Panel B: $\kappa = 50\%$							
T = 240	0.196	0.016	0.046	0.088	0.144	0.386	-0.190
T = 420	0.390	0.033	0.053	0.073	0.112	0.513	-0.123
T = 600	0.588	0.068	0.102	0.127	0.161	0.682	-0.094
Panel C: $\kappa = 37.5\%$							
T = 240	0.514	0.200	0.168	0.174	0.237	0.608	-0.094
T = 420	0.854	0.428	0.277	0.235	0.280	0.754	0.100
T = 600	0.979	0.731	0.484	0.403	0.411	0.907	0.072
Panel D: $\kappa = 25\%$							
T = 240	0.865	0.660	0.500	0.438	0.505	0.748	0.117
T = 420	1.000	0.932	0.770	0.686	0.660	0.871	0.129
T = 600	1.000	0.998	0.964	0.898	0.851	0.947	0.053
Panel E: $\kappa = 0\%$							
T = 240	0.982	0.944	0.864	0.794	0.830	0.807	0.175
T = 420	1.000	0.996	0.979	0.956	0.950	0.862	0.138
T = 600	1.000	0.998	0.992	0.986	0.981	0.926	0.074
Avg. Power	0.781	0.584	0.517	0.488	0.510	0.751	0.030 (4%)

The nominal size of the tests is set at 10%. The table reports the probabilities of rejecting the null hypothesis of symmetric dependence based on 1,000 Monte Carlo simulations. Different values of $\hat{\text{I}}^{o}$ governs the degree of left tail dependence of the underlying data generating process (DGP). When $\kappa = 100\%$, the DGP is a joint normal distribution and the respective rejecting probabilities are empirical sizes. In all other cases, the rejection probabilities are powers. The last column reports power increases when inferences of both tests are based on 399 stationary bootstraps and the exceedance level is set at 0. The average power increase is computed as mean of differences among all the simulation cases considered in this paper.

Table 1.5: Empirical Test Results of Asymmetric Dependence: Size Portfolios

	Entropy-based Test C={0}		HTZ Test C={0}		HTZ Test C={0, 0.5, 1,1.5}	
Portfolios	$S_\rho \times 100$	P-value	Test-stat	P-value	Test-stat	P-value
Size 1	2.027	0.010	4.212	0.040	9.715	0.046
Size 2	1.963	0.000	2.049	0.152	3.281	0.512
Size 3	1.868	0.020	0.937	0.333	1.108	0.893
Size 4	1.689	0.013	0.613	0.434	2.095	0.718
Size 5	1.690	0.030	0.431	0.512	5.015	0.286
Size 6	1.596	0.045	0.234	0.629	3.134	0.536
Size 7	1.477	0.065	0.092	0.761	0.849	0.932
Size 8	1.510	0.085	0.099	0.753	0.146	0.997
Size 9	1.695	0.075	0.005	0.945	0.030	1.000
Size 10	1.511	0.055	0.008	0.930	0.029	1.000

The sample period is from January 1965 to December 2013. P-values of entropy test are computed based on 399 stationary bootstraps. P-values of Hong, Tu, and Zhou (2007) test are computed based on asymptotic Chi-square distribution.

Table 1.6: Empirical Test Results of Asymmetric Dependence: Book-to-Market Portfolios

	Entropy-based Test C={0}		HTZ Test C={0}		C={0, 0.5, 1,1.5}	
Portfolios	$S_\rho \times 100$	P-value	Test-stat	P-value	Test-stat	P-value
BE/ME 1	1.248	0.115	0.023	0.880	0.341	0.987
BE/ME 2	1.208	0.085	0.024	0.876	0.093	0.999
BE/ME 3	1.003	0.263	0.060	0.807	0.066	0.999
BE/ME 4	1.626	0.138	0.064	0.800	1.829	0.767
BE/ME 5	1.610	0.055	0.145	0.703	2.769	0.597
BE/ME 6	1.815	0.025	0.054	0.817	1.099	0.894
BE/ME 7	1.805	0.058	0.082	0.774	0.590	0.964
BE/ME 8	1.571	0.098	0.226	0.634	2.954	0.566
BE/ME 9	2.162	0.005	0.447	0.504	1.667	0.797
BE/ME 10	1.657	0.075	0.805	0.370	2.133	0.711

The sample period is from January 1965 to December 2013. P-values of entropy test are computed based on 399 stationary bootstraps. P-values of Hong, Tu, and Zhou (2007) test are computed based on asymptotic Chi-square distribution.

Table 1.7: Empirical Test Results of Asymmetric Dependence: Momentum Portfolios

	Entropy-based Test C={0}		HTZ Test C=0		HTZ Test C={0, 0.5, 1,1.5}	
Portfolios	$S_\rho \times 100$	P-value	Test-stat	P-value	Test-stat	P-value
L	3.597	0.003	5.191	0.023	6.369	0.173
2	2.572	0.003	2.354	0.125	5.022	0.285
3	2.237	0.018	1.452	0.228	5.407	0.248
4	1.784	0.050	1.008	0.315	5.018	0.285
5	2.155	0.005	0.915	0.339	4.468	0.346
6	1.981	0.003	0.775	0.379	0.921	0.922
7	2.385	0.000	0.717	0.397	3.915	0.418
8	1.959	0.000	1.029	0.311	2.591	0.628
9	2.298	0.000	1.850	0.174	3.507	0.477
W	2.338	0.000	3.329	0.068	13.141	0.011

The sample period is from January 1965 to December 2013. P-values of entropy test are computed based on 399 stationary bootstraps. P-values of Hong, Tu, and Zhou (2007) test are computed based on asymptotic Chi-square distribution.

Chapter 2

Asymmetric Dependence and the Cross-section of Stock Returns

Abstract

This paper examines how non-linear asymmetric dependence between individual stock return and market return is cross-sectionally priced. Motivated by Ang, Chen, and Xing (2006), I construct proxies for the dependence with downside and upside market separately based on non-parametric kernel estimated joint return distribution. Asymmetric dependence is measured using a metric entropy proposed by Granger, Maasoumi, and Racine (2004). Empirically, I find a risk premium (discount) for stocks with high downside (upside) dependence. Asymmetric dependence leaning toward the downside also earns a premium. The risk premia associated with downside dependence and asymmetric dependence are higher than the discount associated with upside dependence. Furthermore, downside dependence premium is almost twice as large as the downside beta premium. The findings suggest that investors' aversion to downside losses are stronger than their attraction to the upside gains.

Keywords: Asymmetric dependence, metric entropy, non-parametric kernel, asset pricing.

JEL Classification: C12, G11, G12, G17

2.1 Introduction

Asymmetric dependence among stock returns, i.e. stocks co-move more strongly when market goes down than when market goes up, have been documented by many prior studies, in the forms of asymmetric covariances, correlations and market betas (see, e.g., Ball and Kothari, 1989; Conrad, Gultekin, and Kaul, 1991; Bekaert and Wu, 2000; Ang and Chen, 2002; Bae, Karolyi, and Stulz, 2003; Ang, Chen, and Xing, 2006). Such asymmetric characteristics of stock returns are important because effective hedging relies on the dependence between assets hedged and financial instruments used. If the dependence structure is varying with the state of the market, portfolio managers may need to worry about the effectiveness of their hedges when they are most needed. Despite the importance, the asset pricing implications of asymmetric dependence on the cross-section of expected stock returns have been less studied in the literature.

Ang, Chen, and Xing (2006) find asymmetric risk premia are associated with downside and upside betas in the cross-section of stock returns. They show that stocks with higher downside betas have on average higher returns, but have mixed evidence on whether higher upside betas are associated with a discount. Since downside and upside betas are highly correlated with market betas (the correlations are above 0.75 as shown in Ang, Chen, and Xing (2006)), i.e. an increase in downside or upside betas are associated with an increase in the market beta, it is hard to distinguish the effects of downside or upside covariation from the overall covariation between stock and the market returns. Alcock and Hatherley (2013) tries to overcome this problem by constructing a beta-invariant asymmetric dependence measure that is a modified J statistic of an asymmetric correlation test proposed by Hong, Tu, and Zhou (2007). Although beta-invariant, their measure still does not capture full dependence structure since it is constructed based on exceedance correlations that can only capture conditional dependence to the second moment (linear dependence).

Under classical Capital Asset Pricing Model (CAPM), it is sufficient to consider only linear correlations (captured by the market beta) between individual stock returns and the market portfolio return. (see Sharpe, 1964; Lintner, 1965). However, more recent studies find supporting evidence that features of the joint distribution of individual stock and market returns beyond the linear correlation also determine the expected stock returns. For example, Harvey and Siddique (2000), Dittmar (2002) among others show that higher order co-moments, such as conditional coskewness and cokurtosis, also play important roles in explaining the cross-sectional returns. The main reason is that returns are assumed to be normally distributed for classical CAPM to hold,[1] while several papers documented that realized stock returns are non-normally distributed (see, e.g., Embrechts, McNeil, and Straumann, 2002; Ang and Chen, 2002). It is well known that the correlation coefficient is only a measure of linear dependence and cannot capture the full dependence structure of non-normal distributions. Some recent papers start to examine cross-sectional asset pricing implications of higher order dependence. Contemporary paper by Chabi-Yo, Ruenzi, and Weigert (2014), in a non-normal distribution framework, uses parametric copula-based tail dependence measure to explain the cross-sectional expected returns. My paper differs from theirs in that they focus on extreme lower tail dependence, or the crash sensitivities of stocks, while, muck like Ang, Chen, and Xing (2006), I focus on the downside and upside dependence when market returns are above or below the mean.

In this paper, I use an entropy-based statistic to empirically measure asymmetric dependence and study its asset pricing implications in the cross-section of expected stock returns. The entropy measure is a modified statistic of an asymmetric dependence test proposed in Chapter 1. Entropy is estimated using empirical probability densities, so it can summarize all the information of a given distribution and capture asymmetric dependence structure existed in all the moments. Ang, Chen, and Xing (2006) shows that under a simplistic representative agent model with disappointment aversion (DA) utility (Gul, 1991) and with certain parameter settings, agents require a premium to hold stocks with strong covariation with the downside market, while are willing to hold stocks with high upside potential at a discount, all else being equal. Motivated from this insight, we expect stocks with stronger

[1] Without normality assumption, CAPM also holds under the assumption of quadratic preferences, which is even less likely to be true in reality. So the violation of normality condition should be held as the major reason for the failure of CAPM.

downside asymmetric dependence, i.e. the dependence with the downside market is stronger than with the upside market, to earn higher average returns, because those stocks are highly risky in the sense that they may incur large loss when the wealth level is low, meanwhile they do not have high upside potential when the market goes up. Furthermore, as pointed out by Ang, Chen, and Xing (2006), the DA utility is kinked at certainty equivalence wealth level, so the higher-order co-moments derived from Taylor expansion, like coskewness and cokurtosis, may not approximate the utility function well globally. This is a theoretical motivation why there may exist asymmetric effects of downside and upside dependence.

In the empirical analysis, I also construct proxies for downside and upside dependence using estimated probabilities that individual stock and market returns both fall below or above the sample means. Using Center for Research in Securities Prices (CRSP) data from 1962 to 2013, I find empirical evidence that stocks with high downside (upside) dependence earn a premium (discount). Both effects are statistically and economically significant after controlling for other known characteristics in cross-sectional Fama and MacBeth (1973) regressions. The value-weighted average return (Carhart (1997) four factor adjusted alpha) of the top quintile portfolio sorted based on downside asymmetric dependence outperforms the lowest quintile portfolio by 12.34% (12.89%) per annum. In Fama-Macbeth (1973) regressions, the premium of downside asymmetric dependence cannot be explained by known characteristics, such as CAPM beta, downside or upside betas, coskewness and cokurtosis, size, book-to-market ratio, past returns and maximum daily return within a month. The downside asymmetric dependence is time-varying and shows limited predictability using its own lag. Yet when using the lagged asymmetric dependence to form a trading strategy, the spread portfolio still earns an average equal-weighted annualized return of 4.5%. The premium is both economically and statistically significant.

The rest of the paper is organized as follows. Section 2.2 introduces a modified entropy-based measure of downside asymmetric dependence. Section 2.3 shows that downside asymmetric dependence is associated with a risk premium contemporaneously using univariate, dependent bivariate portfolio sorting and firm-level cross-sectional regressions, along with a battery of robustness checks. Section 2.4 examines the time-series persistence of downside asymmetric dependence and evaluates whether lagged asymmetric dependence can predict future stock returns cross-sectionally. Section 2.5 concludes.

2.2 Measuring Asymmetric Dependence

2.2.1 Downside Asymmetric Dependence

In finance literature, prior studies utilize exceedance correlations, i.e. the conditional correlations evaluated when both individual stock and the market returns are below or above certain exceedance levels, to construct measures of asymmetry of the joint return distribution. However, it is well known that the correlation coefficient is only a measure of linear dependence and cannot reflect dependence structure beyond the second moment. To overcome this shortcoming, in the first chapter I propose to use entropy to capture asymmetric dependence existed in all the co-moments. Originated from physics and information theory as a measure of uncertainty, entropy has gained increasing popularity in economics and finance. Some important applications of entropy include Sims (2003); Backus, Chernov, and Martin (2011); Hansen (2012). Among most recent notable examples, Cabrales, Gossner, and Serrano (2013) use Shannon's entropy (Shannon, 1948) to quantify the informativeness of a ruin-averse investor's beliefs on the state of nature.

I have shown in Chapter 1 that the entropy test statistic $S_\rho(c)$, defined in equation 1.8, can successfully capture the degree of asymmetric dependence as the entropy-based test demonstrates higher finite sample power than the correlation-based test in Hong, Tu, and Zhou (2007). However, $S_\rho(c)$ is a normalized metric that is always non-negative, so it gives no direction of asymmetry dependence, i.e. it does not indicate whether the dependence is stronger in the downside or upside. In finance, investors are more concerned about the downside risk of an asset (see, e.g., Ang, Chen, and Xing, 2006). Therefore, we need a measure to distinguish the direction of asymmetric dependence.

Graphically, the degree of concentration of return pairs in a given region reflects the degree of dependence of the two variables in the local area. For example, if the points are more concentrated in the third quadrant than in the first quadrant, it indicates stronger dependence during the downside market. A proxy for the direction of asymmetric dependence can be constructed using joint probabilities of return pairs being in each region. The proxy, excessive downside probability (EDP), can be defined as the difference between a lower quadrant probability (LQP) and an upper quadrant probability (UQP). Specifically, LQP and UQP are given by

$$\mathrm{LQP}^c = Pr(\tilde{x} \leq -c,\ \tilde{y} \leq -c) = \int_{-\infty}^{c} \int_{-\infty}^{c} f(\tilde{x}, \tilde{y})\, d\tilde{x} d\tilde{y}, \tag{2.1}$$

$$\mathrm{UQP}^c = Pr(\tilde{x} \geq c,\ \tilde{y} \geq c) = \int_{c}^{+\infty} \int_{c}^{+\infty} f(\tilde{x}, \tilde{y})\, d\tilde{x} d\tilde{y}. \tag{2.2}$$

They measure probabilities of individual stock and market return pairs being both above or below the exceedance level c. When $c = 0$, higher LQP^0 (UQP^0) indicates higher tendency for the stock to co-move with the market below (above) the average levels, and hence is a good proxy for downside (upside) dependence with the market. The EDP is defined as

$$\begin{aligned} \mathrm{EDP}^c &= \mathrm{LQP}^c - \mathrm{UQP}^c \\ &= \textstyle\int_{-\infty}^{c} \int_{-\infty}^{c} [f(\tilde{x}, \tilde{y}) - f(-\tilde{x}, -\tilde{y})]\, d\tilde{x} d\tilde{y}. \end{aligned} \tag{2.3}$$

EDP is a function of exceedance level c. When c is taken to be 0, if $\mathrm{EDP}^0 > 0$, the probability that the asset goes below the mean with the market is greater than the probability that it goes up above the mean with the market, indicating stronger downside dependence. When c equals other values, EDP^c indicates the dependence difference in farther tails. Everything else equal, intuitively, most investors dislike the excessive downside probability defined above. From the viewpoint of utility theory, for example, investors with the disappointment aversion (DA) preference, which is introduced by Gul (1991) and excellently analyzed by Ang, Bekaert, and Liu (2005), weigh outcomes below a certain reference point strictly more heavily than those above it if the DA coefficient is of a usual value less than 1. In other words, the greater the EDP^c, the more they require to be compensated for.

However, the degree of this asymmetric dependence is not fully reflected by EDP^c. On the other hand, as a distance measure of between the original and rotated distributions, $S_\rho(c)$ captures the exact degree of asymmetric dependence. So a measure of downside asymmetric dependence (DownAsy) can be defined as

$$\mathrm{DownAsy}^c = Sign(\mathrm{EDP}^c) S_\rho(c), \tag{2.4}$$

where $Sign(x)$ is a sign function that takes the value of 1 if EDP^c is positive and equals -1 otherwise. It is interesting to exam the asset pricing implications of downside asymmetric dependence measured at the sample means ($c = 0$), as it closely mimics the way how Ang, Chen, and Xing (2006) define downside and upside betas, the conditional linear dependence with the market. The empirical analysis mainly emphasizes on the case of $c = 0$, so the results are directly comparable to Ang, Chen, and Xing (2006). The results of asymmetric dependence measures at farther tails are also reported as robustness checks.

2.2.2 Non-parametric estimation

Empirically, estimating the downside asymmetric dependence measure requires consistent estimation of the unknown joint density and cumulative distribution functions. Similar as in the first chapter, I use the same "Parzen-Rosenblatt" kernel density estimator as in 1.10 and the "product kernel function" given in 1.11 to consistently estimate the density functions in (1.8). $S_\rho(c)$ are then computed via numerical integration.

Besides the joint density functions, I also need to consistently estimate cumulative distribution functions in order to estimate LQP^c, UQP^c, and EDP^c defined in (2.1), (2.2), and (2.3). The cumulative distribution functions can be consistently estimated using either empirical distribution functions or kernel smoothing method suggested by Li, Li, and Racine

(2014). I choose to use kernel estimation method due to several advantages as shown in Li, Li, and Racine (2014).[2]

2.3 Data and Empirical Results

In this section, I introduce data and empirical methodology used in the paper and report the empirical findings.

2.3.1 Data and Research Design

Stock market data are from the CRSP that cover the sample period from January 1962 to December 2013. The data include all common stocks (with share codes of 10 or 11) listed on NYSE, AMEX and NASDAQ. In order to make the trading volume in NASDAQ comparable to NYSE and AMEX, volumes are adjusted based on the way proposed by Gao and Ritter (2010). Turnover ratio is calculated as the adjusted monthly trading volume divided by shares outstanding. Amihud (2002) ratio is also computed using the adjusted trading volumes. Following Acharya and Pedersen (2005), I also normalize the Amihud ratio to adjust for inflation and truncated it at 30 to eliminate the effect of outliers. The detailed steps are given in appendix.

The book value information comes from COMPUSTAT and is supplemented by the hand-collected book value data from Kenneth French's web site.[3] The book-to-market ratio is calculated by the book value of equity (assumed to be available six months after the fiscal year end) divided by current market capitalization. It is truncated at 0.5% percentile and 99.5% percentile to eliminate the effect of extreme values. Following the literature, I take natural logarithm of size, turnover ratio, and book to market ratio before controlling them as firm characteristics.

Following Jegadeesh and Titman (1993), I use returns over past six months to control

[2]Note that the empirical distribution function is a non-smooth step function that jumps up by $1/n$ at each of the n data points. As pointed out by Li, Li, and Racine (2014), the estimate is mechanically equal to 0 (1) at the sample minimum (maximum), while the true population support may not be bounded by the sample minimum and maximum. The problem is more prominent when the sample size is relatively small.

[3]The data are available at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

for the momentum effect. The sample is restricted to stocks with beginning-of-month prices between \$5 and \$1,000 to eliminate stocks whose transaction cost is a huge part of their price and those that have very high prices.[4] I construct β^-, β^+, coskewness and cokurtosis using the definitions given in Ang, Chen, and Xing (2006). Idiosyncratic volatility is calculated as the standard deviation of the CAPM residuals over 12-month horizon. Max is the maximum daily return in a month following the definition of Bali, Cakici, and Whitelaw (2011).

Factor pricing models focus on the contemporaneous risk return relationship. Classical CAPM indicates that stocks that have higher exposure to the market risk earn higher average returns over the same time period. The empirical research design of this paper closely follows Ang, Chen, and Xing (2006); Lewellen and Nagel (2006); Chabi-Yo, Ruenzi, and Weigert (2014) by investigating the contemporaneous relations between the realized risk exposure and realized average returns. This approach may raise some concern that the results may be driven by endogeneity. However, several papers documented that market risk exposures may be time-varying (see,e.g., Fama and French, 1992; Ang and Chen, 2007). In section 2.4, I also find evidence that the downside asymmetric dependence measure is time-varying, since the past DownAsy is not a good predictor of current DownAsy. Following the approach proposed by Ang, Chen, and Xing (2006), the dependence measures (LQP, UQP and DownAsy) are estimated using realized daily return data over overlapping 12-month periods. The estimates are updated monthly. Since the measures are estimated using non-parametric kernel methods that require sufficient data points for reliable estimates, I restrict the sample so that in each stock 12-month combination there are at least 100 daily observations. Furthermore, using 12-month horizon could better capture the time-varying feature of the dependence measures. Very long time intervals may lead to noisy estimates. Other risk measures (β, β^-, β^+, Ivol, Coskew, Cokurt) are estimated using the same way. As advocated by Ang, Chen, and Xing (2006); Lewellen and Nagel (2006), such estimation procedure provides greater statistical power with possible time-varying risk measures.

