From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

---

Doctoral Dissertations, 2011-2026 Economics

2015

Essays on Financial Economics

Van Tassel, Peter

Princeton, NJ : Princeton University

---

http://arks.princeton.edu/ark:/88435/dsp01xk81jn77d

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

ESSAYS ON FINANCIAL ECONOMICS

PETER VAN TASSEL

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF
ECONOMICS
ADVISER: YACINE AÏT-SAHALIA

SEPTEMBER 2015

© Copyright by Peter Van Tassel, 2015.

All Rights Reserved.

Abstract

Asset prices aggregate information and reflect market expectations about real outcomes. In this dissertation, I examine the informational content of prices and investigate the implications for forecasting returns, volatility, and the successful completion of corporate events with applications for popular hedge fund trading strategies.

The first chapter introduces a structural model for stock and option pricing in mergers and acquisitions. I show theoretically and empirically that option prices contain significant content for forecasting deal outcomes. Additionally, I employ my model to study the risks and returns of merger arbitrage strategies. Consistent with the data, my model predicts that merger arbitrage exhibits low volatility and large Sharpe ratios when deals are likely to succeed. To implement this observation, I construct the returns from a buy and hold strategy that overweights deals with a high implied probability of success. The high probability strategy nearly doubles the monthly Sharpe ratio of an equal weighted strategy that invests in all of the active deals in the economy.

The second chapter, which incorporates material from a joint paper with Yacine Aït-Sahalia and Jiangmin Xu, examines the relationship between high frequency machine-readable news and asset prices. Within the trading day, I show that positive news sentiment forecasts high returns and low volatility, and that large quantities of news forecast high volatility and high volumes. In an application of these observations, I use intraday news sentiment to improve the performance of contrarian trading strategies. Additionally, I demonstrate that intraday patterns in the arrival of news are contemporaneous with patterns in realized volatility and volume, and I document examples of large price movements that lead and lag the news.

The third chapter concludes by proposing a new test of dynamic asset pricing models whose expected returns satisfy a conditional beta relationship. The test applies recent developments from the financial econometrics literature to estimate time varying betas

with high frequency data thereby providing a nonparametric alternative to traditional asset pricing tests. Empirically, I find the conditional CAPM is rejected by the data.

Acknowledgements

Thank you to my advisers, family, and friends for ongoing support over the past five years. Your encouragement has made graduate school a rewarding and enjoyable experience.

To my adviser, Yacine Aït-Sahalia, I am greatly indebted. You motivated me, challenged me intellectually, and prepared me to work independently. Thank you for introducing me to new literatures and for encouraging me to explore different research questions. I owe much gratitude for the generous support and kindness you have demonstrated toward me throughout my graduate studies. Additionally, to Jakub Jurek, I would like to express my sincerest thanks. Since I stepped into your classroom, you have provided unwavering encouragement, tremendous guidance, and insightful feedback. I feel privileged to be your student and am thankful for the invaluable mentorship you have provided over the years.

To the rest of my dissertation committee including Markus Brunnermeier, Jakub Kastl, and Ulrich Müller, as well as Kirill Evdokimov, Valentin Haddad, Harrison Hong, José Scheinkman, David Sraer, and Wei Xiong, thank you for many helpful discussions and much sound advice. You have challenged me to become a better financial economist, and I am grateful for having had the opportunity to work with all of you. To my fellow Ph.D. students, especially Chenchuan Li, Delwin Olivan, and Sharon Traiberman, it has been quite the journey. Thank you for the multitude of ways you have provided help and support along the way. I consider myself fortunate to be your colleague and friend, and I look forward to seeing the remarkable places you will go from here. Additionally, I would like to thank Bobray Bordelon, Laura Hedden, Todd Hines, and Karen Neukirchen for enthusiastic and helpful administrative and data support. Also, many thanks are due to the Graduate School and the Department of Economics for the financial support that made my doctoral studies possible.

To my family, this would not have been possible without you. Thanks Mom and Dad, Andrew, and Sarah for always being there. This is for you. Finally, to Sarah, my wife and co-author in life, it was better writing these chapters of our story together. I am excited to see what we write next.

CONTENTS

Abstract iii

Acknowledgments v

Table of Contents vii

List of Tables x

List of Figures xi

1 Merger Options and Risk Arbitrage **1**

1.1 Introduction 2

1.2 Model of Merger Stock and Option Pricing 8

1.2.1 Stock Prices 10

1.2.2 Option Prices 13

1.2.3 Risk Premia 16

1.3 Data Description and Model Estimation 21

1.3.1 Data Description 21

1.3.2 Model Estimation 23

1.3.3 Examples of Estimated Deals 25

1.4 Forecasting Deal Outcomes 27

1.5 Merger Arbitrage Returns 30

1.5.1 Expected Returns, Volatility, and CAPM Beta 32

1.5.2 Path of Expected Merger Arbitrage Returns 35

1.6 Conclusion 37

References . . . 39
Tables . . . 43
Figures . . . 50
1.A Proofs and Derivations . . . 64
1.B Model Extensions and Discussion . . . 69
1.C Model Estimation . . . 75

