Risk and Return in Equity and Options Markets

by

Matthew P. Linn

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Business Administration)
in The University of Michigan
2015

Doctoral Committee:

Professor Tyler G. Shumway, Chair
Professor Robert F. Dittmar
Professor Stefan Nagel
Assistant Professor Christopher D. Williams
Professor Ji Zhu

This thesis is dedicated to Carrie and Hazel. Thank you for all of your love and support.

ACKNOWLEDGEMENTS

I would like to thank all of my committee members, Tyler Shumway, Bob Dittmar, Stefan Nagel, Chris Williams and Ji Zhu for their efforts in making this thesis what it is. I am especially grateful to Tyler Shumway for all of the knowledge he passed along to me as well as the time it required for him to do so. I am thankful to all of the students at Ross who made my time very so enjoyable. I am also thankful to my coauthors Sophie Shive and Shawn Mankad for all of the work they put into joint projects and for teaching me so much along the way. Finally I thank my family: Carrie, Hazel, Ryan and my parents for all of their support along the way.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

ABSTRACT

Risk and Return in Equity and Options Markets

by

Matthew P. Linn

Chair: Tyler Shumway

Option and equity markets are well known to be intimately linked due to the fact that options are contingent claims on underlying equity. A large literature has studied the theoretical link between these markets in terms of relative pricing of options and stocks. While theory can tell us about the relationship between prices of risk in the two markets within the context of a specific model, what we observe in the data rarely fits any single option pricing model with perfect precision. In fact, there seems to be little consensus on a single option pricing model with superior performance above all others. The broad purpose of this thesis is to empirically investigate the risk-return relation in options markets directly, without resorting to the use of option pricing models based upon relative pricing of options in terms of their underlying assets.

Options markets provide a rich cross-section of data with which to study how investors price assets. Option contracts vary across strike prices and times to maturity as well as varying across underlying assets. As a result, options data provides additional and complimentary information beyond the information contained in stocks. Using these facts, in this thesis I empirically investigate the risk-return relationship across stock option, index option and equity markets.

In Chapter I of the thesis I empirically show how to use options data to better estimate the cross-sectional price of market-wide volatility risk. I furthermore compare the price of volatility implicit in the cross-section of stock returns with the price implicit in the cross-section of option returns. In the same chapter I exploit the fact that options can be used to study the term structure properties of risk and return by examining the volatility risk and return tradeoff in options of different times to maturity.

In Chapter II, based upon the paper "Pricing Kernel Monotonicity and Conditional Information," co-authored with Sophie Shive and Tyler Shumway, I use data on index options and the underlying index to extract estimates of stochastic discount factors used by investors to determine prices of assets. We propose a new method for non-parametrically estimating the stochastic discount factor. Our method improves upon existing methods by aligning information sets available to investors at each time in our sample and taking these into consideration in our estimation scheme. Empirical results suggest that this may be the solution to a well known anomaly in the literature whereby non-parametric estimates of the pricing kernel tend to be non-monotonic in market returns.

CHAPTER I

Market-Wide Volatility Price in Options Markets

1.1 Introduction

The role of volatility risk in markets has been intensely studied in the recent literature. Evidence from the cross-section of equity returns suggests a negative price of risk for market-wide volatility, meaning that investors are willing to accept lower expected returns on stocks that hedge increases in market volatility.[1] Evidence from index options also suggests a negative price of volatility risk.[2] Surprisingly however, the volatility risk premium implicit in individual stock options does not appear to coincide with the premium implied by index options.[3] Attempts to cross-sectionally identify a negative price of market-wide volatility risk using stock options have also met with little success.[4] Taken together these results are puzzling, especially when such a tight relationship exists between options and their underlying stocks.

The options market offers an ideal setting in which to study the pricing impact

[1] *Ang et al.* (2006b), *Adrian and Rosenberg* (2008), *Drechsler and Yaron* (2011), *Dittmar and Lundblad* (2014), *Boguth and Kuehn* (2013), *Campbell et al.* (2012) and *Bansal et al.* (2013) study the role of market-wide volatility risk in the cross-section of equity returns.

[2] See *Bakshi and Kapadia* (2003a) and *Coval and Shumway* (2001).

[3] See *Bakshi and Kapadia* (2003b), *Carr and Wu* (2009) and *Driessen et al.* (2009).

[4] Using delta-hedged individual option returns, *Duarte and Jones* (2007) find no significant price of volatility risk *orthogonal* to underlying assets in unconditional models but a significant price in conditional models. *Da and Schaumburg* (2009) and *Di Pietro and Vainberg* (2006) estimate the price of volatility risk in the cross-section of option-implied variance swap returns but find opposite signs for the price of risk. *Driessen et al.* (2009) argue that returns on individual options are largely orthogonal to the part of market-wide volatility that is priced in the cross-section.

of systematic volatility. While far less studied than index options, individual options offer a much richer cross-section with which to study variation in returns because they vary at the firm level in addition to the contract level. Furthermore, option prices depend critically on volatility. Together these facts suggest that using individual stock options data improves the potential of accurately estimating the price of market-wide volatility risk in the cross-section.

While stock options offer a very promising asset class with which to study the price of market-wide volatility and potentially other market-wide risks, relatively little is known about the systematic factors that determine their expected returns. In fact, several papers have offered strong evidence that options are not redundant securities.[5] Coupled with evidence that option returns exhibit some surprising patterns[6] as well as demand-based option pricing,[7] this suggests that returns on options are not determined in exactly the same way as returns on their underlying stocks. Thus, it is important to independently show that market volatility is priced in the cross-section of returns of stock options. If it is not priced in the cross-section of a large class of assets like stock options (as has been suggested in the literature) it would be difficult to make a compelling argument that market-wide volatility is a state factor.

I empirically investigate the price of market-wide volatility risk in both the equity and options markets. Specifically, I empirically address two questions: 1) Is a market-wide volatility factor priced in the cross-section of equity option returns? 2) Is the price of volatility risk the same in the equity and option markets? It is important to distinguish between the systematic risk associated with market-wide volatility and the stock-specific measure of asset volatility, which is often included in models of option prices. I study whether investors are willing to pay a premium for individual stock options that hedge market volatility whereas it is commonly accepted that investors

[5]See for example *Bakshi et al.* (2000), *Buraschi and Jackwerth* (2001) and *Vanden* (2004).
[6]See *Ni* (2008) and *Boyer and Vorkink* (2014)
[7]See *Garleanu et al.* (2009) and *Bollen and Whaley* (2004).

are willing to pay a premium for options whose underlying stocks are volatile. My results show that even though the volatility risk premium extracted from individual stock options data does not appear to be consistent with that of index options, systematic volatility is priced in the cross-section of stock option returns. This supports the notion of volatility as a state factor.

To answer the questions stated above, I first create a new set of option portfolios that are optimally designed to facilitate econometric inference and to identify the price of market volatility. Following *Constantinides et al.* (2013), I adjust the realized returns of each option in order to reduce the effect of contract-level leverage. This paper is the first to apply this leverage adjustment to individual option returns instead of index option returns. The leverage adjustment is econometrically important because it reduces the effect of outliers that arise due to the extreme leverage especially inherent in out-of-the-money options. Furthermore, the adjustment helps to stabilize the stochastic relation between option returns and time-varying risk factors. I also propose a new method of sorting options that results in highly dispersed sensitivity of portfolio returns to market-wide volatility. The combination of forming portfolios of options and leverage-adjusting each option's returns renders standard econometric techniques feasible. This allows me to examine option returns in a manner typical of cross-sectional studies of stock returns as opposed to the highly stylized and non-linear models typically used in the option pricing literature.

Using GMM, I test a wide range of stochastic discount factors (SDFs) while controlling for factors commonly used to explain the cross-section of stock returns.[8] In

[8]I use the Generalized Method of Moments (GMM) in cross-sectional tests because it has several advantages over alternative asset pricing tests when studying option returns. For example, options of different moneyness tend to exhibit different levels of volatility. Thus standard errors from OLS cross-sectional regression cannot be applied to options due to heteroskedasticity of test assets. Furthermore, because the sensitivity of an option to time-varying risk factors can dramatically vary with option-specific parameters, time series regressions used in the first stage of *Fama and MacBeth* (1973) regressions may be very unreliable when using options data. GMM does not rely on a first stage time-series to explicitly estimate betas. In fact applying GMM only requires stationary and ergodic test assets.

addition to augmenting classical linear models with a volatility factor, I also posit SDFs that include factors from the literature that capture tail risk in equity returns. These factors help to disentangle volatility risk from the risk of market downturns, controlling for the well-documented leverage effect whereby market-wide volatility increases when market returns are negative. I show that market-wide volatility is an extremely important and robust risk factor in the cross-section. I then compare estimated prices of risk between the equity and options markets.

I furthermore test the price of market-wide volatility risk using cross-sectional regressions of both index and individual option returns at different time-to-maturity horizons. My results indicate that while volatility risk is significantly priced in the cross section of both index and individual options, the price observed in index option returns is due mostly to short-dated options. The index options actually show a term structure of volatility risk that is decreasing in time-to-maturity. Since options allow us to study prices of risk factors at different horizons as opposed to using the cross-section of stock returns, option returns provide a potentially important tool for analyzing asset prices in the cross-section. I propose a simple simulation to show that leverage-adjusting returns leads to improved cross-sectional tests of linear models typically used in the traditional asset pricing literature.

My results regarding a priced volatility factor align with the argument that market-wide volatility is a state factor. However, I find evidence that the price of volatility risk in the options market is larger in magnitude than in the stock market. This is somewhat surprising given that others have found volatility risk in options to be non-distinguishable from zero or to even take the opposite sign. My results are consistent with the demand-based option pricing theory of *Bollen and Whaley* (2004) and *Garleanu et al.* (2009) whereby intermediaries facing high demand for options charge larger premiums in order to cover positions that cannot be perfectly hedged. As stochastic volatility is a possibly unhedgeable risk that dealers face, my findings

may be the result of equilibrium pricing in the market due to market incompleteness. An alternative explanation is simply that there are limits to arbitrage preventing this apparent mispricing from being arbitraged away. This explanation is consistent with *Figlewski* (1989) who shows that arbitrage opportunities in option markets are costly and often too expensive to exploit in practice. A third explanation is that the two markets are segmented in such a way that market participants who are willing to pay more to hedge volatility invest in options.

The remainder of the paper is organized as follows. Section 2.4 describes the data used in the paper and the construction of factors used in the econometric analysis. Section 1.3 describes the test assets used throughout the paper. Sections 1.4 presents the main results. Section 1.6 provides details of a simulation study demonstrating the merits of leverage-adjusting option returns. Sections 1.5 and 1.7 provide tests of comparisons of prices of risk across different asset classes. Section 1.8 concludes.

1.2 Data

This section describes the data used in the study. I begin by describing the data sources. I then describe the filters used to clean the raw data. Finally, I describe the formation and properties of risk factors used throughout the paper.

1.2.1 Data Sources

Options data for the paper are from the OptionMetrics Ivy DB database. I use equity options for the analysis of the cross-section of option returns. I also use index options on the S&P 500 to construct factors used in the analysis. The OptionMetrics database begins in January 1996 and currently runs through August 2013. Data include daily closing bid and ask quotes, open interest, implied volatility and option greeks. The greeks and implied volatility for European style options on the S&P 500 are computed by OptionMetrics using the standard Black-Scholes-Merton model,

while implied volatilities and greeks for individual options are computed using the *Cox et al.* (1979) binomial tree method. The OptionMetrics security file contains data on the assets underlying each option in the data. These data include closing prices, daily returns and shares outstanding for each underlying stock. For the construction of stock portfolios, I use the entire universe of CRSP stocks over the same time period as the OptionMetrics data.

As is typical in the empirical options literature, I use options data only for S&P 500 firms. This partially eliminates the problem of illiquidity in options data. I follow the convention in the literature and calculate option price estimates by taking the midpoint between closing bid and ask quotes each day. Since the dates I use for monthly holding period returns are not the first and last trading day of a calendar month, I use the daily factor and portfolio data from Kenneth French's website to construct monthly holding period returns for factors and portfolios alike. The risk-free rate I use throughout the paper is also taken from Kenneth French's website.

1.2.2 Data Filters

Option deltas (Δ) measure the sensitivity of on option's price to small movements in the underlying stock. Formally, this is equivalent to defining the delta of an option as the partial derivative of the option price with respect to the price of the underlying stock. For a given underlying stock, the delta of put or call options is a monotonic function of option moneyness. With this logic in mind, I follow the convention in the literature and define option moneyness according to the option's delta as reported by OptionMetrics. Out of the money (OTM), at the money (ATM) and in the money (ITM) puts and calls are defined throughout the paper by the following:

OTM calls:	$0.125 < \Delta \leq 0.375$	OTM puts:	$-0.375 < \Delta \leq -0.125$
ATM calls:	$0.375 < \Delta \leq 0.625$	ATM puts:	$-0.625 < \Delta \leq -0.375$
ITM calls:	$0.625 < \Delta \leq 0.875$	ITM puts:	$-0.875 < \Delta \leq -0.625$.

I follow *Goyal and Saretto* (2009), *Christoffersen et al.* (2011) and *Cao and Han* (2013) among others in my data filtering procedure. First I eliminate options for which the bid price is greater than the ask price or where the bid price is equal to zero. Next I remove all observations for which the bid ask spread is below the minimum tick size. The minimum tick size is \$0.05 for options with bid ask midpoint below \$3.00 and is \$0.10 for options with bid ask midpoint greater than or equal to \$3.00. In order to further reduce the impact of illiquid options, I remove all options with zero open interest. I also remove any options for which the implied volatility or option delta is missing.

Finally, in order to reduce the impact of options that are exercised early, I follow *Frazzini and Pedersen* (2012) by eliminating options that are not likely to be held to maturity. This is done by first calculating each option's intrinsic value $V = (S-K)^+$ for calls and $V = (K-S)^+$ for puts, where K is the option's strike price and S is the price of the underlying stock. I then eliminate all options for which the time value, defined by $\frac{(P-V)}{P}$, is less than 0.05 one month before expiration, where P denotes the price of the option (estimated by the bid-ask midpoint). This final filter tends to remove options that are in the money. In unreported results, I verify that failing to include this final filter does not substantially alter the main results of the paper. Table 1.1 gives summary statistics for the filtered options data.

1.2.3 Option Returns Calculation

Equity options expire on the Saturday following the third Friday of each month. I compute option returns over a holding period beginning the first Monday following an expiration Saturday and ending the third Friday of the following month. Even though

all options in the sample are American and therefore have the option to exercise early, I follow the majority of the literature on option returns and assume all options are held until expiration. The removal of options with low "time value" described above and in *Frazzini and Pedersen* (2012) attempts to remove those options that are likely to be exercised early and not held until the following month's expiration date.

The payoff to the option is calculated using the cumulative adjustment factor to adjust for any stock splits that occur over the holding period. Put and call options' gross returns over the month t are given by

$$R^C_{t+\tau} = max\left\{0, S_{t+\tau}\left(\frac{CAF_{t+\tau}}{CAF_t}\right) - K\right\}\Big/ P_t, \tag{1.1}$$

and

$$R^P_{t+\tau} = max\left\{0, K - S_{t+\tau}\left(\frac{CAF_{t+\tau}}{CAF_t}\right)\right\}\Big/ P_t, \tag{1.2}$$

where τ is the time to maturity.

1.2.4 Factor Construction

Following *Ang et al.* (2006b) and *Chang et al.* (2013), I base my measure of market-wide volatility on the VIX index. Since the VIX exhibits a high level of autocorrelation, innovations in the VIX can simply be estimated by first differences, $\Delta VIX_t = VIX_t - VIX_{t-1}$. Throughout the paper I use VIX/100 because the VIX is quoted in percentages. This way I use a measure of market volatility as opposed to market volatility scaled by 100. Innovations in the VIX are highly negatively correlated with the market factor. This is the well known "Leverage Effect." In order to ensure that the volatility factor I use is not simply picking up negative movements in the market level, I further follow *Chang et al.* (2013) by orthogonalizing innovations in the VIX with respect to market excess returns. This is simply done by regressing ΔVIX on market excess returns and taking the residuals as the orthog-

onalized volatility factor. This orthogonalized measure of innovations in the VIX is the volatility factor referred to throughout the paper.

I construct market-wide jump and volatility-jump factors following *Constantinides et al.* (2013). The jump factor is defined as the sum of all daily returns on the S&P 500 that are below -4% in a given month. Since each month in my sample begins immediately following an option expiration date and ends at the following option expiration date, the jump factor is simply the sum of all daily returns in this time span for which returns are below the -4% threshold. If no such days exist, then the jump factor is zero for the month. Approximately 7% of the months in the sample have a non-zero jump factor. Finally, I include a volatility jump factor which captures large upward jumps in volatility of the market. To construct the volatility jump factor, I take all ATM call options on the S&P 500 and calculate the equally weighted average of implied volatilities over all options between 15 and 45 days to maturity. This gives me a series of daily average implied volatilities of ATM call options. Over each holding period I then take the sum of daily changes in implied volatility for all days in which the change is greater than 0.04. Approximately 29% of months in the sample have non-zero volatility jump.

Downside risk has been proposed as a state variable in the ICAPM and has been shown to perform very well for pricing stocks in *Ang et al.* (2006a) and across a number of additional asset classes including currencies, bonds and commodities in *Lettau et al.* (2013). I follow *Lettau et al.* (2013) by defining a down state to be any month in which market returns are below the mean of monthly returns over the sample period by an amount exceeding one standard deviation of returns over the sample period. The down-state factor is simply equal to returns on the CRSP value-weighted index in periods when the returns are below the down state threshold. In all other months the factor is zero. This gives a factor that is very similar to the jump factor. The main difference between the two is that the jump factor is

computed using daily data to determine when the market has experienced a jump. The magnitude of negative daily returns required to be considered a jump is much more extreme than the one standard deviation measure used to establish a down state. Furthermore, because jumps are defined at a daily frequency, they can more convincingly be considered jumps in the return process as opposed to simply months where the market slowly declines. Approximately 13% of months in the sample have non-zero down-side risk.

Finally I include model-free, implied risk-neutral skewness as a down-side risk. I follow *Bakshi et al.* (2003) to construct a measure of risk-neutral market-wide skewness. I then take innovations of the skewness factor by estimating an ARMA(1,1) model and taking residuals of the estimates. I use these residuals as an additional control for the main tests of volatility risk.

Figure 1.8 shows the time series of each of the volatility, jump, volatility-jump, down-side and skew factors. Panel B shows the orthogonalized volatility factor with the original, non-orthogonal factor in the background. Each of the factors has its largest spike during the recent financial crisis. More recently there are fairly large spikes during the U.S. debt-ceiling crisis in August of 2011. Volatility and volatility-jump experienced very large jumps around the terrorist attack of September 11, 2001. Table 1.6 gives pairwise correlations of the three factors as well as the Fama-French and Momentum factors. The construction of the latter factors are described in the appendix.

1.3 Portfolio Construction and Summary Statistics

In order to study the determinants and behavior of risk premia in the cross-section of option returns I construct 36 portfolios of options that are sorted along three dimensions. The portfolios are constructed in order to give dispersion in mean returns and exposure to changes in the VIX.

1.3.1 Portfolio Construction

I form portfolios of options by first dividing the options into six bins according to type: calls and puts, and three moneyness categories as defined in Section 1.2.2. Within each of these six bins I sort into another six portfolios according to each contract's Black-Scholes-Merton implied volatility premium. For each option k on stock j, I measure the implied volatility premium (IVP) by

$$IVP_{j,k} = \sigma_{j,k}^{BSM} - \sigma_j^{Hist},$$

where $\sigma_{j,k}^{BSM}$ denotes the Black-Scholes-Merton implied volatility extracted from option k's price and σ_j^{Hist} is the historical volatility of the underlying stock. I estimate σ_j^{Hist} from daily returns over the previous year leading up to the beginning of each holding period.

The IVP measure is similar to the sorting measure of *Goyal and Saretto* (2009) but rather than measuring the ratio of implied volatility to historical volatility of the underlying, I take the difference, which represents the premium due to model-implied volatility in excess of historical volatility. Another difference between the way I sort options and the method employed by *Goyal and Saretto* (2009) is that I sort at the contract level as opposed to just taking a single at-the-money option for each underlying stock and comparing the two. This gives my set of portfolios greater dispersion in loadings on innovations in the VIX than does the set of portfolios studied in *Goyal and Saretto* (2009).

To construct a set of equity portfolios, I follow *Ang et al.* (2006b). I use the entire universe of CRSP stocks to double sort stocks according to their loadings on the market excess return and changes in the VIX. On the first day of each holding period I calculate the CAPM betas of each firm over the previous month's daily returns. I only include firms for which CRSP reports returns on every trading day

over the previous month. The stocks are divided into two bins according to their loading on the market factor. Within each bin I then estimate a two factor model with market excess returns and changes in the VIX over the previous month and sort into six portfolios based on loadings on the second factor within each market loading category. This gives a total of twelve portfolios. I choose twelve portfolios so that they can be compared with the twelve ATM option portfolios. I choose to divide first into two market loading bins and then into six VIX innovation portfolios in order to maximize dispersion in loadings on volatility innovations while still double sorting in the manner of *Ang et al.* (2006b). Once the portfolios are formed, they are held for the one month holding period for which value-weighted returns are calculated. At the end of the month, the portfolios are rebalanced.

In unreported results, I find that sorting according to the systematic risk preimium described in *Duan and Wei* (2009) produces similar results to those described in Section 1.4. Furthermore, the results do not appear to be sensitive to the number of portfolios.

1.3.2 Portfolio Returns

Options are levered claims on the underlying stock. As a result of their embedded leverage, they tend to have loadings on systematic risk factors that are much larger than those of the underlying stock. It is very common for options to have market betas up to twenty times that of the underlying. This leverage effect can lead to very skewed returns on options. Highly volatile and skewed distributions are not well suited to estimating linear pricing kernels because a linear SDF is typically not able to capture such extremes. This fact makes linear factor models and the linear stochastic discount factor they imply a poor tool for analyzing raw option returns.

The embedded leverage of options further reduces the effectiveness of standard cross-sectional asset pricing techniques by rendering factor loadings less stable. In the

Black-Scholes-Merton world, loadings of options on any risk factor are approximately equal to the loading of the underlying on the factor scaled by the leverage of the option. The leverage of each option is a function of instantaneous volatility of the underlying which presumably is correlated with volatility of the market. As such the correlation of an option with a risk factor changes with market volatility. This means that even if one forms portfolios of options, the portfolio loadings on risk factors will be sensitive to large changes in volatility. Cross-sectional regressions will thus be sensitive to the instability of portfolio factor loadings.

Forming portfolios of option returns helps to dampen the effect of outliers and thus reduces skewness and excess kurtosis. It also mitigates the problem of the sensitivity of factor loadings to changes in volatility by dampening the effect for those options whose factor loadings are the most sensitive to volatility. Leverage adjusting returns further reduces the effect of each problem. In a world where the Black-Scholes-Merton model holds perfectly, continuously adjusting each option according to its implied leverage will completely solve both problems. As long as the SDF projected onto the space of stock returns can be adequately estimated by a linear model, continuous leverage adjustment renders linear factor models capable of pricing options. Given the well-documented shortcomings of the Black-Scholes-Merton model and the fact that it is impossible to adjust leverage in continuous time, the best we can hope to do with this leverage adjustment is to approximately correct both problems.

The Black-Scholes-Merton implied leverage of an option is given by the elasticity of the option price with respect to the underlying stock's price,

$$\omega_{j,i,t}^{BSM} = \Delta_{j,i,t} \frac{S_{i,t}}{P_{j,i,t}},$$

where $\Delta_{j,i,t}$ is the time t Black-Scholes-Merton option delta for option j on stock i, $S_{i,t}$ is the price of the underlying stock and $P_{j,i,t}$ is the price of the option. Table 1.2 gives

summary statistics for the Black-Scholes-Merton implied leverage of option contracts in the sample. In order to leverage-adjust the returns, I calculate the gross returns to investing $(\omega_{j,i,t}^{BSM})^{-1}$ dollars in option j and $1-(\omega_{j,i,t}^{BSM})^{-1}$ in the risk free rate. Since Δ is negative for puts and positive for calls, this amounts to a short position in puts and long position in calls. Leverage adjusted returns on the individual options are thus a linear combination of the returns on the risk free rate and returns calculated in Equations (1.1) and (1.2). Leverage adjustment is done at the beginning of the holding period, when the position is opened. Thus the leverage-adjusted returns are the returns to a portfolio composed of an option and the risk-free rate where the weight in the option is inversely related to its leverage. Unlike *Constantinides et al.* (2013), I hold the portfolio fixed over time and do not re-adjust leverage as the option's leverage evolves over time. A trading strategy with daily adjustment would incur very high transaction costs since the costs of buying and selling options is generally much higher than the cost incurred when buying and selling more liquid securities. Therefore, in order to replicate a more feasible trading strategy, I create portfolios that do not change over the course of the holding period. The obvious trade off is that these portfolios will not be as free of excess kurtosis and skewness as they would be in the case of daily rebalancing.

The majority of papers in the empirical option pricing literature examine delta-hedged returns in order to study profitability of trading strategies where investors have taken a delta-hedged position in options. The risks whose prices are estimated using delta-hedged option returns like those in *Duarte and Jones* (2007) are risks *orthogonal* to the underlying asset. In this paper I examine the price of *total* volatility risk because this is the risk estimated from the cross-section of equity returns. It is also the risk whose premium is implicitly estimated by looking at the difference between risk-neutral and physical moments of the underlying asset as in *Carr and Wu* (2009) and *Driessen et al.* (2009). Since the purpose of this paper is to resolve the apparent

discrepancy between prices of *total* volatility risk in options and equity, I do not delta hedge option returns.

Finally, to compute portfolio returns within each of the 36 portfolios, I weight the leverage-adjusted returns. In order to facilitate the comparison between the underlying stock returns and portfolios of option returns, I weight the options by the market capitalization of the underlying stock. This is standard practice in the equity pricing literature.

1.3.3 Summary Statistics

Table 1.3 gives summary statistics for the 36 value-weighted option portfolios. Panel A reports the annualized percentage mean returns of each of the 36 portfolios over the 200 months ranging from January 1997 through August 2013. The mean of the call portfolios is increasing in implied volatility risk premium while the mean of the put portfolios tends to decrease progressing from the lowest implied volatility premium, IVP1 to highest IVP6. Recall however that puts have a negative Δ and hence negative leverage, so the put portfolios are actually portfolios of short positions in the option. Therefore, long positions in the put portfolios earn increasing mean returns as a function of IVP. The dispersion in mean returns is much larger for the puts than calls but in all cases except ITM calls, the difference between mean returns in IVP1 and IVP6 is very large. As has been shown in the literature (see e.g. *Coval and Shumway* (2001)), selling puts is very lucrative because investors are willing to pay a premium to use puts as a hedge against large losses, so the large returns in the put portfolios is not surprising.

High levels of returns for puts and decreasing mean put returns as a function of moneyness are consistent with economic theory. The call portfolios however, exhibit increasing mean returns as a function of moneyness. As shown by *Coval and Shumway* (2001), if stock returns are positively correlated with aggregate wealth and investor

utility is increasing and concave, then returns on European call options should be negatively sloped as a function of strike prices. While the options used in this paper are American, I have removed options that are likely to be exercised early so reasoning similar to that in *Coval and Shumway* (2001) should be applicable here. This is not the first paper to document this pattern in mean returns of equity call options; *Ni* (2008) documents this puzzle. She shows that considering only calls on stocks that do not pay dividends and hence should never be exercised early, this pattern still shows up in the data. Furthermore, the pattern is very robust to different measurements of returns and moneyness. The explanation proposed by Ni is that investors in OTM call options have preferences for idiosyncratic skewness for which they are willing to pay a premium in OTM calls.