Except for estimating the risk measures, all the empirical asset pricing analyses are done

[4] As a robustness check, I also repeated the same empirical analyses using stocks with prices in-between \$1 and \$1,000, all the major results remain qualitatively the same.

using CRSP monthly frequency data. After applying the data filters, the number of firms in each month over the sample period ranges from 955 to 4364. In the empirical results to follow, all the dependence measures (LQP, UQP and DownAsy) are evaluated at the sample mean ($c = 0$), except for some cases that are specifically denoted.

Table 2.1 reports time series averages across months of the cross-sectional correlations of main variables, lower quadrant probability (LQP), upper quadrant probability (UQP), downside asymmetric dependence (DownAsy), CAPM beta (β), downside beta (β^-), upside beta (β^+), log of market capitalization (Size), log of book-to-market ratio (Bm), turnover ratio (Turn), normalized Amihud illiquidity measure (Illiq), past six-month return (Mom), idiosyncratic volatility (Ivol), coskewness (Coskew), cokurtosis (Cokurt) and the maximum daily return over the past one month (max), used in this study. At the beginning of each month t, all risk characteristics (LQP, UQP, DownAsy, β, β^- , β^+, Ivol, Coskew, Cokurt) are calculated using daily realized stock and market excess returns over the next 12-month period. Size, Bm, Turn, Illiq, Mom and Max are calculated using information available at the end of month $t - 1$. All the variables are updated monthly. A detailed description of these variables is given in the appendix.

[Insert Table 2.1 about here]

From Table 2.1, we can tell how the new dependence measures are linearly related to traditional variables that have explanatory powers on cross-sectional stock returns. The average correlation between LQP and UQP is relatively modest, at -0.08, which means that the tendency of stock to go up or down with the market may appear independently. This finding also justifies the approach to separately estimate lower and upper quadrant probabilities, which allows for asymmetric dependence in the lower and upper quadrants. As expected, DownAsy has strong positive correlation (0.462) with LQP and negative correlation (-0.515) with UQP, because mechanically the sign of DownAsy coincides with the sign of (LQP$-$UQP). If a variable has similar correlations with LQP and UQP, DownAsy will show little correlation with that variable. Hence we see that DownAsy has almost no correlation with all the other variables. It is more interesting to focus on the correlations with LQP and UQP.

Both LQP and UQP are positively correlated with the CAPM β with correlation coefficients of 0.463 and 0.436 respectively. It is as expected because β captures the linear dependence between individual stock return and market return, while quadrant probabilities measure the general dependence that also captures linear dependence as one component. Stocks with higher β will have high probabilities to be above (below) its sample mean when the market is above (below) average.

On the other hand, β^- and β^+ both have very high positive correlations (around 0.8) with β due to construction. β^- and β^+ are also highly positively correlated with correlation coefficient equal to 0.528. This finding indicates that β^- (β^+) may not be clean measures of downside risk (upside potential). A higher β^- or β^+ is most likely associated with a higher CAPM β.

Interestingly, size is positively correlated with both LQP and UQP with fairly large correlation coefficients of 0.299 and 0.473 respectively. It indicates that excess returns of larger stocks are more likely to be above (below) the sample mean when market is above (below) the average level. Note that UQP increases more strongly with size than LQP, which indicates that larger stocks have less degree of downside asymmetry than small stocks. It is also confirmed by the negative correlation between size and DownAsy. The finding is consistent with Ang and Chen (2002); Hong, Tu, and Zhou (2007), who find that small size portfolios show stronger asymmetric co-movements with the market using formal statistical tests.

LQP and UQP has little correlation with coskewness, but they have high positive correlations with the fourth co-moment, cokurtosis. The findings with coskewness seems odd, but upon scrutiny, it is not surprising. Just like skewness for univariate distribution, coskewness is more related with length of the tails in a joint distribution. LQP and UQP are measured at the sample mean, where the probability mass is more concentrated. Compared to the probability mass at the center, the probability difference at the tails are much less important. In unreported results, I find that LQP and UQP measured at 0.5 and 1 standard deviations away from the sample mean have much higher correlations with coskewness.[5] Cokurtosis measures the fatness of the tails in a given joint distribution. A fatter tail indicates higher probability in that quadrant. It is natural to see that both LQP and UQP are positively correlated with cokurtosis.

2.3.2 Portfolio Sorting

In this subsection, I study the impact of those dependence measures on the cross-section of average stock returns using simple univariate portfolio sorting.

Univariate Portfolio Sorts

At the beginning of each 12-month period at time t, I sort stocks into five quintile portfolios based on their realized LQP, UQP and DownAsy over the next 12 months. The portfolio returns are also computed as the average realized excess returns over the same 12-month period.

Table 2.2 shows the contemporaneous relationship between excess returns and LQP (Panel A), UQP (Panel B), and DownAsy (Panel C). Both equal-weighted and value-weighted excess returns and Carhart (1997) four factor adjusted alphas are reported. The row labeled "High - Low" gives the difference between the returns of portfolio 5 and portfolio 1, with corresponding statistical significance levels. Although I use a 12-month horizon, the quintile portfolios are updated at a monthly frequency. Using overlapping information to compute the returns/alphas is more efficient but the 12-month returns/alphas are autocorrelated by construction. To account for the autocorrelations, I report t-statistics of returns/alphas differences computed using Newey and West (1987) heteroskedasticity and autocorrelation consistent (HAC) standardized errors with 12 lags.[6] The sample covers all U.S. common stocks traded on the NYSE/AMEX/NASDAQ, and the sample period

[5] The results are available upon request.

[6] Although the theoretical number of lags is 11, to follow the practice of Ang, Chen, and Xing (2006), I use 12 lags. Adding one more lag is more conservative and leads to smaller t-statistics than using 11 lags.

is from January 1962 to December 2013, with the last 12-month period starting in January 2013. For robustness checks, I also conduct cross-sectional regression analysis using nonoverlapping yearly periods in the later subsection.

[Insert Table 2.2 about here]

Panel A of Table 2.2 shows an increasing pattern between realized LQP and average annualized returns and Carhart alphas. For value-weighted returns/alphas, the increasing pattern is monotonic. In the second column, Quintile 1 (5) shows an average equal-weighted excess return of -5.89% (16.15%) per annum, and the spread in average excess returns is a 22.03% per annum, with a corresponding Newey-West t-statistic of 10.31. In the fourth column, Quintile 1 (5) shows an average value-weighted excess return of -9.16% (8.89%) per annum, and the value-weighted spread is a 18.05% per annum, statistically significant at the 1% level. Return spread is smaller when more weights are given to larger stocks, but the reduction is not much, at about 4% per annum. The results only consider the effect of one variable, while it is shown that LQP is correlated with other variables that also affect returns, such as CAPM β, size and book-to-market ratio. To account for the effects of market, size (SMB), book-to-market (HML), and momentum (UMD) factors, I calculate the equal-weighted and value-weighted alphas from Carhart (1997) four factor model for the quintile and spread portfolios. The results are listed in the third and fifth columns respectively, with equal-weighted (value-weighted) four factor alpha for the spread portfolio is equal to 19.89% (16.45%). Both alphas are economically large and statistically significant at the 1% level.

Panel B of Table 2.2 shows a decreasing pattern between realized UQP and average annualized returns. For Carhart alphas and value-weighted returns, such decreasing pattern is monotonic. Quintile 1 (5) portfolios have an average equal-weighted excess return of 17.33% (-1.79%) per annum, and the spread in average excess returns is a -19.12% per annum. The return difference is statistically significant at the 1% level and cannot be explained by Carhart (1997) four factor model. The four factor equal-weighted alpha spread is -19.00% per annum. The return to the spread portfolio is much lower, at -8.22% per annum, when the returns are weighted by firm's market capitalization. The four factor value-weighted alpha spread is -9.87% per annum. Both value-weighted returns and alphas are still highly statistically significant, at the 1% level. However, the economic significance is much reduced compared to the equal-weighted case.

As pointed out by Ang, Chen, and Xing (2006), asymmetric market risk exposure is bigger among smaller stocks, so we choose to focus on equal-weighted results. Previous studies on testing for asymmetric correlations (Ang and Chen, 2002; Hong, Tu, and Zhou, 2007) and my work on testing for asymmetric dependence in Chapter 1 all find that smaller stocks tend to co-vary with the market more strongly during the downside market than during the upside market. Such downside asymmetry is statistically significant according to their testing results, but these studies do not find any statistically significant asymmetry in large size portfolios.

The findings indicate that the joint return distribution is more symmetric for larger stocks, which further indicates that the correlations between LQP and UQP should be positive among large stocks.Indeed, among top 10 percentile biggest firms, I find that the time series average correlation between LQP and UQP is 0.31, much larger than its full sample counterpart, -0.08. Such large correlation leads to a reduction in the value-weighted return spread, due to the opposite effects of LQP and UQP on contemporaneous returns. For example, a very big firm with high UQP is sorted into quintile 5 portfolios, but it may also have high LQP. High UQP leads to a lower excess return, but high LQP also leads to a higher excess return, so the combined effect makes the stock to earn higher return than the other stocks in the quintile with similar UQP. Value-weighted results put very large weights on such biggest companies. Thus the average value-weighted return for quintile 5 portfolio is higher than the average equal-weighted return. Similarly, very big stock may have both low LQP and UQP, the combined effect makes the average value-weighted returns for lower quintile portfolios to be smaller than the average equal-weighted returns. This is exactly the case as shown in Panel B. A lower value-weighted return spread is also observed in

Panel A, but empirically the reduction is not as large as in Panel B.

The patterns shown in Panel A and B are similar to the sorting patterns of realized relative β^- (defined as $\beta^- - \beta$) and relative β^+ (defined as $\beta^+ - \beta$) as shown in Ang, Chen, and Xing (2006). The patterns also coincide with the findings documented in Chabi-Yo, Ruenzi, and Weigert (2014), who find a risk premium (discount) for stocks with high extreme lower (upper) tail dependence with the market. Ang, Chen, and Xing (2006) demonstrate that such patterns can be implied by a simplistic representative agent model with DA utility. The DA preferences allow agents to put greater weights on losses than gains. In equilibrium, a representative agent requires a premium to hold stocks with high downside risk, but is willing to hold stocks with high upside potential at a discount, holding other things equal. Empirically, I find that there is a positive premium for high LQP stocks and a negative premium (discount) for stocks with high UQP. The effect is stronger for LQP than for UQP (18.05% v.s. -8.22% for the value-weighted returns).

Given the opposite effects of LQP and UQP on returns and the fact that LQP and UQP are only modestly correlated, we expect the downside asymmetric dependence measure (DownAsy) to be positively associated with returns, since higher DownAsy indicates stronger dependence with the downside market while limited upside potential. Agents dislike this kind of stocks and should require a risk premium to hold them. The risk premium is expected to be larger than that of LQP, because it also combines the effect of UQP. Panel C of Table 2.2 shows average returns for portfolios sorted by DownAsy. We can see a monotonically increasing pattern between realized DownAsy and average annualized returns as well as Carhart alphas. In the second column, Quintile 1 (5) shows an average equal-weighted excess return of -6.96% (21.21%) per annum, and the spread in average excess returns is a 28.17% per annum, which is statistically significant at the 1% level. The equal-weighted four factor alpha for the spread portfolio is 25.58% per annum. Both excess return and alpha are higher compared to those for the spread portfolio sorted by either LQP or UQP. The fourth and fifth columns show value-weighted results. The average value-weighted excess return (alpha) of the spread portfolio is 12.34% (12.89%) per annum, higher than the

UQP return spread, but lower than the LQP spread. The reason is the same as for the UQP sorted portfolios. While the direction and statistical significance of the relationship between the dependence measures and returns hold for both an average stock (equal weighting) or an average dollar (value weighting), the magnitude is smaller with value-weighting.

Since betas are wildly used in the literature as linear dependence measures with the market, for comparison purposes, I also sort stocks into quintile portfolios based on their contemporaneous realized β^- (Panel A), β^+ (Panel B) and $\beta^- - \beta^+$ (Panel C) over 12-month periods. The method and sample used is the same as in Table 2.2. With a longer sample period and with all stocks listed on NYSE/AMEX/NASDAQ, I have got similar findings as Ang, Chen, and Xing (2006). The results are reported in Table 2.3.

[Insert Table 2.3 about here]

Panel A of Table 2.3 shows a monotonically increasing pattern between realized β^- and average annualized returns/alphas. The average equal-weighted excess return of the spread portfolio in average is a 12.23% per annum, which is statistically significant at the 1% level. However, after accounting for Carhart (1997) four factors, the alpha spread is only 5.49% per annum. Although the downside beta does not exactly reflect the exposure to the market factor, the market, size, boot-to-market and momentum factors can still explain more than half of the excess return difference. Compared to the LQP sorted portfolios, the magnitude of downside beta premium is only about half, and the difference in Alpha is even more prominent (5.49% vs. 19.89%). It is clear evidence that the non-linear downside dependence measure, LQP, can better capture the downside risk than the downside beta. While LQP is also estimated using information from the joint distribution of individual stock and market returns, there is no linear structure involved. It may explain why the market factor, combined with the other three factors, fail to explain much of the excess return to the LQP sorted spread portfolio.

Panel B of Table 2.3 shows an increasing pattern between realized β^+ and average annualized returns/alphas. As noted in Ang, Chen, and Xing (2006), the pattern is inconsistent with their model predictions and the reason is due to high correlation between β^+ and CAPM β. They do find that when partially out the β effect, returns to portfolios sorted by realized relative β^+ ($\beta^+ - \beta$) show a decreasing pattern. It indicates that β^+ may not be a clean measure of upside potential. In comparison, the excess return (four factor alpha) to the spread portfolio sorted by UQP is negative and significant both economically and statistically.

Panel C of Table 2.3 shows an increasing pattern in average annualized returns/alphas with increasing realized ($\beta^- - \beta^+$). This measure gauges the effect of downside linear dependence relative to upside linear dependence and can be considered as a linear downside asymmetric dependence measure.[7] Compared to the non-linear downside asymmetric dependence measure, the spread in value-weighted returns is much smaller (6.33% v.s. 12.34% per annum), and a large portion can be explained by Carhart (1997) four factors. The findings suggest that the entropy-based downside asymmetric dependence measure better captures the asymmetry in market risk exposure.

Dependent Portfolio Sorts

The univariate return patterns could be driven by differences in other risk measures or firm characteristics known to affect contemporaneous returns. As shown in Table 2.1, LQP and UQP are correlated with some other variables, such as CAPM β, size and cokurtosis. To see a clearer picture of the composition of the other variables across the LQP and UQP sorted portfolios, Table 2.4 presents summary statistics of the related variables for the stocks sorted into decile portfolios by LQP (Panel A) and UQP (Panel B). Specifically, at the beginning of each month t, I rank all stocks into decile portfolios based on realized LQP and UQP measures over the next 12 months. The table reports for each decile the time-series average across months of the cross-sectional mean values within each month of the same set of variables as appeared in Table 2.1.

[7] Ang, Chen, and Xing (2006) report the portfolio sorting results based on ($\beta^+ - \beta^-$). They find a decreasing pattern with a -7.81% equal-weighted excess return for the spread portfolio.

[Insert Table 2.4 about here]

From Panel A of Table 2.4, we can see that there is enough dispersion in LQP across the deciles, with the smallest being 0.224 and largest being 0.370. As we move from the low LQP to the high LQP decile, all three betas increase monotonically. The pattern may raise some concern that the positive risk premium in Table 2.2 may be driven by higher linear dependence with the market or higher downside beta. We rule out this possibility using dependent portfolio sorts that control for the variations in β in the analysis to follow.

As LQP increases across deciles, firm market capitalization (size) increases and illiquidity (Illiq) decreases, indicating that high LQP stocks tend to be larger and more liquid. This is good news for the univariate results reported in Panel A of Table 2.2, since previous studies have documented that larger (Banz, 1981) and more liquid (Amihud, 2002) stocks tend to earn a return discount, not the return premium observed in the data. The fact that high LQP portfolios contain larger and more liquid stocks but still earn higher average returns works to strengthen the effect of LQP. It is also observed that cokurtosis (cokurt) is increasing with LQP. Dittmar (2002) document that stocks with higher cokurtosis earn higher average returns. Therefore, the premium of LQP may be explained by the difference in cokurtosis across the deciles. It motivates me to do dependent portfolio sorts with LQP and cokurtosis. The book-to-market ratio (Bm) does not show a clear pattern, although the high LQP portfolios seem to have more growth stocks. There is no clear pattern found for other control variables, such as past six-month return (Mom), idiosyncratic volatility (Ivol), coskewness (Coskew), and maximum daily return in the previous month (Max).

Panel B of Table 2.4 shows some interesting patterns for decile portfolios sorted by UQP. As we move from the low UQP to the high UQP decile portfolio, the average across months of the mean UQP of stocks increases from 0.201 in the Decile 1 to 0.361 in Decile 10. Similar as the LQP case, all three betas increase monotonically with UQP. This finding works to strengthen the effect of UQP, because stocks with higher market β or β^{-} tend to have higher average returns instead of the lower average returns indicated by high UQP. Size increases with UQP, which works to weaken the effect of UQP as larger firms tend to earn lower average returns. Illiquidity also decreases for the high UQP deciles, consistent with the fact that these portfolios contain larger stocks. It may also be confounding the effect of UQP, as more liquid stocks tend to have lower returns. The book-to-market ratio decreases with UQP, so it may also explain part of the return discount of high UQP stocks. Past six-month return and cokurtosis show increasing pattern as UQP increases, which strengthens the effect of UQP since stocks with high momentum (Jegadeesh and Titman, 1993) and high cokurtosis (Dittmar, 2002) tend to earn higher average returns instead of the observed lower returns. Idiosyncratic volatility seems to decrease with UQP, but the variation in Ivol is not very large. Other variables, like coskewness and Max, do not show a clear pattern.

Those control variables show almost identical co-movement patterns with either LQP or UQP. Since I use excessive downside probability (EDP), defined as (LQP$-$UQP), to determine the sign of downside asymmetric dependence (DownAsy), the opposite patterns given by LQP and -UQP almost cancel out. There is no clear pattern for any of these control variables, when decile portfolios are formed based on realized DownAsy, which means that the return premium due to DownAsy should not be driven by other known characteristics that affect cross-sectional returns. Therefore, the summary statistics for decile portfolios sorted by DownAsy are not reported.[8]

Motivated by the patterns observed in Table 2.4, I conduct dependent portfolio sorts to explicitly control for the effects of the other stock characteristics that co-vary most with both LQP and UQP, i.e. the CAPM β, size, coskewness, and cokurtosis. I include coskewness in the double sorts mainly due to theoretical consideration, since coskewness is a moment-based measure of asymmetry. Although the linear beta exposure to market and the size effect can be controlled by looking at the Carhart alphas in the univariate portfolio sorts, dependent portfolio sorts can account for some potential nonlinear impact of these control variables.

At each month, I first form quintile portfolios sorted on each of β, size, coskewness, and

[8] The results are available upon request.

cokurtosis, then within each quintile, I further sort stocks into five portfolios based on their realized lower quadrant probability (LQP). The results are reported in Table 2.5. The row labeled "High - Low" reports the difference between the excess returns of portfolio 5 and portfolio 1 in each β, size, coskewness and cokurtosis first-sort quintile with corresponding statistical significance levels. The column labeled "Average" reports the average excess return of stocks in each second-sort quintile. Newey-West (1987) 12-lag adjusted t-statistics are reported in row labeled "t-stat".[9]

[Insert Table 2.5 about here]

Panel A of Table 2.5 reports equal-weighted portfolio excess returns of $\beta\times$ LQP portfolios. Within each β quintile, the return to the high LQP portfolio is larger than the return to the low LQP portfolio. The return spreads are both economically and statistically significant. They range from 24.59% per annum in β quintile 1 to 17.39% per annum in β quintile 5. The average difference in excess returns is 20.57% per annum, only slightly smaller than the return spread in the univariate sorting case. Therefore, market β, although correlated with LQP, can only account for a tiny part of the premium associated with high LQP.

Panel B of Table 2.5 repeats the same analysis as Panel A, with β being replaced by firm size. Within each size quintile, the equal-weighted return to the "How - Low" portfolio is highly significant both economically and statistically, ranging from 32.51% to 14.26% per annum. The return difference decreases as as we move to high size quintile. As mentioned above, this is due to higher correlation between LQP and UQP among larger stocks. It is difficult to purge the effect of UQP from LQP among large size quintile, which leads to a shrunk return spread. Due to the same argument, we also observe similar pattern, among size $\times$ UQP portfolios in Panel B of Table 2.6. Despite the decreasing pattern, the average return difference among five size quintiles is still 24.95% per annum and highly significant statistically. The magnitude is even higher than the return spread in the univariate sorting case, which indicates that size cannot explain the risk premium associated with high LQP

[9] For tables to fit in one page, I only report results for the quintile 5 and quintile 1 second-sort portfolios, and the "High - Low" portfolio. Detailed results are available upon request.

either.

Panel C of Table 2.5 reports equal-weighted portfolio excess returns of coskewness $\times$ LQP portfolios. Empirically, I find coskewness is not much correlated with either downside or upside dependence measures. The reason is that coskewness captures more of the asymmetry in the length of the tails, while LQP and UQP are measured at the sample mean and are less sensitive to the probability difference in the tails.[10] Given the low correlation, we do not expect coskewness to account for the risk premium due to LQP. Within each coskewness quintile, the return of the spread portfolio is large and statistically significant at 1% level, with an average spread of 22.09% per annum. Meanwhile, we can confirm that coskewness has negative impact on returns as documented in Harvey and Siddique (2000).

Panel D of Table 2.5 reports equal-weighted cokurtosis $\times$ LQP double-sorted portfolio excess returns. Within each cokurtosis quintile, the return of the spread portfolio is large and statistically significant at the 1% level, ranging from 29.57% to 10.77% per annum. Although we observe a decreasing return spread as cokurtosis increases, the average spread is 21.89% and is highly significant at the 1% level, indicating that cokurtosis cannot account for the return premium associated with LQP.