2 High Frequency Machine-Readable News and Asset Prices 80
2.1 Introduction . . . 81
2.2 Literature Review . . . 83
2.3 Data Description . . . 87
2.3.1 Thomson Reuters Machine-Readable News . . . 88
2.3.2 Trade and Quote Data . . . 92
2.4 Market Activity Around Intraday News . . . 93
2.4.1 Examples . . . 94
2.4.2 Returns, Volatility, and Volume Around the News . . . 102
2.5 Forecasting Intraday Returns, Volatility, and Volume . . . 111
2.6 News Adjusted Contrarian Trading Strategies . . . 119
2.7 Conclusion . . . 124
References . . . 127
2.A Variable Definitions . . . 133
2.B Tables . . . 136
2.C Figures . . . 139

3 High Frequency Tests of Conditional Linear Factor Models 141
3.1 Introduction . . . 142
3.2 Asset Pricing Tests . . . 146
3.2.1 Infeasible Test . . . 146

3.2.2 Feasible Test 149
3.3 Empirical Application 151
3.4 Conclusion 152
References 153
Tables 156
Figures 157
3.A Proofs and Derivations 158
3.B Noise and Asynchronous Data 165

LIST OF TABLES

1.1 Deal Summary Statistics 43

1.2 Option Summary Statistics 44

1.3 Cash Deal Outcome Regressions 45

1.4 Stock Deal Outcome Regressions 46

1.5 Merger Arbitrage Summary Statistics 47

1.6 Merger Arbitrage Return Regressions 48

1.7 Merger Arbitrage Variance Regressions 49

2.1 News Data Summary Statistics 89

2.2 Stock Market Activity Around Intraday News 105

2.3 Stock Return Summary Statistics After the News 109

2.4 Forecasting Returns, Volatility, and Volume: Panel Regressions 113

2.5 Forecasting Returns: Fama-MacBeth Regressions 117

2.6 Summary Statistics for News Adjusted Contrarian Returns 121

2.7 Summary Statistics and Definitions of Forecasting Variables 136

2.8 Example of Thomson Reuters News Analytics Data 137

2.9 Forecasting Returns, Volatility, and Volume: Panel Regressions II 138

3.1 Conditional CAPM Pricing Errors 156

LIST OF FIGURES

1.1 Model Target Stock Price 50

1.2 Model Target Stock Price Distribution 51

1.3 Model Target Option Prices in a Cash Deal 52

1.4 Model Target Option Prices in a Stock Deal 53

1.5 Model Merger Arbitrage Risk Premia 54

1.6 Model Path of Expected Merger Arbitrage Returns 55

1.7 Google Announces Cash Acquisition of Motorola 56

1.8 Examples of Estimated Cash Deals 57

1.9 Examples of Estimated Stock Deals 58

1.10 Forecasting Cash Deal Outcomes 59

1.11 Option Market Reaction to Cash Deal Announcements 60

1.12 Option Market Reaction to Stock Deal Announcements 61

1.13 Forecasting Deal Outcomes in a Merger Arbitrage Portfolio 62

1.14 The Cross Section of Merger Arbitrage Returns 63

1.15 Model Target Option Comparative Statics 79

2.1 News Arrivals: Intraday and Seasonal Patterns 91

2.2 Example: SEC Charges GS with Securities Fraud 96

2.3 Example: IBM Announces a Share Buyback Program 98

2.4 Example: Verizon Wireless Acquires Alltel 100

2.5 Stock Market Activity Around Intraday News 103

2.6 Stock Returns Around the News: Log Kernel Density Estimate 109
2.7 Patterns in News Arrivals, Realized Variance, and Volume 110
2.8 Forecasting Returns for Individual Stocks 114
2.9 Forecasting Returns: Fama-MacBeth Regressions 118
2.10 News Adjusted Contrarian Returns 122
2.11 Stock Market Activity Around Intraday News from 2010 to 2013 139
2.12 Histograms of Relevance, Sentiment, and Novelty News Variables 140

3.1 Power as a Function of Pricing Error Size and Number of Assets 157
3.2 Conditional CAPM Pricing Errors 157
3.3 Monte Carlo Study: The Impact of Noise and Asynchronous Data 168
3.4 Estimating β: The Impact of Noise and Asynchronous Data 169

CHAPTER 1

MERGER OPTIONS AND RISK ARBITRAGE

This paper introduces a structural model of stock and option pricing to analyze market reactions to deal announcements. The model captures the jump in the target stock price at the announcement and the varied response of target options depending on whether the deal is financed by cash or stock. The analysis illustrates that option prices embed significant predictive content for forecasting deal outcomes that is not subsumed by stock prices. In a sample of 1,409 transactions from 1996 to 2012, the model drives out the explanatory power of target stock prices and obtains an R^2 of 33% in out-of-sample forecasts for cash deals. Additionally, I use the model to study the risks and returns of merger arbitrage strategies. In the model, merger arbitrage exhibits low volatility and large Sharpe ratios when deals are likely to succeed. The model also features expected returns that are hump shaped and volatility that is increasing in event time. These predictions are consistent with the data and support the conclusion that various features of merger arbitrage returns can be interpreted as compensation for deal risk.