Panel B reports annualized return volatility of each value-weighted option portfolio in percent. Volatility is monotonically decreasing in moneyness for the put portfolios. For the call portfolios the pattern is less clear. We also see that volatility is higher for the put portfolios than for the calls. Panels C and D report monthly measures of skewness and kurtosis for each portfolio. As can be seen in Figure 1.8, the put portfolios are negatively skewed while the calls are positively skewed. Furthermore, the magnitude of the skewness is highest in OTM options and tends to decrease monotonically in moneyness. Similarly, kurtosis is largest in OTM options and smallest in the ITM options, with a monotonic, decreasing pattern in moneyness. The purpose of forming leverage-adjusted portfolios of option returns is to reduce excess skewness and kurtosis, thus rendering portfolio returns nearly normally distributed. While the skewness measures are not equal to zero as one would ideally like to have, they are much smaller in magnitude than the skewness of raw option returns. For example, the absolute value of skewness for the empirical distribution of raw returns on all calls and puts used to form the portfolios are on average 4.769 and 6.263 respectively. Return kurtosis is reduced even more dramatically by forming the leverage-adjusted

portfolios. The normal distribution has a kurtosis of 3. The kurtosis of the leverage-adjusted portfolios ranges from 3.763 to 14.615. The kurtosis of raw realized option returns of the options dwarfs that of the leverage-adjusted portfolios. This is most noticeable in the OTM options. The average kurtosis of the empirical distribution of raw returns on OTM calls is 44.297 while that of the OTM puts is 83.697. This means that forming portfolios of leverage adjusted returns reduces kurtosis by nearly 90% in OTM puts and 75% in OTM calls. That is, the shape of the tails of the empirical distribution of the OTM option portfolios is much closer to the that of a normal distribution than are the tails of the empirical distribution of raw returns on OTM options.

Panel E shows the CAPM betas for each portfolio. Recall that the put portfolios are actually short puts. This is why the betas reported for the puts are positive. Betas are monotonically increasing in moneyness for the calls and for the most part decreasing in moneyness for the puts. The betas on the calls are below one while the betas on the puts are mostly above one. Comparing these with the CAPM betas on the stock portfolios shown in Table 1.5 gives an indication of the leverage reduction achieved by leverage adjusting the returns in the option portfolios. It is quite common for options on individual stocks to have Black-Scholes-Merton implied leverages with magnitudes in excess of 20. If an option on a stock has an implied leverage of 20, then in the Black-Scholes-Merton world, for any risk factor, the beta of the option on that risk factor will be 20 times that of the underlying stock. In the case of the put portfolios, the CAPM betas are magnified by roughly 15% above those of the corresponding stock portfolios in Panel A of Table 1.5. In the case of calls, the betas are reduced by about 25% on average. In both cases this suggests a fairly low level of implied leverage in the options.

Panel F of Table 1.3 reports betas on systematic volatility in the two factor model

$$R_{i,t} = \beta_{M,i} MKT_t + \beta_{\Delta VIX,i} \Delta VIX_t + \epsilon_{i,t}, \tag{1.3}$$

where MKT_t denotes time t excess returns on the market and ΔVIX_t denotes first differences in the VIX index. The factors used to proxy for market returns and volatility innovations are formed as described in Section 1.2.4. The volatility betas of call portfolios are much smaller in magnitude than the volatility betas of the put portfolios. Half of the call portfolios betas are statistically significant at the 5% level. On the other hand, all of the volatility betas except that of the ITM IVP6 portfolio are highly significant. The average t-statistic of the put portfolios' volatility betas is -4.33, while that of the call portfolios is only 1.36. The fact that the puts appear to load so much more on the volatility factor suggests that if systematic volatility is indeed priced in equity options, the premium is more likely to be evident in the puts than the calls. Again, since the put portfolios are actually short puts, the loadings on volatility are negative. In both call and put portfolios, the *magnitude* of volatility betas decreases monotonically in moneyness.

As a comparison, in Table 1.4, I include summary statistics for the option portfolios when returns are not leverage adjusted. The extreme volatility is evident in panel B where each portfolios has annualized return volatility of roughly ten times that of the leverage adjusted portfolios. By reducing this volatility and "de-noising" the returns we are able to get more accurate estimates of prices of risk and corresponding stochastic discount factors. While the skewness is not significantly reduced by leverage adjusting, the return kurtosis is. This implies that the tails of the return distributions are much heavier than the normal distribution meaning that linear factor are unlikely to accurately identify prices of risk.

Table 1.5 reports summary statistics for the stock portfolios. The portfolios are

comprised of all CRSP stocks over the 200 months ranging from January 1997 through August 2013. The columns of each panel in the table represent sorts according to betas on market excess return over the previous month of daily data. Rows represent sorts according to loadings on volatility innovations. Panel A reports post formation value-weighted mean returns. For the most part, the post ranking mean returns are higher for the high market beta than the low market beta group. Mean returns to the portfolios are generally decreasing in loadings on the volatility factor as one would expect given that stocks with higher loadings on the VIX act as a hedge agains high volatility states and investors are thus willing to pay a premium for these stocks. The monotonicity in mean returns along the volatility loading dimension is not particularly strong. This is due to the fact that the formation period is only a month long.

Panel B reports annualized percent volatility. There is clear heteroskedasticity between the two market loading bins with the higher market-loading stocks having substantially higher volatility. Skewness is negative for all portfolios and tends to be larger in magnitude for the low market beta stocks than for the high beta stocks. The stock portfolios are less skewed than the option portfolios but the difference is not very dramatic. Similarly, the kurtosis of the stock portfolios is slightly smaller than the option portfolios except in the case of OTM puts where the kurtosis is most extreme. Figure 1.8 plots the histograms of realized returns for each of the six put/call and moneyness bins as well as the realized returns of all puts and all calls separately and all ATM options. Over each is the kernel density estimate of the empirical return distribution of the stock portfolios. One can see from the figure that skewness and variance of the option portfolios is not very different from that of the stocks except perhaps in the case of the OTM calls.

The post ranking CAPM betas of the stock portfolios are much larger for the stocks with large formation period betas suggesting that stocks' covariation with the market is fairly stable. On the other hand, Panel F shows that the post-ranking

volatility betas do not exhibit a clear monotonic pattern. This indicates that at least with the one month formation window, stocks' loadings on innovations in the VIX are less stable.

1.4 Pricing Kernel Estimation

In this section I test a number of specifications of pricing kernels to assess the importance of volatility for the SDF projected onto the space of option returns. Throughout this section I use the Generalized Method of Moments of *Hansen* (1982) and *Hansen and Singleton* (1982) to perform the asset pricing tests. Since the tests combine various portfolios of options as well as stocks, using GMM circumvents any problems that may arise due to heteroskedasticity across asset classes or moneyness-put/call bins that are shown to exist in Tables 1.3 and 1.5. An additional advantage of the GMM methodology over regression-based cross-sectional tests like *Fama and MacBeth* (1973), is the fact that it avoids the error-in-variables problem associated with estimating risk factor loadings in time-series regressions which are subsequently used as independent variables in the cross-sectional regression. This errors in variables problem is particularly glaring in the case of option returns. If one uses individual options as test assets and computes returns to the value of the option at multiple times over the course of the option's lifetime, then any changes in leverage of the option will result in changes in factor loadings in time series regressions. Furthermore, the most liquid options are short dated, meaning that time-series regressions on option returns used in the first step of a procedure like Fama-MacBeth cannot be estimated with a very long time series.

The use of GMM coupled with the option portfolios described in Section 1.3 allows me to circumvent the errors in variables problem. GMM estimation does not require test asset returns to be iid conditional on risk factors. All we need is for our time

series of portfolios to be stationary and ergodic.[9]

1.4.1 GMM specification

In order to investigate the importance of market-wide stochastic volatility in the cross-section of option returns, I apply the GMM methodology of *Hansen and Singleton* (1982) to various specifications of a linear pricing kernel. The specifications include factors commonly used in the empirical asset pricing literature. In this sense, the models used in this paper are directly comparable to some of the most well known reduced form models used to study the cross-section of stock returns. I augment the models with the volatility factor in order to assess the importance of market-wide volatility in the SDF.

In addition to factors studied widely in the classical asset pricing literature, I include factors meant to capture market jump risk and market volatility jump, both of which are commonly included in theoretical option pricing models.[10] I include additional factors meant to capture extreme movements in the market that have been shown to perform well in pricing the cross-section of stock returns. All of these additional factors track extreme movements in the market and are meant to control for the fact that volatility can be difficult to distinguish from downturns in the market level or large changes in the market level.[11]

For each specification of the pricing kernel, I use the two step optimal GMM to estimate the prices of risk associated with each factor. The first stage estimation uses the identity weighting matrix. In the second stage estimation the weighting matrix is set equal to the inverse of the covariance matrix estimated from the first stage.

[9] In an unreported test, all but one of the 36 option portfolios described in Section 1.3 were able to reject non-stationarity at the 1% level using an the Augmented Dickey-Fuller test for non-stationarity. The one portfolio that was not able to reject at the 1% level did reject at the 10% level and the GMM estimation results of this section are not substantially changed by removing this single portfolio.

[10] See for example *Pan* (2002) and *Eraker et al.* (2003).

[11] See *Bates* (2012) for a discussion of difficulties related to disentangling volatility from large changes in market level.

I estimate the weighting matrix using the *Newey and West* (1987) spectral density estimator with 6 lags. As a robustness check I also run the same set of tests with a one-step GMM using the identity weighting matrix and also the one-step GMM using the weighting matrix of *Hansen and Jagannathan* (1997). In both cases the results are similar to those reported in this section. The volatility factor is significant at the 5% level in all specifications with both versions of the single-step GMM and the point estimates are very similar to those obtained with the 2-step GMM.

In each specification of the pricing kernel M, the first N moment restrictions in the GMM test with N test assets are given by

$$\mathbb{E}\left[M_t R_{j,t}\right] - 1 = 0, \tag{1.4}$$

for $j = 1, 2, ..., N$, where $R_{j,t}$ denotes the time t gross return of portfolio j. The final moment condition which is implied by the risk-free rate is given by

$$\mathbb{E}\left[M_t\right] - \frac{1}{R^f} = 0, \tag{1.5}$$

where R^f denotes the risk-free rate.

1.4.2 Linear Pricing Kernels

In this section I restrict our attention to linear pricing kernels of the form

$$M_t = a + b' f_t,$$

where f is a vector of risk factors, b is a fixed vector of prices of risk and a is a constant.

Tables 1.7, 1.8, 1.9 and 1.11 report the results for five specifications of the linear pricing kernel. The first is the single factor model with only the volatility factor. The second and third models are respectively the standard CAPM and the CAPM

augmented with volatility. Model four is the Fama-French/Carhart four factor model and the fifth model is the volatility-augmented version of model four. For each model I report point estimates of the coefficients with t-statistics in parentheses. The final two columns of each table report the J-statistic and associated p-value as well as the Hansen-Jagannathan distance which measures the distance between the implied stochastic discount factor and the set of feasible discount factors.

Table 1.7 reports results of the tests using all 36 option portfolios. The coefficient on the volatility factor is positive and very significant in each specification. A positive coefficient in the SDF implies that investors' marginal rates of substitution are increasing in volatility. This means that investors are willing to pay a premium for assets that covary positively with innovations in volatility. In other words the price of volatility risk is negative. For both the CAPM and the four factor model, adding volatility substantially reduces the J-statistic and the Hansen-Jagannathan distance measure, indicating that the model fits the data much better with the volatility factor than without.

Data filters are implemented to remove illiquid options and I only consider options on S&P 500 constituents in order to avoid results driven by illiquid options. In order to further alleviate any concerns about illiquidity driving the results, I examine just the ATM option portfolios separately as these are the most liquid options according to trading volume. Table 1.8 reports the results which are quite similar to the tests with the full set of option portfolios. The volatility factor is always positive and significant and given the fact that we only have twelve test assets, the significance is very strong. In each specification, the model fit is substantially improved with the addition of the volatility factor.

Table 1.11 reports the pricing kernel estimates for the ATM options and the 12 stock portfolios combined. If volatility is a priced risk factor in the SDF, then the projection of the SDF onto the combined space of stocks and options should also have

a positive, significant coefficient. This is confirmed in Table 1.11. It is worth noting that for the combined stock portfolios and ATM option portfolios, the reduction in J-statistics due to adding the volatility factor are very small. However, the Hansen-Jagannathan distance is substantially reduced. In the case of the SDF projected onto the space of stock returns only, Table 1.9 shows that the fit of the four-factor model improves with the addition of the volatility factor but the two-factor model actually fits worse with the addition of volatility. The volatility coefficient's point estimates for both the stock portfolios as well as the combined stock and ATM option portfolios are well below the point estimates for the full set of option portfolios.

The takeaway from Tables 1.7, 1.8, 1.9, 1.11 is a clearly priced systematic volatility risk factor in option returns. To assess the economic magnitude of the volatility premium one can easily use the coefficient in the SDF to calculate λ^{VOL}, the implied market price of the the volatility risk. λ^{VOL} is equivalent to the prices of risk typically estimated in the second step of Fama-MacBeth regressions. In the case of the full model (model 5), the market price of volatility, λ^{VOL} is equal to -4.13% per month or -62.5% annualized. We can get a sense of how much of the difference in mean returns of the OTM puts and ITM puts is driven by volatility risk by comparing the average volatility betas for each group. For model 5, the average volatility betas for OTM puts and ITM puts are -0.7042 and -0.3482 respectively. Exposure to aggregate volatility therefore accounts for $(-0.70-(-0.35))\times -4.13\% = 1.47\%$ monthly or 19% annualized spread in returns between ITM and OTM puts. For the calls the average OTM beta is 0.238 and the average ITM beta is -0.013. Exposure to aggregate volatility therefore accounts for $(-0.013-0.238)\times -4.13\% = 1.37\%$ monthly or 17.7% annualized spread in returns in the calls. Thus the volatility premium is economically significant as well as statistically significant. It is also worth noting that the implied price of risk, -4.13% per month is 18% larger than the -3.49% price of risk estimated in *Chang et al.* (2013) using stocks.

In an unreported robustness check, I run all of the tests with the same portfolio sorts but weight returns by option open interest rather than stock market capitalization. The results are similar. Volatility is always significant at the 5% level and the point estimates are similar to those reported in Tables 1.7, 1.8, 1.9, 1.11.

For comparison, Table 1.10 shows the results of the at-the-money option portfolios without leverage adjustment. As shown in Table 1.4, the returns of portfolios without leverage adjustment are extremely volatile and heavy tailed. We expect this to reduce the effectiveness of linear models for estimating stochastic discount factors or prices of risk. Table 1.10 reports the GMM estimation results for the option test assets without leverage adjustment. The results show that volatility is not significant. This is consistent with findings in the literature that suggest market-wide volatility may not be priced in the cross-section of individual option returns (see *Driessen et al.* (2009)). In Section 1.6 we verify that leverage adjusting returns can help us estimate price of risk more accurately in the context of linear models.

1.4.3 Exponentially Affine Pricing Kernels

In order to check that the linear form assigned to our pricing kernel is not responsible for the strong significance of the market-wide volatility factor, I test the same set of CAPM and Fama-French-Carhart factors augmented with volatility using an exponentially affine pricing kernel instead of a linear pricing kernel. Whereas standard asset pricing models assume a linearized SDF, the exponentially affine pricing kernel is closer to the kernel derived by hypothesizing a utility function for a representative investor and then solving for the marginal rate of substitution. For an investor with CRRA utility, the SDF can be expressed as

$$M_t = \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma},$$

where C_t denotes time t consumption, γ denotes the coefficient of relative risk aversion and β denotes the investors discount rate. By taking the exponential of the log of the pricing kernel this can be transformed to the exponentially affine form

$$M_{t+1} = e^{log\beta - \gamma log\frac{C_{t+1}}{C_t}}.$$

I use an exponentially affine pricing kernel which assumes a similar form,

$$M_{t+1} = e^{a+b'f_{t+1}}, \tag{1.6}$$

where b is a deterministic vector of coefficients and f is a vector of risk factors. The log-utility CAPM is a special case of the SDF in Equation (1.6) where $a = 0$, $b = -1$, $f = logR^W$ and R^W is return on the wealth portfolio. The exponentially affine framework is better suited for analyzing skewed payoffs like options as it does not rely on linear approximations of the functional form of investors' marginal rates of substitution. Continuous time versions of exponentially affine pricing kernels are commonly used in structural option pricing models, where the factors are typically specific to the underlying asset as opposed to systematic factors.

Tables 1.12, 1.13, 1.14 and 1.15 report the results of GMM tests using the pricing kernel defined in Equation 1.6 with the same set of factors from Tables 1.7, 1.8, 1.9, 1.11.[12] The results again show that market-wide volatility is a significantly priced factor in the cross-section of option returns. The point estimates cannot be directly compared to those in the linear models. However, the volatility factor is estimated to be significantly positive. Table 1.12 reports the results from a one-step GMM estimation where the weighting matrix is set equal to the identity matrix. This greatly reduces the power of the test but is meant to allay any concerns about unstable

[12]I also test the exponentially affine models with the non-orthogonalized volatility factor. I do this because the orthogonalization is linear with respect to market excess returns and I want to be sure that the linear nature of the orthogonalization is not responsible for the results in a nonlinear model. The results for the coefficient on the volatility factor were virtually unchanged.

inversion of the weighting matrix in nonlinear GMM estimation when the number of time series observations is not very large compared to the number of cross-sectional observations.[13] Using all 36 option portfolios, the volatility factor is signifiant at the 5% level for the two models containing the market factor. In the single factor model the volatility factor is only significant at the 10% level. Given that the combination of a single step GMM and a non-linear model substantially reduces the power of the test, the fact that the volatility factor is still significant can be regarded as strong evidence in favor of the volatility factor.

Tables 1.13 and 1.15 give the results of the tests with only the ATM options and the combined portfolios of ATM options and the 12 stock portfolios. In all specifications the volatility factor is very significant. In the case of the test with only ATM options, including the volatility factor drastically reduces the J-statistic and the Hansen-Jagannathan distance, especially in the case of the 4-factor model. The results are not so strong when the ATM options are combined with the 12 stock portfolios in Table 1.15, however the volatility factor is still significant in all specifications, indicating that market-wide volatility plays an important role in the SDF projected onto the joint space of stock and option returns. This holds true despite the fact that for the stock portfolios alone, there is little evidence that the volatility factor is significant in Table 1.14. This is important for two reasons. First, it indicates that we have more power to estimate the role of market-wide volatility in the SDF when using options than using the same number of stocks portfolios. Comparing the 12 ATM option portfolios with 12 stock portfolios sorted in a way that has been the most successful thus far in the literature at showing a significant volatility factor, it is clear that the option portfolios are a more powerful set of test assets. Second, even if the SDF projected onto one space shows the volatility factor to be statistically insignificant, it is entirely possible that the factor is still significantly priced in the

[13]See *Ferson and Foerster* (1994) and *Cochrane* (2005) for discussions about GMM and small sample properties.

SDF. It may just be the case that the space of stock returns is orthogonal to the volatility factor in the SDF while the space of option returns is not orthogonal to the factor. If this is the case, we still expect to find that when estimated from returns on the joint space of stock and option returns, the factor should be significant as we find in Table 1.15.

1.4.4 Pricing Kernels with Tail Risk

As first noted by *Black* (1976), volatility of the market is negatively correlated with the market's level. Table 1.6 shows that in the sample period 1997 through 2013, monthly innovations in the VIX and excess market returns are highly negatively correlated. This is the reason for using orthogonalized VIX innovations in the analysis throughout the paper. More recently *Bates* (2012) discusses the difficulty of separating changes in volatility from jumps. A number of papers have also shown that the risk neutral distribution of stock indices exhibit higher volatility, more negative skewness and have heavier tails than their corresponding physical distributions.[14] This indicates that option prices reflect premia for skewness and kurtosis as well as volatility. Furthermore, *Bates* (2000), *Pan* (2002) and *Eraker et al.* (2003) have shown that jump risk tends to increase during times of higher market volatility. Taken together, all of these empirical regularities suggest that the risk premium attributed to market-wide volatility in our earlier tests may actually be due to fears of tail events. In this section I include additional factors in specifications of the SDF in order to control for the possibility of tail risk driving the significant volatility premium documented thus far. Tables 1.16, 1.17, 1.18 and 1.19 give results of linear models for the SDF with additional factors described in Section 1.2.4.

Tables 1.16 and 1.17 report results for test assets comprised of all 36 option portfolios and the ATM portfolios respectively. The clear result from these two tables is

[14]See *Jackwerth and Rubinstein* (1996), *Jackwerth* (2000) and *Bakshi et al.* (2003).

that volatility risk carries a significant, positive coefficient (and hence a negative price of risk) even when we control for tail risk. While some of the tail-risk factors appear to be significant in a number of the specifications, volatility is the only factor that is significant in all specifications in both tables. In Table 1.16, with all 36 option portfolios as the test assets, downside risk also appears significant and skewness is significant at the 10% level. However, in Table 1.17 where the test assets are the 12 ATM portfolios, neither is significant. This is likely to be at least partially attributable to the fact that we have a small number of test assets and thus less cross-sectional variation. However, volatility is clearly significant even with the small number of test assets and the additional controls for tail risk. It is also worth noting that the jump factor does not appear to be significantly priced even though jumps are often modeled in option returns. However, the jumps included in theoretical option pricing models are jumps in the underlying asset as opposed to market-wide jumps. Of course in the case of index options where the relation between jumps and option prices have been most studied (see for example *Pan* (2002) and *Eraker et al.* (2003)), on cannot distinguish between market-wide risks and risks inherent only in the underlying asset.

Table 1.18 reports the results for the stock portfolios test assets. In this set of tests the volatility factor remains marginally significant at best. This could largely be due to the fact there is a small number of test assets. However, when compared to the 12 ATM option portfolios, it is clear that the volatility factor is much more prominent in the options than in the stock portfolios. In Table 1.19 where stocks and ATM options are the combined test assets, volatility is again very significant. Here skewness and downside risk are also significant.

The results of this section indicate that not only is market-wide volatility a significant risk factor in the cross-section of individual option returns, but it is distinct from market-wide tail risk. Taken together with tests in the previous sections this suggests that volatility is a very robustly priced risk factor in the cross-section.

1.5 Likelihood Ratio-Type Tests

In this section I test whether the prices of risk estimated using options differs from those estimated using the underlying stocks. The tests I use are special cases of those described in *Andrews* (1993). They are also known in the econometrics literature as likelihood ratio-type tests for GMM models. These tests combine stock and options data in restricted and unrestricted GMM tests and compare the resulting objective functions. In this way, the intuition behind the tests is similar to likelihood ratio tests. Of course the difference is that in this setting we have not specified a parametric likelihood function. Here, as in the previous section, I use GMM because I estimate models that simultaneously use stock portfolios and different option portfolios to estimate models. Tables 1.3 and 1.5 demonstrate the need for taking into account possible heteroskedasticity across assets.

Similar to likelihood ratio tests, the GMM likelihood ratio-type test compares the value of an objective function under the null hypothesis to its value under an alternative hypothesis. For the purpose of testing prices of risk in two markets, the comparison is made between models that fix the coefficients on risk factors to be the same in the option and equity pricing kernels and those that relax this assumption. I perform the tests by relaxing the assumption on the volatility factor and comparing the resulting unrestricted GMM objective function to the restricted objective function. Namely, the null hypothesis is

$$H_0 : b^S_{VOL} = b^O_{VOL}$$

where b^S_{VOL} and b^O_{VOL} are the prices of risk in the stock and option markets respectively.

For each proposed model, I estimate the restricted version by pooling stock portfolios together with option portfolios so that the test assets are a mix of the 12 stock

portfolios and 12 ATM option portfolios. The results of estimating the restricted models are given in Table 1.11. For each model I test the restriction by relaxing H_0 and comparing the resulting fit to the corresponding model fit in the restricted model.

Since the tests compare GMM objective functions with and without a linear restriction, one needs to be sure that the difference in objective functions is not driven by the weighting matrix but is driven only by differences due to relaxing the restriction on a given factor. I therefore use the second stage weighting matrix from the restricted model estimation to estimate the unrestricted model in a single step GMM. This also ensures that the test statistic has a well defined asymptotic distribution. In particular, the test statistic has the asymptotic distribution given by

$$LR_{GMM} = T\left(m(\hat{\theta}_R)'\hat{W(\theta_R)}m(\hat{\theta}_R) - m(\hat{\theta}_U)'\hat{W(\theta_R)}m(\hat{\theta}_U)\right) \to \chi^2_1, \qquad (1.7)$$

as $T \to \infty$, where T denotes total number of observations, $\hat{\theta}_R$ and $\hat{\theta}_U$ denote estimated vectors of prices of risk under the restricted and unrestricted models respectively and $m(\hat{\theta}_R)$ and $m(\hat{\theta}_U)$ denote empirical means of moment restrictions under the restricted and unrestricted models.

Table 1.20 gives the test statistics and corresponding p-values for each likelihood ratio-type test. Rows represent the models used for each test. Columns represent the variable whose price of risk is being tested. Of the three baseline models testing the volatility factor, one shows a significant difference between the restricted and unrestricted model at the 5% level and the remaining two give significant test statistics at the the 10% level. This suggests that the price of systematic volatility risk is not necessarily the same in the equity and option markets. Although these results suggest that there may be difference in the prices of volatility risk between the two markets, the difference is likely to fall within no-arbitrage bands as it is well known that no-arbitrage option price ranges can be fairly wide.[15]

[15]See *Figlewski* (1989) for example.

The fact that there is a difference between prices of volatility risk in the stocks and put options is akin to there being a significant price of risk in delta-hedged returns of put options. Whereas delta hedged options look directly at the option with the risk due to the underlying subtracted off, the results here look at the difference in prices estimated from options and stocks separately. These are two similar ways of addressing the same question; Is there significantly priced volatility risk inherent in option contracts that is not due solely to the underlying stock? The fact that I find a positive difference between implied prices of risks suggests the answer is yes. It further provides evidence that options are not redundant securities.

1.6 Simulation

In order to verify that leverage adjusting returns does indeed improve the precision with which we can estimate prices of risk from the cross-section of option returns, I conduct a simple simulation examining the cross-sectional estimates of a single risk price. I propose one of the most simple option settings. I assume that underlying stocks are driven by a single factor. In the case that the single factor is market excess returns, we are in the CAPM world. As this is the most commonly used factor model, I will use this as the setting for my simulation. I assume the price of risk associated with the market factor is fixed over time at $\lambda_M \equiv 2.5$. At each point in time, I assume that the expected excess return on the market is equal to λ_M. The market excess return at time t is given by

$$R^e_{M,t} = \lambda_M + \epsilon^M_t,$$

where ϵ^M_t is distributed normally with mean of zero and a standard deviation set at $\sigma_M \equiv 12\%$. Since excess returns on stock i are assumed to be driven by the stock's

sensitivity to the market factor, the excess returns of stock i are given by

$$R^e_{i,t} = \beta^M_i \lambda_M + \epsilon_{i,t}.$$

In order to apply the standard unconditional linear model, I assume β^M_i is fixed over time for each stock i. I further assume that the betas of each firm are randomly selected from a continuous uniform distribution between .5 and 1.5. The noise term, $\epsilon_{i,t}$ is assumed to be normally distributed with a mean of zero and standard deviation of σ_i where each firms annual return volatility is drawn from a normal distribution with mean 20% and a standard deviation of 5%. With these parameters and distributions for annual returns, I simulate 500 months of returns for 1,000 stocks, assuming that return distributions are the same over each month so that the monthly values are the same for each month making up a year and the aggregate distribution over the 12 months in the annual distribution.

Next, I again assume the simplest setting and propose that option prices are set according to the Black-Scholes-Merton model. For each stock I construct prices of 6 options: an in-the-money call and put, at-the-money call and put and out-of-the money call and put. I set the definition of in, at and out of the money as follows, for puts moneyness is defined as strike price divided by stock price. For calls, moneyness is equal to stock price divided by strike price. I define OTM options as those with moneyness $.925 <= Moneyness < .975$, ATM options are defined by $.975 <= Moneyness < 1.025$, and in the money options are those with $1.025 <= Moneyness < 1.075$. Within each moneyness range, I randomly select the moneyness a given stock's option at time t will be assigned. This is done by randomly drawing from a uniform distribtion between the upper and lower bound of the moneyness. The Black-Scholes-Merton prices are then calculated with an assumed fixed annual risk-free rate of 5%.