Table 2.6 repeats the same exercises as Table 2.5, only replacing LQP by UQP. In general, none of β, size, coskewness or cokurtosis can account for the discount for holding stocks with high UQP. Although if anything, firm market capitalization seems to reduce the negative excess return earned by the UQP spread portfolio. Panel B of Table 2.6 shows that the return spread within size quintile 1 is -19.26%. While within the highest size quintile, the spread is narrowed to -5.48% due to the reason I mentioned above. The average return spread across all size quintiles is -14.89%. The magnitude shrinks compared to -19.12%, the return spread in the univariate sorting case. It is consistent with the finding that the value-weighted UQP return spread is much smaller than the equal-weighted UQP return spread in terms of absolute value. Therefore, size can explain a small part of the return discount due to high UQP, but the unexplained part remains to be quite large. β, coskewness and

[10] When measured at 0.5 and 1 standard deviations away from the sample mean, LQP anf UQP do show much higher correlations with coskewness.

cokurtosis cannot account for any proportion of the return discount due to high UQP.

[Insert Table 2.6 about here]

Finally, in Table 2.7, I report the double-sorting results using the downside asymmetric dependence measure (DownAsy) with β (Panel A), size (Panel B), coskewness (Panel C), and cokurtosis (Panel D). By construction, DownAsy reflects the combined pattern of LQP and UQP. For example, the magnitude of the return spread decreases as we move from small to large size quintile for both LQP and UQP second-sort portfolios. We may expect similar pattern in the size $\times$ DownAsy portfolios. Panel B of Table 2.7 shows that the return spread within size quintile 1 is 30.86% and monotonically decreases to 12.08% in size quintile 5. It is consistent with the findings in the existing literature that small stocks are more exposed to asymmetric downside market risk. The average return difference 23.92%, although is slightly smaller than the return difference in univariate sorts, is still highly significant both economically and statistically. Patterns of return spreads for double-sorted portfolios with β, coskewness and cokurtosis are much similar to the LQP case, as shown in Table 2.5. None of the control variables can largely explain the risk premium earned for holding stocks with high downside asymmetric dependence.

[Insert Table 2.7 about here]

In summary, the results of dependent portfolio sorts provide strong evidence that the risks associated with LQP, UQP and DownAsy are weakly related to size, but clearly are different from risks associated with CAPM β, size, coskewness and cokurtosis. Dependent sorts allow us to control for potential nonlinear impact, but only one other stock characteristic can be controlled for at one time. In the following subsection, I conduct a series of Fama and MacBeth (1973) cross-sectional regressions at the firm level, which allows us to examine the impact of the dependence measures while controlling for many other firm characteristics at the same time.

2.3.3 Fama-Macbeth Regressions

Following several prior studies (see, e.g., Brennan, Chordia, and Subrahmanyam, 1998; Ang, Chen, and Xing, 2006; Ang, Liu, and Schwarz, 2010; Chabi-Yo, Ruenzi, and Weigert, 2014) that test asset pricing models with individual stock data, I run Fama-MacBeth (1973) regressions at the individual stock level over the sample period from January 1962 to December 2013.[11] I regress stock excess returns on realized dependence measures with respect to the market risk (LQP, UQP, and DownAsy), realized betas (β, β^-, and β^+) and other firm characteristics using 12-month rolling periods. Since the regressions are run at monthly frequency with a 12-month horizon, I report t-statistics of the estimated coefficients computed using 12 Newey-West (1987) lags. For each month, the risk characteristics (LQP, UQP, DownAsy, β^-, β^+, Ivol, Coskew, Cokurt) are calculated contemporaneously over the same 12-month period as the excess returns. Log firm size, log book-to-market ratio, turnover ratio, normalized Amihud (2002)illiquidity ratio, past six-month return, and maximum are calculated at the beginning of each month t. All the independent variables are winsorized at the 0.5% and 99.5% levels to avoid some extreme observations driving the results. All the main findings hold no matter I choose to do winsorization or not. Table 2.8 report the regression results with various sets of control variables. For easier interpretation, the second to last column shows time series averages of cross-sectional mean and standard deviation of each independent variable. To test whether the stock characteristics are still significant after taking the effects of commonly used factors into account, I use 12-month Carhart (1997) four factor adjusted excess return as the dependent variable in regressions (8) and (9). The risk-adjusted returns are used by Brennan, Chordia, and Subrahmanyam (1998) to test factor based asset pricing models. This method avoids the errors-in-variables bias in estimating the risk premia of stock characteristics by putting the factor loadings on the

[11] Estimates of risk loadings, such as the realized betas, from individual stock data are less precise than using portfolios as the test assets, which leads to well-known errors-in-variables (EIV) problem. However, Ang, Liu, and Schwarz (2010) argue that with individual stock data, the estimated factor loadings have greater dispersion that reduces the variance of the risk premium estimator and hence is statistically more efficient. Furthermore, Lo and MacKinlay (1990); Lewellen, Nagel, and Shanken (2010) also argue that the method used to form portfolios can lead to very distinct results in asset pricing tests, while using individual stocks as test assets can avoid this arbitrary element in portfolio grouping choice.

left hand side as the dependent variable. The last column reports the change in 12-month Carhart (1997) four factor adjusted excess return given a one standard deviation increase in the respective independent variable based on regressions (8) and (9).

[Insert Table 2.8 about here]

Regression (1) and (2) only include LQP and UQP respectively as the explanatory variable to see the univariate effect. Both variables are highly significant economically and statistically with opposite impacts on returns. A one standard deviation increase in LQP is associated with $2.049 \times 0.042 = 8.6\%$ higher average excess returns per annum. UQP has shown a significantly negative impact and a one standard deviation increase in UQP leads to $1.471 \times 0.047 = 6.9\%$ lower average excess returns per annum.

In regression (3), I include both LQP and UQP as the independent variables to see the joint effects. The positive (negative) coefficient of LQP (UQP) remains unchanged with even higher economic magnitude. Still we can see that UQP has a smaller impact on returns than LQP, indicating that investors show stronger aversion to the downside risk than preference for upside potential. Regression (4) includes only DownAsy as the explanatory variable. Consistent with the findings in univariate portfolio sorts, the impact of downside asymmetric dependence is negative and statistically significant at the 1% level. A one standard deviation increase in DownAsy is associated with $2.443 \times 0.046 = 11.2\%$ higher average excess returns per annum. The impact is economically more significant than the univariate effect of LQP.

In regression (5), I check the effects of β^- and β^+, the linear counterparts of LQP and UQP. With the sample used in this paper, I can confirm the results from Ang, Chen, and Xing (2006) that downside beta earns a risk premium (5.0% per annum), and upside beta earns a discount (-1.3% per annum), both impacts are statistically significant at the 1% level. The economic magnitude of downside beta premium is much higher compared to the upside beta discount. Regression (6) adds a full set of control variables along with LQP and UQP. The results show that the effects of LQP and UQP are still highly significant with similar magnitudes compared to regression (3). The estimated return premium for bearing one standard deviation downside dependence risk is 11.6% per annum. The impact of upside dependence (UQP) is slightly lower, but still earns -6.3% discount with a one standard deviation increase. In comparison, the effect of upside beta is no longer significant, consistent with the findings in Ang, Chen, and Xing (2006). It indicates that the upside beta may not be a good measure of upside risk, as the results are inconsistent with theoretical model prediction.

The findings in regression (6) confirm many patterns that have been documented in the literature. For example, small size stocks and stocks with high book-to-market ratios have high average returns (Fama and French, 1993). Stocks with high past six-month returns earn high average returns during the next 12 months (Jegadeesh and Titman, 1993). Anomaly documented by Ang et al. (2006, 2009) is confirmed that high realized idiosyncratic volatility is associated with low average returns. Less liquid stocks tend to earn lower average returns (Amihud, 2002). Stocks with high coskewness earn low average returns (Harvey and Siddique, 2000) and stocks with positive cokurtosis have high returns (Dittmar, 2002).

In regression (7), I replace the upside and downside dependence measures (LQP and UQP) by the downside asymmetric dependence measure (DownAsy) and include the same set of controls as in regression (6). We can see that the effect of DownAsy is highly significant and the economic magnitude only slightly reduced compared to the univariate regression (4). Interestingly, I find that the effects of Ivol and coskewness are no longer significant statistically after including the entropy-based downside asymmetric dependence measure. Even in regression (6), the effects are not economically significant. The finding suggests that the anomalies due to volatility and coskewness may be explained by the nonlinear dependence with the market risk. However, cokurtosis is still highly significant in both regression (6) and (7).

Regression (8) and (9) use the same controls as in (6) and (7), but replace the dependent variable as Carhart (1997) four factor adjusted return to see whether the characteristics still have explanatory power after accounting for the effects of the four factors. It is clear that the effects of the nonlinear dependence measures are robust even using the risk-adjusted return as the dependent variable. The economic significance of each independent variable is reported in the last column mainly based on regression (9), except for LQP and UQP that are based on regression (8). Among all the explanatory variables, the downside dependence with the market (LQP) has the strongest impact, 9.08% higher adjusted return per annum given a one standard deviation increase. The downside beta, although still negative and significant statistically, has much lower impact (3.68% per annum) on adjusted return, which suggests that LQP is a more accurate measure of downside risk. Downside asymmetric dependence (DownAsy) is also positive and highly significant with 8.59% impact per annum. The upside dependence (UQP) has a significant negative impact of -5.14% per annum on the risk-adjusted return. The magnitude of the discount is much smaller than the risk premium associated with the downside dependence risk or the downside asymmetry risk. The evidence suggests that investors dislike stocks exhibiting strong dependence with the downside market, while prefer stocks with strong upside potential. The aversion to downside risk is stronger compared to the attraction to upside potential.

2.3.4 Robustness Checks

In this subsection, I run a series of Fama-Macbeth (1973) regressions using different weighting schemes, samples and measures of asymmetric dependence at other exceedance levels to check the robustness of the findings in Table 2.8. I use Carhart (1997) four factor adjusted excess return as the dependent variable with the full set of controls in these regressions. The results are reported in Table 2.9.

[Insert Table 2.9 about here]

Regression (1) and (2) report the value-weighted regression results with full set of controls. The weighting variable is firm's market capitalization at the beginning of each month. The regression coefficients now reflect the impacts for each dollar invested. Similar as the findings in the univariate portfolio sorts, the signs and statistical significance of LQP, UQP and DownAsy remain intact, but the economic magnitude of the impacts are reduced for LQP and DownAsy.

Regression (3) and (4) report the regression results when the sample is restricted to NYSE stocks only. Since stocks listed on NYSE tend to be larger size stocks, the findings are similar to value-weighted results.

Regression (5) and (6) report the regression results using non-overlapping yearly observations. Using non-overlapping periods are less efficient statistically, but do not cause the returns to be autocorrelated, so the standard t-statistics are reported. The findings are almost the same as the results using overlapping periods, with only small changes to some coefficients.

Regressions (7) to (10) test whether the impacts of those nonlinear dependence measures still hold when they are evaluated at other exceedance levels, such as 0.5 and 1 standard deviations away from the mean. The measures evaluated at farther tails capture the tendency of a stock to move drastically with large market movements and hence are proxies for joint tail risks. The findings are largely consistent with the previous findings when the measures are evaluated at the sample mean. The only exception is that the effect of UQP^1 is no longer statistically significant as shown in regression (9), indicating that investors's attraction to stocks with high upper tail dependence with the market is not robust. On the other hand, the aversion to downside risk is significant and robust at any exceedance level. Similar findings are also documented by Chabi-Yo, Ruenzi, and Weigert (2014).

2.4 Past Downside Asymmetric Dependence and Future Returns

The empirical results in Section 2.3 demonstrate significant positive relationship between high downside (asymmetric) dependence with the market and the average stock returns over the same period. If the dependence characteristics are stable or predictable over time, then investors can exploit this cross-sectional return relationship and form investable trading strategies based on stocks' asymmetric exposure to the downside risk and upside risk. Since portfolios formed based on contemporaneous DownAsy gives the highest return spread as shown in Table 2.2, in this section, I examine the time-series persistence of downside asymmetric dependence and check if we can predict such asymmetric downside risk exposure in a future period using prior information.

2.4.1 Determinants of Downside Asymmetric Dependence

I explore the determinants of DownAsy using cross-sectional Fama-MacBeth (1973) regressions. Specifically, at each month, I regress realized downside asymmetric dependence (DownAsy) over the next 12-month period on a set of past risk measures and firm characteristics variables including the lagged DownAsy estimated over the previous 12-month period. At the beginning of each month t, the past risk measures (β^-, β^+, Ivol, Coskew, Cokurt) are estimated over the previous 12-month period ($t-12$ to $t-1$). Size, Bm, Turn, Illiq, Mom and Max are calculated at the beginning of the month t. The regression results are in Table 2.10.

[Insert Table 2.10 about here]

We can see that DownAsy is not persistent over time and hard to predict even with all these past variables. When using lagged DownAsy as the only predictor variable, the coefficient 0.058, although highly significant ($t = 6.6$), is far from 1. The corresponding R^2 is lower than 0.01, meaning most variations in current DownAsy cannot be explained by past DownAsy. Size effect is still clear in the predictive setting. Large market capitalization predicts low future downside asymmetric risk. The relationship between current book-to-market ratio and future DownAsy is positive and significant. High current cokurtosis predicts low DownAsy in the future. But the R^2 is only 0.056 even if we put all these past variables as predictors.

An alternative approach is to examine the average 12-month decile portfolio transition matrix, i.e. the average probability $p_{i,j}$ that a stock in decile i during the previous 12-

month period will be in decile j in the next 12-month. In the unreported results, I find that stocks in decile 10 according to lagged 12-month DownAsy have 17.64% chance to be in the same decile over the next 12-month period and the chance of staying in the top 3 deciles is 40.69%. It indicates that stocks with high downside asymmetric risk exposure tend to have slightly higher chance to retain that characteristics over the next 12 months compared to the case when DownAsy is totally random.

The findings indicate that there is limited predictability in a stock's asymmetric exposure to the downside risk and support time-varying risk exposures as suggested in Lewellen and Nagel (2006).

2.4.2 Trading Strategy

Although the predictability is limited, yet it is still interesting to examine whether it is possible to generate abnormal return spreads based on past realized DownAsy. At the beginning of each month, I sort stocks into quintile (1-5) portfolios based on their realized DownAsy over the previous 12 months. Then, I examine equal-weighted average returns of these portfolios over the next 12-month period (Panel A) and over the next one month period (Panel B). The data used in this paper range from January 1962 to December 2013. As I use first 12-month data to estimate the first lagged DownAsy, the first portfolios are formed in January 1963. Then I update those portfolios in a monthly frequency. The results are reported in Table 2.11 below.

[Insert Table 2.11 about here]

Panel A of Table 2.11 shows 12-month holding period returns for portfolios sorted based on lagged DownAsy. Newey-West (1987) standard errors with 12 lags are used to compute the t-statistics (in parentheses) to account for autocorrelations in the 12-month cumulative returns. In the second column, quintile 1 (5) shows an average equal-weighted excess return of 10.63% (13.86%). The spread in average excess returns is a 3.22% per annum, which is statistically significant at the 1% level. To purge any effect due to exposures to systematic risk factors, I regress the returns of each quintile portfolio and the spread portfolio on the market factor, Fama and French (1993) three factors, and Carhart (1997) four factors respectively. The alphas are reported in the third to fifth columns. CAPM alpha spread is at 2.97% per annum, showing that a small part of the premium can be explained by the market factor. After controlling for the size factor (SMB) and the book-to-market factor (HML), alpha increases to 3.47% per annum. Adding the momentum factor (UMD) reduces the alpha spread to only 1.23% per annum that is marginally significant at the 10% level. It indicates that the part of the return based on the trading strategy is due to exposure to the momentum factor.

We find that DownAsy is not persistent over time. During shorter holding period, DownAsy may change less than during the longer period, so in Panel B, I show 1-month holding period returns for portfolios sorted based on lagged DownAsy. Non-overlapping 1-month returns are usually considered to have no autocorrelations, so the standard t-statistics are reported in parentheses. As expected, the trading strategy of investing in high DownAsy stocks and shorting low DownAsy stocks yields an economically significant one month return of 0.37% per month, which amounts to a compounded return premium of 4.53% per annum. The return difference is also statistically significant at the 1% level. The CAPM alpha spread is at 0.34% per month (4.16% per annum). Adding Fama-French factors increase the alpha spread to 0.38% per month (4.66% per annum). When taking the momentum factor into account, the alpha spread decreases to 0.19% per month (2.30% per annum). Exposure to the momentum factor can explain part of the return spread, but still the four factor alpha spread is statistically significant and economically meaningful.

In summary, DownAsy has limited predictability based on past information. It is difficult to exploit the strong contemporaneous relation between downside asymmetry. Although the return to the spread portfolio formed on lagged DownAsy is smaller than the contemporaneous return spread, it is still economically and statistically significant. In comparison, Ang, Chen, and Xing (2006) find that a trading strategy based on past downside beta using all stocks does not yield an economically significant return spread. It also suggests that the nonlinear asymmetric dependence measures can better capture the downside risk than the downside beta does.

2.5 Conclusion

This paper examines whether a stock's nonlinear dependence with the downside and upside market have significant impact on the cross section of stock returns. Using dependence measures constructed with a metric entropy and estimated quadrant probabilities of the joint distribution of stock and the market returns, I find a risk premium (discount) for stocks that are more likely to covary with the market during market declines (rises). The asymmetry between the downside and upside dependence with the market is earns a risk premium as well. The risk premia associated with the downside dependence and downside asymmetric dependence are higher compared to the discount due to upside dependence. The findings suggest that investors' aversion to downside losses are stronger than their attraction to the upside gains.

Fama-Macbeth (1973) regressions show that the contemporaneous impacts of the dependence measures on cross-sectional returns cannot be explained by a set of well-known stock characteristics, such as the market beta (linear exposure to the market risk), downside or upside betas (asymmetric exposure to downside and upside market risk), size and book-to-market effects, illiquidity risk, momentum effect, coskewness, cokurtosis, and stock's lottery feature as captured by the maximum daily return within a month. The estimated cross-sectional excess return premium for bearing downside dependence risk is approximately 11.6% per annum, almost twice as large as the effect of the downside beta (6% as reported in Ang, Chen, and Xing (2006)). The downside premium, downside asymmetry premium and upside discount are robust across a battery of robustness checks. In addition, I also find that the downside dependence and downside asymmetric dependence measures have low cross-sectional correlations with coskewness, since coskewness captures more of the asymmetry in the lengths of the tails. In Fama-Macbeth (1973) regressions, adding the downside and upside dependence measures (or the downside asymmetric dependence measure) along with the downside and upside betas can empirically rule out the effect of coskewness. The finding suggests that exploring the nonlinear dependence with the market factor may help explain some CAPM anomalies.

The downside asymmetric dependence is not persistent over time and shows limited predictability. However, a trading strategy that forms portfolios based on past asymmetric dependence can still earn an average equal-weighted annualized return of 4.5%. Such a premium is both economically and statistically significant. However, a similar trading strategy based on downside beta fails to yield a economically meaningful return spread. All the findings suggest that there are economic gains when going beyond traditional linear dependence measures. Nonlinear dependence measures may better capture the market risk than their linear counterparts. As part of future research, it will be of interest to develop new models to explain observed risk premium in assets that have asymmetric comovement with the market.

2.A Appendix: Variable Definitions

Let us denote a stock i's demeaned daily excess return as $\tilde{r}_{i,d}$, and demeaned daily market excess return as $\tilde{r}_{m,d}$.

CAPM BETA: β is estimated at each month t over the next 12-month, using the following formula

$$\hat{\beta}_{i,t} = \frac{\sum_{d=1}^{d=D_t} \tilde{r}_{i,d}\tilde{r}_{m,d}}{\sum_{d=1}^{d=D_t} \tilde{r}_{m,d}^2}, \tag{2.A–1}$$

where D_t is the number of trading days in a 12-month period starting from month t.

DOWNSIDE and UPSIDE BETAS: Denote the sample average of demeaned daily market excess return during a 12-month period starting from month t as $\hat{\mu}_{m,t}$. Further denote demeaned excess return and demeaned market excess return conditional on market excess return being below (above) $\hat{\mu}_{m,t}$ as $\tilde{r}_{i,d}^-$ ($\tilde{r}_{i,d}^+$) and $\tilde{r}_{m,d}^-$ ($\tilde{r}_{m,d}^+$) respectively. Following the definitions in Ang, Chen, and Xing (2006),

$$\hat{\beta}_{i,t}^- = \frac{\sum_{r_{m,d}<\hat{\mu}_{m,t}} \tilde{r}_{i,d}^-\tilde{r}_{m,d}^-}{\sum_{r_{m,d}<\hat{\mu}_{m,t}} \tilde{r}_{m,d}^{-2}}, \text{ and } \hat{\beta}_{i,t}^+ = \frac{\sum_{r_{m,d}>\hat{\mu}_{m,t}} \tilde{r}_{i,d}^+\tilde{r}_{m,d}^+}{\sum_{r_{m,d}>\hat{\mu}_{m,t}} \tilde{r}_{m,d}^{+2}}. \tag{2.A–2}$$

COSKEWNESS: Following Harvey and Siddique (2000), coskewness of stock i over a 12-month period starting at month t is given by

$$\widehat{\text{coskew}}_{i,t} = \frac{\frac{1}{T}\sum_{d=1}^{d=D_t} \tilde{r}_{i,d}\tilde{r}_{m,d}^2}{\sqrt{\frac{1}{T}\sum_{d=1}^{d=D_t} \tilde{r}_{i,d}^2}\left(\frac{1}{T}\sum_{d=1}^{d=D_t} \tilde{r}_{m,d}^2\right)}, \tag{2.A–3}$$

where D_t is the number of trading days in a 12-month period starting from month t.