1.1 Introduction

Deal failures are a salient feature of mergers and acquisitions. As many as one in four deals fail.[1] In merger arbitrage this constitutes a key risk.[2] For arbitrageurs, the loss in the event of failure is disproportionate to the gain in the event of success. More broadly, obtaining accurate estimates for the probability of deal success is relevant for a wide audience with vested interests in the outcomes of pending transactions including the target and the acquirer as well as their competitors, employees, customers, and bankers.

In this paper, I extract predictive content that forecasts deal outcomes from the reaction of target stock and option prices to deal announcements. In particular, when deals are announced, target stock prices jump and trade at a discount to the offer known as the arbitrage spread.[3] Simultaneously, target option prices respond to changes in the level and the distribution of the target stock price. To analyze how these reactions reflect information about deal risk and the likelihood of deal success, I build a structural asset pricing model. The model captures the jump in the target stock price at the announcement and the varied response of target options depending on whether the deal is financed by cash or stock. I estimate the model in individual deals. To do this, I infer the value of target dividends in the model from target stock prices and then use option pricing errors as additional restrictions

[1] Baker et al. (2012) reports that 25% of offers fail in a sample of 6,462 deals from 1984-2007. In my sample, the unconditional probability of offer failure is 19% in 4,364 deals from 1996 to 2012. Including rival bids does not change this estimate substantially. In particular, there is a 17% probability a target is acquired after an offer is announced in my sample.

[2] Merger arbitrage refers to an event driven trading strategy that attempts to profit through the systematic insurance of deal risk. For a detailed description of merger arbitrage investments see Mitchell and Pulvino (2001). Deals are primarily financed by cash and stock. To capture the arbitrage spread, defined as the difference between the target stock price and the deal payout, investors purchase the target stock in cash deals. In stock deals with fixed exchange ratios, merger arbitrageurs simultaneously purchase the target stock and short the acquirer in an amount equal to the exchange ratio. The exchange ratio in stock deals is defined as the number of acquirer shares to be paid per target share upon consummation. The practitioner use of the word arbitrage is a misnomer in this setting. When deals fail investors potentially suffer a large loss.

[3] In my sample, the median target stock jumps 26.5% after the announcement and trades at a 3.7% spread to the offer. This is consistent with Jetley and Ji (2010) who observe that merger arbitrage spreads have decreased since the early 2000s, in part due to increased assets under management dedicated to merger arbitrage strategies.

that identify model parameters. Empirically, I find the estimated risk neutral probability of success provides accurate forecasts for deal outcomes. To be specific, the estimated structural model exhibits an R^2 of 33% in out-of-sample forecasts for cash deal outcomes. This represents a substantial improvement over the share implied probability which obtains an R^2 of 8%.[4]

In addition to pricing target options and forecasting deal outcomes, I employ my model to study merger arbitrage. Of course, these topics are closely related. The probability of deal success is an important summary statistic for merger arbitrage investments. In the model, merger arbitrage exhibits low volatility when deals are likely to succeed because target stocks place little weight on volatile dividends. Meanwhile, binary deal risk results in expected returns that are hump shaped in the probability of success. It follows that the Sharpe ratio is largest in deals with a high probability of success. Empirically, I confirm the model's predictions. Portfolios that invest in deals with a high probability of success exhibit low volatility and low CAPM betas. Moreover, the monthly Sharpe ratio in a portfolio that overweights deals with a high share implied probability of success is nearly 50% larger than the Sharpe ratio in an equal weighted portfolio. Also, I study the dynamic properties of merger arbitrage in event time.[5] When deal success intensity is calibrated to the hump shaped pattern documented in the literature, the model features expected returns that are hump shaped and volatility that is increasing in event time. This matches the data and emphasizes the point that expected returns are not necessarily proportional to volatility in nonlinear investments.

[4]I define the share implied probability of success following Brown and Raymond (1986) and Samuelson and Rosenthal (1986). It is implied directly from the target stock price, the offer, and a fall back price under the assumption the target stock price equals the offer in the event of success and a fall back price in the event of failure, which is typically taken to be the value of the target stock a day or week before the deal was announced.

[5]Giglio and Shue (2014) demonstrate that deal success intensity is hump shaped in event time whereas the failure intensity is relatively constant. Shleifer and Vishny (2003) and Harford (2005) observe that mergers arrive in waves. In Appendix 1.B I present extensions of the baseline model to accommodate these features of the data.