I run cross-sectional regressions for three different sets of test assets. The first involves running cross-sectional regressions with each of the stocks' six options treated as a continuously traded asset. This way, the first stage, time series regression can be applied to options even though each option only exists for a short window within the simulation. With this resulting panel of options, I next form portfolios by sorting according to the underlying stocks' beta. I form 10 portfolios within each moneyness/put-call category. I then simply equal-weight the portfolio returns. The first set of portfolios involves returns that are not leverage adjusted while the second uses option returns that are de-levered monthly. One difference between the leverage adjustment employed here and the leverage adjustment we employ in the real data is that here we require that prices be determined via Black-Scholes-Merton pricing. This means that at each point in time we can perfectly calculate each option's leverage. However, because I only leverage-adjust my option portfolios monthly instead of instantaneously, this gives an imperfect de-levering as we are bound to have when de-levering real options data.

I simulate this economy 1000 times. The results are given in Table 1.24 and Figure 1.8. Panels A and B of Figure 1.8 give the sampling distribution of estimated prices of market risk in each of the 1000 simulated economies, without and with leverage adjusted returns respectively. The sampling distribution is more accurate and much tighter for the leverage adjusted portfolios than it is for the portfolios without leverage adjustment. When we estimate the price of risk λ_M from the entire panel of options data without forming portfolios, the point estimate of 0.082 is far below the true value of $\lambda_M = 2.5$. The sampling error for the individual options is much smaller because of the fact that the cross-section is so much larger when we don't form portfolios. If one forms portfolios of options without leverage-adjusting, the results improve dramatically. The point estimate $\hat{\lambda}_M = 1.994$ which is much closer to the true price of risk. Clearly forming portfolios is an important way of

reducing the extreme noise present in option returns. This noise may partially be due to the leverage embedded in options due to the non-linear payoff structure of options. The third set of test assets addresses this issue. These test assets are the same as the second set of test assets except that they are de-levered at the beginning of each month. The results shown in Table 1.24 suggest that the point estimates are significantly improved by de-levering. Furthermore, the sampling error is also improved above and beyond just forming portfolios. The sampling error of the point estimates using leverage-adjusted portfolios in each of the 1000 economies is 0.159 as opposed to 0.298 when option returns are not leverage-adjusted. This means that de-levering option returns on top of forming portfolios further improves the efficiency of using options to estimate the price of risk in the cross-section. If the results of this simple model can be extrapolated to more complicated models used to price options for which we have data, this suggests that the prices of risk estimated from portfolios of leverage-adjusted options are a better indicator of the actual prices of risk.

1.7 Volatility Price in Index and Individual Options

A number of papers have documented the disparity in the magnitude of risk premia embedded in measures of volatility from index options and individual options (See for example *Driessen et al.* (2009) and *Bakshi and Kapadia* (2003a)). The literature has found that the so called volatility risk premium, defined as the difference between return volatility under the risk-neutral and physical probability measures, is much larger for index options than for individual options. The generally accepted notion that volatility risk premia measure the premia due to volatility exposure.

In this section I measure the price of volatility risk by examining the cross section of S&P 500 index options and options on the stocks making up the S&P 500. I compare the prices of market-wide volatility risk implicit each set of assets separately. I then compare whether the prices differ. The results point to the price of volatility

risk being consistent between the two classes of assets when we look at options with one month to maturity. This is consistent with a risk-based explanation of the price of market-wide volatility. At the same time the finding is not trivial given the literature on volatility risk premia mentioned above.

Recently *Dew-Becker et al.* (2014) show that the returns to synthetic as well as traded variance swaps on major stock market indices exhibit a strongly downward sloping term structure. More specifically, *Dew-Becker et al.* (2014) find that investors are willing to pay a large premium for short-maturity variance swaps. However, for variance swaps of more than two months to maturity, the additional premium investors are willing to pay is very small. This suggests that investors are only willing to pay a premium for insurance against short term market volatility. In a related paper, *Andries et al.* (2014) show that in a model with stochastic market volatility and investors with horizon-dependent risk aversion, the term structure of the price of volatility in a Heston model exhibits a similar term structure. *Andries et al.* (2014) use S&P 500 index options and the Heston option pricing models to study the term structure of prices of volatility.

In this section I study whether this volatility term structure pattern is evident when applying linear models to de-leveraged option returns. I use both index options and individual options because both *Andries et al.* (2014) as well as *Dew-Becker et al.* (2014) find evidence for downward sloping term structure of volatility risk. Neither of these papers uses the traditional linear model typically used to estimate prices of risk factors from the cross-section of returns. *Constantinides et al.* (2013) do estimate a linear model and find a negative price of volatility risk from the cross-section of returns on index options. However, they use options of maturities 30, 60 and 90 days with various moneyness as their pooled set of test assets. In doing so, they do not look at the possibility that the prices of risk may vary across times-to-maturity.

To be consistent with the methods applied in *Constantinides et al.* (2013) I form

portfolios of index options by by updating the leverage adjustment on a daily basis. following *Constantinides et al.* (2013) I moneyness set target moneyness levels and target times to maturity. I then weight realized, leverage-adjusted option returns using a bivariate Gaussian weighting kernel in target moneyness and target time-to-maturity, with bandwidths of 10 days for time-to-maturity and 0.0125 moneyness. I follow *Constantinides et al.* (2013) by setting target moneyness values at $.9, .925, .95, .975, 1, 1.025, 1.05, 1.075$ and 1.10. Since index options typically have much more trading volume than options on individual stocks, they have a more dense set of strike prices and hence options are available at more moneyness levels for index options than individual options. As such, setting such a wide range of target moneyness values does not leave us with sparsely populated portfolios. The target times to maturity are set at 30 days, 60 days, 90 days, 120 days, 150 days and 180 days.

Next I form my set of test assets from returns on individual options. In order to be most consistent with the test assets used for index option portfolios, I perform leverage adjustment on a daily basis. I break the individual options into moneyness bins in exactly the same way described earlier. I also value weight the returns in each portfolio where value is measured by market capitalization of equity for the underlying firm. In order to further be consistent with the results in *Constantinides et al.* (2013), I estimate the prices of risk associated test assets' sensitivity to risk factors. I perform the standard two-step cross-sectional regression to do so. As in *Constantinides et al.* (2013), I estimate standard errors by bootstrapping the cross-sectional test procedure with 10,000 bootstrap draws from the dates used in the sample. Thus the procedure assumes independent and identical time series distributions but allows for heterskedasticity in the cross-section. Bootstrapping the standard errors easily allows me to estimate standard errors for the difference in estimated price of risk between index option returns and individual option returns.

Tables 1.21 and 1.22 give results of the cross-sectional regressions for index and

individual options respectively. Since close to 99% of the variation in the returns of index option portfolios is explained by the first two principle components of the test asset returns, I only test a two factor model with the market and volatility factors when using index options as test assets. For the individual options I test the two factor model as well as a 5 factor model which includes the addition of SMB, HML and momentum factors. In the individual options, volatility risk, measured by changes in the VIX, carries a significant price of volatility risk for options of all maturities. The price of market-wide volatility risk does not seem to fluctuate much across option maturities. For the index options, however, I find results similar to those described in *Andries et al.* (2014) and *Dew-Becker et al.* (2014). For index options with time to maturity of 1,2 and 3 months, there is not much difference between the prices of volatility extracted from index options and that extracted from individual options.

Since the term structure of volatility risk is more or less flat for the individual options but is decreasing for the index options, I next examine the term structure of the disparity between the price of risk in the two markets. The top panel of Figure 1.8 shows point estimates for the price of volatility risk separately for the index options and the individual option portfolios in the case of the two factor model. I only compare the two factor model results because the index option returns are driven by 2 principle components. In order to have a meaningful comparison, I thus compare the prices from the two factor model for both types of test assets. The bottom panel shows the difference between the point estimates as well as a two standard error confidence band around the point estimates at each maturity. Standard errors are calculated by bootstrapping the cross-sectional regression procedure with 10,000 bootstrap draws. It is clear from the figure that for maturities of 4, 5 and 6 months, the difference in prices is statistically significant. In addition to Figure 1.8, Table 1.23, gives the point estimates and standard errors for differences in the price of volatility risk.

One possible explanation for such a disparity in price of volatility between index

and individual options making up the index is correlation risk. The variance and volatility of the index is composed of two parts: the weighted sum of variances of constituent stocks and the weighted cross-covariances of all pairs of stocks. I follow *Driessen et al.* (2009) who construct a correlation factor based upon the simplifying assumption that all pairwise correlation between stocks is the same regardless of which two stocks we look at. Following this assumption, I construct a correlation factor and use this factor to control for correlation risk. If correlation risk is driving the disparity between prices of volatility risk in index and individual options as suggested in *Driessen et al.* (2009), then the differences in estimated price of risk displayed in Figure 1.8 should disappear once we control for correlation risk. However, Figure 1.8 shows that the significant difference between prices in each of the cross sections still persists even when we control for correlation risk. This further complicates any risk-based explanation for why the term structure of volatility risk appears to be negative when estimated from index options.

1.8 Conclusion

Volatility is generally accepted as playing an important role in determining prices of options. The evidence of a volatility premium in the index options market is well documented. In addition, the growing literature on individual stock option returns is largely comprised of papers examining volatility characteristics and their relation to returns on options. The well documented differences in the volatility and variance risk premia between index options and individual stock options (see *Driessen et al.* (2009) and *Bakshi and Kapadia* (2003a)) suggests that the volatility risk premium inherent in index options may not necessarily translate to a similar premium existing in the cross-section of individual options. In fact, *Duarte and Jones* (2007) find that volatility risk is not significantly priced unconditionally in the cross-section of individual option returns and *Di Pietro and Vainberg* (2006) find volatility risk has the

opposite sign in the cross-section of synthetic variance swaps as in the cross-section of stock returns.

Until now evidence had suggested that market-wide volatility may not be priced in individual stock options. I find that there is strong evidence of a significant market-wide volatility risk factor in the pricing kernel for options on individual stocks. This factor is economically and statistically very significant. My results lend support to recent papers like *Dittmar and Lundblad* (2014), *Boguth and Kuehn* (2013), *Campbell et al.* (2012) and *Bansal et al.* (2013) all of which suggest volatility is a priced state variable in the ICAPM sense. If volatility is a state variable in the ICAPM sense, it should be priced in the cross-section of individual options as well. The results of this paper thus make the make plausible the argument for volatility as a state factor.

I present strong evidence that the term-structure of volatility risk differs between individual options and index options. Namely, individual options embed a negative price of volatility risk that is constant across maturities from one to six months. On the other hand, the term structure of volatility risk in the index options market is downward sloping and at maturities of 4,5 and 6 months the price of volatility risk is significantly greater in magnitude for individual options than it is for index options. I further find evidence that the price of market-wide volatility risk is greater in the the options than in the underlying stocks. This suggests that options are not redundant securities. Furthermore, it suggests that a potential reason for the existence of the option market may be as a market for hedging market-wide volatility risk.

Table 1.1: Options Sample

This table gives the number of option contracts considered in our sample for each of the six call/put and moneyness bins over the 200 month sample from January 1997 through August 2013. There are a total of 599,803 options in the filtered data.

Number of Options

	OTM	ATM	ITM
Calls	93,658	127,423	77,348
Puts	101,925	107,419	92,030

Table 1.2: Option Leverage Estimates

This table gives summary statistics for the Black-Scholes-Merton estimates of leverage in individual stock options in the sample.

Option Leverage

	OTM		ATM		ITM	
	mean	**std dev**	**mean**	**std dev**	**mean**	**std dev**
Calls	18.31	7.33	14.79	6.43	8.16	3.48
Puts	-15.37	6.65	-13.03	6.20	-7.36	3.83

Table 1.3: Summary statistics for 36 value-weighted option portfolios
This table reports summary statistics for each of the 36 value-weighted option portfolios. Columns represent OTM, ATM or ITM calls and puts. Rows represent portfolios sorted by implied volatility premium (IVP) within each moneyness, option-type portfolio; IVP1 denotes portfolio with the smallest implied volatility premium while IVP6 represents the portfolio with largest implied volatility premium. Mean and volatility are reported in terms of annualized returns in percent. Skewness and kurtosis are measures of monthly holding period returns. The sample covers 200 months, from January 1997 through August 2013.

	Calls			Puts			Calls			Puts		
	OTM	ATM	ITM	OTM	ATM	ITM	OTM	ATM	ITM	OTM	ATM	ITM
IVP	**A. Mean (%)**						**B. Volatility (%)**					
IVP1	3.963	5.818	4.799	4.525	5.532	8.654	33.245	25.593	24.323	42.037	34.630	29.701
IVP2	3.190	5.148	6.082	13.427	8.392	8.466	21.402	19.201	17.955	29.903	26.582	22.706
IVP3	4.704	5.437	4.905	16.617	11.619	13.673	19.859	16.802	17.207	29.806	23.699	21.198
IVP4	3.224	6.281	8.046	17.635	14.150	13.560	18.007	17.121	17.703	29.935	25.196	21.541
IVP5	-5.258	4.395	7.737	25.768	22.448	13.171	18.728	20.395	20.362	31.417	26.620	26.335
IVP6	-14.171	-2.688	1.223	33.907	24.271	25.377	22.941	24.516	26.332	35.676	34.025	28.949
	C. Skewness						**D. Kurtosis**					
IVP1	5.777	2.611	1.759	-3.343	-2.330	-1.544	9.153	6.475	4.980	11.108	9.838	6.732
IVP2	2.073	1.291	0.315	-3.144	-2.227	-1.306	9.630	6.677	3.935	10.415	10.161	6.168
IVP3	1.993	0.760	0.163	-3.520	-1.902	-1.374	8.812	4.162	3.763	12.348	8.491	6.034
IVP4	1.083	0.481	0.015	-3.706	-2.085	-1.496	4.008	3.572	4.009	14.615	9.633	7.398
IVP5	1.669	0.925	-0.152	-3.057	-2.210	-1.334	7.677	5.343	4.423	11.579	11.809	7.446
IVP6	1.708	0.775	0.030	-2.248	-1.478	-0.929	7.936	4.741	4.073	7.201	8.092	6.448
	E. CAPM beta						**F. Volatility beta (2 factor model)**					
IVP1	0.874	0.886	0.943	1.589	1.456	1.292	0.545	0.445	0.238	-0.610	-0.394	-0.191
IVP2	0.612	0.726	0.779	1.228	1.179	1.063	0.306	0.180	0.012	-0.616	-0.421	-0.240
IVP3	0.622	0.650	0.766	1.174	1.089	0.988	0.197	0.145	-0.053	-0.663	-0.419	-0.321
IVP4	0.571	0.720	0.816	1.231	1.123	1.009	0.232	0.109	-0.050	-0.697	-0.521	-0.374
IVP5	0.571	0.796	0.913	1.247	1.150	1.130	0.072	0.123	-0.072	-0.851	-0.635	-0.441
IVP6	0.617	0.925	1.073	1.194	1.338	1.180	0.225	0.092	-0.038	-0.720	-0.657	-0.451

Table 1.4: Summary statistics for 36 option portfolios without leverage adjustment
This table reports summary statistics for each of the 36 value-weighted option portfolios without leverage adjusting returns. Columns represent OTM, ATM or ITM calls and puts. Rows represent portfolios sorted by implied volatility premium (IVP) within each moneyness, option-type portfolio; IVP1 denotes portfolio with the smallest implied volatility premium while IVP6 represents the portfolio with largest implied volatility premium. Mean and volatility are reported in terms of annualized returns in percent. Skewness and kurtosis are measures of monthly holding period returns. The sample covers 200 months, from January 1997 through August 2013.

	Calls			Puts			Calls			Puts		
	OTM	ATM	ITM	OTM	ATM	ITM	OTM	ATM	ITM	OTM	ATM	ITM
IVP	**A. Mean (%)**						**B. Volatility (%)**					
IVP1	7.80	56.5	36.8	-31.0	-70.0	-57.4	415.4	294.9	189.3	452.3	344.6	217.3
IVP2	115	89.3	48.3	-87.6	-71.3	-64.5	436.7	278.1	187.1	490.1	342.5	212.8
IVP3	12.4	103	52.6	-79.1	-67.6	-74.8	453.3	275.8	179.3	536.8	331.1	208.9
IVP4	10.2	141	79.0	-71.1	-69.3	-48.2	524.2	309.5	191.5	437.0	327.1	210.0
IVP5	33.7	36.0	29.0	-81.5	-74.3	-55.7	469.6	297.4	194.1	497.7	311.3	200.2
IVP6	20.2	16.7	41.4	-90.1	-89.4	-67.4	497.1	302.7	190.6	392.7	269.3	167.4
	C. Skewness						**D. Kurtosis**					
IVP1	1.946	1.255	0.307	3.422	2.445	1.453	6.837	4.972	2.658	17.844	11.494	5.957
IVP2	1.823	0.770	0.206	2.876	1.959	1.388	6.726	3.127	2.582	11.975	7.001	4.973
IVP3	2.481	0.762	0.083	3.646	1.999	1.376	10.814	3.222	2.571	20.436	7.659	5.374
IVP4	3.097	1.062	0.165	3.020	1.653	1.184	15.095	3.909	2.397	13.469	5.244	4.311
IVP5	2.391	1.228	0.461	2.663	1.746	1.012	10.061	4.831	3.470	10.699	6.199	3.754
IVP6	5.062	1.306	0.704	3.258	1.652	0.741	42.658	5.092	5.027	16.507	6.056	3.148
	E. CAPM beta						**F. Volatility beta (2 factor model)**					
IVP1	11.154	8.840	6.781	-15.927	-13.181	-8.574	3.112	0.790	-0.294	7.258	5.095	2.646
IVP2	12.200	9.840	7.586	-18.522	-14.519	-9.342	4.592	1.553	-0.826	9.537	6.144	3.249
IVP3	11.140	10.413	7.694	-20.098	-14.046	-9.051	2.134	2.289	-0.090	10.381	6.660	3.306
IVP4	10.571	8.720	8.816	11.231	-11.123	-12.009	4.522	2.981	0.582	9.375	5.150	1.472
IVP5	10.571	9.796	8.913	10.247	-12.150	-10.130	6.742	2.879	0.741	7.192	4.030	1.334
IVP6	7.617	8.925	6.073	10.194	-10.338	-9.180	2.784	3.398	1.004	5.654	2.610	1.551

Table 1.5: Summary statistics for stock portfolios

This table reports summary statistics for the stock portfolios formed according to the double sorting procedure. Where the first sort is by β_M, each stock's market beta. The second sort is by $\beta_{\Delta VIX}$, stock loading on changes in the VIX. Mean and volatility are reported in terms of annualized returns in percent. Skewness and kurtosis are measures of monthly holding period returns. The sample includes all CRSP stocks and covers 200 months, from January 1997 through August 2013.

Stock Portfolios

	β_M		β_M	
	(1)	(2)	(1)	(2)
$\beta_{\Delta VIX}$	**A. Mean (%)**		**B. Volatility (%)**	
(1)	11.023	12.499	18.526	30.168
(2)	7.862	7.785	14.912	24.610
(3)	7.987	8.718	14.181	22.403
(4)	6.208	10.955	14.604	24.414
(5)	8.524	11.884	15.761	26.514
(6)	7.494	4.996	22.102	33.240
	C. Skewness		**D. Kurtosis**	
(1)	-0.660	-0.908	4.918	6.458
(2)	-1.078	-1.208	4.936	7.320
(3)	-1.429	-0.861	7.186	7.094
(4)	-1.354	-0.232	7.263	5.967
(5)	-1.126	-0.431	5.801	5.600
(6)	-1.307	-0.416	6.504	5.115
	E. CAPM beta		**F. Volatility beta**	
(1)	0.810	1.417	-0.094	-0.074
(2)	0.679	1.210	-0.082	0.055
(3)	0.654	1.115	-0.097	0.013
(4)	0.661	1.209	-0.158	0.008
(5)	0.716	1.289	-0.076	0.040
(6)	0.945	1.535	-0.101	0.149

Table 1.6: Risk factor correlations

This table presents correlations between the risk factors examined in the paper. Construction of the factors is described in Section 1.2.4. The sample covers 200 months, from January 1997 through August 2013.

Factor Correlations

	MKT	SMB	HML	Mom	VOL	$VOL^{\perp}$	DS	Skew	Jump
MKT	1.000								
SMB	0.304	1.000							
HML	-0.092	-0.148	1.000						
Mom	-0.348	-0.048	-0.305	1.000					
VOL	-0.777	-0.218	-0.002	0.284	1.000				
$VOL^{\perp}$	0.000	0.028	-0.117	0.022	0.630	1.000			
DS	0.755	0.260	0.046	-0.215	-0.693	-0.168	1.000		
Skew	-0.213	-0.079	-0.114	-0.040	0.339	0.275	-0.673	1.000	
Jump	0.480	0.224	0.141	-0.119	-0.456	-0.131	0.674	-0.602	1.000
VJ	-0.501	-0.281	-0.103	0.120	0.502	0.178	-0.621	0.368	-0.700

Table 1.7: Linear GMM Tests with 36 Option Portfolios

This table reports results of GMM tests of linear pricing kernels using all 36 options portfolios as test assets. Each row represents a model and columns represent factors included in the model. The point estimates are reported along with t-statistics in parentheses that are computed using Newey-West adjusted standard errors with 6-month lags. The final two columns give the J-statistic with corresponding asymptotic p-value in [brackets] and the Hansen-Jagannathan measure of distance from the space of valid stochastic discount factors.

All 36 Options Portfolios

	intercept	MKT	SMB	HML	MOM	VOL	Jstat	HJ dist
(1)	0.960					13.915	141.749	0.671
	(40.172)					(5.399)	[0.000]	
(2)	1.014	-0.026					167.332	0.708
	(92.067)	(-2.802)					[0.000]	
(3)	0.962	0.015				14.084	139.063	0.684
	(38.065)	(1.303)				(5.340)	[0.000]	
(4)	1.081	-0.019	0.941	-250.931	9.075		128.420	0.813
	(13.471)	(-0.776)	(0.132)	(-3.244)	(2.666)		[0.000]	
(5)	0.991	0.006	-0.153	-55.246	3.260	17.536	73.384	0.668
	(9.249)	(0.221)	(-0.024)	(-1.068)	(1.577)	(3.660)	[0.000]	

Table 1.8: Linear GMM Tests ATM Portfolios

This table reports results of GMM tests of pricing kernels using the combination of 6 ATM put portfolios and 6 ATM call portfolios as test assets. Each row represents a model and columns represent factors included in the model. The point estimates are reported along with t-statistics in parentheses that are computed using Newey-West adjusted standard errors with 6-month lags. The final two columns give the J-statistic with corresponding asymptotic p-value in [brackets] and the Hansen-Jagannathan measure of distance from the space of valid stochastic discount factors.

ATM calls and puts

	intercept	MKT	SMB	HML	MOM	VOL	Jstat	HJ dist
(1)	1.021					15.982	16.714	0.324
	(23.410)					(3.498)	[0.117]	
(2)	1.010	-0.019					26.991	0.388
	(88.323)	(-1.528)					[0.005]	
(3)	1.020	-0.004				15.718	17.123	0.322
	(23.772)	(-0.212)				(3.364)	[0.072]	
(4)	1.182	0.003	-14.857	-179.253	-0.950		29.922	0.385
	(14.540)	(0.098)	(-1.549)	(-2.348)	(-0.167)		[0.000]	
(5)	0.981	-0.037	2.896	39.551	-6.911	15.278	12.277	0.325
	(15.764)	(-1.452)	(0.374)	(0.456)	(-1.674)	(2.288)	[0.092]	

Table 1.9: Linear GMM Tests for Stocks

This table reports results of GMM tests of pricing kernels using the 12 stock portfolios as test assets. Each row represents a model and columns represent factors included in the model. The point estimates are reported along with t-statistics in parentheses that are computed using Newey-West adjusted standard errors with 6-month lags. The final two columns give the J-statistic with corresponding asymptotic p-value in [brackets] and the Hansen-Jagannathan measure of distance from the space of valid stochastic discount factors.

Stock Portfolios

	intercept	MKT	SMB	HML	MOM	VOL	Jstat	HJ dist
(1)	0.992					12.005	28.838	0.297
	(20.707)					(2.011)	[0.002]	
(2)	1.011	-0.025					25.149	0.306
	(74.416)	(-1.974)					[0.009]	
(3)	1.008	-0.047				11.842	25.139	0.335
	(25.507)	(-2.298)				(1.768)	[0.005]	
(4)	1.061	-0.062	2.107	-35.335	-4.288		17.426	0.363
	(20.830)	(-2.045)	(0.449)	(-0.516)	(-1.284)		[0.026]	
(5)	1.006	-0.024	9.532	139.756	5.472	11.644	9.287	0.312
	(9.490)	(-0.534)	(1.127)	(1.017)	(0.939)	(1.513)	[0.233]	

Table 1.10: Linear GMM Tests for ATM calls and puts without leverage adjustment
This table reports results of GMM tests of pricing kernels using the 12 ATM portfolios without leverage adjustment as test assets. The portfolios are sorted in the same manner as the leverage adjusted portfolios and the weighting is the same. However, the returns are not leverage adjusted. Each row represents a model and columns represent factors included in the model. The point estimates are reported along with t-statistics in parentheses that are computed using Newey-West adjusted standard errors with 6-month lags. The final two columns give the J-statistic with corresponding asymptotic p-value in [brackets] and the Hansen-Jagannathan measure of distance from the space of valid stochastic discount factors.

ATM Portfolios without leverage adjustment

	intercept	MKT	SMB	HML	MOM	VOL	Jstat	HJ dist
(1)	0.934					24.007	15.503	0.520
	(7.302)					(0.861)	[0.041]	
(2)	1.034	-0.023					17.319	0.395
	(34.491)	(-1.346)					[0.003]	
(3)	1.038	-0.024				0.996	15.489	0.398
	(28.314)	(-1.361)				(0.111)	[0.045]	
(4)	1.010	-0.019	2.089	58.815	1.404		12.826	0.401
	(13.530)	(-0.738)	(0.306)	(0.549)	(0.181)		[0.008]	
(5)	0.995	-0.017	1.981	79.316	0.781	3.869	10.396	0.394
	(16.010)	(-0.684)	(0.295)	(0.798)	(0.092)	(0.490)	[0.055]	

Table 1.11: Linear GMM Tests for Combined Stock Portfolios and ATM Options
This table reports results of GMM tests of pricing kernels using the 12 ATM option portfolios combined with the 12 stock portfolios as test assets. Each row represents a model and columns represent factors included in the model. The point estimates are reported along with t-statistics in parentheses that are computed using Newey-West adjusted standard errors with 6-month lags. The final two columns give the J-statistic with corresponding asymptotic p-value in [brackets] and the Hansen-Jagannathan measure of distance from the space of valid stochastic discount factors.

12 ATM Options Portfolios and 12 Stock Portfolios

	intercept	MKT	SMB	HML	MOM	VOL	Jstat	HJ dist
(1)	0.981					8.262	88.983	0.505
	(57.297)					(3.165)	[0.000]	
(2)	1.023	-0.038					84.117	0.560
	(63.870)	(-3.416)					[0.000]	
(3)	0.994	-0.030				9.052	83.737	0.504
	(44.188)	(-2.204)				(3.000)	[0.000]	
(4)	1.121	-0.068	2.004	-182.055	-4.373		74.183	0.607
	(21.060)	(-3.891)	(0.517)	(-3.611)	(-1.868)		[0.000]	
(5)	1.031	-0.046	0.801	-95.068	-1.797	9.460	73.399	0.523
	(22.574)	(-2.588)	(0.234)	(-2.182)	(-0.945)	(2.828)	[0.000]	

Table 1.12: GMM tests 36 option portfolios and Exponentially Affine SDF
This table reports results of GMM tests of exponentially affine pricing kernels using all 36 option portfolios. This is the only table in the paper that reports results for the 1-step GMM with an identity weighting matrix. I use this test instead of the 2-step GMM for this particular test in order to avoid problems associated with multiple-step GMM estimation of non-linear models when the time series of observations is not long compared to the number of test assets. Each row represents a model and columns represent factors included in the model. The point estimates are reported along with t-statistics in parentheses that are computed using Newey-West adjusted standard errors with 6-month lags. The final two columns give the J-statistic with corresponding asymptotic p-value in [brackets] and the Hansen-Jagannathan measure of distance from the space of valid stochastic discount factors.