COKURTOSIS: Cokurtosis of stock i over a 12-month period starting at month t is similarly defined as

$$\widehat{\text{cokurt}}_{i,t} = \frac{\frac{1}{T}\sum_{d=1}^{d=D_t} \tilde{r}_{i,d}\tilde{r}_{m,d}^3}{\sqrt{\frac{1}{T}\sum_{d=1}^{d=D_t} \tilde{r}_{i,d}^2}\left(\frac{1}{T}\sum_{d=1}^{d=D_t} \tilde{r}_{m,d}^2\right)^{3/2}}, \tag{2.A–4}$$

where D_t is the number of trading days in a 12-month period starting from month t.

IDIOSYNCRATIC VOLATILITY: Ivol of stock i at the beginning of each month t is defined as the standard deviation of the CAPM residual series over the next 12 months.

SIZE: Following the existing literature, firm size at each month t is measured using the natural logarithm of the market value of equity at the end of month $t-1$.

BOOK-TO-MARKET: Following Fama and French (1992), a firm's book-to-market ratio in month t is calculated using the market value of equity at the end of December of the last year and the book value of common equity plus balance-sheet deferred taxes for the firm's latest fiscal year ending in the prior calendar year.

MOMENTUM: Following Jegadeesh and Titman (1993), the momentum effect of each stock in month t is measured by the cumulative return over the previous 6 months, with the previous one month skipped, i.e. the cumulative return from month $t-7$ to month $t-2$.

TURNOVER: Turnover ratio is calculated monthly as the adjusted monthly trading volume divided by shares outstanding.

ILLIQUIDITY: Following Amihud (2002), the proxy for the stock illiquidity is from normalizing $L_{i,t} = |r_{i,t}|/dv_{i,t}$. It is the ratio of absolute change of price $r_{i,t}$ to the dollar trading volume $dv_{i,t}$ for stock i at day t. The monthly illiquidity ratios are the daily average of the illiquidity ratio for each stock. To get an accurate estimate of monthly Amihud ratio, we drop the months for stocks if the number of the monthly observations is smaller than 15. Following Acharya and Pedersen (2005), we also normalize the Amihud ratio to adjust for inflation and truncated it at 30 to eliminate the effect of outliers (the stocks with transaction cost larger than 30% of the price).

$$\mathrm{ILLIQ}_{i,t} = \min\left(0.25 + 0.3L_{i,t} \times \frac{\text{capitalization of market portfolio}_{t-1}}{\text{capitalization of market portfolio}_{July1962}},\ 30\right) \quad (2.A–5)$$

MAXIMUM: Following Bali, Cakici, and Whitelaw (2011), Max of stock i at month t is defined as the maximum daily excess return within that month.

References

Acharya, V. V., and L. H. Pedersen. 2005. Asset pricing with liquidity risk. *Journal of Financial Economics* 77:375–410.

Alcock, J., and A. Hatherley. 2013. Characterising the asymmetric dependence premium. *Available at SSRN 2476803* .

Amihud, Y. 2002. Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets* 5:31–56.

Ang, A., G. Bekaert, and J. Liu. 2005. Why stocks may disappoint. *Journal of Financial Economics* 76:471–508.

Ang, A., and J. Chen. 2002. Asymmetric correlations of equity portfolios. *Journal of Financial Economics* 63:443–94.

———. 2007. Capm over the long run: 1926–2001. *Journal of Empirical Finance* 14:1–40.

Ang, A., J. Chen, and Y. Xing. 2006. Downside risk. *Review of Financial Studies* 19:1191–239.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang. 2006. The cross-section of volatility and expected returns. *The Journal of Finance* 61:259–99.

———. 2009. High idiosyncratic volatility and low returns: International and further us evidence. *Journal of Financial Economics* 91:1–23.

Ang, A., J. Liu, and K. Schwarz. 2010. Using stocks or portfolios in tests of factor models. *Working Paper, Columbia University and University of Pennsylvania* .

Backus, D., M. Chernov, and I. Martin. 2011. Disasters implied by equity index options. *The Journal of Finance* 66:1969–2012.

Bae, K.-H., G. A. Karolyi, and R. M. Stulz. 2003. A new approach to measuring financial contagion. *Review of Financial Studies* 16:717–63.

Bali, T. G., N. Cakici, and R. F. Whitelaw. 2011. Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics* 99:427–46.

Ball, R., and S. Kothari. 1989. Nonstationary expected returns: Implications for tests of market efficiency and serial correlation in returns. *Journal of Financial Economics* 25:51–74.

Banz, R. W. 1981. The relationship between return and market value of common stocks. *Journal of Financial Economics* 9:3–18.

Bekaert, G., and G. Wu. 2000. Asymmetric volatility and risk in equity markets. *Review of Financial Studies* 13:1–42.

Brennan, M. J., T. Chordia, and A. Subrahmanyam. 1998. Alternative factor specifications, security characteristics, and the cross-section of expected stock returns. *Journal of Financial Economics* 49:345–73.

Cabrales, A., O. Gossner, and R. Serrano. 2013. Entropy and the value of information for investors. *The American Economic Review* 103:360–77.

Carhart, M. M. 1997. On persistence in mutual fund performance. *The Journal of finance* 52:57–82.

Chabi-Yo, F., S. Ruenzi, and F. Weigert. 2014. Crash sensitivity and the cross-section of expected stock returns. *Available at SSRN 2011746* .

Conrad, J., M. N. Gultekin, and G. Kaul. 1991. Asymmetric predictability of conditional variances. *Review of Financial Studies* 4:597–622.

Dittmar, R. F. 2002. Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns. *The Journal of Finance* 57:369–403.

Embrechts, P., A. McNeil, and D. Straumann. 2002. Correlation and dependence in risk management: properties and pitfalls. *Risk management: value at risk and beyond* 176–223.

Fama, E. F., and K. R. French. 1992. The cross-section of expected stock returns. *The Journal of Finance* 47:427–65.

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

Fama, E. F., and J. D. MacBeth. 1973. Risk, return, and equilibrium: Empirical tests. *The Journal of Political Economy* 81:607–36.

Gao, X., and J. R. Ritter. 2010. The marketing of seasoned equity offerings. *Journal of Financial Economics* 97:33–52.

Granger, C., E. Maasoumi, and J. Racine. 2004. A dependence metric for possibly nonlinear processes. *Journal of Time Series Analysis* 25:649–69.

Gul, F. 1991. A theory of disappointment aversion. *Econometrica* 59:667–86.

Hansen, L. P. 2012. Dynamic valuation decomposition within stochastic economies. *Econometrica* 80:911–67.

Harvey, C. R., and A. Siddique. 2000. Conditional skewness in asset pricing tests. *The Journal of Finance* 55:1263–95.

Hong, Y., J. Tu, and G. Zhou. 2007. Asymmetries in stock returns: Statistical tests and economic evaluation. *Review of Financial Studies* 20:1547–81.

Jegadeesh, N., and S. Titman. 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance* 48:65–91.

Lewellen, J., and S. Nagel. 2006. The conditional capm does not explain asset-pricing anomalies. *Journal of Financial Economics* 82:289–314.

Lewellen, J., S. Nagel, and J. Shanken. 2010. A skeptical appraisal of asset pricing tests. *Journal of Financial Economics* 96:175–94.

Li, C., H. Li, and J. S. Racine. 2014. Cross-validated mixed datatype bandwidth selection for nonparametric cumulative distribution/survivor functions. *McMaster University, Working paper* .

Lintner, J. 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics* 47:13–37.

Lo, A. W., and A. C. MacKinlay. 1990. Data-snooping biases in tests of financial asset pricing models. *Review of Financial Studies* 3:431–67.

Newey, W. K., and K. D. West. 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55:703–08.

Shannon, C. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27:379–423.

Sharpe, W. F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance* 19:425–42.

Sims, C. A. 2003. Implications of rational inattention. *Journal of Monetary Economics* 50:665–90.

Table 2.1: Time series average of cross-sectional correlations

	LQP^0	UQP^0	$DownAsy^0$	β	β^-	β^+	Size	Bm	Turn	Illiq	Mom	Ivol	Coskew	Cokurt	Max
LQP^0	1														
UQP^0	-0.08	1													
$DownAsy^0$	0.462	-0.515	1												
β	0.463	0.436	-0.029	1											
β^-	0.3	0.273	-0.036	0.8	1										
β^+	0.356	0.34	-0.011	0.797	0.528	1									
Size	0.299	0.473	-0.079	0.222	0.097	0.233	1								
Bm	-0.071	-0.207	0.066	-0.259	-0.212	-0.202	-0.28	1							
Turn	0.164	0.2	-0.032	0.451	0.382	0.331	0.109	-0.201	1						
Illiq	-0.297	-0.405	0.048	-0.337	-0.233	-0.298	-0.637	0.256	-0.29	1					
Mom	0.009	0.051	-0.019	0.114	0.119	0.072	0.012	-0.237	0.153	-0.061	1				
Ivol	-0.091	-0.203	0.029	0.308	0.336	0.17	-0.525	-0.058	0.232	0.357	0.042	1			
Coskew	0.033	0.024	0.038	0.013	-0.343	0.388	0.038	-0.01	-0.025	0.01	-0.04	0.007	1		
Cokurt	0.453	0.497	-0.059	0.687	0.577	0.647	0.523	-0.203	0.239	-0.472	0.075	-0.234	0.013	1	
Max	0.01	-0.041	0.009	0.277	0.273	0.178	-0.259	-0.1	0.335	0.129	0.024	0.498	-0.001	-0.021	1

The table reports the averages across months of the cross-sectional correlations of main variables used in this study: lower quadrant probability (LQP), upper quadrant probability (UQP), downside asymmetric dependence (DownAsy), CAPM beta (β), downside beta (β^-), upside beta (β^+), log of market capitalization (Size), log of book-to-market ratio (Bm), turnover ratio (Turn), normalized Amihud illiquidity measure (Illiq), past six-month return (Mom), idiosyncratic volatility (Ivol), coskewness (Coskew), cokurtosis (Cokurt) and the maximum daily return over the past one month (max). At the beginning of each month t, all risk characteristics (LQP, UQP, DownAsy, β, β^- , β^+, Ivol, Coskew, Cokurt) are calculated using daily realized stock and market excess returns over the next 12-month period. Size, Bm, Turn, Illiq, Mom and Max are calculated using information available at the end of month $t-1$. A detailed description of these variables is summarized in the appendix. The sample covers all U.S. common stocks traded on the NYSE/AMEX/NASDAQ, and the sample period is from January 1962 to December 2013, with the last 12-month period starting in January 2013.

Table 2.2: Univariate Portfolio Sorts: Dependence Measures

Panel A: Lower Quadrant Probability

Portfolio	EW Return	Carhart-Alpha	VW Return	Carhart-Alpha
1 Low	-5.89%	-11.92%***	-9.16%	-13.85%***
2	11.68%	-1.42%	1.37%	-5.99%***
3	17.23%	5.59%***	6.26%	0.19%
4	18.11%	8.39%***	7.84%	1.40%**
5 High	16.15%	7.97%***	8.89%	2.60%***
High - Low	22.03%***	19.89%***	18.05%***	16.45%***
t-stat	(10.31)	(8.56)	(8.11)	(7.72)

Panel B: Upper Quadrant Probability

Portfolio	EW Return	Carhart-Alpha	VW Return	Carhart-Alpha
1 Low	17.33%	10.18%***	12.46%	8.02%***
2	18.74%	8.66%***	10.86%	4.72%***
3	15.61%	3.25%***	9.08%	2.13%***
4	8.61%	-3.11%***	6.30%	-0.42%
5 High	-1.79%	-8.82%***	4.24%	-1.85%***
High - Low	-19.12%***	-19.00%***	-8.22%***	-9.87%***
t-stat	(-12.55)	(-12.06)	(-4.09)	(-5.79)

Panel C: Downside Asymmetric Dependence

Portfolio	EW Return	Carhart-Alpha	VW Return	Carhart-Alpha
1 Low	-6.96%	-12.80%***	-1.56%	-7.54%***
2	8.37%	-3.88%***	5.17%	-1.86%***
3	15.82%	4.08%***	8.38%	1.66%***
4	20.27%	9.89%***	10.61%	4.40%***
5 High	21.21%	12.78%***	10.78%	5.35%***
High - Low	28.17%***	25.58%***	12.34%***	12.89%***
t-stat	(21.45)	(16.54)	(9.46)	(8.84)

This table reports both equal-weighted and value-weighted average annualized returns and Carhart's (1997) four factor alphas of stock portfolios sorted by contemporaneous lower quadrant probability, upper quadrant probability and downside asymmetric dependence evaluated at the mean (exceedance $c = 0$). In each month, we rank stocks into quintile (1-5) portfolios based on the next 12 month realized measures and report the average excess returns over the same 12 months for each portfolio. The row labeled "High - Low" reports the difference between the returns of portfolio 5 and portfolio 1, with corresponding statistical significance levels. The sample covers all U.S. common stocks traded on the NYSE/AMEX/NASDAQ, and the sample period is from January 1962 to December 2013, with the last 12-month period starting in January 2013. Since the 12-month returns are computed using overlapping periods, the t-statistics computed using Newey-West (1987) heteroskedastic-robust standard errors with 12 lags are reported in parentheses. *, ** and *** indicate significance levels at 0.1, 0.05 and 0.01 respectively.

Table 2.3: Univariate Portfolio Sorts: Beta Measures

Panel A: Downside Beta (β^-)

Portfolio	EW Return	Carhart-Alpha	VW Return	Carhart-Alpha
1 Low	7.36%	0.33%	4.36%	-0.28%
2	8.54%	0.73%	5.02%	0.20%
3	10.08%	0.66%	6.45%	0.05%
4	12.94%	1.74%**	8.63%	0.33%
5 High	19.60%	5.82%***	11.24%	1.31%
High - Low	12.23%***	5.49%***	6.88%**	1.59%
t-stat	(4.08)	(2.62)	(2.40)	(0.73)

Panel B: Upside Beta (β^+)

Portfolio	EW Return	Carhart-Alpha	VW Return	Carhart-Alpha
1 Low	10.41%	1.55%*	5.71%	-0.54%
2	10.53%	1.90%**	6.31%	0.18%
3	11.02%	2.03%***	6.46%	1.03%*
4	11.91%	1.47%**	6.38%	0.04%
5 High	13.87%	1.86%*	6.60%	-2.04%
High - Low	3.46%	0.30%	0.89%	-1.50%
t-stat	(1.58)	(0.19)	(0.38)	(-0.68)

Panel C: Downside Beta - Upside Beta ($\beta^- - \beta^+$)

Portfolio	EW Return	Carhart-Alpha	VW Return	Carhart-Alpha
1 Low	7.67%	-1.32%	3.16%	-2.15%**
2	9.61%	1.15%*	6.69%	1.12%**
3	11.06%	2.08%***	7.30%	1.36%***
4	13.12%	2.90%***	8.59%	0.77%
5 High	16.80%	4.11%***	9.49%	-0.17%
High - Low	9.13%***	5.43%***	6.33%***	1.99%
t-stat	(6.87)	(4.88)	(3.24)	(1.00)

This table reports both equal-weighted and value-weighted average annualized excess returns and Carhart's (1997) four factor alphas of stock portfolios sorted by contemporaneous β^-, β^+ and $\beta^- - \beta^+$. In each month, we rank stocks into quintile (1-5) portfolios based on the next 12 month realized beta measures and report the average excess returns over the same 12 months for each portfolio. The row labeled "High - Low" reports the difference between the returns of portfolio 5 and portfolio 1, with corresponding statistical significance levels. The sample covers all U.S. common stocks traded on the NYSE/AMEX/NASDAQ, and the sample period is from January 1962 to December 2013, with the last 12-month period starting in January 2013. Since the 12-month returns are computed using overlapping periods, the t-statistics computed using Newey-West (1987) heteroskedastic-robust standard errors with 12 lags are reported in parentheses. *, ** and *** indicate significance levels at 0.1, 0.05 and 0.01 respectively.

Table 2.4: Summary Statistics

Panel A: Summary statistics for decile portfolios sorted by LQP

Decile	LQP^0	β	β^-	β^+	Size	Bm	Turn	Illiq	Mom	Ivol	Coskew	Cokurt	Max
1 Low	0.224	0.370	0.531	0.311	4.060	-0.310	0.028	13.366	0.065	0.026	-0.079	0.910	0.052
2	0.265	0.611	0.777	0.526	4.765	-0.498	0.052	8.532	0.094	0.026	-0.090	1.380	0.053
3	0.288	0.743	0.908	0.653	5.175	-0.583	0.066	5.940	0.111	0.025	-0.099	1.668	0.052
4	0.302	0.831	0.984	0.744	5.355	-0.601	0.071	4.805	0.113	0.024	-0.103	1.852	0.052
5	0.314	0.897	1.033	0.817	5.448	-0.600	0.073	4.304	0.111	0.024	-0.101	1.980	0.052
6	0.323	0.955	1.068	0.876	5.514	-0.587	0.075	4.064	0.106	0.024	-0.097	2.068	0.052
7	0.332	1.019	1.107	0.940	5.567	-0.578	0.077	3.895	0.102	0.024	-0.094	2.161	0.053
8	0.341	1.085	1.147	1.002	5.625	-0.569	0.080	3.755	0.098	0.024	-0.091	2.245	0.054
9	0.352	1.158	1.181	1.071	5.700	-0.547	0.082	3.650	0.092	0.024	-0.087	2.327	0.054
10 High	0.370	1.266	1.203	1.184	5.904	-0.531	0.080	3.343	0.084	0.024	-0.073	2.464	0.055

Panel B: Summary statistics for decile portfolios sorted by UQP

Decile	UQP^0	β	β^-	β^+	Size	Bm	Turn	Illiq	Mom	Ivol	Coskew	Cokurt	Max
1 Low	0.201	0.412	0.622	0.336	3.704	-0.158	0.026	15.098	0.049	0.029	-0.081	0.914	0.054
2	0.234	0.624	0.797	0.524	4.232	-0.325	0.045	10.706	0.075	0.028	-0.093	1.325	0.055
3	0.254	0.753	0.901	0.662	4.701	-0.437	0.059	7.369	0.090	0.026	-0.095	1.606	0.054
4	0.270	0.843	0.975	0.754	5.054	-0.511	0.069	5.409	0.098	0.025	-0.096	1.800	0.054
5	0.284	0.914	1.030	0.827	5.311	-0.561	0.073	4.386	0.104	0.024	-0.097	1.940	0.053
6	0.297	0.982	1.079	0.902	5.539	-0.605	0.078	3.569	0.110	0.024	-0.094	2.070	0.053
7	0.309	1.043	1.121	0.969	5.769	-0.650	0.082	3.007	0.113	0.023	-0.092	2.199	0.053
8	0.323	1.081	1.138	1.014	5.952	-0.674	0.083	2.744	0.113	0.022	-0.090	2.301	0.052
9	0.337	1.119	1.144	1.054	6.106	-0.690	0.083	2.678	0.109	0.022	-0.087	2.383	0.051
10 High	0.361	1.151	1.125	1.071	6.327	-0.692	0.077	2.722	0.103	0.021	-0.082	2.474	0.050

This table shows summary statistics of the main variables for decile portfolios formed every month based on realized lower quadrant probability (Panel A) and upper quadrant probability (Panel B). The table reports for each decile the average across the months in the sample of the cross-sectional mean values within each month of the variables–CAPM β, downside beta (β^-), upside beta (β^+), log of market capitalization (Size), log of book-to-market ratio (Bm), turnover ratio (Turn), normalized Amihud illiquidity measure (Illiq), past six-month return (Mom), idiosyncratic volatility (Ivol), coskewness (Coskew), cokurtosis (Cokurt) and the maximum daily return over the past one month (max). At the beginning of each month t, all risk characteristics (LQP, UQP, β, β^-, β^+, Ivol, Coskew, Cokurt) are calculated using daily realized excess returns over the next 12-month period. Size, Bm, Turn, Illiq, Mom and Max are calculated using information available at the end of month $t-1$. The sample covers all U.S. common stocks traded on the NYSE/AMEX/NASDAQ, and the sample period is from January 1962 to December 2013, with the last 12-month period starting in January 2013.