This paper contributes to the literature in several respects. To my knowledge, it is the first structural model that presents a reconciled view of target stock and option pricing during the periods before and after deal announcements. Valuation is performed by computing the discounted expected value of dividends paid by the target firm under the risk neutral measure, similar to the approach in Merton (1974) and Leland (1994). In contrast to existing studies that have focused on option pricing in the post-announcement period, my approach can be used to interpret the reaction of option prices to deal announcements. This produces new insights. For example, the reaction of target options depends on a number of deal characteristics including the value of the offer, the probability of success, the type of financing, and the volatility of the acquirer stock. To be specific, consider a cash deal. In the model, the target stock price is a weighted average of dividends and cash. Deal announcements shift weight from dividends to cash, which tends to decrease volatility. At the same time, there is a risk after the announcement that a deal failure could result in a large loss for target shareholders. When cash deals are announced, the combination of these effects results in a decrease in the implied volatility of at-the-money options and the emergence of a volatility smile. By way of contrast, in a stock financed deal, the reaction of target options depends on the relationship between the volatility of target dividends and the volatility of the acquirer stock. When the acquirer stock is less volatile than target dividends, stock deal announcements coincide with a decrease in implied volatility, which is similar to cash deals. If the acquirer stock is more volatile than target dividends, stock deal announcements can precipitate an increase in implied volatility. The different empirical reactions in cash and stock deals suggests that market participants are attuned to nuanced deal characteristics and incorporate this knowledge in option valuation. In turn, option prices reflect market expectations of deal outcomes under the pricing measure.

Related papers that study merger options include Barraclough et al. (2013), Bester et al. (2013), Cao et al. (2005), and Subramanian (2004).[6] One of the key differences in my

[6]Barraclough et al. (2013) illustrate that options can enhance the measurement of merger wealth effects by providing additional equations to identify latent firm values in the event of deal success or failure. This allows

paper is the emphasis on studying the joint reaction of target stock and option prices to deal announcements. Additionally, a conceptual difference in my approach is that deal durations are random. This contrasts existing studies which assume mergers will fail or succeed on a fixed date known in advance by market participants. In practice, the amount of time until consummation is typically unknown and subject to significant uncertainty at the announcement date, which is inconsistent with the assumption of a fixed effective date.[7] Random deal durations circumvent this problem and introduce several theoretical contributions. For example, my approach facilitates the analysis of dynamic properties of target stock and option prices. I exploit this feature of the model to study the expected returns and volatility of merger arbitrage investments in event time. Additionally, for stock deals, my model predicts that the volatility of the target stock and its correlation with the acquirer depend on both the probability of deal success and the volatility of the acquirer stock. This differs from existing models and is consistent with the empirical reaction of target option prices to deal announcements. Another advantage is that closed form option prices can be derived for any maturity. Finally, my estimation strategy only uses stock and option data from the window surrounding deal announcements.[8] My analysis of cash deals is similar

the authors to decompose estimates of merger wealth effects into the potential gains from synergy versus the news revealed about the target and the acquirer when deals are announced. An important difference in my model is that I account for risk aversion and discounting. This allows me to capture the fact that merger arbitrage exhibits positive excess returns relative to standard risk models, a feature which is absent in Barraclough et al. (2013) where target stocks earn no risk premium. Subramanian (2004) studies merger options in stock deals while Bester et al. (2013) studies merger options in cash deals. They present evidence that option prices in the post-announcement period forecast deal outcomes in a sample of 26 stock deals and 422 cash deals respectively. Cao et al. (2005) investigates the increase in stock and option volumes prior to takeover announcements and argues that announcement returns are increasing in pre-announcement call volume imbalances.

[7]An exception of interest for the models with fixed effective dates can be found in tender offers. In a tender offer target shareholders can exchange their stock at a specific time for securities from the acquirer outlined in the term sheet filed with the SEC under the Williams Act.

[8]By way of contrast, the assumption of a fixed effective date in Subramanian (2004) implies that target and acquirer stocks are perfectly correlated with identical volatility while deals are pending, regardless of the probability of success. In Bester et al. (2013) closed form prices are available for maturities after the effective date. Before the effective date it is necessary to integrate over both the latent variables. Additionally, Bester et al. (2013) employs a Markov Chain Monte Carlo algorithm to estimate their model which relies on the full sample of stock and option prices between the announcement and the resolution of uncertainty. This complicates the interpretation of the forecasting ability of their model versus the target stock. To compare the explanatory power of my model, Bester et al. (2013) reports an R^2 of 12% using the first tenth of the data after deals are announced in 422 cash deals. This roughly corresponds to my post-announcement window

to Bester et al. (2013) in that both models deliver a volatility smile for post-announcement target options and a kink in option prices when the strike price equals the offer.