36 Options Exponentially Affine SDF

	intercept	MKT	SMB	HML	MOM	VOL	Jstat	HJ dist
(1)	-0.068					10.578	164.526	0.669
	-(1.026)					(1.878)	[0.000]	
(2)	-0.000	-0.035					167.256	0.680
	(-0.045)	(-1.601)					[0.000]	
(3)	-0.063	-0.002				10.230	163.140	0.678
	(-1.010)	(-0.072)				(1.966)	[0.000]	
(4)	-0.075	-0.012	1.034	-57.061	4.521		187.225	0.935
	(-1.114)	(-0.516)	(0.119)	(-0.560)	(1.110)		[0.000]	
(5)	-0.259	-0.001	11.357	68.002	2.679	15.457	127.853	0.694
	(-1.573)	(-0.036)	(1.196)	(0.962)	(0.774)	(2.748)	[0.000]	

Table 1.13: GMM tests ATM option portfolios and Exponentially Affine SDF
This table reports results of GMM tests of exponentially affine pricing kernels using portfolios the 12 ATM option portfolios. Each row represents a model and columns represent factors included in the model. The point estimates are reported along with t-statistics in parentheses that are computed using Newey-West adjusted standard errors with 6-month lags. The final two columns give the J-statistic with corresponding asymptotic p-value in [brackets] and the Hansen-Jagannathan measure of distance from the space of valid stochastic discount factors.

ATM Options Exponentially Affine SDF

	intercept	MKT	SMB	HML	MOM	VOL	Jstat	HJ dist
(1)	-0.089					12.518	19.560	0.330
	-(1.767)					(2.999)	[0.052]	
(2)	0.004	-0.017					26.480	0.386
	(0.677)	(-1.492)					[0.006]	
(3)	-0.085	-0.002				12.219	19.700	0.329
	(-1.705)	(-0.117)				(2.918)	[0.032]	
(4)	-0.341	-0.036	-22.754	-303.736	-12.031		23.305	0.458
	(-1.165)	(-1.767)	(-2.478)	(-4.172)	(-2.989)		[0.003]	
(5)	-0.505	0.002	-22.518	-160.048	-11.553	16.424	8.434	0.378
	(-1.655)	(0.071)	(-2.272)	(-1.304)	(-2.069)	(2.379)	[0.296]	

Table 1.14: GMM tests 12 Stock Portfolios and Exponentially Affine SDF
This table reports results of GMM tests of exponentially affine pricing kernels using the 12 stock portfolios. Each row represents a model and columns represent factors included in the model. The point estimates are reported along with t-statistics in parentheses that are computed using Newey-West adjusted standard errors with 6-month lags. The final two columns give the J-statistic with corresponding asymptotic p-value in [brackets] and the Hansen-Jagannathan measure of distance from the space of valid stochastic discount factors.

12 Stock Portfolios

	intercept	MKT	SMB	HML	MOM	VOL	Jstat	HJ dist
(1)	-0.013					-3.241	35.386	0.297
	(-0.246)					(-0.232)	[0.006]	
(2)	-0.001	-0.010					26.088	0.310
	(-0.151)	(-0.632)					[0.073]	
(3)	-0.033	-0.010				7.535	25.454	0.347
	(-0.420)	(-0.585)				(0.814)	[0.062]	
(4)	-0.001	-0.019	2.300	-8.952	-0.538		25.046	0.367
	(-0.028)	(-0.896)	(0.448)	(-0.123)	(-0.185)		[0.034]	
(5)	-0.136	-0.001	2.086	59.796	2.487	12.436	22.345	0.452
	(-0.871)	(-0.048)	(0.361)	(0.859)	(0.650)	(1.445)	[0.050]	

Table 1.15: GMM Tests with Exponentially Affine SDF

This table reports results of GMM tests of exponentially affine pricing kernels using portfolios the 12 ATM option portfolios combined with the 12 stock portfolios. Each row represents a model and columns represent factors included in the model. The point estimates are reported along with t-statistics in parentheses that are computed using Newey-West adjusted standard errors with 6-month lags. The final two columns give the J-statistic with corresponding asymptotic p-value in [brackets] and the Hansen-Jagannathan measure of distance from the space of valid stochastic discount factors.

12 ATM Option Portfolios and 12 Stock Portfolios

	intercept	MKT	SMB	HML	MOM	VOL	Jstat	HJ dist
(1)	-0.053					6.301	85.832	0.516
	(-2.947)					(2.184)	[0.000]	
(2)	0.004	-0.031					86.080	0.554
	(0.643)	(-2.984)					[0.000]	
(3)	-0.046	-0.019				6.150	83.037	0.512
	(-2.757)	(-1.733)				(2.092)	[0.000]	
(4)	-0.036	-0.043	4.907	-120.448	-2.315		81.892	0.636
	(-0.852)	(-3.061)	(0.958)	(-2.619)	(-1.116)		[0.000]	
(5)	-0.105	-0.031	4.972	-97.279	-1.535	8.349	67.327	0.556
	(-2.146)	(-2.053)	(0.875)	(-1.794)	(-0.659)	(2.665)	[0.000]	

Table 1.16: Linear GMM tests with Tail Risk

This table reports results of GMM tests of pricing kernels using all 36 options portfolios as test assets. Each row represents a model and columns represent factors included in the model. The point estimates are reported along with t-statistics in parentheses that are computed using Newey-West adjusted standard errors with 6-month lags. The final two columns give the J-statistic with corresponding asymptotic p-value in [brackets] and the Hansen-Jagannathan measure of distance from the space of valid stochastic discount factors.

All 36 Options Portfolios

	intercept	MKT	SMB	HML	MOM	VOL	DS	SKEW	JUMP	VOL JUMP	Jstat	HJ dist
(1)	1.069	-0.052				19.435	13.801				63.047	0.670
	(17.045)	(-1.965)				(5.520)	(3.454)				[0.001]	
(2)	0.985	-0.021				13.326		-6.255			59.437	0.685
	(11.144)	(-0.866)				(3.033)		(-1.709)			[0.003]	
(3)	1.013	-0.008				17.739			0.645		91.861	0.686
	(19.489)	(-0.491)				(5.570)			(0.606)		[0.000]	
(4)	1.115	-0.020				17.256				-6.920	95.473	0.686
	(18.961)	(-1.076)				(5.710)				(-3.107)	[0.000]	
(5)	1.126	-0.066	-2.818	8.808	2.838	21.913	13.933				43.186	0.633
	(10.086)	(-1.805)	(-0.481)	(0.144)	(1.573)	(3.937)	(2.450)				[0.056]	
(6)	0.835	-0.008	-5.807	15.057	3.514	17.502		-7.773			48.943	0.699
	(9.206)	(-1.213)	(-1.280)	(0.154)	(1.569)	(3.064)		(-1.850)			[0.141]	
(7)	1.093	-0.011	1.002	-67.713	2.490	18.722			0.655		63.720	0.683
	(8.573)	(-0.355)	(0.154)	(-1.176)	(0.824)	(3.239)			(0.424)		[0.000]	
(8)	1.146	-0.029	1.671	-9.296	2.904	16.517				-5.846	49.337	0.647
	(13.055)	(-0.999)	(0.286)	(-0.162)	(1.461)	(3.711)				(-1.386)	[0.015]	

Table 1.17: Linear GMM Tests with Tail Risk

This table reports results of GMM tests of pricing kernels using the combination of 6 ATM put portfolios and 6 ATM call portfolios as test assets. Each row represents a model and columns represent factors included in the model. The point estimates are reported along with t-statistics in parentheses that are computed using Newey-West adjusted standard errors with 6-month lags. The final two columns give the J-statistic with corresponding asymptotic p-value in [brackets] and the Hansen-Jagannathan measure of distance from the space of valid stochastic discount factors.

ATM Portfolios

	intercept	MKT	SMB	HML	MOM	VOL	DS	SKEW	JUMP	VOL JUMP	Jstat	HJ dist
(1)	1.136	-0.047				19.843	10.239				15.288	0.291
	(7.690)	(-1.011)				(2.245)	(1.081)				[0.083]	
(2)	0.983	-0.003				17.005		-0.318			19.472	0.290
	(10.972)	(-0.140)				(2.688)		(-0.144)			[0.021]	
(3)	0.994	-0.010				16.040			-5.008		17.597	0.325
	(18.490)	(-0.689)				(2.409)			(-2.132)		[0.057]	
(4)	1.129	-0.023				14.137				-3.710	16.065	0.315
	(10.510)	(-1.041)				(3.062)				(-0.986)	[0.066]	
(5)	1.174	-0.084	5.997	72.356	-1.979	24.059	15.124				8.504	0.293
	(6.358)	(-1.492)	(0.649)	(0.612)	(-0.302)	(2.108)	(1.260)				[0.203]	
(6)	0.975	-0.005	-0.120	-2.903	3.374	24.108		-4.972			9.871	0.298
	(10.597)	(-0.233)	(-0.010)	(-0.033)	(0.515)	(2.550)		(-1.870)			[0.218]	
(7)	0.933	-0.023	3.650	13.343	-9.402	11.971			-0.888		16.309	0.319
	(10.392)	(-0.877)	(0.395)	(0.152)	(-2.419)	(1.935)			(-0.592)		[0.012]	
(8)	0.990	-0.042	4.680	63.669	-6.074	15.034				-0.268	11.417	0.336
	(7.529)	(-1.435)	(0.527)	(0.701)	(-1.402)	(2.096)				(-0.063)	[0.076]	

Table 1.18: Linear GMM Tests with Tail Risk

This table reports results of GMM tests of pricing kernels using 12 stock portfolios as test assets. Each row represents a model and columns represent factors included in the model. The point estimates are reported along with t-statistics in parentheses that are computed using Newey-West adjusted standard errors with 6-month lags. The final two columns give the J-statistic with corresponding asymptotic p-value in [brackets] and the Hansen-Jagannathan measure of distance from the space of valid stochastic discount factors.

12 Stock Portfolios

	intercept	MKT	SMB	HML	MOM	VOL	DS	SKEW	JUMP	VOL JUMP	Jstat	HJ dist
(1)	1.025	-0.049				13.116	4.340				19.652	0.259
	(11.794)	(-1.611)				(2.077)	(0.675)				[0.186]	
(2)	0.958	-0.036				13.527		-2.353			20.648	0.383
	(19.566)	(-2.044)				(1.831)		(-1.094)			[0.148]	
(3)	0.991	-0.039				15.546			0.445		20.668	0.188
	(21.655)	(-2.209)				(2.159)			(0.448)		[0.148]	
(4)	1.001	-0.042				12.859				-1.497	19.939	0.276
	(9.096)	(-1.954)				(1.874)				(-0.386)	[0.174]	
(5)	1.022	-0.056	0.716	5.890	0.515	13.858	4.026				18.971	0.269
	(6.188)	(-0.860)	(0.174)	(0.127)	(0.163)	(2.438)	(0.379)				[0.089]	
(6)	1.009	-0.054	-4.181	-61.357	-4.002	9.870		-5.858			18.032	0.229
	(12.897)	(-1.762)	(-0.716)	(-0.732)	(-0.857)	(1.213)		(-1.282)			[0.115]	
(7)	0.981	-0.038	-0.086	-8.364	0.245	11.648			0.369		19.014	0.209
	(14.028)	(-1.264)	(-0.021)	(-0.212)	(0.103)	(1.663)			(0.293)		[0.088]	
(8)	0.969	-0.033	-0.773	-6.840	-1.276	10.114				-0.749	18.379	0.281
	(2.504)	(-0.626)	(-0.074)	(-0.058)	(-0.322	(1.595)				(-0.065)	[0.105]	

Table 1.19: Linear GMM Tests with Tail Risk

This table reports results of GMM tests of pricing kernels using the 12 ATM option portfolios combined with the 12 stock portfolios as test assets. Each row represents a model and columns represent factors included in the model. The point estimates are reported along with t-statistics in parentheses that are computed using Newey-West adjusted standard errors with 6-month lags. The final two columns give the J-statistic with corresponding asymptotic p-value in [brackets] and the Hansen-Jagannathan distance measure.

12 ATM Options Portfolios and 12 Stock Portfolios

	intercept	MKT	SMB	HML	MOM	VOL	DS	SKEW	JUMP	VOL JUMP	Jstat	HJ dist
(1)	1.116	-0.070				15.024	14.377				67.400	0.488
	(14.535)	(-2.592)				(3.819)	(2.772)				[0.000]	
(2)	0.914	-0.018				16.228		-6.837			60.283	0.503
	(26.498)	(-1.401)				(4.171)		(-2.990)			[0.000]	
(3)	1.055	-0.046				10.216			0.417		91.981	0.512
	(14.670)	(-2.069)				(2.575)			(0.262)		[0.000]	
(4)	0.986	-0.036				8.664				-1.638	71.371	0.499
	(11.851)	(-2.298)				(2.812)				(-0.532)	[0.000]	
(5)	1.098	-0.077	5.106	-154.740	0.127	12.847	13.863				49.322	0.565
	(8.849)	(-2.008)	(0.867)	(-2.738)	(0.044)	(2.822)	(1.928)				[0.000]	
(6)	0.881	-0.028	0.848	-127.291	-0.673	15.976		-7.858			45.738	0.537
	(16.309)	(-1.959)	(0.240)	(-2.294)	(-0.247)	(4.210)		(-3.808)			[0.000]	
(7)	1.121	-0.067	1.725	-144.936	-3.695	7.742			0.404		71.455	0.554
	(13.555)	(-2.581)	(0.347)	(-2.613)	(-1.338)	(1.836)			(0.214)		[0.000]	
(8)	1.008	-0.052	-2.301	-96.950	-3.626	8.493				-0.476	61.052	0.532
	(10.099)	(-2.677)	(-0.937)	(-2.348)	(-1.860)	(2.403)				(-0.111)	[0.000]	

Table 1.20: GMM Likelihood Ratio-type tests

This table reports results of GMM-based likelihood ratio-type tests of restricting individual factors to be the same in both the stock SDF and the SDF estimated from option portfolios. Each pair of numbers represents a test where the model is estimated first under the restriction that the prices of risk for all risk factors in each model are the same for stocks and call options. This corresponds to the results in Table 1.11. This restriction is relaxed for volatility factor to estimate the unrestricted model. The values in the table are the test statistic and corresponding p-values in brackets. The null hypothesis is that the price of risk for the volatility factor are the same in options and stocks. The alternative is that the price of risk differs between the two markets.

Likelihood Ratio tests

One Factor Model		**Two Factor Model**		**Five Factor Model**	
Test Statistic	p-value	Test Statistic	p-value	Test Statistic	p-value
3.557	0.059	3.002	0.083	4.698	0.030

Table 1.21: Cross-sectional regressions of S&P 500 Index Options
This table reports results of cross-sectional regressions using portfolios of S&P 500 index options as test assets. The test assets include nine moneyness portfolios for calls 9 moneyness portfolios for puts. Each row represents a different time to maturity of options in the portfolios. The regressions are run separately for each time to maturity from one to six months. Columns give point estimates and corresponding standard errors for prices of risk associated with factors in the model. The point estimates are reported along with bootstrapped standard errors in parentheses that are computed using 10,000 bootstrap iterations. The final column gives the adjusted R^2 for the cross-sectional regression.

S&P 500 Index Option Portfolios

	intercept	MKT	VOL	adjusted R^2
1 month	0.04	-2.85	-4.57	0.41
	(0.01)	(1.77)	(1.28)	
2 month	0.04	-2.65	-3.89	0.39
	(0.01)	(1.88)	(1.00)	
3 month	0.02	-2.13	-2.62	0.33
	(0.01)	(1.14)	(0.74)	
4 month	0.01	-0.99	-1.65	0.35
	(0.01)	(1.15)	(0.76)	
5 month	0.01	-0.74	-1.29	0.29
	(0.01)	(1.29)	(0.85)	
6 month	0.01	1.02	-1.27	0.31
	(0.01)	(1.43)	(1.31)	

Table 1.22: Cross-sectional regressions of individual options portfolios
This table reports results of cross-sectional regressions using portfolios of S&P 500 index options as test assets. The test assets include nine moneyness portfolios for calls and 9 moneyness portfolios for puts. Each row represents a different time to maturity of options in the portfolios. The regressions are run separately for each time to maturity from one to six months. Columns give point estimates and corresponding standard errors for prices of risk associated with factors in the model. The point estimates are reported along with bootstrapped standard errors in parentheses that are computed using 10,000 bootstrap iterations. The final column gives the adjusted R^2 for the cross-sectional regression.

Individual Option Portfolios

	intercept	MKT	SMB	HML	MOM	VOL	Adjusted R^2
1 month	0.03	-2.22				-4.82	0.46
	(0.01)	(0.73)				(1.03)	
	-0.03	2.06	-9.72	2.19	-7.19	-4.91	0.47
	(0.02)	(1.61)	(2.11)	(2.17)	(3.13)	(1.24)	
2 month	0.02	-2.00				-3.83	0.43
	(0.01)	(0.61)				(0.84)	
	-0.00	-0.43	-5.14	-1.94	-5.46	-4.46	0.43
	(0.01)	(1.05)	(1.29)	(1.80)	(1.81)	(1.12)	
3 month	0.02	-1.24				-3.17	0.32
	(0.01)	(0.67)				(0.97)	
	0.02	-1.22	-0.24	2.04	1.61	-2.04	0.41
	(0.01)	(1.26)	(1.44)	(1.52)	(1.64)	(1.35)	
4 month	0.02	-1.65				-4.82	0.39
	(0.01)	(0.64)				(0.99)	
	0.02	-2.09	-1.43	1.14	-1.04	-4.91	0.43
	(0.01)	(0.91)	(1.07)	(1.71)	(1.64)	(1.25)	
5 month	0.01	-1.06				-3.61	0.42
	(0.01)	(0.62)				(0.72)	
	-0.00	-0.27	-1.44	0.60	1.19	-2.72	0.32
	(0.01)	(0.85)	(1.08)	(1.31)	(1.68)	(0.95)	
6 month	0.02	-1.33				-6.75	0.30
	(0.01)	(0.78)				(1.26)	
	0.01	-0.63	-0.47	1.01	5.89	-4.92	0.32
	(0.01)	(1.08)	(1.51)	(1.71)	(2.49)	(1.63)	

Table 1.23: Difference in Volatility Prices: Index vs. Individual Options
This table reports results of estimates of the difference between prices of volatility risk. The point estimates correspond to the difference between the point estimates for the price associated with first differences in the VIX extracted from the cross-sectional regressions of Tables 1.22 and 1.21. The standard errors are calculated from 10,000 bootstrap draws of the cross-sectional regressions.

Difference in volatility prices

	vol diff	Standard Errors
1 month	-0.25	(1.17)
2 month	0.05	(1.08)
3 month	-0.55	(0.98)
4 month	-3.17	(1.08)
5 month	-2.32	(0.93)
6 month	-5.49	(1.66)

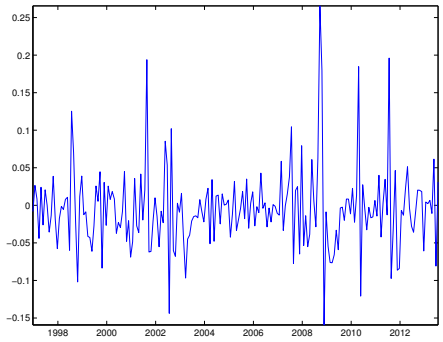
Table 1.24: Simulation Parameters
Distribution of parameters used in the simulation to study the effect of leverage adjusting option returns. Simulation results base on 1,000 underlying stocks each with 500 months of returns. Each stock has six options: one each of in-the-money calls and puts, at-the-money calls and puts and out-of-the money calls and puts.

Simulation Paramters

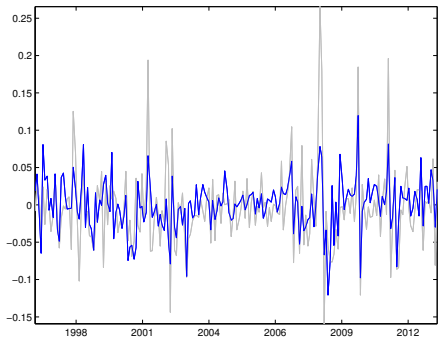
λ_M	σ_M	ϵ_M	β_i	σ_i	$\epsilon_{i,t}$
2.5	12%	$N(0, \sigma_M)$	$U[.5, 1.5]$	$N(20\%, 5\%)$	$N(0, \sigma_i)$

Simulation Results

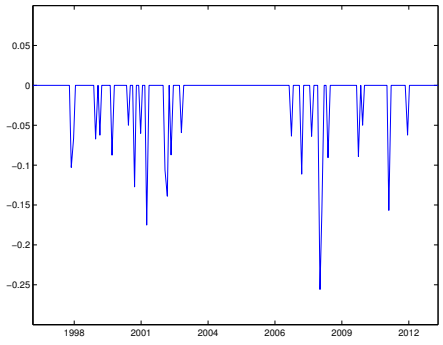
	$\hat{\lambda}_M$	sampling error
individual options	0.082	0.0136
portfolios of options	1.994	0.298
leverage-adjusted portfolios of options	2.374	0.159



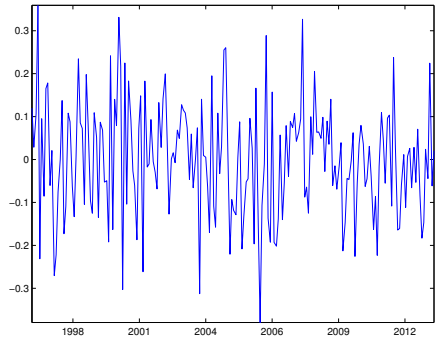
A) Volatility (innovations)



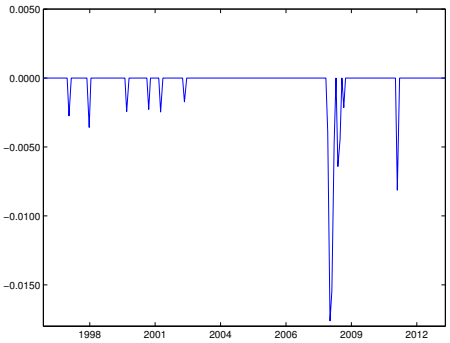
B) Orthogonal Volatility (innovations)



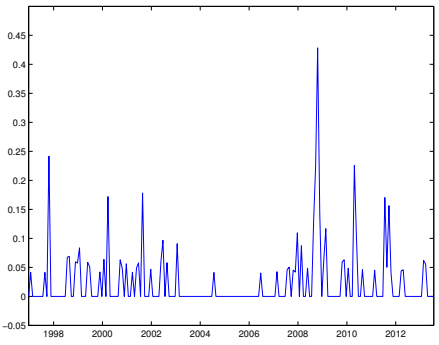
C) Down Side



D) Model-Free Skew (innovations)



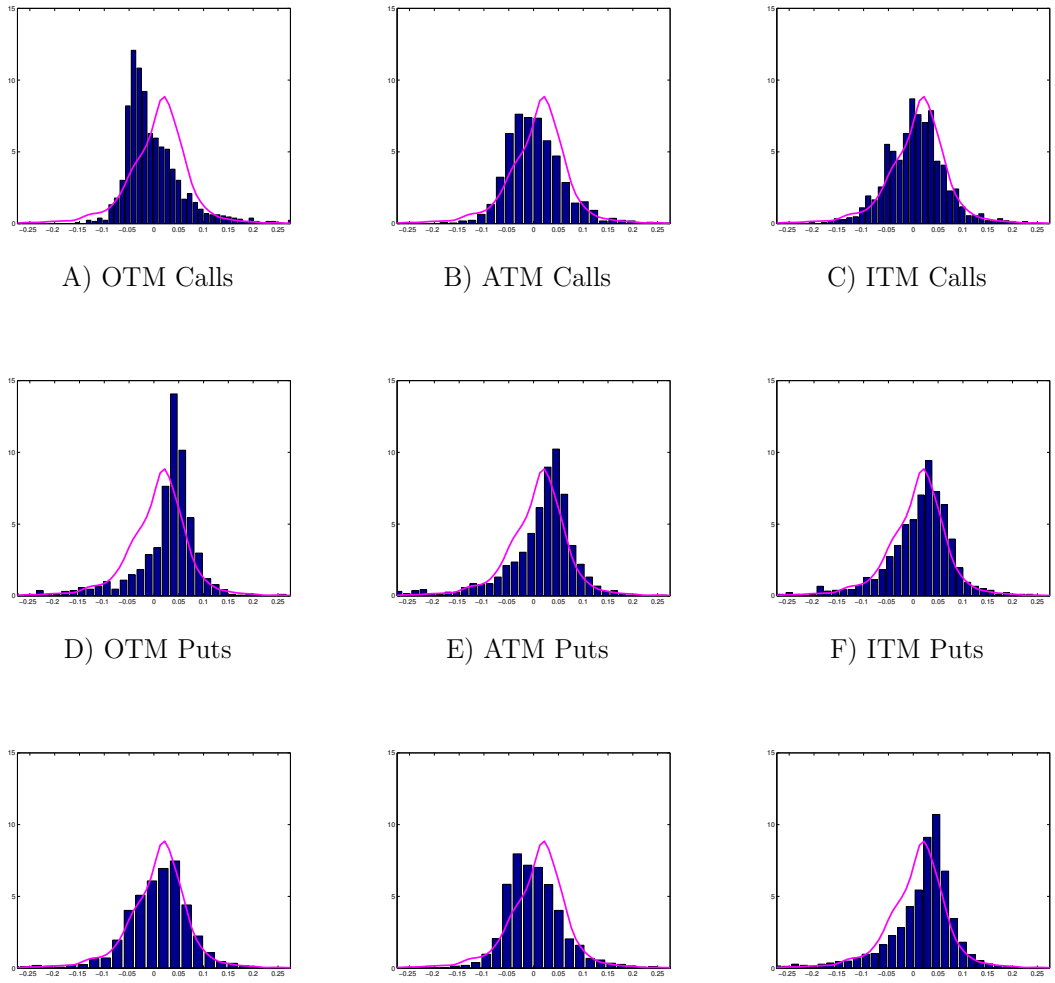
E) Jump



F) Volatility Jump

Figure 1.1: Factors

Panel A plots innovations in the VIX. Panel B plots the time series of residuals from regressing VIX innovations on market excess returns (MKT). This is the orthogonalized volatility factor used in tests throughout the paper. The time series plotted in each Panels C, D, E and F represent tail risk factors.



A) OTM Calls B) ATM Calls C) ITM Calls

D) OTM Puts E) ATM Puts F) ITM Puts

G) ATM Calls and Puts H) All calls I) All Puts

Figure 1.2: Empirical densities of moneyness, put/call portfolios
Panels A-F plot the empirical densities of OTM Calls, ATM Calls, ITM Calls, OTM Puts, ATM Puts and ITM Puts respectively. The horizontal axis measures monthly returns and the vertical axis measures density of the distribution. Each panel has a kernel density estimate of the realized returns for the 12 stock portfolios overlaying the empirical density for comparision. Each empirical density in panels A-F is composed of 1,200 observed returns; 200 monthly holding period returns from each of the 6 implied volatility premium portfolios within a given moneyness-put/call portfolio. Panel G plots the combined ATM calls and ATM puts. Panels H and I plot the empirical densities of all call portfolios and all put portfolios respectively, across all three moneyness bins.