Table 2.5: Dependent Portfolio Sorts: Lower Quadrant Probability (LQP)

Panel A: Beta (β) and LQP

Portfolio	1 Low β	2	3	4	5 High β	Average
1 Low LQP	-9.21%	-7.66%	-6.72%	-4.55%	1.57%	-5.32%
5 High LQP	15.39%	13.97%	13.87%	14.10%	18.96%	15.26%
High - Low	24.59%***	21.63%***	20.59%***	18.65%***	17.39%***	20.57%***
t-stat	(20.45)	(13.34)	(11.20)	(9.17)	(7.56)	(13.40)

Panel B: Size and LQP

Portfolio	1 Low size	2	3	4	5 High size	Average
1 Low LQP	-6.27%	-9.23%	-8.44%	-6.24%	-2.81%	-6.60%
5 High LQP	26.23%	20.83%	18.03%	15.23%	11.45%	18.36%
High - Low	32.51%***	30.07%***	26.47%***	21.47%***	14.26%***	24.95%***
t-stat	(17.92)	(14.05)	(11.87)	(10.07)	(7.26)	(14.39)

Panel C: Coskewness and LQP

Portfolio	1 Low coskew	2	3	4	5 High coskew	Average
1 Low LQP	-1.75%	-3.68%	-6.36%	-7.46%	-7.51%	-5.35%
5 High LQP	20.89%	19.59%	17.03%	14.58%	11.59%	16.73%
High - Low	22.64%***	23.27%***	23.39%***	22.04%***	19.10%***	22.09%***
t-stat	(9.88)	(11.24)	(11.55)	(11.09)	(9.03)	(11.44)

Panel D: Cokurtosis and LQP

Portfolio	1 Low cokurt	2	3	4	5 High cokurt	Average
1 Low LQP	-10.16%	-9.08%	-6.00%	-2.31%	4.79%	-4.55%
5 High LQP	19.36%	18.04%	17.28%	16.44%	15.56%	17.33%
High - Low	29.57%***	27.12%***	23.28%***	18.75%***	10.77%***	21.89%***
t-stat	(23.96)	(18.16)	(12.31)	(8.82)	(5.00)	(14.24)

This table reports equal-weighted average annualized excess returns of portfolios double-sorted by realized lower quadrant probability (LQP) and realized CAPM β (Panel A), firm market capitalization (Panel B), realized coskewness (Panel C) and realized cokurtosis (Panel D), respectively. LQP is evaluated at the sample mean. For each month, we compute LQP, β, coskewness and cokurtosis using daily realized stock and market excess returns over the next 12 months. Size is computed at the beginning of each month using information at the end of previous month. First, we form quintile portfolios sorted on β, size, coskewness and cokurtosis respectively. Then, we rank stocks within each first-sort quintile into additional quintiles based on LQP. The row labeled "High - Low" reports the difference between the returns of portfolio 5 and portfolio 1 in each β, size, coskewness and cokurtosis first-sort quintile with corresponding statistical significance levels. The column labeled "Average" reports the average return of stocks in each second-sort quintile. The sample covers all U.S. common stocks traded on the NYSE/AMEX/NASDAQ, and the sample period is from January 1962 to December 2013, with the last 12-month period starting in January 2013. Newey-West (1987) 12-lag adjusted t-statistics are reported in parentheses. *, ** and *** indicate significance levels at 0.1, 0.05 and 0.01 respectively.

Table 2.6: Dependent Portfolio Sorts: Upper Quadrant Probability (UQP)

Panel A: Beta (β) and UQP

Portfolio	1 Low β	2	3	4	5 High β	Average
1 Low UQP	16.26%	19.60%	20.44%	22.55%	26.54%	21.08%
5 High UQP	-7.82%	-7.16%	-5.59%	-2.60%	4.72%	-3.69%
High - Low	-24.37%***	-26.76%***	-26.03%***	-25.16%***	-21.82%***	-24.80%***
t-stat	(-18.61)	(-20.05)	(-19.63)	(-14.02)	(-10.04)	(-17.71)

Panel B: Size and UQP

Portfolio	1 Low size	2	3	4	5 High size	Average
1 Low UQP	19.37%	16.04%	14.73%	12.06%	9.27%	14.29%
5 High UQP	0.11%	-3.66%	-2.65%	-0.58%	3.79%	-0.60%
High - Low	-19.26%***	-19.70%***	-17.38%***	-12.63%***	-5.48%***	-14.89%***
t-stat	(-7.83)	(-9.12)	(-8.54)	(-6.69)	(-3.11)	(-8.57)

Panel C: Coskewness and UQP

Portfolio	1 Low coskew	2	3	4	5 High coskew	Average
1 Low UQP	19.44%	19.11%	17.44%	16.14%	14.58%	17.34%
5 High UQP	3.71%	0.55%	-2.13%	-4.38%	-6.59%	-1.77%
High - Low	-15.72%***	-18.56%***	-19.57%***	-20.52%***	-21.17%***	-19.11%***
t-stat	(-9.12)	(-11.36)	(-11.95)	(-13.42)	(-13.26)	(-12.79)

Panel D: Cokurtosis and UQP

Portfolio	1 Low cokurt	2	3	4	5 High cokurt	Average
1 Low UQP	17.55%	18.51%	18.48%	18.55%	18.40%	18.30%
5 High UQP	-10.86%	-11.32%	-7.95%	-2.89%	6.23%	-5.36%
High - Low	-28.77%***	-29.84%***	-26.43%***	-21.43%***	-12.17%***	-23.67%***
t-stat	(-23.40)	(-18.91)	(-15.57)	(-14.16)	(-8.81)	(-18.53)

This table reports equal-weighted average annualized excess returns of portfolios double-sorted by realized upper quadrant probability (UQP) and realized CAPM β (Panel A), firm market capitalization (Panel B), realized coskewness (Panel C) and realized cokurtosis (Panel D), respectively. UQP is evaluated at the sample mean. For each month, we compute LQP, β, coskewness and cokurtosis using daily realized stock and market excess returns over the next 12 months. Size is computed at the beginning of each month using information at the end of previous month. First, we form quintile portfolios sorted on β, size, coskewness and cokurtosis respectively. Then, we rank stocks within each first-sort quintile into additional quintiles based on LQP. The row labeled "High - Low" reports the difference between the returns of portfolio 5 and portfolio 1 in each β, size, coskewness and cokurtosis first-sort quintile with corresponding statistical significance levels. The column labeled "Average" reports the average return of stocks in each second-sort quintile. The sample covers all U.S. common stocks traded on the NYSE/AMEX/NASDAQ, and the sample period is from January 1962 to December 2013, with the last 12-month period starting in January 2013. Newey-West (1987) 12-lag adjusted t-statistics are reported in parentheses. *, ** and *** indicate significance levels at 0.1, 0.05 and 0.01 respectively.

Table 2.7: Dependent Portfolio Sorts: Downside Asymmetric Dependence (DownAsy)

Panel A: Beta (β) and DownAsy

Portfolio	1 Low β	2	3	4	5 High β	Average
1 Low DownAsy	-8.55%	-7.77%	-7.46%	-6.28%	-0.28%	-6.07%
5 High DownAsy	16.90%	20.31%	21.77%	24.74%	28.68%	22.48%
High - Low	25.45%***	28.08%***	29.22%***	31.01%***	28.96%***	28.55%***
t-stat	(24.86)	(26.14)	(21.46)	(18.83)	(14.89)	(23.35)

Panel B: Size and DownAsy

Portfolio	1 Low size	2	3	4	5 High size	Average
1 Low DownAsy	-4.74%	-8.18%	-7.24%	-4.92%	-1.13%	-5.24%
5 High DownAsy	26.12%	22.47%	19.21%	14.63%	10.95%	18.68%
High - Low	30.86%***	30.65%***	26.46%***	19.54%***	12.08%***	23.92%***
t-stat	(22.50)	(17.87)	(15.04)	(12.84)	(11.33)	(18.85)

Panel C: Coskewness and DownAsy

Portfolio	1 Low coskew	2	3	4	5 High coskew	Average
1 Low DownAsy	-1.97%	-5.25%	-7.58%	-8.83%	-9.49%	-6.62%
5 High DownAsy	24.89%	24.07%	21.63%	19.31%	16.17%	21.21%
High - Low	26.86%***	29.32%***	29.20%***	28.14%***	25.65%***	27.83%***
t-stat	(16.23)	(21.83)	(21.48)	(21.59)	(19.87)	(22.36)

Panel D: Cokurtosis and DownAsy

Portfolio	1 Low cokurt	2	3	4	5 High cokurt	Average
1 Low DownAsy	-10.28%	-10.61%	-7.89%	-4.00%	3.69%	-5.82%
5 High DownAsy	19.65%	21.83%	22.20%	21.94%	20.46%	21.22%
High - Low	29.93%***	32.44%***	30.09%***	25.94%***	16.77%***	27.03%***
t-stat	(30.12)	(24.50)	(18.90)	(14.53)	(10.44)	(22.06)

This table reports equal-weighted average annualized excess returns of portfolios double-sorted by realized downside asymmetric dependence (DownAsy) and realized CAPM β (Panel A), firm market capitalization (Panel B), realized coskewness (Panel C) and realized cokurtosis (Panel D), respectively. DownAsy is evaluated at the sample mean. For each month, we compute LQP, β, coskewness and cokurtosis using daily realized stock and market excess returns over the next 12 months. Size is computed at the beginning of each month using information at the end of previous month. First, we form quintile portfolios sorted on β, size, coskewness and cokurtosis respectively. Then, we rank stocks within each first-sort quintile into additional quintiles based on LQP. The row labeled "High - Low" reports the difference between the returns of portfolio 5 and portfolio 1 in each β, size, coskewness and cokurtosis first-sort quintile with corresponding statistical significance levels. The column labeled "Average" reports the average return of stocks in each second-sort quintile. The sample covers all U.S. common stocks traded on the NYSE/AMEX/NASDAQ, and the sample period is from January 1962 to December 2013, with the last 12-month period starting in January 2013. Newey-West (1987) 12-lag adjusted t-statistics are reported in parentheses. *, ** and *** indicate significance levels at 0.1, 0.05 and 0.01 respectively.

Table 2.8: Firm-Level Cross-Sectional Return Regressions

	(1) Return	(2) Return	(3) Return	(4) Return	(5) Return	(6) Return	(7) Return	(8) 4-fct R_{adj}	(9) 4-fct R_{adj}	Mean (Std Dev)	Econ Sig
LQP^0	2.049***		2.738***			2.755***		2.162***		0.311	9.08%
	(11.36)		(12.45)			(15.36)		(15.73)		(0.042)	
UQP^0		-1.471***	-1.706***			-1.339***		-1.093***		0.287	-5.14%
		(-12.68)	(-6.84)			(-4.27)		(-4.19)		(0.047)	
$DownAsy^0$				2.443***			2.330***		1.867***	0.013	8.59%
				(11.25)			(11.10)		(11.19)	(0.046)	
β^-					0.071***	0.071***	0.070***	0.052***	0.052***	0.993	3.68%
					(5.14)	(5.95)	(5.60)	(4.98)	(4.64)	(0.708)	
β^+					-0.017***	-0.009	-0.008	-0.006	-0.005	0.811	-0.39%
					(-2.65)	(-1.23)	(-0.99)	(-0.82)	(-0.63)	(0.783)	
Size						-0.031***	-0.025***	-0.022***	-0.017***	5.265	-2.88%
						(-7.02)	(-6.33)	(-6.31)	(-5.51)	(1.693)	
Bm						0.013**	0.015***	-0.004	-0.002	-0.530	-0.16%
						(2.29)	(2.59)	(-0.88)	(-0.49)	(0.775)	
Turn						-0.470***	-0.423***	-0.383***	-0.349***	0.071	-2.62%
						(-6.65)	(-6.49)	(-5.00)	(-4.65)	(0.075)	
Illiq						0.003***	0.003***	0.002***	0.002***	5.838	1.72%
						(5.68)	(6.46)	(5.07)	(5.85)	(8.591)	
Mom						0.057***	0.056***	0.005	0.003	0.096	0.08%
						(4.01)	(3.83)	(0.34)	(0.26)	(0.277)	
Ivol						-1.869**	-0.845	-1.116*	-0.314	0.024	-0.31%
						(-2.17)	(-0.91)	(-1.73)	(-0.46)	(0.010)	
Coskew						-0.046*	-0.039	-0.037	-0.031	-0.091	-0.51%
						(-1.72)	(-1.36)	(-1.37)	(-1.20)	(0.163)	
Cokurt						0.029***	0.061***	0.024***	0.047***	1.902	5.33%
						(4.00)	(7.49)	(3.25)	(6.10)	(1.133)	
Max						-0.045	-0.032	-0.326***	-0.315***	0.054	-1.13%
						(-0.65)	(-0.43)	(-5.29)	(-4.87)	(0.036)	
Constant	-0.517***	0.539***	-0.211***	0.096***	0.055***	-0.191***	0.130***	-0.212***	0.027		
	(-9.52)	(17.38)	(-3.48)	(4.83)	(2.98)	(-3.36)	(4.63)	(-4.76)	(1.17)		
Obs	1,307,423	1,307,423	1,307,423	1,307,423	1,307,423	1,307,423	1,307,423	1,307,423	1,307,423		
R^2	0.082	0.065	0.131	0.067	0.061	0.248	0.202	0.146	0.116		

This table gives the results of multivariate Fama-MacBeth (1973) regressions. 12-month firm-level excess returns over the risk-free rate are regressed on LQP, UQP, downside asymmetric dependence (DownAsy) all evaluated at the sample mean and a set of other explanatory variables that have been shown to affect cross-sectional stock returns by previous studies. At the beginning of each month t, all risk characteristics (LQP, UQP, DownAsy, β^-, β^+, Ivol, Coskew, Cokurt) are calculated using daily realized stock excess returns and market returns over the following 12-month period. The dependent return variables are computed contemporaneously over the same period. Size, Bm, Turn, Illiq, Mom and Max are calculated at the beginning of each month using information available at the end of month $t-1$. The second to last column displays the time series averages of cross-sectional mean and standard deviation of each independent variable. The last column reports the change in 12-month Carhart (1997) four factor adjusted excess returns for a one standard deviation increase in the respective independent variable based on regressions (8) and (9). The dependent variable is the 12-month excess return in model (1-7) and Carhart (1997) four factor adjusted return in model (8-9). The adjusted returns are calculated following the method suggested by Brennan, Chordia, and Subrahmanyam (1998). The sample includes all U.S. common stocks traded on the NYSE/AMEX/NASDAQ, and the sample period is from January 1962 to December 2013, with the last 12-month period starting in January 2013. Newey-West (1987) 12-lag adjusted t-statistics are reported in parentheses. *, ** and *** indicate significance levels at 0.1, 0.05 and 0.01 respectively.

Table 2.9: Firm-Level Cross-Sectional Return Regressions: Robustness Checks

	Value-weighted		NYSE only		Nonoverlapping		Other Exceedance levels			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
LQP^0	1.555***		1.918***		2.323***					
	(13.77)		(16.78)		(14.19)					
UQP^0	-1.198***		-1.331***		-1.184***					
	(-5.85)		(-5.10)		(-4.29)					
$DownAsy^0$		1.157***		1.705***		1.935***				
		(10.31)		(10.72)		(10.37)				
β^-	0.052***	0.049***	0.063***	0.061***	0.050***	0.049***	0.006	0.032**	0.031***	0.062***
	(4.13)	(3.34)	(5.92)	(5.21)	(3.61)	(3.39)	(0.47)	(2.50)	(2.65)	(4.16)
β^+	-0.002	0.003	-0.017**	-0.014*	-0.004	-0.002	0.026***	0.018**	0.000	0.018**
	(-0.16)	(0.27)	(-2.07)	(-1.78)	(-0.48)	(-0.17)	(3.48)	(2.52)	(0.01)	(2.04)
Size	-0.021***	-0.017***	-0.020***	-0.016***	-0.026***	-0.022***	-0.023***	-0.019***	-0.025***	-0.025***
	(-6.35)	(-5.28)	(-6.50)	(-5.49)	(-7.14)	(-6.30)	(-6.61)	(-6.26)	(-7.07)	(-6.33)
Bm	-0.020***	-0.020***	-0.009**	-0.007	-0.001	0.000	-0.002	-0.000	0.001	-0.004
	(-3.63)	(-3.60)	(-2.09)	(-1.55)	(-0.26)	(0.08)	(-0.37)	(-0.10)	(0.14)	(-0.73)
Turn	-0.204**	-0.173*	-0.314***	-0.295***	-0.599***	-0.552***	-0.408***	-0.365***	-0.401***	-0.311***
	(-2.12)	(-1.85)	(-3.90)	(-3.75)	(-4.87)	(-4.72)	(-5.31)	(-4.76)	(-5.09)	(-4.00)
Illiq	0.004***	0.005***	0.003***	0.004***	0.001**	0.002***	0.002***	0.002***	0.002***	0.005***
	(3.51)	(4.49)	(3.81)	(5.27)	(2.40)	(2.98)	(5.11)	(5.66)	(4.70)	(4.77)
Mom	0.024	0.019	0.000	-0.002	-0.023	-0.024	0.004	0.002	-0.000	0.002
	(1.56)	(1.19)	(0.02)	(-0.10)	(-0.99)	(-1.07)	(0.33)	(0.14)	(-0.00)	(0.11)
Ivol	-6.892***	-6.707***	-3.160***	-2.249***	-0.093	0.634	-0.192	-0.895	0.300	-2.616**
	(-7.81)	(-7.24)	(-3.86)	(-2.72)	(-0.10)	(0.68)	(-0.27)	(-1.19)	(0.40)	(-2.36)
Coskew	-0.070*	-0.054	-0.027	-0.017	-0.022	-0.013	-0.080***	-0.043	0.016	-0.002
	(-1.78)	(-1.45)	(-0.74)	(-0.53)	(-0.59)	(-0.32)	(-2.88)	(-1.54)	(0.62)	(-0.06)
Cokurt	0.007	0.014	0.020**	0.033***	0.032***	0.056***	0.034***	0.043***	0.027***	0.031***
	(0.62)	(1.35)	(2.48)	(3.96)	(3.29)	(6.09)	(4.06)	(5.32)	(3.10)	(3.44)
Max	-0.082	-0.034	-0.251***	-0.236***	-0.536***	-0.520***	-0.341***	-0.298***	-0.343***	-0.311***
	(-0.94)	(-0.37)	(-3.83)	(-3.47)	(-6.12)	(-6.03)	(-5.21)	(-4.39)	(-5.21)	(-4.03)
$LQP^{0.5}$							2.715***			
							(16.68)			
$UQP^{0.5}$							-1.766***			
							(-9.31)			
$DownAsy^{0.5}$								0.432***		
								(11.59)		
LQP^1									2.459***	
									(10.03)	
UQP^1									0.064	
									(0.22)	
$DownAsy^1$										0.090***
										(5.02)
Constant	0.083	0.141***	-0.044	0.063**	-0.215***	0.049*	-0.018	0.058**	0.020	0.113***
	(1.37)	(4.48)	(-0.82)	(2.36)	(-3.79)	(1.76)	(-0.70)	(2.28)	(0.78)	(3.14)
Obs	1,307,423	1,307,423	702,879	702,879	108,675	108,675	1,307,423	1,301,597	1,307,423	823,063
R^2	0.203	0.171	0.162	0.127	0.151	0.120	0.096	0.087	0.087	0.105

This table reports the results of a battery of multivariate Fama-MacBeth (1973) regressions under different specifications for robustness checks. Carhart (1997) four factor adjusted returns is used as the dependent variable in all the 10 regressions. Model 1 and 2 report the value-weighted regression results with full set of controls and the weighting variable is firm's market capitalization. Model 3 and 4 report the regression results with the same specification but the sample is restricted to NYSE stocks only. Model 5 and 6 report the regression results using non-overlapping yearly observations. Model 7 to 10 report the regression results with the same set of controls, but lower quadrant probability (LQP), upper quadrant probability (UQP) and downside asymmetric dependence (DownAsy) are evaluated at other exceedance levels, i.e. 0.5 and 1 standard deviations away from the mean. The sample includes all U.S. common stocks traded on the NYSE/AMEX/NASDAQ, and the sample period is from January 1962 to December 2013, with the last 12-month period starting in January 2013. Newey-West (1987) 12-lag adjusted t-statistics are reported in parentheses, except for model 5 and 6 where standard t-statistics are reported. *, ** and *** indicate significance levels at 0.1, 0.05 and 0.01 respectively.

Table 2.10: Determinants of Downside Asymmetric Dependence

Past Variables												
$DownAsy^0$	β^-	β^+	Size	Bm	Turn	Illiq	Mom	Ivol	Coskew	Cokurt	Max	R^2
0.058*** (6.60)												0.009
	-0.003*** (-3.82)											0.009
		-0.003*** (-4.04)										0.008
			-0.002*** (-5.30)									0.017
				0.005*** (5.72)								0.013
					-0.05*** (-3.68)							0.006
						0.000*** (3.21)						0.012
							-0.003** (-2.50)					0.004
								0.066 (0.96)				0.013
									0.002 (1.24)			0.004
										-0.004*** (-4.34)		0.013
											-0.014 (-1.08)	0.006
0.046*** (5.52)	0.000 (0.31)	-0.001*** (-2.83)	-0.001*** (-4.09)	0.003*** (5.56)	-0.027*** (-3.77)	-0.000** (-2.50)	-0.003*** (-3.43)	0.052 (0.78)	0.011*** (3.43)	-0.002** (-2.09)	-0.019*** (-5.10)	0.056

The table reports the results of Fama-MacBeth (1973) regressions of realized downside asymmetric dependence (DownAsy) over a 12-month period on a set of past firm characteristics and risk measure variables including the 12-month lagged DownAsy. At the beginning of each month t, the past risk measures (β^-, β^+, Ivol, Coskew, Cokurt) are estimated over the previous 12 months ($t-12$ to $t-1$) that does not overlap with the current 12-month period when the dependent variable is evaluated. Size, Bm, Turn, Illiq, Mom and Max are calculated at the beginning of the month t. The sample includes all U.S. common stocks traded on the NYSE/AMEX/NASDAQ, and the sample period is from January 1963 to December 2013, with the last 12-month period starting in January 2013. Newey-West (1987) 12-lag adjusted t-statistics are reported in parentheses. *, ** and *** indicate significance levels at 0.1, 0.05 and 0.01 respectively.