Additionally, my paper contributes to the existing literature that studies the risks and returns of merger arbitrage including Mitchell and Pulvino (2001), Baker and Savaşoglu (2002), and Giglio and Shue (2014), and, at a broader level, to studies that have examined the risk-return trade-off in nonlinear investments including examples such as Coval and Shumway (2001) who study expected option returns, Aït-Sahalia et al. (2009) who solve portfolio choice problems in the presence of jumps, and Jurek and Stafford (2015) who analyze hedge fund returns. With respect to merger arbitrage, Mitchell and Pulvino (2001) observe that merger arbitrage strategies exhibit larger CAPM betas in economic states for which the market suffers a large loss. In my model, merger arbitrage has a larger CAPM beta when deals are likely to fail because target stocks place more weight on dividends that are correlated with the market portfolio when the probability of success is low. Also, my model is similar to Baker and Savaşoglu (2002) in the respect that binary deal risk results in expected returns that are hump shaped in the probability of success. Finally, an extension of my baseline model incorporates the observation in Giglio and Shue (2014) that deal success intensity is hump shaped in event time. When deal risk is priced, the model predicts that expected returns are hump shaped and that volatility is increasing in event time, which is consistent with the data.

I now discuss the theoretical intuition for why option prices outperform the target stock price in forecasting deal outcomes and explain why I elect to study market reactions to merger announcements from the perspective of a structural model as opposed to a nonparametric or reduced form approach.[9] First, note that the share implied probability of success from

which uses up to 10 calendar days of data after the announcement. By focusing on the change in option prices before and after the announcement my model obtains an R^2 of 34% in a sample of 805 deals.

[9]The literature highlights the target stock price Brown and Raymond (1986); Samuelson and Rosenthal (1986), the attitude of the acquirer Schwert (2000); Walkling (1985), and the relation of the offer to the fifty-two week high of the target stock price Baker et al. (2012) as variables that forecast deal outcomes. In particular, hostile acquirers are less likely to initiate successful transactions as targets can employ a number of defense tactics to prevent outsiders from gaining control, including staggered boards, supermajority rules, fair price clauses, employee stock ownership plans, and differential voting rights. The relevance of acquirer

Brown and Raymond (1986) and Samuelson and Rosenthal (1986) is computed by making the unrealistic assumption that the distribution of the target stock price has only two outcomes, the offer in the event of success and a fall back price in the event of failure. In general, the target stock price is not a sufficient statistic for the risk neutral probability of success. By way of contrast, option prices can provide a model free estimate for the risk neutral probability of success. To understand this result, recall that Arrow Debreu state prices may be obtained from options following the insight of Breeden and Litzenberger (1978). Assuming that interest rates are constant, this is equivalent to knowing the risk neutral distribution of the target stock price. In deals with fixed payouts, such as cash deals or stock deals with a floating exchange ratio, target shareholders receive a fixed payment if the deal succeeds. This creates a discontinuity in the cumulative distribution function of the target stock price at the value of the offer. Option prices reveal the size of the discontinuity which equals the risk neutral probability a deal will succeed prior to a particular maturity. This information cannot be recovered from the target stock price.

The challenge for implementing a nonparametric approach is two-fold. First, obtaining the cumulative distribution function of the target stock requires differentiating the option pricing function with respect to the strike price. In practice, exchange-traded options are observed at fixed strikes. There are not typically a large number of option prices at strikes surrounding the acquirer's offer. Additionally, the nonparametric approach is not applicable in stock deals with fixed exchange ratios. The key difference in stock deals is that target shareholders receive a fixed amount of acquirer shares, known as the exchange ratio, but a random amount of cash, which depends on the value of the acquirer stock price upon consummation. Abstracting from funding costs, investors capture the arbitrage spread in

attitude in forecasting deal outcomes is limited by the fact that the frequency of hostile transactions has decreased since the 1980s. Andrade et al. (2001) report that only 14% of deals in the 1980s and 4% of deals in the 1990s involved a hostile offer. In my sample only 1.5% of offers are flagged as hostile by SDC. Baker et al. (2012) argues that peak prices serve as important reference points in merger negotiations. They find offers larger than the fifty-two week high of the target stock are more likely to succeed, which highlights a channel for financial prices to influence capital reallocation. In comparison, Edmans et al. (2012) illustrate that decreases in target stock prices increase the likelihood of takeover.

stock deals by simultaneously purchasing the target stock and selling the acquirer short in an amount equal to the exchange ratio. In this case, the distribution of the merger arbitrage investment depends on the joint distribution of the target stock and acquirer stock. One can obtain model free univariate distributions of the target stock price and acquirer stock price from option prices, but not the joint distribution. This contrasts cash deals for which options are traded directly on the target stock which is equivalent to the merger arbitrage investment. In particular, the discussion suggests that option prices will be more informative about deal outcomes in cash deals as opposed to stock deals. Indeed, I find this to be true empirically. To circumvent the issues noted above, I build a structural model to analyze market reactions to deal announcements.

The remainder of the paper proceeds as follows. Section 1.2 presents the model. Section 1.3 describes the data and model estimation. Section 1.4 discusses forecasting deal outcomes, and Section 1.5 presents the analysis of merger arbitrage returns. Section 1.6 concludes.