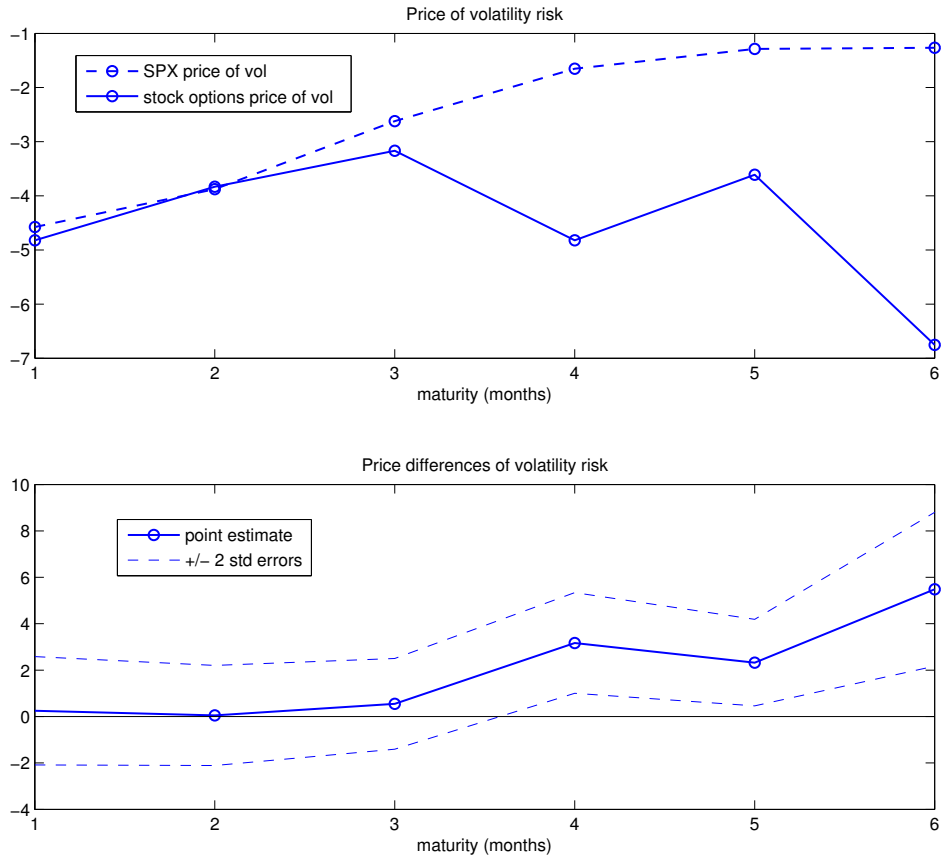


Figure 1.3: Term Structure of the Price of Volatility

Panel A shows the prices of volatility risk estimated from portfolios of both individual options as well as S&P 500 index options. The vertical axis measures the price while the horizontal axis measures time to maturity of the test assets. Panel B plots the difference in the estimated prices of volatility risk. The solid line represents point estimate and the dashed line shows confidence intervals of plus and minus two standard errors as computed from 10,000 bootstrap iterations.

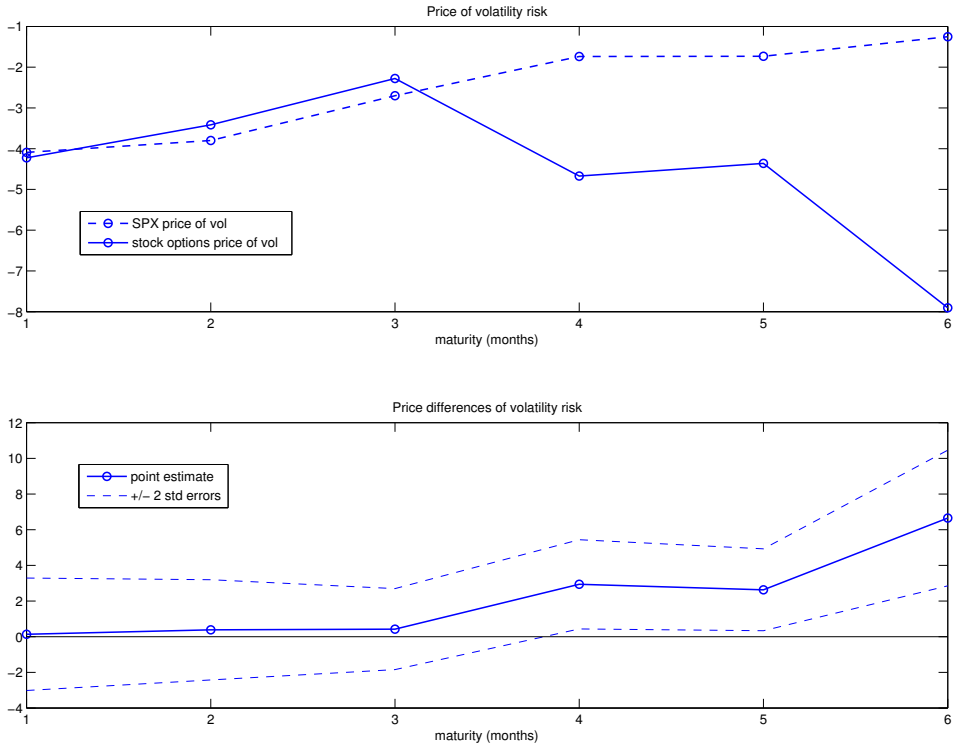
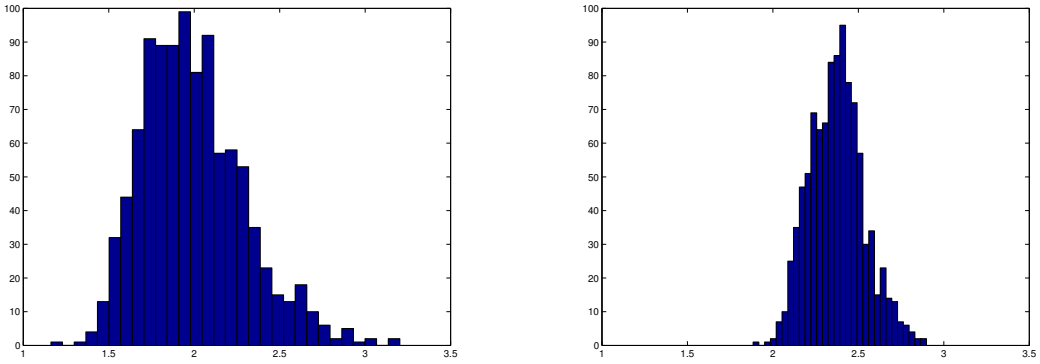


Figure 1.4: Term Structure of the Price of Volatility Controlling for Correlation Risk
This figure shows the differences of prices of volatility risk extracted from index options versus individual options at different maturities.



A) Portfolios without leverage adjustment B) Portfolios with leverage adjustment

Figure 1.5: Simulation sampling distribution
This figure shows the sampling distribution of estimated price of the factor risk in the simulation. The true price is $\lambda_M \equiv 2.5$. Panel A shows the sampling distribution for portfolios of options without leverage adjusted returns. Panel B shows the sampling distribution for portfolios of options with monthly leverage adjustment.

CHAPTER II

Pricing Kernel Monotonicity and Conditional Information

2.1 Introduction

It is well known that the absence of arbitrage implies the existence of a positive pricing kernel, or stochastic discount factor (SDF), that prices all assets. Almost all models of the tradeoff between risk and return specify a pricing kernel that decreases monotonically with the quality of the state of the world. The state of the world is often modeled as a function of the change in aggregate wealth, which is measured by the return on a broad stock market index. A number of researchers combine index option data with historical returns to estimate the pricing kernel nonparametrically, but the kernels they estimate are generally not monotonic functions of the market return. We argue that many of the methods used to estimate the pricing kernel compare a forward-looking, conditional risk-neutral density estimated with option prices to a backward-looking, essentially unconditional physical density estimated with historical returns. We propose a new, completely nonparametric pricing kernel estimator that explicitly accounts for the fact that option prices should reflect all information available. The new estimator suggests that the pricing kernel is a monotonic function of stock market return realizations.

Since the pricing kernel summarizes the attitudes of economic agents about risk, understanding its behavior is one of the primary goals of asset pricing. The research on SDF estimation from option data starts with *Jackwerth* (2000) and *Ait-Sahalia and Lo* (2000), which exploit the relation between option prices and the risk-neutral density. The risk-neutral density is proportional to the SDF multiplied by the (physical) density of the underlying asset. *Breeden and Litzenberger* (1978) show that the second derivative of the price of a call option with respect to the strike price is proportional to the risk-neutral density. Both *Jackwerth* (2000) and *Ait-Sahalia and Lo* (2000) cleverly use this fact to estimate the risk-neutral density with market index option prices for different strike prices and then they divide the resulting risk-neutral density by a nonparametric estimate of the physical density based on historical return data. The resulting ratio of densities is what we refer to as the "classic" nonparametric SDF estimator. Existing research has found that it is typically a decreasing function of the market return over much of its range, but it is also often increasing over part of its range. Many other researchers apply similar techniques, though sometimes with important improvements, and also find that the SDF appears to be a nonmonotonic function. More recent papers in this literature include *Rosenberg and Engle* (2002), *Chaudhuri and Schroder* (2009), *Audrino and Meier* (2012), *Härdle et al.* (2014), *Beare and Schmidt* (2013), *Bakshi and Chabi-Yo* (2013), and *Song and Xiu* (2014). In related research, *Bakshi et al.* (2010) find that average index option returns in several countries are consistent with a U-shaped pricing kernel, but the noise in average returns makes it difficult for them to draw strong conclusions. One paper that does not appear to find an upward sloping kernel is *Barone-Adesi et al.* (2008). Using data from January 2002 to December 2004 and adjusting the variance of the physical distribution using a GARCH model, they find a pricing kernel that appears to be decreasing.

If the pricing kernel is truly increasing in some range of aggregate wealth, then

the marginal value of a dollar is higher when markets rise than when they fall over that range. For financial economists, this is extremely counterintuitive. Even with a multidimensional state vector, it is difficult to see how a higher realized value of the market portfolio could be systematically worse than a lower one. A non-monotonic pricing kernel is so surprising that it has been coined the "implied risk aversion puzzle" or the "pricing kernel puzzle" in the literature that has developed to explain it. *Ziegler* (2007) attributes it to differences in beliefs among agents about the mean and variance of expected returns. *Polkovnichenko and Zhao* (2013) postulate a model with rank-dependent utility to explain the puzzle. *Barone-Adesi et al.* (2013) explain the puzzle with overconfidence, and *Grith et al.* (2013) propose the heterogeneity of investor reference points.

Another set of explanations for the puzzle relies on state dependence, generally with higher moments as additional factors. *Chabi-Yo et al.* (2007) identify latent factors as a probable cause and propose a parametric option pricing model that can generate upward slopes. *Christoffersen et al.* (2013) and *Song and Xiu* (2014) propose models that include volatility as a factor.

One criticism of the empirical papers that find nonmonotonicity is that they do not agree on the location of nonmonotonicities. In *Ait-Sahalia and Lo* (2000) and *Audrino and Meier* (2012) for example, monotonicities appear in the center, while *Christoffersen et al.* (2013) and *Bakshi et al.* (2010) find a U-shaped kernel. Another criticism of almost all of the empirical papers that find nonmonotonicity is that they compare a conditional risk-neutral density to an essentially unconditional physical density. Since option prices, like all market-determined prices, are discounted expectations of future cash flows conditional on all information available, the risk-neutral density estimated from option prices is a conditional density. In our data, most of the moments of the estimated risk-neutral densities change substantially from one month to the next. Since there is no widely accepted method to nonparametrically

estimate physical densities conditional on all available information, common practice is to rely on the use of a rolling window of historical data to make the physical density conditional. Of course, this is not really comparable to using forward-looking option prices to back out market expectations. In fact, given that from one period to the next, the nonparametric estimate of the physical density may only change because of the inclusion of one new observation and the exclusion of one old one, the estimated physical density can often be considered almost unconditional. At times when the conditional risk-neutral density has a higher variance, skewness, kurtosis or other moment than the estimated physical density, the ratio of the two densities can easily display nonmonotonicity.

To demonstrate the problem caused by failing to account for conditional information in the denominator of the SDF, we give two examples of how nonmonotonicity can arise in an estimated pricing kernel implied by a misspecified Black-Scholes model. The first example shows that in a simple single-period setting we can get nonmonotonic ratios of risk-neutral densities to physical densities if we allow the variances of the two to differ as they may when we compare conditional and unconditional densities. In the second example we simulate data from the misspecified Black-Scholes model in order to show that using a rolling window to estimate the physical density while using strictly conditional estimates of the risk-neutral density can lead to nonmonotonic and inaccurate estimators.

We propose a new method that avoids comparing conditional risk-neutral densities to historical data, and creates an estimate that is fully conditional on all moments of the forward-looking distributions. Our method exploits the insight that, at any given time, the conditional density of the future market return is only the density for that particular return realization. We can think of the observations we have as a series of risk-neutral densities accompanied by a corresponding series of return realizations, with each period's risk-neutral density being different and with only one realization

available for each density. Given these data, we can integrate each of the risk-neutral densities up to their corresponding realizations to obtain a set of realized CDF values. If the risk-neutral density is the same as the physical density, the resulting CDF values will be uniformly distributed. To the extent that the empirical distribution of the CDF values is not uniform, we can use the distribution of the CDF values to identify the pricing kernel. This is the intuition behind our pricing kernel estimator. In simulations we find that our method substantially outperforms the classic method in recovering the SDF that generated the data.

To estimate the SDF, we use monthly S&P 500 and FTSE 100 index option data to nonparametrically estimate risk-neutral densities in the standard fashion, following *Figlewski* (2008) with slight improvements. We then assume a stable but flexible unconditional SDF function, which we model with a spline estimator. Finally, we estimate the spline, identifying the model with the fact that integrating the inverse of the SDF times the risk-neutral density up to each realized value should produce a set of cumulants that are uniformly distributed. In using this fact to identify our model, we follow *Bliss and Panigirtzoglou* (2005), who use the same fact to estimate implied risk aversion coefficients parametrically. We use a bootstrapping procedure to estimate confidence bounds for our nonparametric SDF. We refer to our method as a Conditional Density Integration (CDI) method.

We estimate risk-neutral densities from option prices and physical densities from historical returns, and find that these two sets of densities have surprisingly different characteristics. Furthermore, when we (incorrectly) follow the classic procedure by dividing our risk-neutral densities by our physical densities, we also find implied pricing kernels that are nonmonotonic. These nonmonotonic pricing kernels are very sensitive to how the physical densities are estimated, which suggests they are not econometrically robust. However, when we properly account for the conditional nature of the risk-neutral densities estimated from option prices by using the CDI estimator, the

resulting pricing kernel estimate is monotonically decreasing.

In the next section of the paper, we discuss both the classic estimation method and our new CDI method in detail. We also motivate our estimation method theoretically and show that the misspecified Black-Scholes model produces a nonmonotonic SDF. In Section 3 we report the results of simulations designed to compare the performance of the CDI method to that of the classic method. Section 4 describes the data that we use for our tests, and Section 5 reports our primary results. Section 6 concludes.

2.2 Estimating the SDF

The new CDI method we use to derive an estimate of the stochastic discount factor that properly accounts for conditional information is perhaps the biggest contribution of our paper, so we describe it in detail in this section. Our CDI method allows an econometrician to better account for the information set available to investors at the time investment decisions are made. We carefully explain how this is achieved. We also discuss the classic nonparametric approach to estimating the SDF, point out its shortcomings and discuss how these can lead to economically implausible pricing kernels. In Section 2.5, we apply the estimation procedures described here and show that the proposed econometric method has the potential to solve the risk aversion puzzle.

2.2.1 Classic Method

The classic nonparametric method of estimating the SDF of *Jackwerth* (2000) and *Ait-Sahalia and Lo* (2000) relies on a well known result from probability theory known as the Radon-Nikodym Theorem.[1] The theorem implies that if $\mathbb{F}^Q$ and $\mathbb{F}^P$ are measures induced by the risk-neutral and physical cumulative distribution functions,

[1]See *Billingsley* (2012), for example.

the SDF can be expressed as

$$m_{t,t+s} = e^{-rs} \frac{d\mathbb{F}^Q}{d\mathbb{F}^P}, \tag{2.1}$$

a change of measure between two *conditional* probability measures where each probability is conditional upon the same information set, $\mathcal{F}_t$.

Furthermore, a corollary to this theorem states that if probability measures $\mathbb{P}$ and $\mathbb{Q}$ are equivalent measures, then the Radon-Nikodym derivative of $\mathbb{P}$ with respect to $\mathbb{Q}$ is equal to the reciprocal of the Radon-Nikodym derivative of $\mathbb{Q}$ with respect to $\mathbb{P}$,

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \left(\frac{d\mathbb{P}}{d\mathbb{Q}} \right)^{-1}. \tag{2.2}$$

Furthermore, if both $\mathbb{Q}$ and $\mathbb{P}$ are equivalent to dx, then

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{d\mathbb{Q}}{dx} \Big/ \frac{d\mathbb{P}}{dx}. \tag{2.3}$$

The corollary allows one to express the Radon-Nikodym derivative as a ratio of two derivatives. This corrolary is implicitly invoked in the method we refer to as the classic nonparametric method of SDF estimation. The classic approach relies on the fact that the SDF is proportional to the Radon-Nikodym derivative of the risk-neutral distribution with respect the physical distribution. Furthermore, the method relies on the fact that for sufficiently well behaved distributions, the Radon-Nikodym derivative in question is simply the ratio of the risk-neutral density, $\frac{d\mathbb{F}^Q}{dx}$ to the physical density, $\frac{d\mathbb{F}^P}{dx}$. This fact allows econometricians to estimate the SDF by estimating the risk-neutral and physical densities separately and then taking the ratio of the densities.

Since the classic nonparametric method relies on estimation of the Radon-Nikodym derivative via Equation (2.3), it reduces to estimating the two densities separately. Theoretically, the densities in the numerator and denominator of the Radon-Nikodym

derivative in Equation (2.3) are conditional densities; they take into account investors' beliefs at the time of investment, conditional on all information available, $\mathcal{F}_t$. As such, we ideally should take care to estimate the densities in a conditional, forward-looking manner. For estimation of the numerator, one typically relies on the result of *Breeden and Litzenberger* (1978), that $\frac{d\mathbb{F}^Q}{dK} = e^{rT}\frac{\partial^2 C}{\partial K^2}$, where C represents the option price, K represents strike prices and $\frac{d\mathbb{F}^Q}{dK}$ represents the risk-neutral density over possible realizations of the underlying. Since options data typically allow us to observe option prices with a number of strike prices K, we are able to estimate the derivative $\frac{d\mathbb{F}^Q}{dK}$ over a collection of points K. Various techniques for estimating or interpolating values of the density between observed strike prices have been proposed in the literature. This gives an estimate of the risk-neutral density which is forward-looking and hence conditional in nature.

On the other hand, there are no known methods for estimating $d\mathbb{F}_t^P$, the time t physical density, in a forward-looking manner, taking into account the information investors base their investment decisions on *at time t*. In previous studies, $d\mathbb{F}_t^P$ has been estimated by smoothing or averaging past realized returns. In order to make the estimates reflect a conditional rather than unconditional density, a rolling window is typically used to estimate the physical density. This approach clearly leaves much to be desired. Nonparametric estimates require large amounts of data, thus forcing recent data, even if it accurately reflects beliefs about the future, to be a small part of the estimated density.

In effect, the classic method of nonparametrically estimating the SDF implicitly assumes that physical probability measures and their corresponding densities are stable over time, or that the conditional densities are the same as unconditional densities. The assumption of stable physical densities and distributions is widely believed to be implausible. The method of *Breeden and Litzenberger* (1978) for estimating conditional, risk-neutral densities reveals that their time series is not stable. We

characterize the risk-neutral densities implied by our option price data in Table 1, which is discussed in Section 2.5. If the risk-neutral densities are not stable over time, it is implausible that physical densities are.

To investigate whether comparing a conditional density to an unconditional density can cause non-monotonicity in practice, we calculate implied pricing kernels under Black-Scholes assumptions, but with a slightly higher risk-neutral than physical variance. Our example is motivated by the fact that the risk-neutral density can change significantly from period to period while the estimated physical density will typically be more stable. In some periods, the risk-neutral density may have a higher variance than the physical, while in other periods it may have a lower variance. In Panel A of Figure 1, we plot the physical and risk-neutral densities under the assumption that returns are lognormally distributed and have parameter values that correspond to our risk-neutral sample moments. Panel B plots the corresponding pricing kernel function, which is monotonically decreasing in market returns. In Panels C and D, we plot the densities and pricing kernel under the assumption that the variance of the risk-neutral density is slightly higher than that of the physical density, changing the (monthly) σ parameter from 5.26% to 5.50% percent. The pricing kernel in Panel D starts at a very high level and is first decreasing and then increasing, reflecting a pattern often found in prior work. This example only allows the second moment to differ across these densities. In typical pricing kernel estimation, all the moments of the estimated risk-neutral density can, in principle, vary from period to period while the estimated physical density, based on historical data, is relatively stable. This shows that if the estimated physical density does not change to reflect new information as much as the risk-neutral density does, the corresponding estimated pricing kernel can be increasing over some range. This problem is inherently present in all of the nonparametric pricing kernel estimators based on option prices that we are familiar with.

In the remainder of this section, we discuss our conditional density integration method in detail. We begin with an in depth description of how we estimate the risk-neutral densities of the market's beliefs about one-month returns on both the S&P 500 and FTSE 100 indices. Our estimation technique draws from many existing methods, but it most closely follows *Figlewski* (2008). Next, we discuss how we use these densities to estimate the SDF using the CDI method as well as the classic method. It is important to note that, in both cases, we use the same risk-neutral densities. This way, when we discuss our empirical results in Section 2.5, we are able to ensure that the differences in the results come from differences in accounting for conditioning information as opposed to differences in the risk-neutral densities used in the estimation.

2.2.2 Estimating Risk-Neutral Densities

In order to estimate the stochastic discount factor over the horizon spanned by the OptionMetrics data, we first estimate monthly risk-neutral densities following the method outlined in *Figlewski* (2008), with a few modifications that we describe below. Each month, for the options data with best bids (or last prices when bids are not available) exceeding \$3/8, we fit a fourth degree spline to implied volatilities associated with each observed strike price. This is done by placing a single knot at the close price on the day the option is traded, with the remainder of the required knots placed at the minimum and maximum strike prices within our sample. This creates a continuous curve in the implied volatility space. We then convert the implied volatility curve back to the price space by inverting the transformation used to obtain implied volatilities. With the given prices we apply the result of *Breeden and Litzenberger* (1978), that $\frac{d\mathbb{F}^Q}{dK} = e^{rT}\frac{\partial^2 C}{\partial K^2}$, where $\mathbb{F}^Q$ represents the risk-neutral CDF and $\frac{d\mathbb{F}^Q}{dK}$ represents the density over prices, K. Because we smooth implied volatilities our estimation procedure always results in reasonable density functions with positive

values.

The practice of removing options data with very small prices is standard in the options literature as options with extremely low prices tend to provide misleading data because they are so far out of the money. While extremely small prices can often give rise to misleading data, leaving them out of our data poses a problem as well because our estimated densities are often left truncated in the tails, especially in the upper tail because far out-of-the-money call options are relatively thinly traded. The densities obtained by taking second derivatives over strike prices will often look like that in Figure 2.2. We refer to this part of the density as the *truncated* density. It is clear from the figure that truncating the data in our sample can potentially cause us to miss out on a large portion of the density. We circumvent this problem by applying the method of *Figlewski* (2008) to estimate the tails of the risk-neutral distributions in our sample.

The tail estimation method relies on results from *Pickands III* (1975) and *Balkema and De Haan* (1974) both of which show that for an independent, identically distributed sequence of random variables, the conditional distribution given that the variable exceeds some threshold approaches a generalized Pareto distribution as the specified threshold becomes large. Following the logic of this result, we find the parameters from a generalized Pareto distribution that give the closest match to the truncated risk-neutral density. By pasting the resulting generalized Pareto distribution onto the truncated risk-neutral density, we complete the estimation of the entire density.[2]

[2]Our method differs slightly from that of *Figlewski* (2008), which uses a generalized extreme value distribution rather than a generalized Pareto distribution to estimate the tails of the risk-neutral density. The use of generalized extreme value distribution comes from similar theory of statistics of extremes. The Fisher-Tippett theorem (see for example *Embrechts et al.* (1997)) states that the sample maximum of an independent, identically distributed sequence of random variables approaches a generalized extreme value distribution as the sample size approaches infinity. However, since we are looking at matching the tail of the distribution beyond some extreme point determined by our data, we feel that an application of the results in *Pickands III* (1975) and *Balkema and De Haan* (1974) is most appropriate. So we use a generalized Pareto distribution as opposed to a generalized extreme value distribution when estimating the tails of the risk-neutral densities. In more recent

The generalized Pareto distribution is characterized by three parameters: a location parameter, a scale parameter and a shape parameter. In order to fit the tail distribution, we choose three points on each side of the truncated distribution. With these three points, we then find the three parameter values of the generalized Pareto distribution that lies closest to the truncated distribution at the three points. By choosing three points, we are able to identify the three parameter values. We do this for each tail of the distribution. While *Figlewski* (2008) only uses two points for each tail and imposes the additional constraint that the area under the curve must equal one, we find that the optimization gives smoother transitions between the truncated density and the tails if we do not include the constraint on the area. Instead, we match three points in each tail and then normalize our estimate to ensure that the area of the density is equal to one. In most cases, this normalization does not change the curve estimation much at all as the tail matching itself gives densities whose area is nearly equal to one. In the few cases where the normalization has much impact, imposing a constraint on the area in the tail-matching optimization results in awkward kinks in the density which are clearly just an artifact of the optimization and its constraints.

In a small number of cases, the truncated part of the distribution does not go far enough into the tail of the distribution to allow the tail matching procedure to fit well. This happens when the upper end of the central distribution, which is determined by our data, does not extend far enough past the peak of the distribution. In these cases, we interpolate the implied volatility curve to larger return values using cubic spline interpolation. The resulting implied volatility curve is then transformed back to the option price space so that we can take the second derivative to obtain the truncated part of the risk-neutral distribution. This extends the truncated part of the distribution just far enough that the tail matching procedure gives a meaningful

work Figlewski also adopts the generalized Pareto distribution.

upper tail.

We use risk-neutral densities estimated with this method to calculate the SDF using both the classic nonparametric method and our new CDI method. Using the same set of risk-neutral densities, the classic nonparametric method yields nonmonotonic SDF estimates but the CDI method produces monotonic estimates. Thus, our method of estimating risk-neutral densities does not seem to drive the monotonicity result that we find.

2.2.3 Standard Approach to Estimating Physical Densities

Once we have the forward-looking, risk-neutral densities, we can proceed with estimating the stochastic discount factor. For the classic method, which relies on Equation (2.3), we are left to estimate the physical densities corresponding to each of the risk-neutral densities. As described above, until now there has been no known way to estimate the physical density in a forward-looking manner, and the solution proposed in the literature is to use a rolling window of data to nonparametrically estimate the physical densities. We use a Gaussian kernel density estimator with a rolling window. To obtain a conditional estimate, it is best to use as short a window as possible without compromising the integrity of the kernel estimator.

As discussed earlier, in theory, the physical and risk-neutral densities should have the same support. Empirically, using a rolling window of data to estimate the kernel density often results in estimates of the physical density with different (machine measurable) support from the risk-neutral density for the same period. This is itself a sign that there is a problem with the estimation procedure. This is a result of improperly matching conditional information in the numerator and denominator of the Radon-Nikodym derivative. If, for instance, previous returns within the rolling window tend to be low but recently the market received news suggesting high returns in the future, then the upper tail of the forward looking risk-neutral density may

have support beyond the range of positive support for the physical density estimate. Similarly, we observe instances where the physical density has wider support than the risk-neutral density. In practice, when this happens, we need to truncate the densities such that they have the same region of positive support, to avoid dividing a positive density by zero for some returns. To avoid this problem, for each date, we estimate the pricing kernel over the range between the maximum of the lower bounds of support for the densities and the minimum of the upper bound.