Table 2.11: Trading Strategy Based on Past Downside Asymmetric Dependence (DownAsy)

Panel A: Portfolios Returns with 12-month Holding Period

Portfolio	Return	CAPM-Alpha	FF-Alpha	Carhart-Alpha
1 Low	10.63%	4.89%***	0.51%	1.38%**
2	9.66%	3.57%***	0.48%	1.02%*
3	10.28%	4.16%***	1.05%*	1.17%**
4	11.81%	5.47%***	2.15%***	1.47%***
5 High	13.86%	7.86%***	3.98%***	2.61%***
High - Low	3.22%***	2.97%***	3.47%***	1.23%*
t-stat	(3.73)	(3.35)	(3.52)	(1.78)

Panel B: Portfolios Returns with 1-month Holding Period

Portfolio	Return	CAPM-Alpha	FF-Alpha	Carhart-Alpha
1 Low	0.65%	0.15%	-0.12%*	0.02%
2	0.75%	0.17%	-0.03%	0.06%
3	0.85%	0.27%***	0.07%*	0.12%***
4	0.94%	0.36%***	0.16%***	0.16%***
5 High	1.03%	0.49%***	0.25%***	0.21%***
High - Low	0.37%***	0.34%***	0.38%***	0.19%**
t-stat	(4.39)	(4.04)	(4.61)	(2.47)

This table reports equal-weighted average returns and alphas of stock portfolios sorted by past DownAsy evaluated at the sample mean. In each month, we rank stocks into quintile (1-5) portfolios based on the past 12-month realized DownAsy. we report the average excess returns/alphas over the next 12 months for each portfolio in Panel A and the average excess returns/alphas over the next 1 month in Panel B. The row labeled "High - Low" reports the difference between the returns/alphas of portfolio 5 and portfolio 1, with corresponding statistical significance levels. The sample covers all U.S. common stocks traded on the NYSE/AMEX/NASDAQ, and the sample period is from January 1963 to December 2013, with the last 12-month period starting in January 2013. The t-statistics computed using Newey-West (1987) heteroskedastic-robust standard errors with 12 lags are reported in parentheses for Panel A and standard t-statistics in parentheses for Panel B. *, ** and *** indicate significance levels at 0.1, 0.05 and 0.01 respectively.

Chapter 3

The Gap Between the Conditional Wage Distributions of Incumbents and the Newly Hired Employees: Decomposition and Uniform Ordering (joint with Esfandiar Maasoumi and Melinda Pitts)

Abstract

We examine the cardinal gap between wage distributions of the incumbents and newly hired workers based on entropic distances which are well defined welfare theoretic measures. Decomposition of several effects is achieved by identifying several counterfactual distributions of different groups. These go beyond the usual Oaxaca-Blinder decompositions at the (linear) conditional means. Much like quantiles, these entropic distances are well defined inferential objects and functions whose statistical properties have recently been developed. Going beyond these strong rankings and distances, we consider weak uniform ranking of these wage outcomes based on statistical tests for stochastic dominance. The empirical analysis is focused on employees with at least 35 hours of work in the 1996-2012 monthly Current Population Survey (CPS). Among others, we find incumbent workers enjoy a better distribution of wages, but the attribution of the gap to wage inequality and human capital characteristics varies between quantiles. For instance, highly paid new workers are mainly due to human capital components, and in some years, even better wage structure.

Keywords: Wage gap, metric entropy distance, stochastic dominance, wage distributions, counterfactual analysis, inequality, labor markets.

JEL Classification: I31, C43

3.1 Introduction

Wage differentials among different types of workers, e.g. the gender earnings gap, wage differences between immigrants and natives, etc., have drawn much attention from labor economists and policy makers. There is an extensive literature on labor market outcomes, much of it focused on the analysis of the wage gap at the mean, median and other quantiles of the wage distribution. More recently, techniques have been provided for identifying entire distributions and general function of the distributions. These techniques provide the backdrop for the current paper's approach. One central object of inference in this paper is a summary measure of the "distance" between the entire distributions of interest. Our proposed summary measure makes clear that all other measures of the gap between two distributions are special, and all imply and are implied by well defined welfare functions. Seen in this light, comparison at the mean, median, or any particular quantile would appear to place too much weight on a part of the population, or too equal a weight everywhere. For example, Blau and Kahn (2006) documented the slowing convergence of the gender gap at the mean, median and 90th percentile levels. Albrecht, Björklund, and Vroman (2003) looked at wages differentials at different parts of the distribution to see whether the gender gap is larger in the upper tail than in the lower tail of the wage distribution due to a "glass ceiling" effect in Sweden. Kampkötter and Sliwka (2011)investigated average wage differences between newly hired and incumbent employees. While these focused examinations are informative and useful, recent papers have examined the wage differentials at the entire distribution level. For example, Maasoumi and Wang (2013) employed a metric entropy measure proposed by Granger, Maasoumi, and Racine (2004) to examine the gender wage gap based on the metric distance between two distributions. The measure is the metric member of the Generalized Entropy class of measures with very credible welfare theoretic foundations.

All measures of the gap provide strong ranking of outcome distributions since they are based on implicit "cardinal" welfare or weighting functions. They are inevitably subjective even though some are less extreme than others. In view of this, we explore weak uniform rankings based on the concept of stochastic dominance which allow assessments over entire classes of welfare functions. We do so by rigorous statistical tests for various orders of dominance.

A key issue of interest is about decomposition of observed gaps and rankings in order to identify the factors that underlie the overall wage differentials. Specifically, are those differentials associated with inequality or discrimination in the wage structure, or are they due to human capital composition effect. The classic decomposition method is due to Oaxaca (1973) and Blinder (1973). It is a regression-based method focusing on linear conditional mean decomposition. One major limitation of the Oaxaca-Blinder procedure discussed by Barsky et al. (2002) is that the decomposition provides consistent estimates of the structure and composition effect only under the assumption that the conditional expectation is linear. As advocated in DiNardo, Fortin, and Lemieux (1996), we take an alternative non-parametric decomposition approach based on propensity score reweighting methods. The key advantage of this reweighting approach is that it identifies the counterfactual distribution under less restrictive assumptions and hence can easily be applied to more general distributional statistics, rather than the simple mean and quantiles.

Several recent papers, e.g. Firpo, Fortin, and Lemieux (2007), Maasoumi and Wang (2013), have applied this reweighting method for wage gap decompositions. Following the recent approach, this paper decomposes the wage gap between newly hired and incumbent employees across the entire distribution. The wage differences between newly hired and incumbent employees is a less studied topic in labor economics. The seminal work of Doeringer and Piore (1985) provided a theoretical foundation in this area, claiming that the incumbent wage could partially be determined by internal labor markets. Following the work of Baker, Gibbs, and Holmstrom (1994), many empirical studies investigated the wage structure of the internal labor markets. But very few studies have been done to examine the difference between the wage structure of the internal labor markets and that of the external labor markets. Studying the differential is very important because it sheds light on how much external market forces could determine the wage formation within firms. It could also potentially serve as an indicator of competitiveness of labor markets, since wage differentials between new hires and incumbents with identical characteristics should not exist in perfectly competitive labor markets.

This paper's analysis focuses on a sample of employed workers. As we do not address the issue of selection into the labor market, this work is only generalizable to the work force and not the population as a whole. The plan of the rest of the paper is as follows. In section 3.2, the decomposition and counterfactual approach are explained. Subsection 3.2.1 introduces the idea of decomposition with a general distributional function. Subsection 3.2.2 discusses the Oaxaca-Blinder decomposition that employs linear conditional expectation as the functional form. Section 3.3 first introduces a metric entropy measure and its welfare implications, and then discusses the empirical and analytical methodologies in details, i.e. the stochastic dominance tests and the propensity score reweighting method used to identify counterfactual distributions. Section 3.4 explains how to construct the linked CPS monthly data set used in the paper. Section 3.5 gives the results of the stochastic dominance tests and counterfactual analysis. The conclusion is in Section 3.6.

3.2 The Decomposition Problem

A key question of interest in this paper is how to decompose the distributional wage gap between incumbent and newly hired employees into a composition effect, corresponding to differences in the covariates between the two groups, and a wage structure effect corresponding to differences in the return to the covariates. In this section, we present a general theoretical framework illustrating the decomposition at the distributional level. We also link this decomposition to the more popular Oaxaca-Blinder decomposition method. We then propose to apply an entropy metric, a distributional statistic that could summarize differences between two distributions, to measure the structure and composition effects, and present its welfare function underpinnings.

3.2.1 Decomposition with General Distributional Function

The outcome variable of interest is the log hourly wage. We have two groups of workers, the incumbent group denoted as group 0 and the new hire group denoted as group 1. Let $ln(w^0)$ and $ln(w^1)$ denote the log wages of incumbent and newly hired employees, respectively. We observe a random sample of $N = N_0 + N_1$ workers. N_0 denotes the sample size of incumbents and N_1 is the sample size of the newly hired employees. Let $F_0(y) \equiv Pr[ln(w^0) \leq y]$ represents the cumulative distribution function (CDF) of $ln(w^0)$ and $f_0(y)$ is the corresponding probability density function (PDF). The same notations apply to the log wages of newly hired employees.

The wage structure of the incumbent group is denoted by g_0 and that of the newly hired group is denoted by g_1. Individual wages are determined non-parametrically by both observed characteristics X_i and unobserved characteristics ε_i via the unknown wage structure functions g_d,

$$ln(w_i^D) = g_D(X_i, \varepsilon_i) \qquad for\ D = 0, 1 \tag{3.1}$$

This non-parametric approach avoids imposing distributional assumptions or specific functional forms, which allow for very flexible interactions among X_i and ε_i. We only assume that $(ln(w), X, D)$ have some unknown joint distribution. Under such specification, the wage differential is assumed to be associated to two primary sources: (1) differences in observed human capital characteristics X_i (e.g. education, age, etc.), and unobserved human capital characteristics ε_i (e.g. innate ability). However, under the unconfoundedness assumption elaborated in the next section, the composition effect only comes from differences in X_i and differences in the wage structures, $g_D(\cdot)$.

With observed data, we can identify the conditional distribution of a new hire's log hourly wage, $ln(w^1)|X, D = 1 \overset{d}{\sim} F_{1|X}$, and the conditional distribution of the incumbent's log hourly wage, $ln(w^0)|X, D = 0 \overset{d}{\sim} F_{0|X}$. With certain further assumptions discussed later, we are able to identify the conditional counterfactual distribution of $ln(w^0)|X, D =$

$1 \overset{d}{\sim} F_{C|X}$ using the aforementioned propensity score reweighting method. The conditional counterfactual distribution $F_{C|X}$ is the wage distribution that would have been observed under the wage structure of group 0, but with the distribution of observed and unobserved characteristics of group 1. Accordingly, the unconditional (on X) distributions are denoted as F_1, F_0, and F_c. We analyze the distributional wage gap between groups 0 and 1 using some distributional function. Following Firpo, Fortin, and Lemieux (2007), we denote ν as a function of the conditional joint distribution of $\left(ln(w^1), ln(w^0)\right)|D$, i.e. $\nu : F_\nu \to \mathbb{R}$, where F_ν belongs to a class of distribution functions that satisfy $\forall\ F \in F_\nu$ and $\parallel \nu(F) \parallel < +\infty$. Under this specification, the distributional wage gap between two groups can be written in terms of ν:

$$\triangle_O^\nu = \nu\left(F_1\right) - \nu\left(F_0\right) = \nu_1 - \nu_0 \tag{3.2}$$

We can then further decompose equation 3.2 into two parts, given that X is not evenly distributed across the two groups:

$$\triangle_O^\nu = \left(\nu_1 - \nu_C\right) + \left(\nu_C - \nu_0\right) = \triangle_S^\nu + \triangle_X^\nu \tag{3.3}$$

where the first term $\triangle_S^\nu$ reflects the wage structure effect, meaning the effect caused by changing $g_1\left(\cdot,\cdot\right)$ to $g_0\left(\cdot,\cdot\right)$ while holding characteristics $\left(X,\varepsilon\right)|D = 1$ constant. The other term $\triangle_X^\nu$ indicates the composition effect, which is the effect from changing the distribution of characteristics from $\left(X,\varepsilon\right)|D = 1$ to $\left(X,\varepsilon\right)|D = 1$, while keeping the "wage structure" $g_0\left(\cdot,\cdot\right)$ constant.

3.2.2 Oaxaca-Blinder Decomposition as a Special Case

With such settings, we can include Oaxaca-Blinder decomposition as a special case, where the ν function is the mathematical expectation $\mathbb{E}$. Under the assumption that the conditional expectation takes linear form, we have

$$\mathbb{E}\left[ln(w_i^D)|X\right] \equiv X_i\beta_D, \quad for\ D = 0,1 \tag{3.4}$$

Then the expected wage gap between the "treated" and untreated group, $\triangle_O^\mu$, can be written as

$$\begin{aligned}
\triangle_O^\mu &= \mathbb{E}_x[\mathbb{E}[(ln(w)|X, D = 1)] - \mathbb{E}_x[\mathbb{E}[(ln(w)|X, D = 0)] \\
&= \mathbb{E}[ln(w)|D = 1] - \mathbb{E}[ln(w)|D = 0] \\
&= \mathbb{E}[X|D = 1]\beta_1 - \mathbb{E}[X|D = 0]\beta_0 \\
&= \mathbb{E}[X|D = 1](\beta_1 - \beta_0) + (\mathbb{E}[X|D = 1] - \mathbb{E}[X|D = 0])\beta_0 \\
&\equiv \triangle_S^\mu + \triangle_X^\mu
\end{aligned}$$

The second line follows from the Law of Iterated Expectations. Note that the decomposition in the fourth line uses group 1 as the base group. The counterfactual outcome indicates the mean wage that would have been observed under the wage structure of group 0, but with X from group 1, can directly be computed by $\mathbb{E}[X|D = 1]\beta_0$, which is the counterpart of ν_C in equation 3.3. $\triangle_S^\mu$ is the mean wage structure effect and accordingly $\triangle_X^\mu$ stands for the mean composition effect. Oaxaca-Blinder decomposition is very appealing empirically due to its ease of estimation and interpretation. However, as Barsky et al. (2002) pointed out, consistent estimates of both effects rely on the assumption of the linear structure, which is restrictive. Moreover, Kline (2011) showed that the counterfactual mean identified by the Oaxaca-Blinder method constitutes a propensity score reweighting estimator based upon a linear model for the conditional odds of being treated. Therefore, Oaxaca-Blinder decomposition is indeed a special linear case of propensity score reweighting method. By applying the reweighting method generally, we impose less structure and hence lead to more robust inference.

3.3 Empirical Methodology

3.3.1 A Metric Entropy Measure of the Wage Gap

A comparison of means is implicitly based on a welfare/weighting function that is additive and attaches equal weight to each wage earner. Among others, this implicit welfare function imposes infinite substitutability. Assessment at the median, or any other quantile is justified by even more radical welfare weighting schemes. To overcome these limitations we choose more general distributional functions that could summarize information along the whole distribution. Several commonly used information-based entropy measures such as Shannon's entropy and Kullback-Leibler relative entropy are good candidates for such distributional functions. They are well analyzed in the field of income inequality where the corresponding welfare functions are identified. For instance, an axiomatic approach to "ideal" inequality measures, equivalently welfare functions, or risk averse utility functions, renders the class of Generalized Entropy as ideal. Further additive decomposition requirements render Shannon's entropy, and Theil's measures of inequality as "best". (For example, see Bourguignon (1979), Shorrocks (1978), and Maasoumi (1986)) Inequality measures are divergence measures between any distribution and a uniform (rectangular) size distribution representing perfect equality. The latter is eliminated when the difference between the "inequalities" of two wage distributions is computed. However, entropy divergence measures are generally not metric since they violate the triangular inequality. Hence they are not proper measures of distance. This paper uses a metric entropy measure S_ρ proposed by Granger, Maasoumi, and Racine (2004) as the specific distributional ν function, which is a normalization of the "Bhattacharya-Matusita-Hellinger" measure of distance. It is the one member of the Generalized Entropy family that is a metric. It is given by

$$S_\rho = \frac{1}{2}\int_{-\infty}^{\infty}(f_1^{\frac{1}{2}} - f_0^{\frac{1}{2}})^2 dy \tag{3.5}$$

This measure has several desirable properties: 1. it is well defined for both continuous and discrete variables;[1] 2. it is normalized to 0 if two distributions are equal and lies between 0 and 1; 3. it satisfies the properties of a metric and hence is a true measure of distance; and 4. it is invariant under continuous and strictly increasing transformation on the underlying variables. Note that the natural log of earnings is used through out this paper. Since the logarithm is a strictly increasing function, the findings of this paper are invariant whether using the raw wages or the log form. Following Granger, Maasoumi, and Racine (2004) and Maasoumi and Racine (2002), we consider a kernel based nonparametric implementation of the entropy measure shown in equation 3.5.

3.3.2 Stochastic Dominance

Using S_ρ as the distributional distance measure, we can estimate the distance between original wage distribution and counterfactual wage distribution, and thus the distributional wage structure and composition effects. However, this analysis is still subjective as it would reflect the social welfare based on the generalized entropy function.

In order to compare the different wage distributions robustly, and relative to large classes of welfare functions, we need to examine Stochastic Dominance rankings. First order Stochastic Dominance corresponds to a class (denoted as U_1) of all (increasing) von Neumann-Morgenstern type of social welfare functions u such that welfare is increasing in wages (i.e. $u' > 0$), and the second order Stochastic Dominance test corresponds to the class of social welfare functions in U_1 such that $u'' \leq 0$ (i.e. concavity), denoted as U_2. Concavity implies an aversion to higher dispersion (or inequality) of wages across workers. In this paper, we focus on the one-dimensional social welfare function of only earnings.

Case 1. First Order Dominance: Incumbent employee wage distribution First Order Stochastically Dominates newly hired employee wage distribution (denoted as $ln(w^0)$ FSD $ln(w^1)$) if and only if

1. $E[u(ln(w^0))] \geq E[u(ln(w^1))]$ for all $u \in U_1$ with strict inequality for some u;
2. Or, $F_0(y) \leq F_1(y)$ for all y with strict inequality for some y.

[1] For discrete variables, $S_\rho = \frac{1}{2}\sum(p_1^{1/2} - p_0^{1/2})$.

Case 2. Second Order Dominance: Incumbent wage distribution Second Order Stochastically Dominates newly hired employee wage distribution (denoted as $ln(w^0)$ SSD $ln(w^1)$) if and only if

1. $E[u(ln(w^0))] \geq E[u(ln(w^1))]$ for all $u \in U_2$ with strict inequality for some u;
2. Or, $\int_{-\infty}^{y} F_0(t)dt \leq \int_{-\infty}^{y} F_1(t)dt$ for all y with strict inequality for some y.

The stochastic dominance tests used in this paper are based on a generalized Kolmogorov-Smirnov test as discussed in Linton, Maasoumi, and Whang (2005). The test statistics for FSD and SSD are given by

$$d = \sqrt{\frac{N_0 N_1}{N_0 + N_1}} \min\{\sup[F_0(y) - F_1(y)],\ \sup[F_1(y) - F_0(y)]\} \quad (3.6)$$

$$s = \sqrt{\frac{N_0 N_1}{N_0 + N_1}} \min\{\sup \int_{-\infty}^{y} [F_0(t) - F_1(t)]dt,\ \sup \int_{-\infty}^{y} [F_1(t) - F_0(t)]dt\} \quad (3.7)$$

When we report the empirical test results in Section 5, we denote $\sup[F_0(y) - F_1(y)]$ as $d_{1,max}$ and $\sup[F_1(y) - F_0(y)]$ as $d_{2,max}$. We report both $d_{1,max}$ and $d_{2,max}$ along with the test statistic d for clarity of interpretation. $s_{1,max}$ and $s_{2,max}$ are similarly defined. In empirical applications, the CDFs are replaced with their empirical counterparts. The empirical CDFs are given by $\widehat{F_d(y)} = \frac{1}{N_d} \sum_{i=1}^{N_d} I(ln(w_i^d) \leq y),\ d = 0, 1,$ where $I(\cdot)$ is an indicator function. The underlying distribution of the test statistics are generally unknown and depend on the data. Following Maasoumi and Heshmati (2000), simple bootstrap technique based on 199 replications are employed to obtain the empirical distribution of the test statistics.

3.3.3 Identification of the Counterfactual Distributions

The fundamental question this paper addresses is to identify the wage structure and composition effects through the identification of counterfactual wage distributions, and determine which effect dominates the wage differential. We consider the following counterfactual situation: holding the human capital characteristics of the newly hired workers constant, if we change their wage structure to the wage structure of the incumbents, would the counterfactual wage distribution be different from the original one? If so, would the counterfactual wage distribution stochastically dominate the original one in terms of welfare? If we find such dominance in the first or second order, we would conclude that the wage structures are different and the internal wage structure is better. Similarly, we could also check whether the wage gap is due to differences in human capital characteristics by changing the distribution of the newly hired employees' characteristics to that of the incumbents, holding their wage structure unchanged. To conduct such counterfactual analysis, we follow the propensity score reweighting methods as discussed in Firpo (2007) to identify the counterfactual distributions mentioned above. Simple bootstrap with replacement is applied to obtain the statistical significance of the dominance tests. Specifically, we want to identify the distributions of the following two counterfactual outcomes:

$$ln(w_i^{c1}) = g_0(X_i, \varepsilon_i)|D = 1 \qquad \text{(counterfactual outcome \#1)} \tag{3.8}$$

$$ln(w_i^{c2}) = g_1(X_i, \varepsilon_i)|D = 0 \qquad \text{(counterfactual outcome \#2)} \tag{3.9}$$

The benchmark outcome we considered is the conditional wage distribution of the new hires $ln(w_i^1) = g_1(X_{i1}, \varepsilon_{i1})$. The counterfactual outcome #1, $ln(w_i^{c1})$, indicates the hypothetical unobserved wage of the newly hired employees if they were paid under the incumbent wage structure. Comparing the benchmark wage $ln(w_i^1)$ to $ln(w_i^{c1})$ using the S_ρ measure would yield the wage structure effect $\triangle_S^\nu$ as in equation 3.3. Comparing distributional distance between $ln(w_i^{c1})$ and the incumbent wage $ln(w_i^0)$ would give us the composition effect, denoted by $\triangle_X^\nu$ in equation 3.3. However, that would require using $ln(w_i^0)$ as the benchmark. For ease of interpretation, we choose to use $ln(w_i^1)$ as the benchmark when identifying both effects and construct counterfactual outcome #2 $ln(w_i^{c2})$, which indicates the hypothetical wage of the newly hired employees if they had the characteristics of the incumbents. With the new hire wage as the benchmark, S_ρ measure of $ln(w_i^1) - ln(w_i^{c2})$

gives the composition effect that are reported in empirical applications in Section 5.