1.2 Model of Merger Stock and Option Pricing

Consider a continuous time economy in which a firm may become the target of a merger or acquisition. In the absence of arbitrage, I assume there exists an equivalent martingale measure under which dividends follow

$$\frac{dD_t}{D_t} = g(s_t)dt + \sigma(s_t)dW_t \tag{1.1}$$

where W_t is a standard Brownian motion.[10] The risk free rate is assumed to be a constant $r \geq 0$ and the firm is assumed to reside in one of three states $s_t \in \{\mathcal{P}, \mathcal{T}, \mathcal{A}\}$ referred to as the pre-target, target, and acquired states. The state impacts the dividend process as follows. If $s_t \in \{\mathcal{P}, \mathcal{T}\}$ the firm has not been acquired and dividends evolve according to

[10]Formally I fix a probability space $(\Omega, \mathcal{F}, \mathbb{Q})$ where $(\mathcal{F}_t) = \{\mathcal{F}_t : t \geq 0\}$ is a filtration assumed to satisfy the usual conditions. For technical details on conditions that imply the existence of an equivalent martingale measure, for example the assumption of no approximate arbitrage, see Chapter 6 in Duffie (2010).

a geometric Brownian motion with growth rate g and idiosyncratic volatility $\sigma \geq 0$. The acquired state is absorbing. When $s_t \in \mathcal{A}$ the growth rate and volatility are set to zero and the target firm stops paying dividends.[11]

When deals succeed target shareholders receive a liquidating payment from the acquirer in the form of cash V and acquirer shares ϕ. In cash deals the payout is independent from the acquirer stock price. For stock and hybrid deals the acquirer stock price is modeled as

$$\frac{dP_{A,t}}{P_{A,t}} = (r - q_A)\, dt + \sigma_A dW_{A,t} \tag{1.2}$$

where $q_A > 0$ is the dividend yield and $W_{A,t}$ is a standard Brownian whose instantaneous correlation with W_t equals ρ. Implicitly, I assume the value of the acquirer is unrelated to the deal outcome. This approximation is based on the observation that acquirers tend to be much larger than their targets. As such, gains or losses from an individual deal may have a minor impact on the value of the acquirer. Empirically, the assumption is supported by the observation that targets experience large and positive announcement returns whereas acquirer announcement returns are closer to zero (Andrade et al. (2001)), which is consistent with the free rider problem outlined in Grossman and Hart (1980).

Transitions across states are assumed to follow a continuous time homogeneous Markov chain whose generator is defined as Λ. Informally the risk neutral probability of transitioning from state i to state j over a short interval Δ is approximately $\lambda_{ij}\Delta$.

$$\Lambda = \begin{bmatrix} -\lambda & \lambda & 0 \\ \lambda_F & -\lambda_F - \lambda_S & \lambda_S \\ 0 & 0 & 0 \end{bmatrix} \tag{1.3}$$

[11] This can be modeled formally as $dD_t/D_{t-} = g(s_t)dt + \sigma(s_t)dW_t + J(s_{t-}, s_t)dM_t$ where M_t is a counting process equal to one when the firm transitions $dM_t = \mathbf{1}[s_{t-} \neq s_t]$. The assumption that dividend payments are set to zero when the firm is acquired is equivalent to $J(\mathcal{T}, \mathcal{A}) = -1$. The model can also be extended to allow $J(\mathcal{P}, \mathcal{T}) = z$ and $J(\mathcal{T}, \mathcal{P}) = -z/(1+z)$ so that deal announcements and failures have a symmetric and opposite impact on target dividends. Similarly, the impact of the deal on the acquirer can be modeled as $dP_{A,t}/P_{A,t-} = (r - q_A)\, dt + \sigma_A dW_{A,t} + J_A(s_{t-}, s_t)dM_t$ where $J_A(\mathcal{T}, \mathcal{A}) = z_S$, $J_A(\mathcal{P}, \mathcal{T}) = z_A$, and $J_A(\mathcal{T}, \mathcal{P}) = -z_A/(1+z_A)$. This allows for the possibility that the deal might create or destroy value for the acquiring firm.

Naturally one may interpret deal announcements as transitions from the pre-target state to the target state which occur with intensity λ. Once a firm is in the target state the pending deal will fail with intensity λ_F and succeed with intensity λ_S.

Several extensions of the baseline specification are also discussed. Note that the number of states has been selected to keep the model as parsimonious as possible. In particular, the pre-target and target states allow for asset pricing in the periods before and after deals are announced. The acquired state allows deals to succeed. An implication of the current setup is that firms eventually become acquired with probability one. To relax this feature one can easily add an absorbing non-target state $\mathcal{N}$ in which shareholders receive target dividends and no mergers occur. Rival bids may also be modeled by enlarging the number of states. Additionally, the baseline model abstracts from the observation that mergers tend to arrive in waves (Shleifer and Vishny (2003), Harford (2005)) and the fact that success intensity is hump shaped in event time (Giglio and Shue (2014)). To accommodate these features of the data I present extensions of the baseline model in Appendix 1.A that allow $\lambda_S(t)$ to be a deterministic function of time and $\lambda(t)$ to be random and correlated with target dividends.

1.2.1 Stock Prices

The derivation of the main results is included in Appendix 1.A. Stock prices are obtained by solving a system of risk neutral pricing partial differential equations that apply in the pre-target and target states. The proposition below documents the stock price. The corollary provides the stock price when the pre-announcement intensity decreases to zero, which may be interpreted as the case with no pre-announcement anticipation of the deal.