2.2.4 CDI Approach

In order to estimate the SDF with option prices observed over a period of time, we need to make a stationarity assumption for the SDF. We assume the following:

Assumption II.1. *The stochastic discount factor over our sample period is stationary up to a rate of time discount factor $e^{-r_t\tau}$, where r_t is the risk free rate at time t and τ is the duration of the payoff period over which the SDF is discounting.*

While this is a very common assumption in empirical asset pricing, it probably merits a little extra discussion in this context. It is equivalent to the assumption that the ratio of risk-neutral to physical densities is stable over time. This is a fairly plausible assumption if one believes that the representative investor's preferences are relatively stable over time, since investor preferences are responsible for the difference between the risk-neutral and physical densities. This assumption is consistent with the empirical finding that risk-neutral and physical densities change over time, but it requires that the two vary together. Mathematically, the assumption reduces to stability of $\frac{d\mathbb{F}_t^Q}{d\mathbb{F}_t^P}$, as opposed to stability of $d\mathbb{F}_t^P$. Our assumption is also the key identifying assumption made in *Bliss and Panigirtzoglou* (2005), where it is argued that this is a more plausible assumption than the assumption that is implicitly required for the classical estimator of the stochastic discount factor.

If we do not take our stationarity assumption to be literally true, our estimate of the SDF can be interpreted as an average or unconditional SDF over our sample period. Many of the researchers who apply the classic nonparametric SDF estimation method report an average SDF, and our estimate can easily be compared to theirs. While it would be nice to be able to identify variation in the pricing kernel over time, we argue that there simply is not enough information in the data to do this consistently. To be able to estimate a pricing kernel month by month, we would need a convincing way to estimate conditional physical densities. In the absence of such a method, at least we know that we can estimate the average SDF correctly.

Our identification strategy relies on several well known properties from statistics and probability theory. The first of these properties, which is central to our method, allows us to circumvent the need for estimating the physical densities corresponding to each of the risk-neutral densities. The property is given in the following proposition:

Proposition II.2. *For any continuous random variable, X with CDF $\mathbb{F}_x$, the random variable defined by $\mathbb{F}_x(X)$ is uniformly distributed on the interval* $[0, 1]$,

$$\mathbb{F}_x(X) \sim U[0, 1]. \tag{2.4}$$

We let $\mathbb{F}_t^P$ be the unobserved probability measure representing investors' aggregate beliefs about returns on the S&P 500 under the physical measure at time t and let returns over the subsequent period be given by X_t. Now it follows from Proposition II.2, that

$$\int_{-\infty}^{X_t} d\mathbb{F}_t^P(x) \sim U[0, 1]. \tag{2.5}$$

Since there are no known methods for estimating $d\mathbb{F}_t^P$ in a forward-looking manner, estimating Equation (2.5) directly from the data is not a simple task. It would presumably require obtaining a long time series of past realizations of ex-dividend

returns.[3] One would then have to find a way to use these returns to estimate forward looking beliefs about returns under the physical measure. As discussed earlier, this method would require something beyond simply smoothing a long time series of *past* returns, since that does not do a good job of estimating the *current* beliefs held by the market. In order to circumvent this problem, we make use of the fact that we do have forward looking estimates of market beliefs about future returns under the risk-neutral measure.

We express Equation (2.5) in terms of the risk-neutral densities estimated using our generalized Pareto distribution tail matching procedure. Let $d\mathbb{F}_t^Q$ be the time t risk-neutral probability measure and let $\frac{d\mathbb{F}_t^P}{d\mathbb{F}_t^Q}$ denote the Radon-Nikodym derivative of time t physical distribution with respect to time t risk-neutral distribution. Then

$$\int_{-\infty}^{X_t} d\mathbb{F}_t^P = \int_{-\infty}^{X_t} \frac{d\mathbb{F}_t^P}{d\mathbb{F}_t^Q} d\mathbb{F}_t^Q = \int_{-\infty}^{X_t} \left(\frac{d\mathbb{F}_t^Q}{d\mathbb{F}_t^P} \right)^{-1} d\mathbb{F}_t^Q \sim U[0,1], \tag{2.6}$$

where the first equality in Equation (2.6) follows from Theorem **??** and the second equality follows from Corollary **??**.

Since we can estimate the risk-neutral densities and we observe realized returns over the periods corresponding to each density, it only remains to estimate the random variable $\left(\frac{d\mathbb{F}_t^Q}{d\mathbb{F}_t^P} \right)^{-1}$, which is proportional to the inverse of the stochastic discount factor. Therefore, by estimating $\left(\frac{d\mathbb{F}_t^Q}{d\mathbb{F}_t^P} \right)^{-1}$, we have essentially estimated the stochastic discount factor. It is important, however, that we first establish uniqueness of the random variable we attempt to estimate. The following proposition ensures that there is such a unique random variable.

Proposition II.3. *For any equivalent measures $\mathbb{Q}$ and $\mathbb{P}$ on $\mathbb{R}$ with random variable*

[3] We use percentage changes in market value because option payoffs are based on the market value of the S&P 500 at expiration. This amounts to shifting the cum-dividend return density to the left, but a stable dividend yield does not affect the shape of the SDF.

$X \sim \mathbb{P}$, there exists a unique (a.s. $\mathbb{Q}$) non-negative function $g : \mathbb{R} \to \mathbb{R}_+$ such that

$$\int_{-\infty}^{X} g(y) d\mathbb{Q}(y) \sim U[0,1]. \tag{2.7}$$

A proof of this proposition appears in the Appendix.

The function denoted g in Proposition II.3 is similar to the Radon-Nikodym term in Equation (2.6), the main difference being that in Equation (2.7), the region of integration is itself random. So the Radon-Nikodym Theorem is not directly applicable here. The functional form of g defines a random variable in Proposition II.3 because it is evaluated at possible values of the random outcome. We can think of inputs to the function g as values the random variable X can take. The outcomes of the random variable depend upon $\omega \in \Omega$ the probability space determining returns, $X = X(\omega)$. As such, the integral with respect to $d\mathbb{Q}$ can be interpreted as the integral with respect to the measure $\mathbb{Q}(\{\omega : X(\omega) \in dy\})$. In this way, $g(y) = \phi(\{\omega : X(\omega) \in dy\})$, where ϕ is a mapping from Ω to the non-negative real line, $\phi : \Omega \to \mathbb{R}_+$. So $g(y)$ represents possible realizations of the random variable $g(X(\omega)) = \phi(\omega)$. We will let g denote the inverse of the SDF up to a rate of time discount $e^{r_t \tau}$, where r_t is the risk-free rate at time t and τ is the time to expiration of the option. Our estimation procedure will focus on estimating g.

Proposition II.3 establishes uniqueness of the function g that transforms the integral with respect to measure $\mathbb{Q}$, to a specific distribution. This is similar to the statement of the Radon-Nikodym Theorem. The function g, mapping realizations of returns to non-negative values is itself a random variable, much the same as the Radon-Nikodym derivative. The difference is that here we have a random domain $(-\infty, X]$. We thus estimate the functional form of g that maps random outcome of percentage changes in the S&P 500 to the unique kernel that transforms the integral in Equation (2.7) to the uniform distribution.

It is interesting to note that our methodology is similar to one that would examine the discrepancy of a time-series average fitted returns on butterfly spreads for each return realization with a uniform distribution. We believe the CDI method is superior to this one for several reasons. First, our method of estimating the left tails of the risk-neutral distribution is more palatable than one that would assume left and right end-points for the series of butterfly spread returns. Second, butterfly spread returns are highly non-normal due to the large mass at zero payoffs, and averages of these returns could be unstable. Last, we want to follow a method that is more comparable to the existing literature so that a comparison of results is possible along multiple dimensions.

2.2.5 CDI Approach Estimation and Inference

Our goal is to estimate the SDF in a way that reflects investors' beliefs as accurately as possible. For this reason, we do not impose any parametric restriction on the form of the stochastic discount factor. Instead, we use a cubic spline to obtain nonparametric estimates of the inverse SDF. Since any real valued function can be reproduced by a cubic spline of infinite order, this is a completely model-free estimation procedure. We use finite order cubic B-splines to approximate the function g. We use cubic B-splines as opposed to polynomials because they offer more flexibility in estimating functional forms. The use of splines of order b requires that we first choose the placement of knots which will determine the bases to be used for estimation purposes. We simply use equally spaced knots over our range of returns. The minimum of the range is set to the minimum value for which our estimated risk-neutral densities, over all months in the sample, have a positive (machine measurable) support. The maximum of the range is the maximum realized return within our sample. This range corresponds to the values over which the integral in Equation (2.7) is taken, once we replace $-\infty$ with the minimum value for which $d\mathbb{F}^Q$ has positive support.

The cubic B-spline of order b is a linear combination of b basis functions,

$$g(y) \approx \sum_{j=1}^{b} \theta_j B_j(y),$$

where $B_j(\cdot)$ denotes the jth basis function of the spline. Using this approximation to the function g, we can also approximate the integral in Equation (2.6) as a linear combination of integrals,

$$\int_{-\infty}^{X} g(y) d\mathbb{F}^Q(y) \approx \sum_{j=1}^{b} \theta_j \int_{-\infty}^{X} B_j(y) d\mathbb{F}^Q(y). \tag{2.8}$$

Since we have a linear function in θ, our estimated function $\hat{g}$ is given by

$$\hat{g} = A\hat{\theta}, \tag{2.9}$$

where $\theta = (\theta_1, ..., \theta_b)'$ and $\hat{\theta} = (\hat{\theta}_1, ..., \hat{\theta}_b)'$. A is our data matrix which is expressed in terms of risk-neutral distributions estimated from options data, realized S&P 500 index returns corresponding to each risk-neutral distribution, denoted X_t and the spline basis functions. We can formally represent the data matrix $A \in \mathbb{R}^{T \times b}$ by

$$A_{i,j} = \int_{-\infty}^{X_i} B_j(y) d\mathbb{F}_i^Q(y), \qquad i = 1, ..., T;, \ j = 1, ..., b, \tag{2.10}$$

where T represents the number of monthly estimates of $\mathbb{F}^Q$ available and b is the number of basis functions included in our estimated spline approximation of g.

Since we will be using non-overlapping data on monthly options from OptionMetrics which only goes back as far as 1996 for the S&P 500 and 2002 for the FTSE, as described in Section 2.4, our sample is not extremely large. For this reason, we use a GMM type optimization with only the first stage optimization. This has been shown to perform best when one does not have extremely large data sets with which

to perform GMM estimation (see for example *Hayashi* (2000)). In order to make the best use of the data available to us, we optimally choose model parameters b and m in order to balance the trade off between the number of moment restrictions and the number of parameters to be estimated. A larger number of spline basis functions, b, corresponds to a more flexible and accurate spline approximation of the function g. However, increasing the number of basis functions requires that we increase the number of moment restrictions in our estimation because identification of θ requires that the number of moment restrictions be at least as large as the dimension of θ, $b \leq m$. Arbitrarily increasing the number of moment restrictions, on the other hand, decreases our degrees of freedom in estimating θ, resulting in data limitations. Our goal is to make the best possible use of the finite data sample available to us by letting the data determine the optimal values of b and m.

To estimate θ, we solve the first stage GMM optimization,

$$\hat{\theta} = \underset{\theta \in \mathbb{R}^b}{\operatorname{argmin}} \sum_{j=1}^{m} \left(\sum_{t=1}^{T} \left(\underbrace{\sum_{j=1}^{b} \theta_j \int_{-\infty}^{X_t} B_j(y) d\mathbb{F}_t^Q(y)}_{\hat{g}(\theta)} \right)^j - \frac{1}{j+1} \right)^2, \tag{2.11}$$

where we use the fact that the jth moment of the uniform distribution over the unit interval is equal to $\frac{1}{j+1}$ and we use the the first m moments in estimating the vector θ. It is important to note that the solution to Equation (2.11) is found by minimizing over $\mathbb{R}^b$, in other words, we place no restrictions on our estimate of θ.

Once we have the estimated $\hat{\theta}$, it is straight forward to estimate g. We simply need to plug $\hat{\theta}$ into Equation (2.9) to obtain our estimate for g, the inverse of the Radon-Nikodym derivative, $\frac{d\mathbb{F}_t^Q}{d\mathbb{F}_t^P}$, for all t. By Corollary **??**, $\frac{d\mathbb{F}_t^Q}{d\mathbb{F}_t^P} = \frac{1}{\hat{g}}$ for all t. So our estimated SDF is given by $e^{-r_t \tau} \frac{1}{\hat{g}(X)}$, where r_t denotes the risk free rate at time t, τ represents time to maturity of time t index options on the S&P 500 index and X_t

denotes returns on the S&P 500 index. This can be re-expressed as

$$m_{t,t+\tau}(X) = e^{-r_t\tau} M(X),$$

where $M(x) \equiv \frac{1}{g(x)}$.

Since we assume that the function g is time-invariant, it follows that M and $\hat{M} = \frac{1}{g}$ are also time invariant. Since $\hat{M}$ is time invariant, the SDF will be time invariant up to the term $e^{-r_t\tau}$. The value of $e^{-r_t\tau}$ is also very stable over our sample period. So the SDF does not vary substantially over our sample under our set of assumptions. We focus only on the estimation of $\hat{M}$ because the time discount factor $e^{-r_t\tau}$ does not tell us anything about investors' preferences over states of the world and returns on market indices. In Section 2.5, we will discuss our empirical results based upon estimates of $M(x)$, as described above.

For the purposes of inference, we calculate pointwise confidence intervals for the estimated SDF. We resample with replacement from the set of rows of the data matrix in Equation (2.10). This is equivalent to sampling with replacement from the set of dates associated with each risk-neutral density we estimate. For each sample, we can re-calculate the SDF estimate using the CDI method. We then calculate the accelerated bias-corrected (BCa) percentile bootstrap confidence intervals as described in *Efron and Tibshirani* (1993). This gives us a virtual continuum of pointwise confidence intervals if we take a fine partition of the return space. However, as is the case with most nonparametric methods, in order to get a very tight confidence interval, a large amount of data is needed.

2.2.6 Model Selection

In order to estimate θ, we use a GMM type estimation to match the resulting estimate to the moments of the uniform distribution over the unit interval as in

Equation (2.11). This requires that we choose the number of moment restrictions m as well as b, the dimension of θ. As we do throughout the paper, we wish to impose as little structure as possible on the estimation. This allows us to estimate the SDF in a manner we feel best approximates the market's beliefs and risk preferences, which determine the SDF. In keeping with this goal, we optimally choose the m and b according to our data and we place no restrictions on θ in our estimation.

Our model selection criterion for determining b and m uses the Cramer-von Mises statistic[4] which is a common nonparametric criterion for determining the goodness of fit of an estimated distribution. The Cramer-von Mises statistic compares an estimated distribution to a target distribution (uniform in our case) by comparing the corresponding CDFs, $\hat{\mathbb{F}}$ and $\mathbb{F}_U$ respectively. Here $\hat{\mathbb{F}}$ is the empirical distribution function. A small Cramer-von Mises statistic implies a good fit while larger statistics imply poor fit. The statistic is given by

$$CvM = \int_{x=-\infty}^{\infty} (\hat{\mathbb{F}}(x) - \mathbb{F}_U(x))^2 d\mathbb{F}_U(x).$$

In the case of the uniform distribution over the unit interval, we can express this as

$$CvM = \int_{x=0}^{1} (\hat{\mathbb{F}}(x) - x)^2 dx.$$

While we choose the model based solely on the value of the Cramer-von Mises statistic, this doesn't necessarily tell us how well our optimal model transforms the

[4]We use the Cramer-von Mises statistics as our criterion because it minimizes the mean-squared distance between CDFs as opposed to the Kolmogorov-Smirnov statistic, which minimizes the maximum distance between two CDFs,

$$KS = \sup_{x \in \mathbb{R}} |\hat{\mathbb{F}}(x) - \mathbb{F}_U(x)|.$$

This amounts to choosing the estimate which minimizes the difference over the entire range of values in a mean-squared sense, as opposed to choosing the statistic which minimizes the size of the largest error.

data to match the uniform distribution. We also calculate the p-values corresponding to the null hypothesis that the estimated distribution is the same as the hypothesized distribution. We calculate p-values base upon simulated outcomes as opposed to asymptotic distributions. This gives us a sense of exactly how well our model selection and subsequent optimization perform given our finite sample size.

We refer to optimal selection of b and m as model selection, and we will use the optimal model to estimate θ and hence $\hat{g}$ as well as the SDF. In order to optimally select our model, we examine goodness of fit of our estimated CDF with the uniform CDF. Our estimated CDF is given by the empirical CDF corresponding to the estimated vector $\hat{\theta}$ for a given combination of b and m,

$$\hat{\mathbb{F}}_{b,m}(x) = \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\left(\sum_{j=1}^{b} \hat{\theta}_j \int_{-\infty}^{X_t} B_j(y) d\mathbb{F}_t^Q(y) \leq x\right), \tag{2.12}$$

where $\mathbb{1}(E)$ represents the indicator function taking value 1 in the where event E is true and the value zero otherwise.

We evaluate Equation (2.12) with the estimated parameter vectors and then compare the Cramer-von Mises statistics for each, keeping in mind that in order for θ to be identified requires that $b \leq m$. That is, the number of moment restrictions must be at least as large as the dimension of the vector to be estimated, θ. The smallest Cramer-von Mises statistic corresponds to the model for which the CDI procedure transforms the data to a distribution closest to the uniform distribution. We refer to this as the optimal model.

2.3 Simulation

This section examines the efficacy of the CDI method in sample sizes typical of those in the empirical literature on pricing kernel estimation, and contrasts this with the efficacy of the classical estimator in the same sample. We extend the example

described in Section 2.2.1 from a static, single period setting to a multiperiod setting with data comparable to that which we observe in the S&P 500 and FTSE 100 data. By simulating data with a known SDF, we can observe how accurately each is able to estimate the true SDF. Our simulated data assumes underlying index returns are distributed log-normally as is the case in the Black-Scholes world, but note that the CDI method is more general and does not make this assumption. We choose parameters of the distribution to fit the data generated by our risk-neutral S&P 500 densities.

We begin by defining an SDF that will be used to generate our data. As we have done throughout the paper, we refer to the SDF as the Radon-Nikodym derivative of the risk-neutral with respect to the physical measure and we ignore the rate of time discount factor. To be consistent with our data and Assumption II.1, we assume that the stationary SDF in the economy is given by the SDF in Panel B of Figure 2.6. This is the SDF resulting from taking the ratio of the (risk-neutral) log-normal density with location parameter $\mu_Q = 0.00011$ and scale parameter $\sigma_Q = 0.0526$ and the (physical) log-normal density with location parameter $\mu_P = 0.0040$ and scale parameter $\sigma_P = 0.0526$. As described in Section 2.2.1, these parameters are chosen to match the average of the monthly distributions corresponding to those (annualized) values given in Panel A of Table 2.1. Notice that we have set $\sigma_Q = \sigma_P$ to be consistent with the Black-Scholes model. As in the Black-Scholes model, the location parameters μ_q and μ_p differ.

The S&P 500 risk-neutral densities described in Table 2.1 are time varying and it is generally accepted that both σ_P and σ_Q are time varying (but equal to each other such that the pricing kernel is stable). We fit our series of S&P 500 monthly variances described in Table 2.1 to an Ornstein-Uhlenbeck process. This is done by simply taking the variance of each risk-neutral density estimated using the method described in Section 2.2.2, and maximizing the likelihood function to estimate the parameters of

the Ornstein-Uhlenbeck process being fit to the series of variances. With the resulting estimated parameters of the process, we simulate a series of N risk-neutral variances. Along with the fixed location parameter μ_Q and the assumption of log-normality, this variance process gives us a series of N risk-neutral densities. Both the CDI method and the classical method use these densities to recover the SDF estimates. Once we have the risk-neutral densities we can use the true stochastic discount factor to get the physical densities corresponding to each risk-neutral density. Recall that $d\mathbb{F}_t^P = \left(\frac{d\mathbb{F}_t^Q}{d\mathbb{F}_t^P}\right)^{-1} d\mathbb{F}_t^Q$. We use this fact to get the physical densities corresponding to each risk-neutral density. We then take a single random draw from each of the physical densities in the series. This is done by first recovering the CDF, $\mathbb{F}_t^P$ from each physical density $d\mathbb{F}_t^P$. Next we generate a series of draws from a uniform distribution over the unit interval, $u_t \sim U[0,1]$, for $t \in \{1,2,...,N\}$. Draws from the physical density $d\mathbb{F}_t^P$ are given by $(\mathbb{F}_t^P)^{-1}(u_t)$ which has exactly the distribution of our physical density $d\mathbb{F}_t^P$. Each of these draws from the physical distribution correspond to the realized monthly returns we observe in the data. Now we have a series $\left(d\mathbb{F}_t^Q, X_t\right)$ for $t \in \{1,2,...,N\}$, where X_t represents the time t realization of a draw from the time t physical density $d\mathbb{F}_t^P$. Since the physical density and the true SDF are unobservable to the econometrician, this series of risk-neutral densities and single realizations from physical densities replicates the data that is available to the econometrician.

With the series $\left(d\mathbb{F}_t^Q, X_t\right)$, we estimate true SDF using both the CDI method and the classical method. We show results of both estimation procedures for $N = 200, 500,$ and $1,000$. By comparing these estimates we can see how well each of the methods performs with small data samples. In particular, comparing the two methods allows us to see how estimates can be affected when comparing forward-looking estimates with backward-looking estimates. We use a 60 period rolling window of realized returns X_t to compute kernel density estimates of the physical densities which are unknown to the econometrician. The results of the simulations for both esti-

mators are shown in Figure 2.3. Panel A shows that for all values of N, the CDI estimator does a very good job of recovering the true SDF. While the smallest data sample recovers the true SDF fairly well over the range $[.95, 1.05]$, outside of the range $[0.95, 1.05]$, the CDI estimator veers away from the true SDF when $N = 200$. This is hardly surprising given that there are relatively few realized observations outside this range. For $N = 500$ and $N = 1,000$, the CDI estimator does a very good job of recovering the true SDF over the entire range depicted, $[0.9, 1.1]$. This is made possible by the fact that larger samples have a larger number of observations near both 0.9 and 1.1, allowing the spline to accurately estimate the SDF near those values of returns.

Panel B shows the results of the simulation performed for the classical method. It is clear from the figure that none of the estimates are able to recover the true SDF with any accuracy. The estimates resulting from $N = 200$ and $N = 1,000$ simulated months exhibit extreme non-monotonicity and do not come close to recovering the true SDF. The estimate when $N = 500$ does far better than the other two estimates using the classical method. However, if we compare the classical method with $N = 500$ to the poorest performing CDI estimator, that with $N = 200$, it is clear that the the poorest performing CDI estimate significantly outperforms the best estimate using the classical method. Figure 2.3 shows that the CDI method performs very well while the classical method performs poorly.[5] The reason is that the CDI method properly accounts for conditional information whereas the classical method uses the ratio of a forward-looking estimate to a backward-looking estimate, thus failing to take account of conditional information.

[5]Work such as *Audrino and Meier* (2012) and *Beare and Schmidt* (2013) improve on the classical approach, and their methods may produce better results.

2.4 Data

We start with daily S&P 500 and FTSE 100 data from OptionMetrics. For the S&P 500 index options, price midpoints are available from September, 1996 through December, 2012, for a total of 196 months. For the FTSE data, closing prices are available from January 2002 through July 2013. Prior to 2006, FTSE data was collected from the exchange directly. After 2006, Optionmetrics began receiving tick data with more limited availability until 2007. As a consequence, several months are unavailable in 2006 and 2007 and we are left with 121 total months of data. We use options with one month to maturity, giving a non-overlapping time-series of options prices. This non-overlapping data allow us to obtain independent observations for beliefs about the coming month and an independent realization of returns. Using monthly rather than higher frequency data does not cause a significant loss of information for our analysis because we only have one option expiration per month.

We also use OptionMetrics implied volatilities for each strike price at each date in our set. We remove data for which there is no available implied volatility as these violate static no arbitrage conditions. We wind up using put prices for relatively low strike prices, call prices for relatively high strike prices and weighted averages for intermediate strike prices. We use a logistic function that is centered at the closing index value with a volatility parameter that is half of the range of observable option prices to determine the relative weights of puts and calls when both prices are observable. Using open interest to calculate the weighted average gives almost exactly the same result, but the logistic function is slightly smoother.

We obtain S&P 500 closing prices for monthly trading dates and for option expiration dates from CRSP, and closing FTSE 100 values from OptionMetrics Europe. To estimate the SDF with the classic procedure, we also use prices from up to ten years prior to the start of our OptionMetrics sample for our rolling window estimations of the physical density. Finally, we calculate the risk-free rate from continuously

compounded yields on secondary market 3-month Treasury Bills. This data is from the Federal Reserve report H.15.

2.5 Results

In this section, we present the results of our estimation described in Section 2.2, using the data described in Section 2.4. We compare CDI results with the results obtained by using the classic nonparametric method over the same sample period. We argue that our estimation procedure results in economically plausible SDFs, unlike the classic method, which does not properly account for conditional information and suggests the existence of a pricing kernel puzzle. Throughout this section, it is important to recall that the risk-neutral densities used for estimation of the SDF with the classic method are the same densities used for the CDI method. This allows us to compare the methods consistently.

Table 2.1 presents sample averages of the mean, variance, skewness and kurtosis associated with both the risk-neutral and physical densities estimated for each of the 196 months from September, 1996 through December 2012 for the S&P 500 and the 121 available months from January 2002 to July, 2013 for the FTSE 100. The physical densities described in Table 1 are estimated with a kernel density method using the past 60 months of index returns. Looking first at the means of both the risk-neutral and physical densities, we see that the average means are about the same, but the physical density means are much more variable than the risk-neutral density means.

Theory dictates that the expected value of the risk-neutral density should equal r_t for all t. The average of the annualized expected return associated with the estimated risk-neutral S&P 500 densities is 2.76% with a sample standard deviation of 0.97%. This is remarkably close to the value we obtain when we plug in the mean value for r_t over our sample period, $\bar{r} = 2.64\%$. Of course, this is not exactly the correct comparison to make, as one would want to compare $e^{r_t \tau}$ with the expected value

of each risk-neutral density in our sample. We calculate the absolute value of this difference for each month in our sample. The mean absolute monthly difference is 0.18% with a standard deviation of 0.17%. This suggests that our estimation procedure does very well in terms of matching the risk-free rate. This is rather remarkable given that our estimation does not constrain the mean of the distributions in any way. It is interesting to note that even during the crisis, the risk-neutral densities have means that are close to the risk-free rate. The risk-neutral annualized mean returns for the S&P 500 index on September 18th and October 23rd of 2008 are estimated to be -2.81% and 7.14%, respectively. The estimated risk-neutral annualized mean returns on September 17th and October 22nd of 2008 for the FTSE 100 are -5.7% and 13%. It may be that the risk-neutral means are generally close to the risk-free rate because most option traders use some variant of the Black-Scholes model, which sets the risk-neutral mean equal to the risk-free rate.

Considering next the annualized standard deviations of risk-neutral and physical densities, the risk-neutral densities have higher average standard deviations than the physical densities for both indices. Their standard deviations are also much more variable than those of the physical densities. This difference is presumably driven by the conditional nature of the risk-neutral densities. When investors believe the market will be volatile in the future, this belief is immediately reflected by the risk-neutral density. However, the kernel density estimator used in the classic procedure smoothes out any extreme returns and has no way to incorporate investors' beliefs. For the S&P 500, the estimated risk-neutral annualized standard deviations for September 18th and October 23rd of 2008 are 61% and 77%, respectively. The corresponding values for the physical density are 12.96% and 16.4%. The FTSE 100 risk-neutral densities on September 17th and October 22nd of 2008 have annualized standard deviations of 38% and 56%, also much higher than the estimates under our rolling window physical density estimates which have annualized standard deviations of 18%

for both days. While the physical densities certainly respond to the extreme returns during the financial crisis, their response is much smaller than the response of the risk-neutral densities.