Firpo (2007) proved that under certain assumptions, such counterfactual distributions are identified. Following Firpo (2007) and Firpo, Fortin, and Lemieux (2007), we make similar assumptions:

1. Unconfoundedness: Suppose (Y, D, X) have a joint distribution, where Y is the outcome and D is a dummy indicating treatment: (Y_1, Y_0) and D are jointly independent conditional on $X = x$.

2. Common support: For all $x \in X$, $0 < Pr\{D = 1|X = x\} := p(x) < 1$.

Assumption 1 means that fixing the values of observable human capital characteristics X, the distribution of the wage outcome or the error term ε is independent of whether one is incumbent or newly hired. Assumption 2 rules out the possibilities that some specific x belongs only to either one of the two worker groups and hence such x can predict the probability of being treated perfectly.

The counterfactual distribution of $ln(w_i^{c1})$ could be identified by the following propensity score reweighting methods discussed in Firpo (2007).

$$F_{c1} = \mathbb{E}[\omega_{c1}(D_1, X) \cdot I[(ln(w_i) \leq y)] \tag{3.10}$$

where $\omega_{c1}(D_1, X) = \left(\frac{p_1(x)}{1-p_1(x)}\right)\left(\frac{1-D_1}{p_1}\right)$, D_1 is a treatment dummy variable taking the value of 1 for incumbent employees, $p_1(x) = Pr\{D_1 = 1|X = x\}$ is the propensity score, $p_1 = Pr\{D_1 = 1\} = \mathbb{E}[p_1(X)]$ is the marginal probability of being treated. In practice, we will estimate the propensity score parametrically using a logit model.[2] Applying the weights $\omega_{c1}(D_1, X)$ gives us the counterfactual distribution of $ln(w_i^{c1})$. Identifying the distribution of counterfactual outcome $ln(w_i^{c2})$, F_{c2}, is similar, but we need to take newly hired employees as the treated group. Let D_2 be the treatment dummy taking the value of 1 for new hires.

[2]Nonparametric kernel regression can also be used to estimate the propensity score, which allows more flexible dependence relations among independent and dependent variables.

$$F_{c2} = \mathbb{E}\left[\omega_{c2}(D_2, X) \cdot I[(ln(w_i) \leq y)\right] \tag{3.11}$$

where $\omega_{c2}(D_2, X) = \left(\frac{p_2(x)}{1-p_2(x)}\right)\left(\frac{1-D_2}{p_2}\right)$, $p_2(x)$ and p_2 are similarly defined as in the previous case.[3] Applying the weights $\omega_{c2}(D_2, X)$ gives us the counterfactual distribution of $ln(w_i^{c2})$. Once we identify the counterfactual distributions of interest, we can then perform stochastic dominance tests to compare those counterfactual distributions with the original distribution.

3.4 Data

The data used in this paper come from 1996-2012 monthly Current Population Survey (CPS). The monthly CPS is a survey of a probability sample of housing units. Although the CPS is designed to be a cross-sectional survey, it does not survey a completely new set of housing units every month. The sample is divided into eight representative rotation groups. Therefore, a typical housing unit in the sample is interviewed in 8 different months, given no attrition during survey period. If a housing unit is randomly selected into monthly CPS for the first time, it will be interviewed for four consecutive months, followed by an 8-month break, and then be surveyed for another four consecutive months. The rotation group could be identified by the CPS variable "month in sample" (MIS).

The CPS sample design actually allows us to longitudinally link a household in sample over 8 different months. Following methods as discussed in Madrian and Lefgren (1999), we conducted one-month matching for all the eligible rotation groups (MIS = 2-4 or 6-8) in each monthly sample, i.e. linking those eligible subsamples with their previous month observations.[4] In our sample, the matching rate for those eligible groups is over 90 percent on average.[5] Using this longitudinally linked data set, we could identify the incumbent and

[3]Note that under such setting, $p_1(x) + p_2(x) = 1$ and $p_1 + p_2 = 1$.

[4]In 1995, the Census made some changes to CPS sample ID variable, which leads to very poor matching rates for that year, so we chose 1996 as the starting year to circumvent the problem.

[5]One shortcoming of this linked data is that we can only follow workers who remain in the same household. Thus any new hire that moved in order to take a new job could not be matched in this data set.

newly hired employees. Since we are interested in the wage differentials, we first restrict our sample to individuals of working age, i.e. those who aged from 18 to 64. Then we keep those who remained full-time employed (35 hours per week or above) in both month $t-1$ and month t. Among those full-time workers, we define incumbent workers to be those who stayed with the same employer from month $t-1$ to t. The newly hired employees are defined to be those who changed their employer from month $t-1$ and t, i.e. workers that switched to a new job with a new employer at time t.[6]

Following the literature (e.g. Maasoumi and Wang (2013)), we use the log of hourly wages, measured by an individual's weekly wage income divided by the number of hours worked per week. Note that, as we mentioned above, the metric entropy measure of the wage differential and stochastic dominance tests are invariant to the logarithm transformation, while many conventional measures are not. The observed human capital variables used in the counterfactual analysis include age, age squared, gender, education (five education groups: less than high school, high school, some college, college, graduate), marital status, ethnicity and region (Northeast, Midwest, South and West). Occupation variable is grouped into three categories: high-skill (managerial and professional occupations); medium-skill (technician, technical production, sales, and administrative support occupations); and low-skill(other occupations such as maintenance, construction, and farming occupations).

3.5 Results

3.5.1 Baseline Analysis

Trend of the Wage Differential between Internal and External Labor Markets

Table 3.1 shows various measures of the log wage differences between the incumbent and newly hired employees, i.e. $ln(w^0) - ln(w^1)$. The second column in the table reports the distributional measure of the wage gap S_ρ. Since S_ρ is a normalized metric taking on values between 0 and 1, for easy interpretation we report the original results multiplied by 100

[6] We also exclude those with hourly wage less than or equal to 1 dollar, because those extremely low wages are likely be due to misreporting.

throughout the paper. Under the null hypothesis of no difference between incumbent and new hire wage distributions, we calculate the statistical significance of the $S_\rho$measure using 199 simple bootstrap replications. The p values are reported in the third column. The other columns in Table 3.1 report conventional measures (e.g. mean, median and quantiles at various levels) of earning differentials commonly used in the literature. We can see that both the traditional measures and the metric entropy measure S_ρ imply that there exists wage differentials between those two groups of workers for all the years in sample. The distributional distances are statistically significant at 5% in 2011 and at 1% in remaining years. The mean differences and all the quantile differences except for the 90th quantile in 2010, are all positive, clearly showing wage gaps that in favor of the incumbent employees. However, it is hard to tell a clear trend over time for any of these measures and even harder to tell whether our new measure shows a different pattern of the time trend from other traditional measures.

[Insert Table 3.1 about here]

The S_ρ measure and other conventional measures are not directly comparable. Thus, to enable easy comparisons, we normalize all these measures by setting the value in the year of 1996 to 100 and computing the normalized values. The plot of these normalized values of S_ρ, mean, median, 25th and 75th percentiles in Figure 3.1. As shown in the graph, other than the 75th percentile, the traditional measures display similar time trends as the S_ρ entropy measure. In order to check how the wage differentials fit with macro business cycles, we plot the recession periods with shaded vertical bars in the figure. During the sample period, Mar 2001 to Nov 2001 and Dec 2007 to Jun 2009 are considered as recession periods by the NBER. Since our measures are computed at yearly frequency, we roughly pick 2001, 2008 and 2009 as the recession years and the three years are indicated by the shaded bars in Figure 1. The line plots do not show very clear cyclical patterns, but all measures, except the 75th percentile of wage differentials do seem to increase during the recent great recession period from 2008 to 2009. During the great recession the level of payroll employment fell by 5.4%, more than four times the employment decline faced in the 2001 recession.[7] Many firms reduced or halted hiring, which reduced the bargaining power of those job seekers. So the newly hired employees may have had to accept lower wages, which increases the wage gaps between the incumber and the newly hired workers.

[Insert Figure 3.1 about here]

Stochastic Dominance Test Results

As discussed above, these measures of the gender gap could not give a clear ranking of the earnings distributions in terms of social welfare. Therefore, in Table 3.2 we present the stochastic dominance test results. The second column labeled Observed Ranking details if the distributions can be ranked in either the first or second order, where FSD is short for First-order Stochastic Dominance and SSD stands for Second-order Stochastic Dominance. The columns labeled $Pr[d \leq 0]$ and $Pr[s \leq 0]$ report the probabilities of the test statistics (of the first and second order dominance tests respectively) to be non-positive based on the simple bootstrap with replacement for 199 replications. The probability serves a similar role as p-values in any hypothesis test, but the interpretation is reversed. For example, if we observe FSD (SSD) and $Pr[d \leq 0]$ ($Pr[s \leq 0]$) is 0.95, then it means that the test statistic is statistically significant at 5% level (p-value=0.05).

[Insert Table 3.2 about here]

From Table 3.2, we can see that the wage distribution of incumbents lies predominantly to the right of the wage distribution of new hires, meaning that incumbent workers enjoy higher level of wages. For all the years in sample, we find stochastic dominance relations either in the first or second order. In 4 out of 17 years (1996, 2004, 2007 and 2008), we find the wage distribution of incumbent workers to empirically dominates, in a first-order sense, the wage distribution among newly hired workers, but such dominance relation is not statistically significant in any of the 4 years. For the remaining years, highly significant

[7] Authors' calculation; Source: BLS, Haver Analytics

second-order dominance is found in the years of 1997, 2000-2002 and 2008, with confidence level greater than 0.95. This suggests that any worker with a social welfare function in the class U_2 (increasing and concave in wage) would prefer the incumbent distribution to the new hire distribution in those 5 years. Such dominance ranking is only possible when we account for an aversion to higher dispersion in the welfare criteria. This finding is quite interesting because those significant second-order dominance cases mainly occurred around recession periods (2001 and 2008 are recession years). Second-order dominance indicates that starting from the very left tail of the wage distribution, incumbent workers are better paid than newly hired workers at most quantiles. This is in line with the findings of Oreopoulos, Von Wachter, and Heisz (2006), which finds that young graduates entering the labor market in a recession suffer significant initial earnings losses. SSD also suggests that at the far right tail of the wage distribution, some newly hired workers could be paid better than their incumbent counterparts. One possible explanation could be the differences in human capital characteristics. Those who managed to find highly paid jobs during a recession may have very strong human capital characteristics. We will further test this hypothesis using counterfactual analysis in a latter section.

3.5.2 Counterfactual Analysis

Table 3.3 reports the estimated wage structure effect, i.e. the wage gap caused by the inequality in the pay structure. Metric entropy and traditional measures of the log wage differences between the newly hired employees and their counterfactual outcome #1, i.e. $ln(w^1) - ln(w^{c1})$, are presented. The p values of S_ρ measure are calculated using the same bootstrap method as applied in Table 3.1. From Table 3.3, we can see that most means and quantiles in almost all years except for 2011 are negative, which means that the counterfactual wages under the incumbent's wage structure while keeping new employee characteristics unchanged are generally better than actual wages those new hires earn. The distributional distance measured by S_ρ is smaller and less significant than the distance between $ln(w^0)$ and $ln(w^{c1})$, as it only reflects the wage structure effect.

[Insert Table 3.3 about here]

Table 3.4 reports dominance test results of the actual wage distribution of the new hires versus the counterfactual wage distribution #1. Recall that this comparison identifies the difference of the wage structures between the external and internal labor markets. Any finding of stochastic dominance indicates the inequality in the pay structure instead of the differences in human capital characteristics. We find the counterfactual wage distribution #1 SSD the original wage distribution of the new hires for all the years in sample, except for the year of 2011, which means that if the newly hired workers were paid under incumbent wage structure, such outcomes are preferred at least for those with social welfare functions in the class of U_2. As indicated by the bootstrapped probabilities, those dominance relations are statistically significant in 1997, 2006 and 2008, with confidence level greater than 0.9 and are close to significant in 2000 and 2001. Second-order dominance indicates that such findings holds mainly at the lower tail of the wage distribution, while at the upper tail, the wage structure of those new hires may actually be better than those of the incumbents, so such counterfactual wages may be lower than their actual wages for those highly paid new employees. To further test this finding, in the following subsection we divide our sample into two sub groups, higher and lower wage groups, and conducted counterfactual analysis respectively.

[Insert Table 3.4 about here]

Table 3.5 reports the estimated composition effect, i.e. the wage gap caused by the differences human capital characteristics. S_ρ and conventional measures of the log wage differences between the newly hired employees and their counterfactual outcome #2, i.e. $ln(w^1) - ln(w^{c2})$, are reported. From the table we can see that all the means and quantiles in all the years in the sample are negative, which indicates that the counterfactual wages under the incumbent characteristics while keeping new hire's wage structure unchanged are generally better than actual wages of the new hires. We conclude that the differences in human capital characteristics between the incumbents and new hires also contributed to the their wage gap. The distributional distance measured by S_ρ is a little smaller and less significant than those reported in Table 3.3, which means that the estimated composition effect is smaller compared to the estimated wage structure effect.

[Insert Table 3.5 about here]

Table 3.6 reports the stochastic dominance test results from the comparison between the actual wage distribution of the new hires versus the counterfactual wage distribution #2. Note that this comparison identifies the wage gap caused by differences in human capital characteristics. As shown in the table, we find the counterfactual wage distribution #2 FSD the actual new hire wage distribution in all year. However, such first-order dominance relations are not statistically significant. FSD always indicates SSD, but those second-order dominance relations are largely insignificant as well. Hence we have found some evidence for differences in human capital characteristics, but the evidence is not quite strong. The data seem to tell us that even though there is some difference in human capital between incumbent and newly hired workers, such a difference is not large enough to be statistically meaningful.

[Insert Table 3.6 about here]

3.5.3 Counterfactual Analysis of Different Wage Groups

In this section, we report the findings of counterfactual analysis for different wage groups. We used the weighted median wage of our sample, \$18.5 per hour, as the cut-off point. Higher wage group consists of workers with wages above the median, and the rest are in the lower wage group.

Counterfactual Analysis of Higher Wage Group

We conducted the two kinds of counterfactual analysis again for the higher wage workers. The findings are reported in Tables 3.7 and 3.8. In line with Table 3.6, Table 3.8 also indicates a first-order distributional wage premium of human capital characteristics in favor of the incumbents. But still, among higher paid workers, such an edge is also not statistically significant, neither is the second-order dominance relation significant for any year in sample. We have our most interesting findings in Table 3.7, which summarize the wage gap caused by inequality in wage structure for higher paid workers.

[Insert Table 3.7 about here]

In many years, we find that the newly hired worker's wage distribution and the counterfactual wage distribution # 1 are generally unrankable. However, we do find second-order dominance relations in the years of 1997, 2000, 2001, 2003, 2008, 2010 and 2011. More notably, the dominance relations reversed direction. The wage distribution of newly hired workers empirically dominates, in a second-order sense, the counterfactual wage distribution # 1. Although they are largely statistically insignificant, the reverse of the dominance relations, to some degree, confirmed our hypothesis that certain highly paid new workers actually enjoyed a better wage structure than their incumbent counterparts, the so called "new hire premium" in the literature. For workers with a social welfare function in the class U_2, the counterfactual case that replace new hire's wage structure with that of incumber workers, while keeping their characteristics constant, would actually make those new hires worse off.

[Insert Table 3.8 about here]

Counterfactual Analysis of Lower Wage Group

The results of counterfactual analysis for the lower paid group are reported in Table 3.9 and Table 3.10. Table 3.10 shows similar results as that in Table 3.6, indicating better human capital characteristics among incumbent workers with hourly wage lower than $18.5.

[Insert Table 3.9 about here]

Table 3.9 reports the stochastic dominance test results between the original wage distribution of the new hires and the counterfactual wage distribution #1 among lower paid group. We have some interesting findings here. In the years of 1996, 1997, 2000-2003 and 2005-2009, the counterfactual wage distribution #1 empirically dominates, in a first-order sense, the wage distribution of newly hired workers. In the years of 1998, 2004, 2010 and 2011, the counterfactual wage distribution #1 empirically dominates, in a second-order sense, the wage distribution of newly hired workers. First-order dominance relation is largely insignificant, but in 1997 and 2008, the second-order dominance relations are statistically significant, with p-values less than 0.1. The findings indicate that for lower wage workers, the counterfactual wage distribution #1 are preferred compared to the actual new hire wage distribution for workers with a social welfare function in the class of U_2 in both years. The significant dominance relation in 2008, provides a strong evidence that during the recent great recession year, lower wage new hired workers suffer from a much worse pay structure than that of the incumbents. It is an indicator showing that the external labor market deteriorates much more than the internal labor market during the recent recession.

[Insert Table 3.10 about here]

3.6 Conclusion

This paper employs a distribution based entropy metric to measure the wage differentials between incumbent and newly hired employees. The entropy measure incorporates the differences at the entire distribution level and thus gives a better picture on wage comparison. We also use stochastic dominance tests to rank those wage distributions based on social welfare. We find that the incumbent workers are generally paid better than the newly hired worker in any year from 1996 to 2012. Further counterfactual analysis shows that the wage gap could be attributed to both the inequality in wage structures and the differences in human capital characteristics, depending on a worker's wage level. For highly paid new workers, the wage gap mainly comes from the differences in human capital characteristics and those new hires tend to enjoy a better wage structure than the incumbents in certain years. For lower paid new workers, the wage differential comes from both gap in human capital characteristics and the inequality in wage structure. Especially in the recent recession year 2008, those lower wage new hires suffer more from the significantly worse wage structure than that of the incumbents.

References

Albrecht, J., A. Björklund, and S. Vroman. 2003. Is there a glass ceiling in sweden? *Journal of Labor Economics* 21:145–77.

Baker, G., M. Gibbs, and B. Holmstrom. 1994. The internal economics of the firm: Evidence from personnel data. *The Quarterly Journal of Economics* 109:881–919.

Barsky, R., J. Bound, K. K. Charles, and J. P. Lupton. 2002. Accounting for the black–white wealth gap: a nonparametric approach. *Journal of the American Statistical Association* 97:663–73.

Blau, F. D., and L. M. Kahn. 2006. The us gender pay gap in the 1990s: Slowing convergence. *Industrial & Labor Relations Review* 60:45–66.

Blinder, A. S. 1973. Wage discrimination: reduced form and structural estimates. *Journal of Human Resources* 8:436–55.

Bourguignon, F. 1979. Decomposable income inequality measures. *Econometrica: Journal of the Econometric Society* 47:901–20.

DiNardo, J., N. M. Fortin, and T. Lemieux. 1996. Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica* 64:1001–44.

Doeringer, P., and M. Piore. 1985. *Internal labor markets and manpower analysis.* ME Sharpe (Armonk, NY).

Firpo, S. 2007. Efficient semiparametric estimation of quantile treatment effects. *Econometrica* 75:259–76.

Firpo, S., N. Fortin, and T. Lemieux. 2007. Decomposing wage distributions using recentered influence function regressions. *University of British Columbia, Working Paper* .

Granger, C., E. Maasoumi, and J. Racine. 2004. A dependence metric for possibly nonlinear processes. *Journal of Time Series Analysis* 25:649–69.

Kampkötter, P., and D. Sliwka. 2011. Wage premia for newly hired employees: theory and evidence. *IZA discussion paper* .

Kline, P. 2011. Oaxaca-blinder as a reweighting estimator. *The American Economic Review* 101:532–37.

Linton, O., E. Maasoumi, and Y. Whang. 2005. Consistent testing for stochastic dominance under general sampling schemes. *The Review of Economic Studies* 72:735–65.

Maasoumi, E. 1986. The measurement and decomposition of multi-dimensional inequality. *Econometrica: Journal of the Econometric Society* 54:991–7.

Maasoumi, E., and A. Heshmati. 2000. Stochastic dominance amongst swedish income distributions. *Econometric Reviews* 19:287–320.

Maasoumi, E., and J. Racine. 2002. Entropy and predictability of stock market returns. *Journal of Econometrics* 107:291–312.

Maasoumi, E., and L. Wang. 2013. The gender earnings gap: Measurement and analysis. *Emory University, Working Paper* .

Madrian, B. C., and L. Lefgren. 1999. A note on longitudinally matching current population survey (cps) respondents. Technical Working Paper Series 247, National Bureau of Economic Research.

Oaxaca, R. 1973. Male-female wage differentials in urban labor markets. *International Economic Review* 14:693–709.

Oreopoulos, P., T. Von Wachter, and A. Heisz. 2006. The short- and long-term career effects of graduating in a recession: Hysteresis and heterogeneity in the market for college graduates. Working Paper Series 12159, National Bureau of Economic Research.

Shorrocks, A. F. 1978. The measurement of mobility. *Econometrica: Journal of the Econometric Society* 46:1013–24.