Proposition 1 ***Target Stock Prices:*** *Consider a firm whose dividend process follows equation 1.1 and transitions across states according to a Markov chain with generator Λ given in equation 1.3. Upon consummation of a successful transaction target shareholders receive V in cash and ϕ shares of acquirer stock which evolves according to equation 1.2. If*

$r - g > 0$ the share price is given below. The coefficients are provided in the Appendix.

$$P\left(D_t, P_{A,t}, s_t\right) = \begin{cases} A_0 D_t + A_1 V + A_2 P_{A,t} & s_t \in \mathcal{P} \\ B_0 D_t + B_1 V + B_2 P_{A,t} & s_t \in \mathcal{T} \end{cases} \tag{1.4}$$

Corollary 1 *: As the deal announcement intensity decreases to zero the price in the pre-target state converges to the Gordon growth formula and the price in the target state converges to the discounted expected payoff under the risk neutral measure when deal failures are absorbing*

$$\lim_{\lambda \to 0} P_t = \begin{cases} \frac{D_t}{r-g} & s_t \in \mathcal{P} \\ \frac{\lambda_F + r - g}{\lambda_S + \lambda_F + r - g} \frac{D_t}{r-g} + \frac{\lambda_S}{\lambda_S + \lambda_F + r} V + \frac{\lambda_S}{\lambda_S + \lambda_F + q_A} \phi P_{A,t} & s_t \in \mathcal{T} \end{cases} \tag{1.5}$$

Define the risk neutral probability of deal success as $p_Q = \frac{\lambda_S}{\lambda_S + \lambda_F}$ and risk neutral expected deal duration as $\tau_Q = \frac{1}{\lambda_S + \lambda_F}$. Then in the target state the price may be rewritten as

$$P_t = \frac{1 - p_Q + (r-g)\,\tau_Q}{1 + (r-g)\,\tau_Q} \frac{D_t}{r-g} + \frac{p_Q}{1 + r\tau_Q} V + \frac{p_Q}{1 + q_A \tau_Q} \phi P_{A,t} \tag{1.6}$$

Target stock prices follow a modified Gordon Growth formula. In addition to target dividends, the extra terms correspond to the value of cash and acquirer stock that will be paid to shareholders upon the consummation of a successful deal. The coefficients vary depending on whether the firm is in the pre-target or target state. Further intuition may be gleaned from the corollary. As the announcement intensity decreases to zero the stock price in the pre-target state converges to the net present value of dividends under the risk neutral measure. In other words, the Gordon growth formula is approximately correct in the pre-target state when deal announcements are infrequent. Similarly, the price in the target state is approximately the risk neutral probability of deal success multiplied by the deal payout plus the risk neutral probability of deal failure multiplied by the net present

value of target dividends. The corollary properly accounts for discounting and the dividends received between the announcement and the resolution of deal uncertainty. The proposition generalizes to the case when deal announcement intensity is positive.

The stock price is illustrated in Figure 1.1. Salient features of the model are highlighted in the plot. The three lines correspond to the price as a function of the fundamental D_t in the pre-target, target, and acquired states. The offer $V + \phi P_{A,t}$ is held constant. When the firm transitions between states the price exhibits a discontinuous jump between the different lines on the plot. For example, when the firm transitions from the pre-target state to the target state, interpreted as a deal announcement, the stock price will exhibit a large increase if the offer is greater than the pre-target price. In this manner the model captures the empirical observation that target stock announcement returns are large and positive. Another feature of the plot is that the slope with respect to dividends is smaller in the target state versus the pre-target state. In particular, when deals succeed shareholders receive a liquidating payment from the acquirer and receive no further dividends from the target firm. The implication is that stock prices are less sensitive to changes in the level of the fundamental while the firm is in the target state. This is consistent with the corollary above which shows that increasing the risk neutral probability of deal success decreases the slope of the target stock price with respect to dividends. The change in the slope when deals are announced has implications for option pricing that will be discussed shortly. As a preview, in cash deals, the price is a weighted average of the fundamental and cash. Since cash has zero volatility, there is less diffusive risk in the target state as a result of the lower slope with respect to target dividends. The implication for option pricing is that deal announcements precipitate a decrease in the implied volatility of at-the-money options. Moreover, the fact that the slope is decreasing in the risk neutral probability of success implies that deals with higher probabilities of success p_Q will exhibit a larger decrease in the implied volatility of at-the-money options. For stock deals the price is a weighted average of the fundamental and the acquirer stock. In this case, diffusive risk may decrease or increase when deals are announced which depends on how the

volatility of the acquirer stock relates to the volatility of the target fundamental dividend process.