The monthly skewness and kurtosis values are quite different for risk-neutral densities than they are for physical densities. The results on these higher moments combined with those for the means and standard deviations suggest that using a smoothing method to estimate the conditional physical densities is misguided. As discussed earlier, the implicit assumption made in order to use rolling window estimates for the physical densities is that the physical densities are stable over time. In our data, neither the physical nor the risk-neutral densities appear stable over time. Furthermore, if the pricing kernel is stationary then the physical and the risk-neutral densities should be related to each other. In fact, in a Black-Scholes world, the variance, skewness and kurtosis of the risk-neutral density are equal to those of the physical density. However, in our data the moments of the risk-neutral densities are not very close to those of the physical densities. Even using models which forecast variances (e.g. *Rosenberg and Engle* (2002)) will likely fail to miss variation in skewness or kurtosis. This highlights a major advantage of the CDI method over existing methods.

2.5.1 Classic Method Results

We first present the results of estimating the average of a series of estimated SDFs using the classic nonparametric method similar to those of *Jackwerth* (2000) and *Ait-Sahalia and Lo* (2000). We should point out that while our classic method estimates are similar to those of other papers, they are not exactly the same as any particular paper. We use monthly data over a longer time span than most other papers, and other papers often have slightly different ways to model the SDF. Nevertheless, our classic method results should be very similar to those of other papers. For both the

FTSE 100 and the S&P 500 data, we use the same risk-neutral densities that are used in the CDI method. These risk-neutral densities are estimated using the procedure described in Section 2.2.2, and an example of a risk-neutral density estimate appears in Figure 2. We then estimate the corresponding physical densities using a Gaussian kernel density estimator based upon a rolling window of past returns. We use a bandwidth of $h = n^{-\frac{1}{5}} \times \sigma_{data}$, where σ_{data} denotes the standard deviation of all the data used in the kernel estimation for all time periods. The results do not seem to vary much with different choices of h. When using the kernel density estimator, there is a trade off between the number of data points available and the temporal proximity of the data points. A larger number of data points improves the mechanical estimation of the kernel density estimator, but does not solve the real problem, which is the use of backward-looking data to estimate conditional beliefs. By taking realized returns further back, we are using older, possibly irrelevant data as far as investors' time t decision making is concerned.

Figures 2.4 and 2.5 present estimation results using the classic nonparametric method. The panels of Figures 2.4 and 2.5 use different window lengths when calculating the physical densities of returns. In all panels, the same general pattern appears but significant variations arise across different window lengths. The SDF is sharply decreasing over states with low returns before displaying nonmonotonicity and sometimes gradual increasing as returns increase. In both figures the four panels look similar over lower returns, while there is some variation across the panels as returns increase. We are not able to estimate the mean SDF with any precision for gross index returns outside of the range of 0.9 to 1.1. Even though index realizations of 0.9 (-10% change) are rare, they do exist and we would like to be able to identify the form of the pricing kernel at such low return values. As we look toward larger returns, in the S&P 500 panels we see a portion of the estimated SDF that is increasing in returns between 0.95 and 1.0. We also see at least one bump that appears for short

rolling windows but not for long windows. The FTSE 100 estimates appear almost flat for some window lengths, and again bumps appear and disappear as the window length changes. It is surprising how much these classic estimates vary as we change the window length. An estimator that changes our inference about nonmonotonicity as we alter the window length for estimating physical densities does not seem very robust.

The figures include pointwise 95% bootstrapped confidence intervals. Since we use a rolling window of historical data to estimate the physical densities, we are able to obtain tighter confidence intervals than we will using the CDI method, which does not use a window of previous returns. Accordingly, the intervals become tighter as we increase the length of the rolling window for both the FTSE 100 and the S&P 500 estimates. The confidence intervals are in fact tight enough so that in every panel in both Figures 2.4 and 2.5, we are able to obtain statistically significant non-monotonicity. We define a non-monotonicity to be statistically significant in the estimated SDF if at any point on the returns (horizontal) axis, the lower confidence bound exceeds the upper bound of any confidence interval at a lower level of returns. For example, in each panel of Figure 2.4, the lower confidence bound at 1.02 on the returns axis exceeds the upper confidence bound at 0.98. Therefore the estimates exhibit a statistically significant non-monotonicity. As one would expect, using a longer window of returns allows us to identify non-monotonicity at higher confidence levels. In Panel A of Figure 2.4, the non-monotonicity is just significant at the 95% level. However, as we increase the length of the rolling windows used in our estimates, the confidence intervals become tighter and the non-monotonicities are more pronounced and thus are significant at even higher levels of confidence.

2.5.2 CDI Results

The upward sloping portions of the SDF in Figures 2.4 and 2.5 cannot be easily reconciled with standard economic theory of risk averse investors, and similar estimates in the literature have perpetuated the pricing kernel puzzle. The remainder of the paper investigates whether properly accounting for investors' information sets can eliminate the non-monotonicities of estimated SDFs as functions of the index returns.

In order to simultaneously select the optimal model and estimate θ, we evaluate Equation (2.12) for different numbers of moment restrictions and spline bases and then compare the Cramer-von Mises statistics for each of the 1081 combinations of b and m satisfying $5 \leq b \leq m \leq 50$. The smallest Cramer-von Mises statistic occurs when $b = m = 9$, for both the FTSE and S&P data, with values of 0.00016 (p = 0.976) and 0.00047 (p = 0.836), respectively. This means that the optimal model we choose will solve Equation (2.11) when using the first nine moment restrictions of the $U[0, 1]$ distribution to estimate the coefficients for a spline with nine bases. The null hypothesis of each associated goodness of fit test is that the estimated distribution comes from the hypothesized distribution of $U[0, 1]$.

We also calculate the Cramer-von Mises statistic and corresponding p-value for our data in the case of no transformation. These statistics indicate the form the results would take if we did not transform the data by estimating a pricing kernel. More specifically, the case of no transformation means that we take $g(y) \equiv 1$ in Equation (2.8). So the non-transformed data we use to calculate the Cramer-von Mises statistic is given by the vector V with

$$V_i = \int_{-\infty}^{X_i} d\mathbb{F}_i^Q(y), i = 1, ...T.$$

For the S&P 500 the untransformed data produce a statistic of 0.00057 (p = 0.534), while for the FTSE 100 the statistic is 0.0012 (p = 0.481). These numbers

imply that our estimation procedure succeeds in transforming the S&P 500 data to a $U[0,1]$ sample quite well. We are not able to fit the FTSE data to the uniform distribution quite as well as we can the S&P data. We can also see from the results that even prior to our estimation, the data are not statistically different from $U[0,1]$ at accepted significance levels. These results should not be considered a formal test comparing the transformed model to the non-transformed data. That being said, our transformation does appear to improve the fit and according to the Cramer-von Mises criterion the fit is very good for the S&P data. For the FTSE data, the fit is not quite as strong but is still good. Figure 2.6 displays histograms of our data before and after the transformation. Panels A and B clearly show the Cramer-von Mises results for the S&P 500 are confirmed. The transformed S&P 500 data appears very close to a uniform distribution over the unit interval and it does appear more uniform than the non-transformed data. Panels C and D, on the other hand, show that we are not able to fit the uniform distribution of with the FTSE data nearly as well as we can with the S&P data. Furthermore, the histograms in Panels C and D do not visually display the improvement in fit suggested by the Cramer-von Mises statistics. This is simply due to the fact that the histogram with fairly thick bars is not always a good indication of fit. Both the Cramer-von Mises results and Table 2.2, which we discuss below, show a significant improvement in fit from the non-transformed to the transformed FTSE data. The vertical axis in Figure 2.6 counts the number of data points falling within each bin as opposed to the density, which is simply a normalization of the count.

We focus on the functional form of the inverse of the function $\hat{g}$ whose estimation is described in Section 2.2. Below, we plot the estimated functional form of $\hat{M}(x) = \frac{1}{\hat{g}(x)}$ which we will refer to as the SDF since $e^{-r_t\tau}$ is approximately equal to one for our entire sample. Furthermore, multiplying $M(x)$ by a constant will not change the qualitative aspects of the SDF we are attempting to capture.

It is easily seen from Figure 2.7 that the SDF estimated with the CDI approach is a downward sloping function of S&P 500 index realizations. Figure 2.8 shows the estimated SDF for the FTSE data is downward sloping over the returns ranging from 0.88 to 1.03, but is upward sloping at returns larger than 1.03. However, there are relatively few observed returns larger than 1.05 in the FTSE data set. As a result, our nonparametric estimator is bound to be imprecise at larger values of index returns. The SDF estimates based on the FTSE data look similar to the $N = 200$ estimates in Panel A of Figure 2.3. This could suggest that the true SDF is actually downward sloping everywhere while our estimate shows non-monotonicity in the right tail only as a result of insufficient data. In order to investigate whether there are indeed non-monotonicities in the SDF, we need to determine whether the non-monotonicity of the estimated SDF is statistically significant. We include bootstrap confidence intervals based on 20,000 resamples in Figures 2.7 and 2.8. In virtually all forms of non-parametric estimation, an extremely large set of data is required for one to achieve tight confidence intervals. Since options data does not go back very far, we don't have many extreme observed returns within the time series of realized returns corresponding to the options data. As a result, confidence intervals for our estimates are not very tight at the extreme ends of the estimated SDFs. It can be seen in Figures 2.7 and 2.8 that the pointwise 95% confidence intervals for the SDF are not very tight in regions that correspond to far out-of-the-money options. This is to be expected as we have only 196 months worth of S&P data and 121 months for the FTSE data.

We note that the estimated SDF based on the S&P 500 data, which has 33% more observations than the FTSE data, is clearly downward sloping and the pointwise confidence intervals, while wide at certain points, do not allow us to reject monotonicity. Furthermore, the confidence intervals are rather tight between 0.95 and 1.05, a region where many previous studies have found the SDF to be increasing. Our estimated

downward sloping M is in agreement with mainstream financial and economic theory that risk averse investors' marginal rates of substitution should be downward sloping as a function of states of the world. While the FTSE 100 SDF appears upward sloping in the region of large positive returns, the 95% confidence intervals show that this non-monotonicity is not statistically significant. Thus, our evidence suggests that avoiding the mixture of forward-looking and historical data is a solution to the pricing kernel puzzle.

Since the CDI method is related to the estimation method of *Bliss and Panigirtzoglou* (2005), we report results of the Berkowitz test, which is the main test used in *Bliss and Panigirtzoglou* (2005) to assess parametric estimates of the risk aversion function. The test involves two separate likelihood ratio tests. The first, with a test statistic denoted LR_3 is a joint test of the hypothesis that our observed cumulants, $\int_{-\infty}^{X_t} \hat{g}(y) d\mathbb{F}_t^Q(y)$, $t = 1, 2, ..., T$, are i.i.d. and are uniformly distributed over the interval $[0, 1]$. The LR_3 test statistic is distributed χ_3^2 asymptotically. The second likelihood ratio test, with test statistic LR_1, tests the null hypothesis that our observations are iid. The LR_1 statistic follows a χ_1^2 asymptotic distribution. The two likelihood ratio tests are complementary in that if we reject the joint test based upon LR_3, but we do not reject the test of independence based upon LR_1, then it must be the case that we reject the null hypothesis of a uniform distribution. Rejecting the hypothesis of a uniform distribution after the transformation would mean that we do not have the correct SDF, whose inverse transforms our data to a uniform distribution. The results of the Berkowitz test are given in Table 2.2. We report the results of the test for both the transformed data as well as the non-transformed data, $\int_{-\infty}^{X_t} d\mathbb{F}_t^Q(y)$, $t = 1, 2, ..., T$.

We can see in Panel A of Table 2.2, that for the untransformed S&P 500 data, we can reject the joint hypothesis at the 90% confidence level, with a p-value of 0.0732. This result, along with the fact that we cannot reject the test of independence, implies

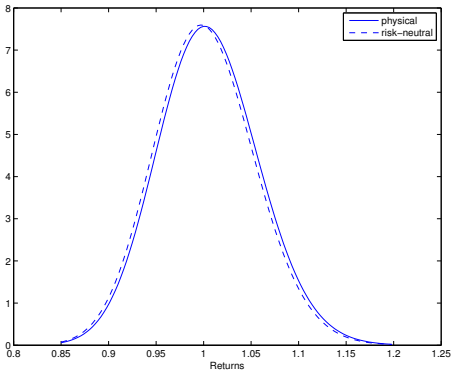
that the non-transformed data cannot be rejected as independent but we can reject the hypothesis of a uniform distribution. On the other hand, the transformed data has a p-value of 0.8777 for the joint test, confirming the results of the Cramer-von Mises statistics and suggesting that the transformation gives a valid SDF. Panel B of Table 2.2 shows that the transformation of the FTSE data is not able to match the uniform distribution as well as that of the S&P data. Again the LR_1 statistics for both the non-transformed data and the transformed data are small enough that the the corresponding p-values are 0.9348 and 0.8544 respectively. This means that the data appear to be convincingly independent. However, the LR_3 statistics of 6.4839 and 2.4112 with corresponding p-values of 0.0903 and 0.4916 suggest that we can reject the uniform distribution of the non-transformed FTSE 100 data but we cannot reject the uniform distribution for the transformed data. However, the p-value of 0.4916 corresponding to the LR_3 statistic in Panel B does not suggest that we have a very great fit of the data to the uniform distribution over the unit interval.

2.6 Conclusion

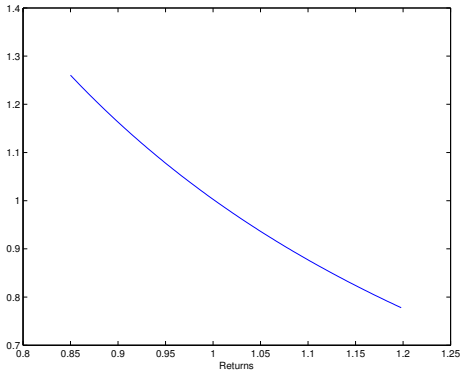
The pricing kernel puzzle is the finding that the stochastic discount factor implied by option prices and historical returns data is not monotonically decreasing in market returns. We argue that this finding is an artifact of econometric technique, driven particularly by comparing two estimates of densities that condition on different information sets. We propose a new nonparametric pricing kernel estimator that properly reflects all the information that option investors use when they set option prices. Our estimator outperforms the classical method in simulations. In S&P 500 and FTSE index option data, our estimator suggests that the pricing kernel is monotonically decreasing in market returns.

It is important to confirm that the stochastic discount factor is monontonically decreasing in market returns because a discount factor that increases in returns over

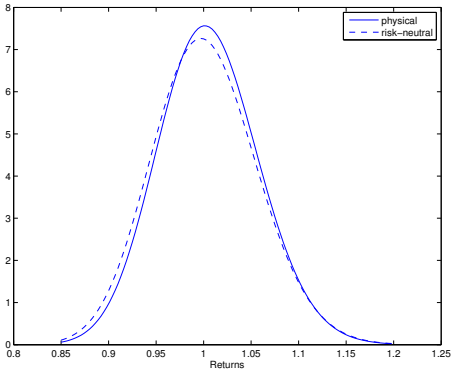
some range implies that the representative agent prefers lower returns (or higher risk) over that range. It is unnatural to think of the representative agent exhibiting risk-loving behavior over any range of market returns. Explaining the pricing kernel puzzle therefore lends credence to standard risk and return theory.



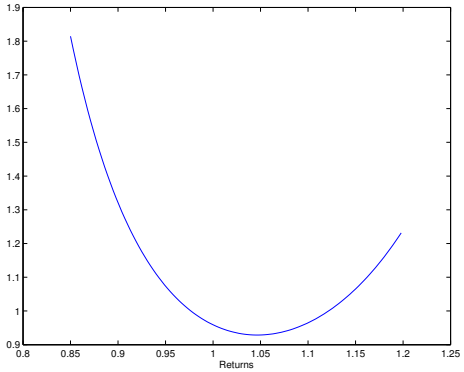
A) log-normal densities $\sigma_P = \sigma_Q$



B) SDF corresponding to panel A



C) log-normal densities $\sigma_P < \sigma_Q$



D) SDF corresponding to panel C

Figure 2.1: Black-Scholes-implied densities

Panel A plots the log-normal risk-neutral (dashed) and physical (solid) densities that arise under the Black-Scholes model. We choose location parameters to match those of our samples for monthly returns. The physical location parameter is thus set to $\mu_P = 0.0040$ and the risk-neutral location parameter is set equal to $\mu_Q = 0.00011$. Under the Black-Scholes model, both distributions have the same scale parameter, σ, so we set these both equal to the scale parameter for our sample of (monthly physical) returns, $\sigma_P = \sigma_Q = 0.0526$. Panel B plots the SDF corresponding to the densities in Panel A. In Panel C we slightly increase σ_Q to $\sigma_Q = 0.055$, and in Panel D we plot the corresponding SDF.

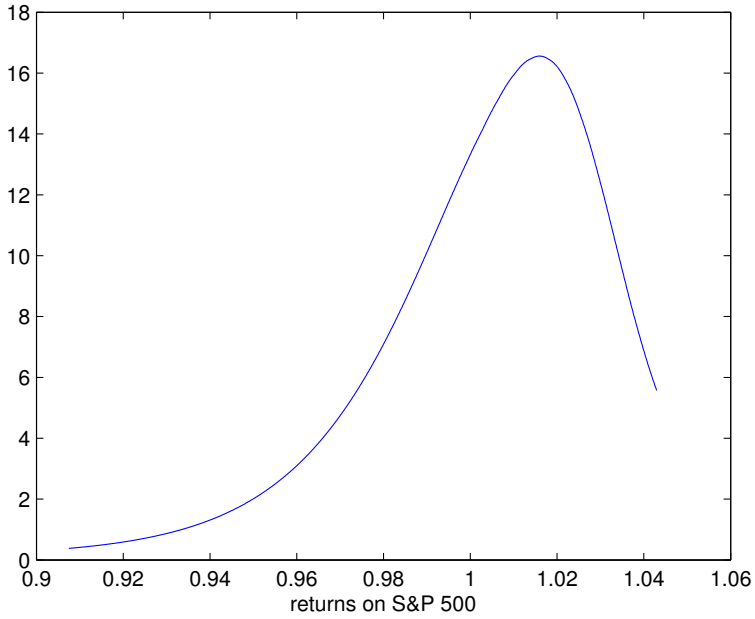
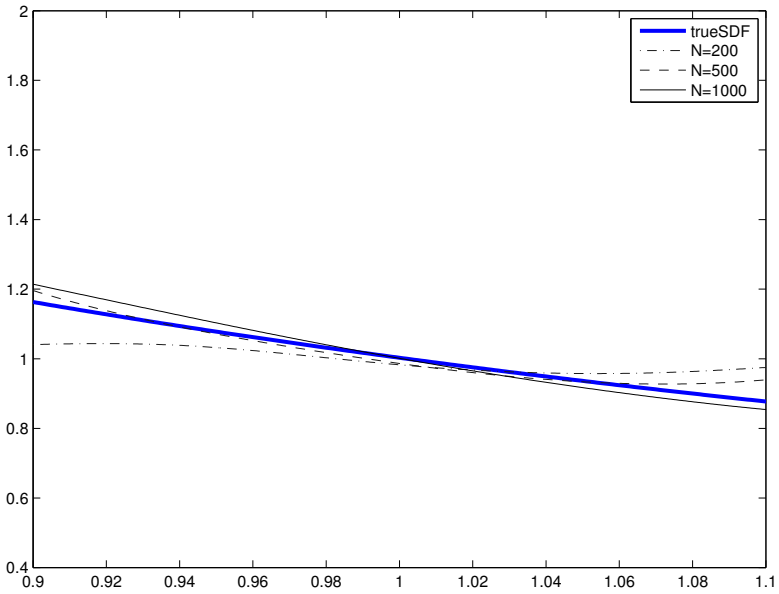
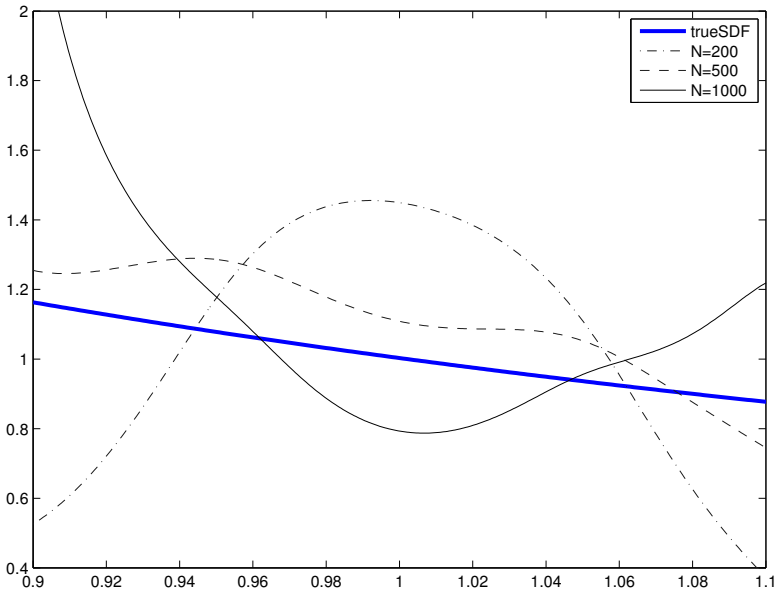


Figure 2.2: Example risk-neutral density
This risk-neutral density was estimated using option prices from April 20, 2006 with best bids exceeding \$3/8. For April 20, 2006, there are 43 valid option prices which we use, corresponding to 37 unique strike prices. Each month we use option prices to estimate a risk-neutral density like this one. We estimate the tails of the distribution by matching a generalized Pareto Distribution to the slope of the density very close to where we can no longer estimate it. The method for estimating the risk-neutral densities is described in detail in Section 2.2.2.

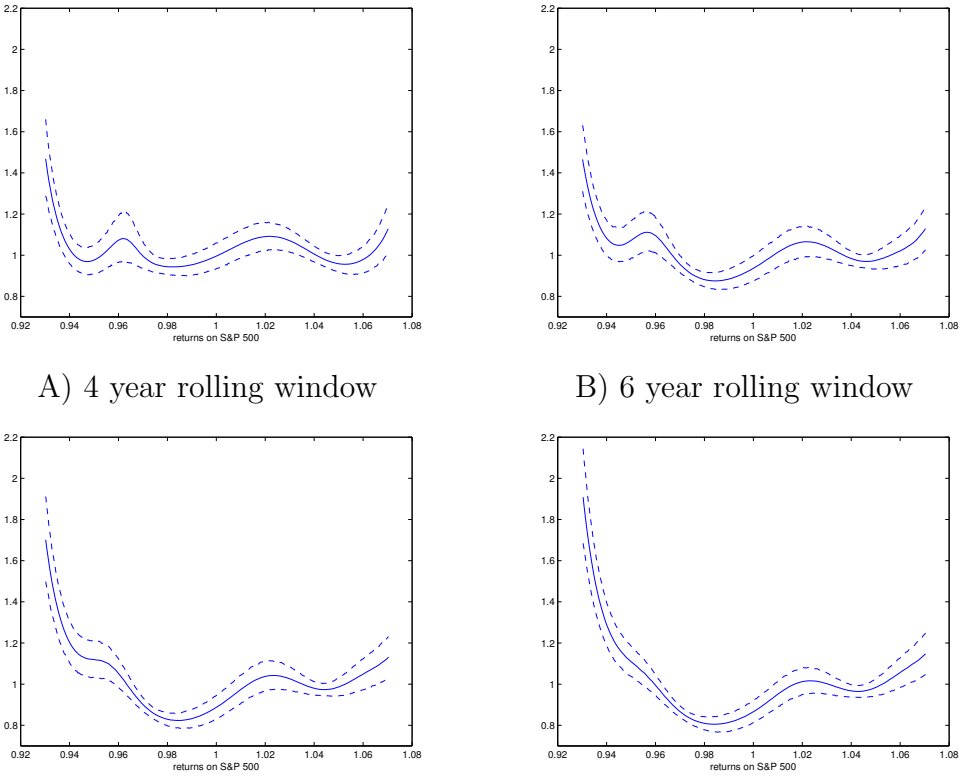


A)CDI estimates



B) Classical method estimates

Figure 2.3: Estimated and true SDFs from simulations
The side by side plots compare the performance of the CDI method and classical method of non-parametric estimates of the SDF. Our simulated data is generated using the true SDF depicted by the bold line in each panel. The estimates of each method are depicted with the true SDF.



A) 4 year rolling window

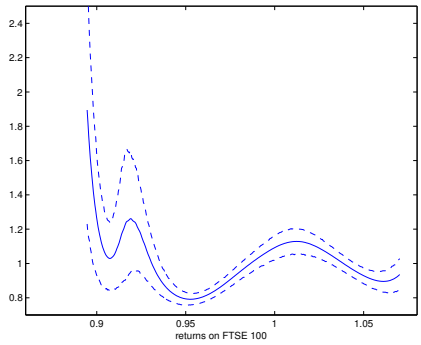
B) 6 year rolling window

C) 8 year rolling window

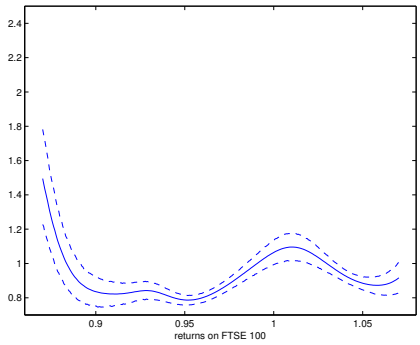
D) 10 year rolling window

Figure 2.4: Estimated SDFs using classic procedure: S&P 500

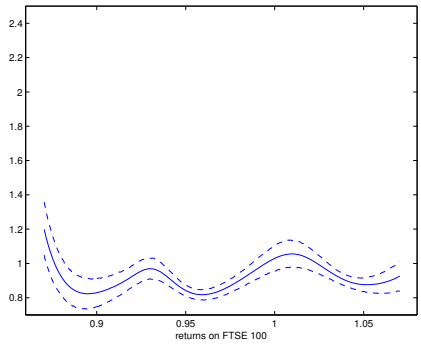
Version of the classic nonparametric estimates of the stochastic discount factor as the average of monthly SDF estimates with pointwise bootstrap 95% confidence intervals. Each monthly SDF is the ratio of a risk-neutral density to a physical density estimate of returns on the S&P 500 index. Each panel represents the resulting estimate when a different widow is used to estimate the physical density using a Gaussian kernel estimator.



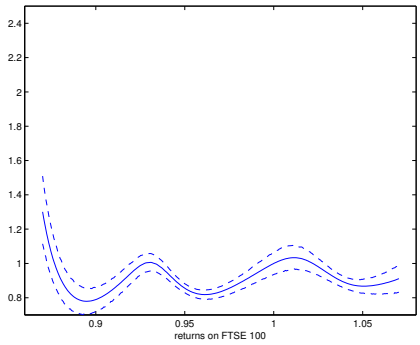
A) 4 year rolling window



B) 6 year rolling window

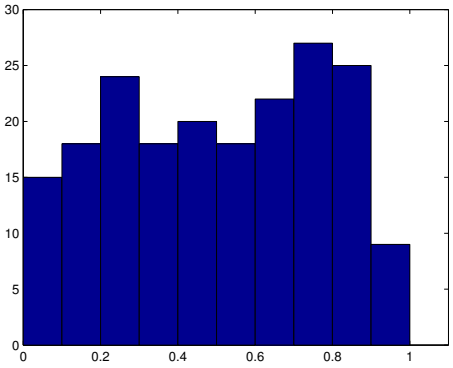


C) 8 year rolling window

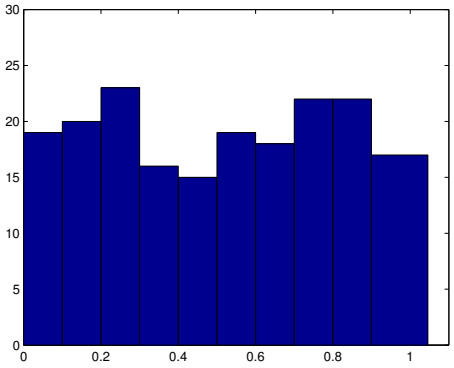


D) 10 year rolling window

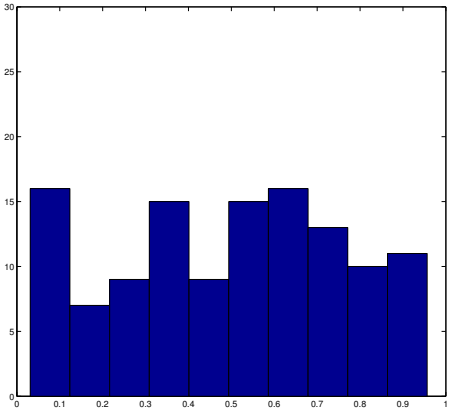
Figure 2.5: Estimated SDFs using classic procedure: FTSE 100
Version of the classic nonparametric estimates of the stochastic discount factor as the average of monthly SDF estimates with pointwise bootstrap 95% confidence intervals. Each monthly SDF is the ratio of a risk-neutral density to a physical density estimate of returns on the FTSE 100 index. Each panel represents the resulting estimate when a different widow is used to estimate the physical density using a Guassian kernel estimator.