Figure 3.1: Time Trend of Wage Differential with Business Cycle Indicator

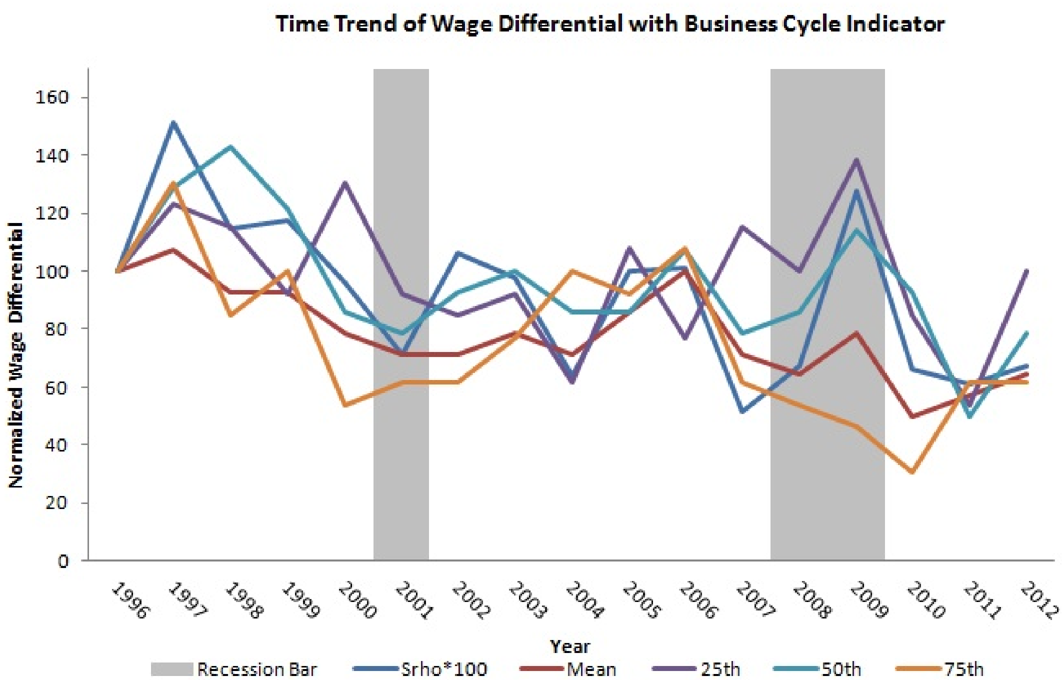


Table 3.1: Entropy Measures of Wage Differentials

Year	$S_\rho \times 100$	p of S_ρ	mean	10th	25th	50th	75th	90th
1996	0.97	(0.00)	0.14	0.09	0.13	0.14	0.13	0.11
1997	1.47	(0.00)	0.15	0.13	0.16	0.18	0.17	0.08
1998	1.11	(0.00)	0.13	0.08	0.15	0.20	0.11	0.04
1999	1.14	(0.00)	0.13	0.15	0.12	0.17	0.13	0.07
2000	0.93	(0.00)	0.11	0.07	0.17	0.12	0.07	0.07
2001	0.69	(0.00)	0.10	0.07	0.12	0.11	0.08	0.04
2002	1.03	(0.00)	0.10	0.11	0.11	0.13	0.08	0.04
2003	0.95	(0.00)	0.11	0.13	0.12	0.14	0.10	0.04
2004	0.62	(0.00)	0.10	0.06	0.08	0.12	0.13	0.07
2005	0.97	(0.00)	0.12	0.06	0.14	0.12	0.12	0.12
2006	0.98	(0.00)	0.14	0.09	0.10	0.15	0.14	0.14
2007	0.50	(0.00)	0.10	0.09	0.15	0.11	0.08	0.06
2008	0.65	(0.01)	0.09	0.06	0.13	0.12	0.07	0.03
2009	1.24	(0.00)	0.11	0.11	0.18	0.16	0.06	0.09
2010	0.64	(0.00)	0.07	0.10	0.11	0.13	0.04	-0.08
2011	0.59	(0.03)	0.08	0.05	0.07	0.07	0.08	0.09
2012	0.65	(0.00)	0.09	0.06	0.13	0.11	0.08	0.05

Notes: Columns (2)-(3) report metric entropy measure of distributional distance and its p values respectively. The p values are obtained from 199 simple bootstrap under the null hypothesis of no difference between incumbent and new hire wage distributions.

Table 3.2: Stochastic Dominance Test Results

Year	Obrank	$d_{1,max}$	$d_{2,max}$	d	$\Pr(d \leq 0)$	$s_{1,max}$	$s_{2,max}$	s	$\Pr(d \leq 0)$
1996	FSD	-0.03	5.15	-0.03	0.30	-0.04	104.05	-0.04	0.94
1997	SSD	0.06	6.30	0.06	0.25	-0.05	110.07	-0.05	0.99
1998	SSD	0.11	5.47	0.11	0.04	-0.05	98.14	-0.05	0.87
1999	SSD	0.06	5.78	0.06	0.19	-0.05	93.93	-0.05	0.81
2000	SSD	0.11	4.96	0.11	0.05	-0.06	86.00	-0.06	0.96
2001	SSD	0.01	4.39	0.01	0.43	-0.06	80.49	-0.06	0.97
2002	SSD	0.24	4.48	0.24	0.01	-0.06	76.70	-0.06	0.99
2003	SSD	0.07	4.20	0.07	0.14	-0.05	71.31	-0.05	0.87
2004	FSD	-0.06	4.27	-0.06	0.41	-0.06	68.43	-0.06	0.79
2005	SSD	0.09	4.54	0.09	0.18	-0.09	73.32	-0.09	0.92
2006	SSD	0.08	4.62	0.08	0.21	-0.09	95.11	-0.09	0.94
2007	FSD	-0.06	3.88	-0.06	0.55	-0.07	67.61	-0.07	0.87
2008	FSD	-0.03	4.16	-0.03	0.48	-0.08	55.20	-0.08	0.98
2009	SSD	0.38	3.56	0.38	0.00	-0.04	61.14	-0.04	0.76
2010	SSD	0.45	2.76	0.45	0.01	-0.06	47.83	-0.06	0.93
2011	SSD	0.07	2.13	0.07	0.09	-0.07	35.03	-0.07	0.54
2012	SSD	0.01	3.34	0.01	0.21	-0.06	47.96	-0.06	0.64

Table 3.3: Measures of Differences between New Hire and New Hire counterfactual #1 Distributions

Year	$S_\rho \times 100$	p of S_ρ	mean	10th	25th	50th	75th	90th
1996	0.34	(.03)	-0.06	-0.05	-0.07	-0.06	-0.05	-0.02
1997	0.77	(.00)	-0.07	-0.09	-0.11	-0.10	-0.08	0.00
1998	0.47	(.01)	-0.05	-0.04	-0.06	-0.09	-0.03	0.04
1999	0.47	(.01)	-0.06	-0.07	-0.06	-0.09	-0.04	0.00
2000	0.46	(.01)	-0.03	-0.07	-0.11	-0.04	0.00	0.04
2001	0.32	(.05)	-0.03	-0.00	-0.04	-0.04	-0.01	0.05
2002	0.47	(.00)	-0.03	-0.07	-0.08	-0.05	0.00	0.00
2003	0.38	(.06)	-0.02	-0.08	-0.05	-0.05	0.00	0.05
2004	0.23	(.25)	-0.03	-0.05	-0.03	-0.05	-0.04	0.03
2005	0.36	(.04)	-0.04	-0.06	0.00	-0.04	-0.03	-0.03
2006	0.28	(.17)	-0.03	0.00	-0.04	-0.07	-0.02	-0.00
2007	0.19	(.58)	-0.03	-0.05	-0.10	-0.03	-0.02	0.03
2008	0.37	(.17)	-0.03	-0.06	-0.07	-0.05	0.00	0.07
2009	0.63	(.00)	-0.04	-0.07	-0.09	-0.08	0.00	-0.00
2010	0.40	(.11)	-0.00	-0.05	-0.04	-0.03	0.04	0.12
2011	0.47	(.00)	0.01	0.00	0.00	0.04	0.04	0.01
2012	0.33	(.06)	-0.03	-0.02	-0.04	-0.03	0.01	0.00

Notes: Columns (2)-(3) report metric entropy measure of distributional distance and its p values respectively. The p values are obtained from 199 simple bootstrap under the null hypothesis of no difference between the new hire and their counterfactual #1 wage distributions.

Table 3.4: Stochastic Dominance Test Results, Counterfactual #1

Year	Obrank	$d_{1,max}$	$d_{2,max}$	d	$\Pr(d \leq 0)$	$s_{1,max}$	$s_{2,max}$	s	$\Pr(d \leq 0)$
1996	SSD	14.22	0.58	0.58	0.05	240.39	-0.28	-0.28	0.85
1997	SSD	17.43	0.73	0.73	0.03	261.86	-0.25	-0.25	0.96
1998	SSD	16.37	2.39	2.39	0.01	206.92	-0.25	-0.25	0.71
1999	SSD	15.41	1.31	1.31	0.01	201.00	-0.46	-0.46	0.60
2000	SSD	11.39	1.92	1.92	0.00	156.57	-0.34	-0.34	0.88
2001	SSD	8.80	2.64	2.64	0.02	132.64	-0.30	-0.30	0.89
2002	SSD	9.60	2.66	2.66	0.00	136.40	-0.26	-0.26	0.81
2003	SSD	9.20	2.70	2.70	0.00	123.87	-0.26	-0.26	0.72
2004	SSD	8.55	1.16	1.16	0.01	109.38	-0.32	-0.32	0.61
2005	SSD	10.93	1.43	1.43	0.01	143.21	-0.34	-0.34	0.77
2006	SSD	12.29	1.38	1.38	0.01	159.84	-0.45	-0.45	0.91
2007	SSD	9.16	1.75	1.75	0.03	104.39	-0.33	-0.33	0.73
2008	SSD	11.21	3.73	3.73	0.00	113.81	-0.43	-0.43	0.94
2009	SSD	11.15	3.16	3.16	0.00	155.81	-0.28	-0.28	0.63
2010	SSD	5.33	7.22	5.33	0.00	98.05	-0.29	-0.29	0.46
2011	None	3.39	4.66	3.39	0.00	29.92	25.71	25.71	0.16
2012	SSD	8.42	0.85	0.85	0.00	80.50	-0.22	-0.22	0.43

Table 3.5: Measures of Differences between New Hire and New Hire counterfactual #2 Distributions

Year	$S_\rho \times 100$	p of S_ρ	mean	10th	25th	50th	75th	90th
1996	0.27	(0.53)	-0.08	-0.04	-0.07	-0.07	-0.08	-0.08
1997	0.28	(0.59)	-0.08	-0.09	-0.07	-0.10	-0.14	-0.08
1998	0.29	(0.64)	-0.08	-0.04	-0.06	-0.13	-0.10	-0.06
1999	0.28	(0.44)	-0.08	-0.07	-0.03	-0.09	-0.12	-0.11
2000	0.23	(0.48)	-0.08	-0.06	-0.11	-0.05	-0.09	-0.08
2001	0.24	(0.08)	-0.08	0.00	-0.05	-0.10	-0.09	-0.08
2002	0.28	(0.42)	-0.09	-0.05	-0.11	-0.10	-0.10	-0.10
2003	0.23	(0.50)	-0.07	-0.07	-0.05	-0.09	-0.10	-0.07
2004	0.27	(0.66)	-0.08	-0.06	-0.03	-0.11	-0.12	-0.08
2005	0.26	(0.14)	-0.08	-0.06	-0.01	-0.07	-0.10	-0.11
2006	0.29	(0.05)	-0.09	-0.03	-0.04	-0.10	-0.11	-0.12
2007	0.26	(0.26)	-0.08	-0.05	-0.10	-0.09	-0.10	-0.07
2008	0.20	(0.97)	-0.07	-0.03	-0.08	-0.10	-0.11	-0.06
2009	0.25	(0.68)	-0.09	-0.07	-0.09	-0.11	-0.11	-0.12
2010	0.19	(0.80)	-0.08	-0.05	-0.07	-0.09	-0.10	-0.09
2011	0.19	(0.98)	-0.07	-0.04	-0.05	-0.08	-0.09	-0.10
2012	0.20	(0.51)	-0.07	-0.00	-0.06	-0.10	-0.09	-0.10

Columns (2)-(3) report metric entropy measure of distributional distance and its p values respectively. The p values are obtained from 199 simple bootstrap under the null hypothesis of no difference between between the new hire and their counterfactual #2 wage distributions.

Table 3.6: Stochastic Dominance Test Results, Counterfactual #2

Year	Obrank	$d_{1,max}$	$d_{2,max}$	d	$\Pr(d \leq 0)$	$s_{1,max}$	$s_{2,max}$	s	$\Pr(d \leq 0)$
1996	FSD	14.50	-0.30	-0.30	0.21	303.47	-0.30	-0.30	0.47
1997	FSD	14.80	-0.29	-0.29	0.18	303.02	-0.29	-0.29	0.38
1998	FSD	15.24	-0.26	-0.26	0.16	340.35	-0.30	-0.30	0.48
1999	FSD	14.62	-0.45	-0.45	0.21	293.55	-0.45	-0.45	0.51
2000	FSD	13.78	-0.25	-0.25	0.14	285.84	-0.37	-0.37	0.51
2001	FSD	14.62	-0.29	-0.29	0.17	295.95	-0.29	-0.29	0.52
2002	FSD	15.84	-0.33	-0.33	0.36	344.21	-0.41	-0.41	0.77
2003	FSD	13.85	-0.26	-0.26	0.15	301.62	-0.26	-0.26	0.43
2004	FSD	15.42	-0.25	-0.25	0.25	274.34	-0.32	-0.32	0.87
2005	FSD	15.22	-0.22	-0.22	0.31	255.04	-0.27	-0.27	0.58
2006	FSD	15.64	-0.26	-0.26	0.34	375.20	-0.26	-0.26	0.81
2007	FSD	15.40	-0.25	-0.25	0.10	281.80	-0.56	-0.56	0.53
2008	FSD	13.37	-0.34	-0.34	0.11	220.90	-0.46	-0.46	0.47
2009	FSD	13.66	-0.28	-0.28	0.20	297.15	-0.28	-0.28	0.54
2010	FSD	12.28	-0.22	-0.22	0.21	284.13	-0.22	-0.22	0.51
2011	FSD	11.90	-0.39	-0.39	0.22	244.62	-0.42	-0.42	0.59
2012	FSD	12.49	-0.25	-0.25	0.33	216.75	-0.25	-0.25	0.71

Table 3.7: Stochastic Dominance Test Results for High Wage Workers, Counterfactual #1

Year	Obrank	$d_{1,max}$	$d_{2,max}$	d	$\Pr(d \leq 0)$	$s_{1,max}$	$s_{2,max}$	s	$\Pr(d \leq 0)$
1996	None	2.09	3.17	2.09	0.00	0.26	40.10	0.26	0.26
1997	SSD	1.85	3.10	1.85	0.10	-0.30	49.34	-0.30	0.65
1998	None	0.80	6.16	0.80	0.06	0.80	118.62	0.80	0.18
1999	None	0.34	6.09	0.34	0.03	0.14	79.75	0.14	0.38
2000	SSD	1.05	3.50	1.05	0.15	-0.65	80.04	-0.65	0.47
2001	SSD	1.38	5.59	1.38	0.00	-0.39	100.11	-0.39	0.51
2002	None	0.23	4.42	0.23	0.07	0.23	118.22	0.23	0.23
2003	SSD	0.46	4.28	0.46	0.00	-0.61	102.91	-0.61	0.46
2004	None	1.72	5.78	1.72	0.00	3.52	30.94	3.52	0.16
2005	None	0.76	4.63	0.76	0.02	0.66	56.92	0.66	0.14
2006	None	2.59	2.61	2.59	0.00	11.49	38.30	11.49	0.23
2007	None	5.66	2.16	2.16	0.02	21.11	0.63	0.63	0.18
2008	SSD	0.44	7.88	0.44	0.03	-0.18	104.73	-0.18	0.59
2009	None	4.27	6.15	4.27	0.02	10.85	121.45	10.85	0.14
2010	SSD	0.33	11.37	0.33	0.00	-0.83	315.47	-0.83	0.82
2011	SSD	2.01	2.83	2.01	0.00	-0.35	18.44	-0.35	0.30
2012	None	4.34	1.83	1.83	0.00	15.58	0.55	0.55	0.13

Table 3.8: Stochastic Dominance Test Results for High Wage Workers, Counterfactual #2

Year	Obrank	$d_{1,max}$	$d_{2,max}$	d	$\Pr(d \leq 0)$	$s_{1,max}$	$s_{2,max}$	s	$\Pr(d \leq 0)$
1996	FSD	14.50	-0.30	-0.30	0.18	303.47	-0.30	-0.30	0.39
1997	FSD	14.80	-0.29	-0.29	0.12	303.02	-0.29	-0.29	0.37
1998	FSD	15.24	-0.26	-0.26	0.12	340.35	-0.30	-0.30	0.41
1999	FSD	14.62	-0.45	-0.45	0.21	293.55	-0.45	-0.45	0.42
2000	FSD	13.78	-0.25	-0.25	0.16	285.84	-0.37	-0.37	0.54
2001	FSD	14.62	-0.29	-0.29	0.19	295.95	-0.29	-0.29	0.42
2002	FSD	15.84	-0.33	-0.33	0.32	344.21	-0.41	-0.41	0.76
2003	FSD	13.85	-0.26	-0.26	0.19	301.62	-0.26	-0.26	0.49
2004	FSD	15.42	-0.25	-0.25	0.32	274.34	-0.32	-0.32	0.81
2005	FSD	15.22	-0.22	-0.22	0.36	255.04	-0.27	-0.27	0.66
2006	FSD	15.64	-0.26	-0.26	0.34	375.20	-0.26	-0.26	0.75
2007	FSD	15.40	-0.25	-0.25	0.12	281.80	-0.56	-0.56	0.41
2008	FSD	13.37	-0.34	-0.34	0.11	220.90	-0.46	-0.46	0.60
2009	FSD	13.66	-0.28	-0.28	0.15	297.15	-0.28	-0.28	0.49
2010	FSD	12.28	-0.22	-0.22	0.15	284.13	-0.22	-0.22	0.45
2011	FSD	11.90	-0.39	-0.39	0.22	244.62	-0.42	-0.42	0.59
2012	FSD	12.49	-0.25	-0.25	0.20	216.75	-0.25	-0.25	0.60

Table 3.9: Stochastic Dominance Test Results for Lower Wage Workers, Counterfactual #1

Year	Obrank	$d_{1,max}$	$d_{2,max}$	d	$\Pr(d \leq 0)$	$s_{1,max}$	$s_{2,max}$	s	$\Pr(d \leq 0)$
1996	FSD	13.34	-0.13	-0.13	0.15	365.28	-0.22	-0.22	0.84
1997	FSD	18.51	-0.22	-0.22	0.83	544.61	-0.26	-0.26	0.92
1998	SSD	18.00	0.47	0.47	0.04	402.38	-0.20	-0.20	0.70
1999	None	15.13	0.19	0.19	0.13	388.87	0.64	0.64	0.21
2000	FSD	13.52	-0.17	-0.17	0.58	394.86	-0.23	-0.23	0.71
2001	FSD	10.85	-0.23	-0.23	0.55	287.80	-0.25	-0.25	0.84
2002	FSD	11.11	-0.26	-0.26	0.46	327.00	-0.29	-0.29	0.81
2003	FSD	10.37	-0.01	-0.01	0.29	293.57	-0.20	-0.20	0.68
2004	SSD	8.33	0.90	0.90	0.04	198.47	-0.23	-0.23	0.31
2005	FSD	8.71	-0.26	-0.26	0.11	202.16	-0.26	-0.26	0.40
2006	FSD	12.60	-0.31	-0.31	0.38	243.86	-0.52	-0.52	0.84
2007	FSD	10.71	-0.15	-0.15	0.17	222.28	-0.24	-0.24	0.36
2008	FSD	10.11	-0.03	-0.03	0.35	246.93	-0.17	-0.17	0.90
2009	FSD	10.45	-0.01	-0.01	0.34	268.62	-0.19	-0.19	0.71
2010	SSD	6.42	0.03	0.03	0.15	200.86	-0.21	-0.21	0.72
2011	SSD	6.46	1.14	1.14	0.01	148.40	-0.19	-0.19	0.39
2012	None	8.91	0.64	0.64	0.01	177.41	0.77	0.77	0.28

Table 3.10: Stochastic Dominance Test Results for Lower Wage Workers, Counterfactual #2

Year	Obrank	$d_{1,max}$	$d_{2,max}$	d	$\Pr(d \leq 0)$	$s_{1,max}$	$s_{2,max}$	s	$\Pr(d \leq 0)$
1996	FSD	14.50	-0.30	-0.30	0.11	303.47	-0.30	-0.30	0.46
1997	FSD	14.80	-0.29	-0.29	0.10	303.02	-0.29	-0.29	0.36
1998	FSD	15.24	-0.26	-0.26	0.22	340.35	-0.30	-0.30	0.53
1999	FSD	14.62	-0.45	-0.45	0.20	293.55	-0.45	-0.45	0.46
2000	FSD	13.78	-0.25	-0.25	0.13	285.84	-0.37	-0.37	0.46
2001	FSD	14.62	-0.29	-0.29	0.22	295.95	-0.29	-0.29	0.62
2002	FSD	15.84	-0.33	-0.33	0.31	344.21	-0.41	-0.41	0.76
2003	FSD	13.85	-0.26	-0.26	0.12	301.62	-0.26	-0.26	0.48
2004	FSD	15.42	-0.25	-0.25	0.31	274.34	-0.32	-0.32	0.85
2005	FSD	15.22	-0.22	-0.22	0.30	255.04	-0.27	-0.27	0.65
2006	FSD	15.64	-0.26	-0.26	0.46	375.20	-0.26	-0.26	0.77
2007	FSD	15.40	-0.25	-0.25	0.11	281.80	-0.56	-0.56	0.41
2008	FSD	13.37	-0.34	-0.34	0.09	220.90	-0.46	-0.46	0.46
2009	FSD	13.66	-0.28	-0.28	0.19	297.15	-0.28	-0.28	0.55
2010	FSD	12.28	-0.22	-0.22	0.19	284.13	-0.22	-0.22	0.54
2011	FSD	11.90	-0.39	-0.39	0.17	244.62	-0.42	-0.42	0.65
2012	FSD	12.49	-0.25	-0.25	0.15	216.75	-0.25	-0.25	0.62