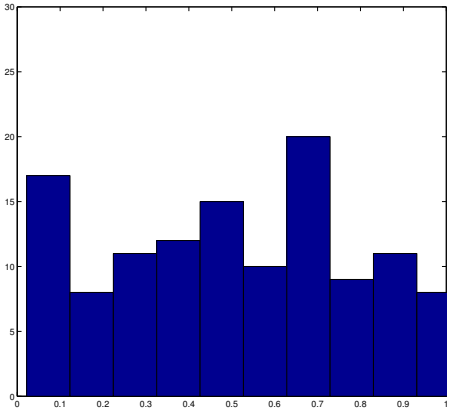
A) No pricing kernel S&P 500



B) Optimal pricing kernel S&P 500 ($b = m = 9$)



C) No pricing kernel FTSE 100



D) Optimal pricing kernel FTSE 100 ($b = m = 9$)

Figure 2.6: Histograms of cumulants with and without a pricing kernel
The histograms plotted in Panels A and C are estimates of the density of the cumulants that result from integrating risk-neutral densities up to their corresponding realized values of the S&P 500 and FTSE 100 data respectively, or $\int_{-\infty}^{X_t} d\mathbb{F}_t^Q(y), \quad t = 1, 2, ..., T$. If the pricing kernel is constant (or there is no compensation for risk) then we would expect this histogram to be close to a uniform [0,1] density. The histograms in Panels B and D are estimates of the density of corresponding cumulants resulting from our CDI estimation method. Specifically, it is a histogram of $\int_{-\infty}^{X_t} \hat{g}(y) d\mathbb{F}_t^Q(y), \quad t = 1, 2, ..., T$, where $\hat{g}(y)$ is the CDI estimate of the inverse of the pricing kernel. The fact that the histogram in Panel B appears to be approximately uniformly [0,1] distributed shows that the CDI pricing kernel fits the S&P 500 data very well. The histogram in Panel D shows that the CDI pricing kernel fits the FTSE 100 data only moderately well.

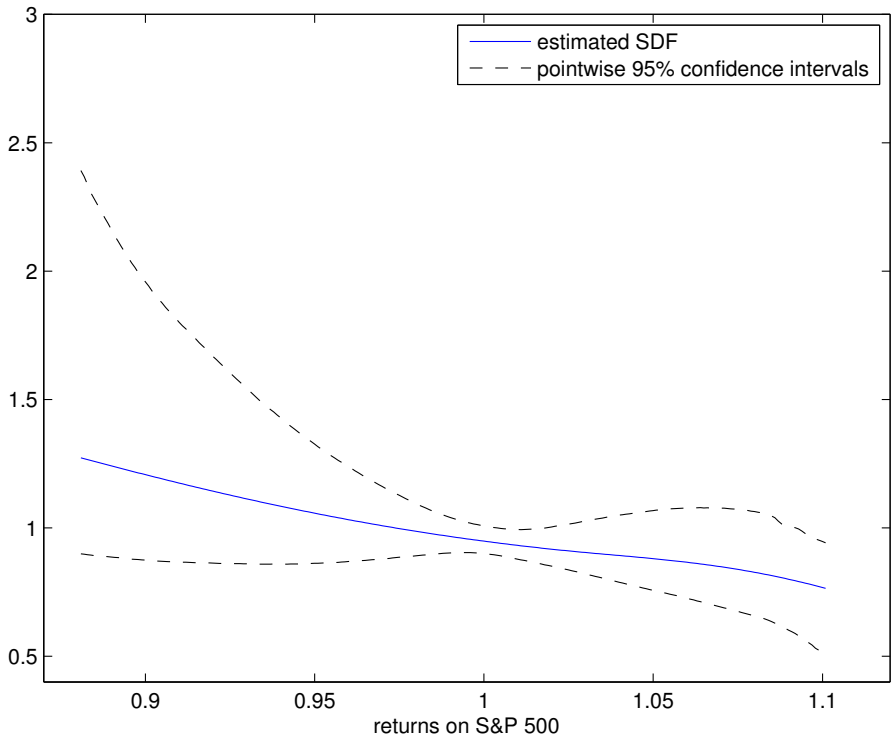


Figure 2.7: Estimated stochastic discount factor using CDI method: S&P 500
The result of our CDI estimation of the pricing kernel for the S&P 500 is plotted above. It is clearly monontonically decreasing on the interval over which we can estimate it with some precision. The CDI method estimates the pricing kernel by matching the moments of the distribution of the cumulants,

$$\int_{-\infty}^{X_t} \hat{g}(y) d\mathbb{F}_t^Q(y), \quad t = 1, 2, ..., T,$$

to the moments of the uniform distribution by nonparametrically estimating the function $g(\cdot)$. The SDF in this formulation is actually the inverse of $g(\cdot)$, so that is what we plot above. 95% confidence intervals, which are plotted with dashed lines, are based on 20,000 bootstrap iterations of the CDI method, sampling our set of dates with replacement.

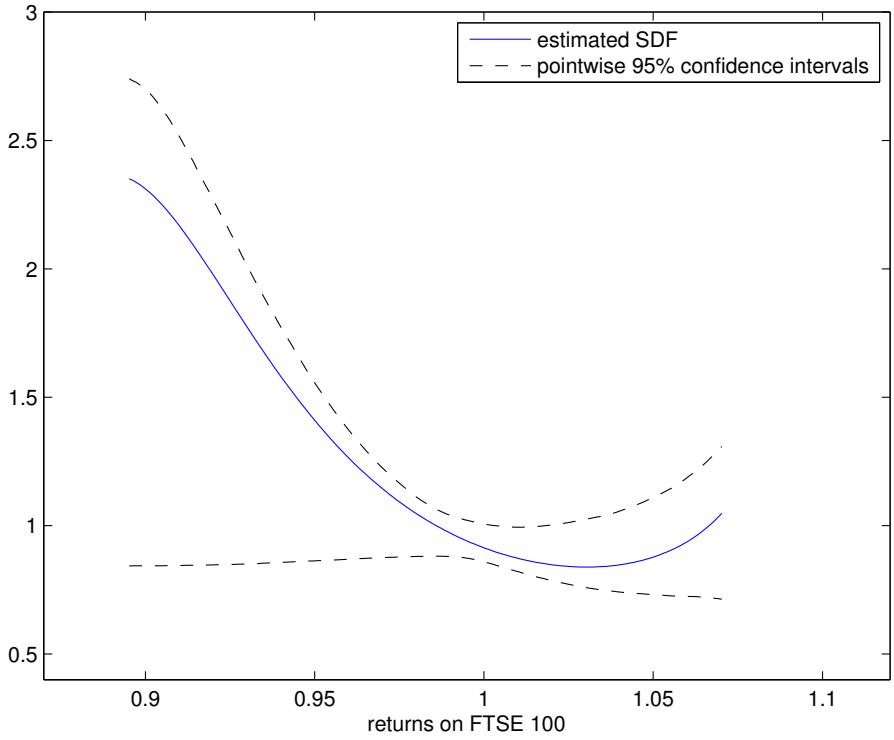


Figure 2.8: Estimated stochastic discount factor using CDI method: FTSE 100
The result of our CDI estimation of the pricing kernel for the FTSE 100 is plotted above. The estimate exhibits some non-monotonicity at the end of the interval over which we can estimate it with some precision. The non-monotonicity is not statistically significant according to the 95% bootstrapped confidence intervals. The CDI method estimates the pricing kernel by matching the moments of the distribution of the cumulants,

$$\int_{-\infty}^{X_t} \hat{g}(y) d\mathbb{F}_t^Q(y), \quad t = 1, 2, ..., T,$$

to the moments of the uniform distribution by nonparametrically estimating the function $g(\cdot)$. The SDF in this formulation is actually the inverse of $g(\cdot)$, so that is what we plot above. 95% confidence intervals, which are plotted with dashed lines, are based on 20,000 bootstrap iterations of the CDI method, sampling our set of dates with replacement.

Table 2.1: Summary statistics for risk-neutral densities

For each of the months in our sample (196 months: from September 1996 through December 2012 for S&P 500 data. 121 months: from January 2002 to December 2012 for FTSE 100 data), we estimate both a risk-neutral density based on option prices and a physical density based on historical data. The physical densities are estimated with a Gaussian kernel density estimator using 60 months of past returns, and the risk-neutral densities are estimated as described in Section 2.2.2. This table reports summary statistics on the moments of these densities. The table reports both sample averages and sample standard deviations of the first four centralized moments in terms of returns: mean, standard deviation, skewness and kurtosis. The average means and standard deviations are annualized to ease interpretation.

Panel A: S&P 500

	Risk-Neutral Densities from Options Prices			
	Annualized Mean Ret	Annualized Standard Dev	Monthly Skewness	Monthly Kurtosis
Sample average	2.76%	22.98%	-1.1814	6.2283
Sample standard deviation	0.97%	24.00%	0.4844	2.1343
	Physical Densities from 60 months of Historical Data			
	Annualized Mean Ret	Annualized Standard Dev	Monthly Skewness	Monthly Kurtosis
Sample average	6.48%	18.33%	-0.4661	4.1178
Sample standard deviation	2.29%	11.49%	0.3537	1.4569

Panel B: FTSE 100

	Risk-Neutral Densities from Options Prices			
	Annualized Mean Ret	Annualized Standard Dev	Monthly Skewness	Monthly Kurtosis
Sample average	3.57%	21.43 %	-1.0365	7.9168
Sample standard deviation	4.01%	9.00%	0.6857	4.1581
	Physical Densities from 60 months of Historical Data			
	Annualized Mean Ret	Annualized Standard Dev	Monthly Skewness	Monthly Kurtosis
Sample average	3.45%	16.71 %	-0.5273	3.4905
Sample standard deviation	3.08%	1.10%	0.1047	0.2389

Table 2.2: Berkowitz statistics and p-values

The first line reports likelihood ratio test statistics and corresponding p-values for Berkowitz tests of the transformed data using the optimal model, $b = m = 9$. The second line reports likelihood ratio test statistics and corresponding p-values for Berkowitz tests of the non-transformed data, or the data without a pricing kernel. The LR_3 statistic tests the joint hypothesis that data is iid and $U[0,1]$. The LR_1 statistic tests the hypothesis that the data are independent. Rejection based upon the LR_3 statistic can come from the data not being independent or the data not being uniformly distributed. If we reject base upon the LR_3 statistic but fail to reject based upon the LR_1 statistic, this implies that the data does a poor job fitting the $U[0,1]$ distribution.

Panel A: S&P 500

Model	LR_3	p-value	LR_1	p-value
Optimal model ($b = m = 9$)	0.6808	0.8777	0.0054	0.9412
No pricing kernel	6.9593	0.0732	0.0447	0.8326

Panel B: FTSE 100

Model	LR_3	p-value	LR_1	p-value
Optimal model ($b = m = 9$)	2.4112	0.4916	0.0337	0.8544
No pricing kernel	6.4839	0.0903	0.0067	0.9348

APPENDICES

APPENDIX A

Proofs

Proof of Proposition II.3: We first prove existence. We can apply the Radon-Nikodym Theorem on the probability space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, where $\mathcal{B}(\mathbb{R})$ is the Borel σ-field generated on $\mathbb{R}$. Then by the Radon-Nikodym Theorem, there exists $(a.s\ \mathbb{Q})$ unique random variable $\frac{d\mathbb{P}}{d\mathbb{Q}}$ such that

$$\mathbb{P}((-\infty, x]) = \mathbb{P}(x) = \int_{-\infty}^{X_t} \frac{d\mathbb{P}}{d\mathbb{Q}}(y) d\mathbb{Q}(y) \quad \forall x \in \mathbb{R}. \tag{A.1}$$

Now if we define $\mathcal{G}_t(X_t)$ by

$$\mathcal{G}_t(X_t) := \int_{-\infty}^{X_t} g(y) d\mathbb{Q}(y), \tag{A.2}$$

we know from Proposition II.2, that if we take $g(y) = \frac{d\mathbb{P}}{d\mathbb{Q}}(y)$, then we have $\mathcal{G}(X) \sim U[0,1]$. This establishes existence.

Next we establish uniqueness. Since we can only show almost sure $(\mathbb{Q})$ uniqueness, we reduce the space in question by removing all $\mathbb{Q}-$null sets. Call this reduced space over the real line $\mathbb{R}'$. Since g is non-negative, the function $\mathcal{G}$ uniquely determines where g must be zero over $\mathcal{B}(\mathbb{R}')$. So any functions satisfying the criteria of the

proposition must take the value zero over the exact same subsets of $\mathcal{B}(\mathbb{R}')$. Now it only remains to show that over the sets where $g \neq 0$, the functional form is unique. Let $\mathcal{N}$ denote the set in $\mathcal{B}(\mathbb{R}')$ where $g > 0$. Over this set, the function $\mathcal{G}$ is invertible because $g > 0$.

Suppose there is another function g' satisfying Equation (A.2) over $\mathcal{N}$. Define $\mathcal{G}'_t$ as

$$\mathcal{G}'_t(X_t) \equiv \int_{-\infty}^{X_t} g'(y) d\mathbb{Q}(y),$$

where, by our assumption on g', we know $\mathcal{G}'(X) \sim U[0,1]$. Since $\mathcal{G}$ and $\mathcal{G}'$ are invertible over $\mathcal{N}$, we know that on the restricted domain, for a fixed x,

$$\mathbb{P}(\mathcal{G}'(X) \leq x) = \mathbb{P}(X \leq \mathcal{G}'^{-1}(x))$$

and

$$\mathbb{P}(\mathcal{G}(X) \leq x) = \mathbb{P}(X \leq \mathcal{G}^{-1}(x)).$$

Since P and Q are equivalent by assumption, and $\mathcal{N}$ does not contain any $\mathbb{Q}$-null sets, it follows that $\mathcal{N}$ does not contain any $\mathbb{P}$-null sets. This implies that $\mathbb{P}(X \leq \cdot)$ is a strictly increasing function and hence

$$\mathcal{G}'^{-1}(x) = \mathcal{G}^{-1}(x)$$

for a fixed x. It follows that for deterministic sets E (e.g. $E = (-\infty, x]$)

$$\int_E g'(y) d\mathbb{Q}(y) = \int_E g(y) d\mathbb{Q}(y) \quad \forall E \subset \mathcal{B}(\mathcal{N}). \tag{A.3}$$

Now we can apply the Radon-Nikodym Theorem on $(\mathcal{N}, \mathcal{B}(\mathcal{N}), \mathbb{Q})$. From Equation (A.3), the Radon-Nikodym Theorem implies $g' = g$ *a.s.* $\mathbb{Q}$ on $\mathcal{N}$. Since the values of g and g' must be zero on non-null subsets of $\mathcal{N}^c$, we have that $g' = g$ *a.s.* $\mathbb{Q}$ and

hence g is unique ($a.s.$ $\mathbb{Q}$). ■

APPENDIX B

Correlation Factor

Construction of Correlation Factor B: Since the volatility of the index is comprised of a weighted average of volatilities of individual constituents of the index as well as the weighted pairwise covariances of constituents, we can express the volatility of the index as

$$\sigma_{I,t}^2 = \sum \omega_{i,t}^2 \sigma_{i,t}^2 + 2 \sum_{i \neq j} \omega_i \omega_j \sigma_{i,t} \sigma_{j,t} \rho_{ij,t}, \tag{B.1}$$

where $\sigma_{I,t}^2$ denotes index variance at time t, $\sigma_{i,t}^2$ denotes time t variance of constituent firm i, ω_i denotes the index weight assigned to constituent i and $\rho_{ij,t}$ denotes the pairwise correlation between constituents i and j at time t.

Under the simplifying assumption that $\rho_{ij,t} \equiv \rho_t \quad \forall t$, we can rearrange Equation (B.1) to express correlation as the following:

$$\rho_t = \frac{\sigma_I^2 - \sum \omega_i^2 \sigma_i^2}{2 \sum_{i \neq j} \omega_i \omega_j \sigma_i \sigma_j}.$$

Under this representation, I estimate σ_I and σ_i for each constituent by taking the average of the Black-Scholes-Merton implied volatility of at the money calls and put

options for the particular firm or index. Thus, at each time t I have an estimate of the average pairwise correlation between stocks in the S&P 500.

BIBLIOGRAPHY

BIBLIOGRAPHY

Adrian, T., and J. Rosenberg (2008), Stock returns and volatility: Pricing the short-run and long-run components of market risk, *The Journal of Finance*, *63*(6), 2997–3030.

Ait-Sahalia, Y., and A. W. Lo (2000), Nonparametric risk management and implied risk aversion, *Journal of Econometrics*, *94*(1-2), 9–51.

Andrews, D. W. (1993), Tests for parameter instability and structural change with unknown change point, *Econometrica: Journal of the Econometric Society*, pp. 821–856.

Andries, M., T. M. Eisenbach, and M. C. Schmalz (2014), Asset pricing with horizon-dependent risk aversion, *FRB of New York Staff Report*, (703).

Ang, A., J. Chen, and Y. Xing (2006a), Downside risk, *Review of Financial Studies*, *19*(4), 1191–1239.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006b), The cross-section of volatility and expected returns, *The Journal of Finance*, *61*(1), 259–299.

Audrino, F., and P. Meier (2012), Empirical pricing kernel estimation using a functional gradient descent algorithm based on splines, *Working paper*.

Bakshi, G., and F. Chabi-Yo (2013), Variance bounds on the permanent and transitory components of stochastic discount factors, *Forthcoming, Journal of Financial Economics*.

Bakshi, G., and N. Kapadia (2003a), Delta-hedged gains and the negative market volatility risk premium, *Review of Financial Studies*, *16*(2), 527–566.

Bakshi, G., and N. Kapadia (2003b), Volatility risk premiums embedded in individual equity options: Some new insights, *The Journal of Derivatives*, *11*(1), 45–54.

Bakshi, G., C. Cao, and Z. Chen (2000), Do call prices and the underlying stock always move in the same direction?, *Review of Financial Studies*, *13*(3), 549–584.

Bakshi, G., N. Kapadia, and D. Madan (2003), Stock return characteristics, skew laws, and the differential pricing of individual equity options, *Review of Financial Studies*, *16*(1), 101–143.

Bakshi, G., D. Madan, and G. Panayotov (2010), Returns of claims on the upside and the viability of u-shaped pricing kernels, *Journal of Financial Economics*, *97*, 130–154.

Balkema, A. A., and L. De Haan (1974), Residual life time at great age, *The Annals of Probability*, pp. 792–804.

Bansal, R., D. Kiku, I. Shaliastovich, and A. Yaron (2013), Volatility, the macroeconomy, and asset prices, *The Journal of Finance*.

Barone-Adesi, G., R. F. Engle, and L. Mancini (2008), A garch option pricing model with filtered historical simulation, *Review of Financial Studies*, *21*, 1223–1258.

Barone-Adesi, G., L. Mancini, and H. Shefrin (2013), A tale of two investors: Estimating optimism and overconfidence, *Working paper*.

Bates, D. S. (2000), Post-'87 crash fears in the s&p 500 futures option market, *Journal of Econometrics*, *94*(1-2), 181–238.

Bates, D. S. (2012), Us stock market crash risk, 1926–2010, *Journal of Financial Economics*, *105*(2), 229–259.

Beare, B. K., and L. Schmidt (2013), An empirical test of pricing kernel monotonicity, *Working paper*.

Billingsley, P. (2012), *Probability and Measure*, John Wiley & Sons.

Black, F. (1976), {Studies of stock price volatility changes}, *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, pp. 177–181.

Bliss, R. R., and N. Panigirtzoglou (2005), Option-implied risk aversion estimates, *The Journal of Finance*, *59*(1), 407–446.

Boguth, O., and L.-A. Kuehn (2013), Consumption volatility risk, *The Journal of Finance*, *68*(6), 2589–2615.

Bollen, N. P., and R. E. Whaley (2004), Does net buying pressure affect the shape of implied volatility functions?, *The Journal of Finance*, *59*(2), 711–753.

Boyer, B. H., and K. Vorkink (2014), Stock options as lotteries, *The Journal of Finance*.

Breeden, D. T., and R. H. Litzenberger (1978), Prices of state-contingent claims implicit in option prices, *Journal of Business*, pp. 621–651.

Buraschi, A., and J. Jackwerth (2001), The price of a smile: Hedging and spanning in option markets, *Review of Financial Studies*, *14*(2), 495–527.

Campbell, J. Y., S. Giglio, C. Polk, and R. Turley (2012), An intertemporal capm with stochastic volatility, *Tech. rep.*, National Bureau of Economic Research.

Cao, J., and B. Han (2013), Cross section of option returns and idiosyncratic stock volatility, *Journal of Financial Economics*, *108*(1), 231–249.

Carr, P., and L. Wu (2009), Variance risk premiums, *Review of Financial Studies*, *22*(3), 1311–1341.

Chabi-Yo, F., R. Garcia, and E. Renault (2007), State dependence can explain the risk aversion puzzle, *Review of Financial Studies*, *21*(2), 973–1011.

Chang, B. Y., P. Christoffersen, and K. Jacobs (2013), Market skewness risk and the cross section of stock returns, *Journal of Financial Economics*, *107*(1), 46–68.

Chaudhuri, R., and M. Schroder (2009), Monotonicity of the stochastic discount factor and expected option returns, *Working paper (SSRN 1344061)*.

Christoffersen, P., R. Goyenko, K. Jacobs, and M. Karoui (2011), Illiquidity premia in the equity options market, *Available at SSRN 1784868*.

Christoffersen, P., S. Heston, and K. Jacobs (2013), Capturing option anomalies with a variance-dependent pricing kernel, *Review of Financial Studies*, *26*, 1962–2006.

Cochrane, J. (2005), *Financial markets and the real economy*, Now Publishers Inc.

Constantinides, G. M., J. C. Jackwerth, and A. Savov (2013), The puzzle of index option returns, *Review of Asset Pricing Studies*, p. rat004.

Coval, J. D., and T. Shumway (2001), Expected option returns, *The journal of Finance*, *56*(3), 983–1009.

Cox, J. C., S. A. Ross, and M. Rubinstein (1979), Option pricing: A simplified approach, *Journal of financial Economics*, *7*(3), 229–263.

Da, Z., and E. Schaumburg (2009), The pricing of volatility risk across asset classes and the fama-french factors, *Tech. rep.*, Working paper, Northwestern University.

Dew-Becker, I., S. Giglio, A. Le, and M. Rodriguez (2014), The price of variance risk, *Fama-Miller Working Paper*.

Di Pietro, V., and G. Vainberg (2006), Systematic variance risk and firm characteristics in the equity options market, *Available at SSRN 858304*.

Dittmar, R., and C. Lundblad (2014), A simple consumption-based asset pricing model and the cross-section of equity returns, *Manuscript*.

Drechsler, I., and A. Yaron (2011), What's vol got to do with it, *Review of Financial Studies*, *24*(1), 1–45.

Driessen, J., P. J. Maenhout, and G. Vilkov (2009), The price of correlation risk: Evidence from equity options, *The Journal of Finance*, *64*(3), 1377–1406.

Duan, J.-C., and J. Wei (2009), Systematic risk and the price structure of individual equity options, *Review of Financial Studies*, *22*(5), 1981–2006.

Duarte, J., and C. S. Jones (2007), The price of market volatility risk, *Unpublished working paper, University of Southern California*.

Efron, B., and R. J. Tibshirani (1993), *An introduction to the bootstrap*, *Monographs on Statistics and Applied Probability*, vol. 57, 1 ed., Chapman and Hall, New York.

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997), *Modelling extremal events: for insurance and finance*, *Stochastic Modelling and Applied Probability*, vol. 33, Springer Verlag.

Eraker, B., M. Johannes, and N. Polson (2003), The impact of jumps in volatility and returns, *The Journal of Finance*, *58*(3), 1269–1300.

Fama, E. F., and J. D. MacBeth (1973), Risk, return, and equilibrium: Empirical tests, *The Journal of Political Economy*, pp. 607–636.

Ferson, W. E., and S. R. Foerster (1994), Finite sample properties of the generalized method of moments in tests of conditional asset pricing models, *Journal of Financial Economics*, *36*(1), 29–55.

Figlewski, S. (1989), Options arbitrage in imperfect markets, *The Journal of Finance*, *44*(5), 1289–1311.

Figlewski, S. (2008), *Volatility and Time Series Econometrics: Essays in Honor of Robert Engle*, chap. 8. Estimating the Implied Risk Neutral Density for the U.S. Market Portfolio, Oxford University Press, Oxford, UK.

Frazzini, A., and L. H. Pedersen (2012), Embedded leverage, *Tech. rep.*, National Bureau of Economic Research.

Garleanu, N., L. H. Pedersen, and A. M. Poteshman (2009), Demand-based option pricing, *Review of Financial Studies*, *22*(10), 4259–4299.

Goyal, A., and A. Saretto (2009), Cross-section of option returns and volatility, *Journal of Financial Economics*, *94*(2), 310–326.

Grith, M., W. K. Härdle, and V. Krätschmer (2013), Reference dependent preferences and the epk puzzle, *Working paper*.

Hansen, L. P. (1982), Large sample properties of generalized method of moments estimators, *Econometrica: Journal of the Econometric Society*, pp. 1029–1054.

Hansen, L. P., and R. Jagannathan (1997), Assessing specification errors in stochastic discount factor models, *The Journal of Finance*, *52*(2), 557–590.

Hansen, L. P., and K. J. Singleton (1982), Generalized instrumental variables estimation of nonlinear rational expectations models, *Econometrica: Journal of the Econometric Society*, pp. 1269–1286.

Härdle, W., Y. Okhrin, and W. Wang (2014), Uniform confidence bands for pricing kernels, *Journal of Financial Econometrics*, *12*, 1–38.

Hayashi, F. (2000), *Econometrics*, vol. 1, 60–69 pp., Princeton University Press.

Jackwerth, J. C. (2000), Recovering risk aversion from option prices and realized returns, *Review of Financial Studies*, *13*(2), 433–451.

Jackwerth, J. C., and M. Rubinstein (1996), Recovering probability distributions from option prices, *The Journal of Finance*, *51*(5), 1611–1631.

Lettau, M., M. Maggiori, and M. Weber (2013), Conditional risk premia in currency markets and other asset classes, *Tech. rep.*, National Bureau of Economic Research.

Newey, W. K., and K. D. West (1987), Hypothesis testing with efficient method of moments estimation, *International Economic Review*, pp. 777–787.

Ni, S. X. (2008), Stock option returns: a puzzle, *Available at SSRN 1259703*.

Pan, J. (2002), The jump-risk premia implicit in options: Evidence from an integrated time-series study, *Journal of financial economics*, *63*(1), 3–50.

Pickands III, J. (1975), Statistical inference using extreme order statistics, *the Annals of Statistics*, *3*, 119–131.

Polkovnichenko, V., and F. Zhao (2013), Probability weighting functions implied in options prices, *Journal of Financial Economics*, *107*, 580–609.

Rosenberg, J. V., and R. F. Engle (2002), Empirical pricing kernels, *Journal of Financial Economics*, *64*(3), 341–372.

Song, Z., and D. Xiu (2014), A tale of two option markets: Pricing kernels and volatility risk, *Working Paper*.

Vanden, J. M. (2004), Options trading and the capm, *Review of Financial Studies*, *17*(1), 207–238.

Ziegler, A. (2007), Why does implied risk aversion smile?, *Review of Financial Studies*, *20*(3), 859–904.