Figure 1.2 presents a different perspective on how deal announcements impact the target stock price by illustrating the risk neutral distribution of the target stock over a four month horizon. When a deal is announced, the mass in the probability density function shifts to the right corresponding to a discontinuous increase in the stock price. Beyond this level effect, the plot highlights the additional impact a deal announcement has on the distribution. The change in the distribution between the pre-target and target states is particularly stark for cash deals. After the announcement the distribution of the stock price for cash deals more closely resembles a mixture. The probability density function has two peaks which correspond to the mode of the distribution in the pre-target and target states. Cash deals also feature a discontinuity in the cumulative distribution function equal to the risk neutral probability a deal will succeed prior to the horizon being reached. This occurs in cash deals as shareholders receive a fixed amount if the deal succeeds. For stock deals the change in the distribution is less severe when the deal is announced. In other words, the mixture of lognormal distributions appears approximately lognormal. Of course, it is important to note that stock deals provides a generalization of cash deals. As the acquirer volatility decreases in stock deals, the distribution of the target stock price will more closely resemble the distribution in a cash deal.

1.2.2 Option Prices

I price target options by computing the discounted expected payoff under the risk neutral measure. The payoff is broken into three terms corresponding to the value of the option conditioned on the firm being in the pre-target, target, and acquired states which are multiplied by the risk neutral probability of the firm being in each of these states at maturity. To that end, it is necessary to compute the risk neutral transition probabilities between states. One approach to obtain the transition probabilities is to compute the matrix

exponential of the generator which can be performed numerically. Alternatively, I derive closed form solutions for the transition probabilities in the proposition below. Closed form solutions are helpful in the empirical application for which the model is estimated using a large number of option prices.

Proposition 2 ***Target Transition Probabilities:*** *The risk neutral transition probabilities between deal states assuming s_t follows a continuous time Markov chain with generator Λ are given by*

$$Q(\tau) = \begin{bmatrix} q_{11}(\tau) & q_{12}(\tau) & q_{13}(\tau) \\ q_{21}(\tau) & q_{22}(\tau) & q_{23}(\tau) \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{aligned} q_{11}(\tau) &= \tfrac{\lambda}{v_2 - v_1}\left[\left(\tfrac{v_2}{\lambda}+1\right)e^{v_1\tau} - \left(\tfrac{v_1}{\lambda}+1\right)e^{v_2\tau}\right] \\ q_{12}(\tau) &= \tfrac{\lambda}{v_2 - v_1}\left[e^{v_2\tau} - e^{v_1\tau}\right] \\ q_{21}(\tau) &= \tfrac{\lambda}{v_2 - v_1}\left[\left(\tfrac{v_2}{\lambda}+1\right)\left(\tfrac{v_1}{\lambda}+1\right)\left(e^{v_1\tau} - e^{v_2\tau}\right)\right] \\ q_{22}(\tau) &= \tfrac{\lambda}{v_2 - v_1}\left[\left(\tfrac{v_2}{\lambda}+1\right)e^{v_2\tau} - \left(\tfrac{v_1}{\lambda}+1\right)e^{v_1\tau}\right] \end{aligned} \tag{1.7}$$

where $v_1 < v_2 < 0$ are equal to $\frac{-[\lambda+\lambda_F+\lambda_S] \pm \sqrt{(\lambda+\lambda_F+\lambda_S)^2 - 4\lambda\lambda_S}}{2}$ and $q_{13}(\tau) = 1 - q_{11}(\tau) - q_{12}(\tau)$, $q_{23}(\tau) = 1 - q_{21}(\tau) - q_{22}(\tau)$.

The proposition below presents European put prices and the corollary provides simplified formulas for cash deals. In cash deals the option price conditioned on the state of the firm at maturity may be derived using the Black and Scholes (1973) formula by adjusting the strike price for the value of the offer. The key difference in stock deals is that the target stock price equals the sum of two lognormal random variables. For a sufficiently high realization of the acquirer stock price a put option will be out of the money in a stock deal. This value corresponds to the upper bound of integration in the proposition. Thus, one can obtain option prices in a stock deal by integrating the option pricing formula in cash deals over realizations of the acquirer stock price below the upper bound of integration. Alternatively,

one can compute option prices numerically using the cumulative distribution function of the target stock price which is presented in Appendix 1.A. Call option prices can also be derived in closed form or obtained via put-call parity.

Proposition 3 ***Target Option Prices:*** *The price of a European put option with strike K and maturity τ equals*

$$F\left(D_t, P_{A,t}, s_t\right) = \begin{cases} q_{11}(\tau)F_{\mathcal{P}} + q_{12}(\tau)F_{\mathcal{T}} + q_{13}(\tau)F_{\mathcal{A}} & s \in \mathcal{P} \\ q_{21}(\tau)F_{\mathcal{P}} + q_{22}(\tau)F_{\mathcal{T}} + q_{23}(\tau)F_{\mathcal{A}} & s \in \mathcal{T} \end{cases} \tag{1.8}$$

Let $P_{BS}(S, K, \tau, r, q, \sigma)$ denote the put price in the Black-Scholes model. Then we have the following result,

$$\begin{aligned} F_{\mathcal{P}} &= \int_{-\infty}^{x_A^*} P_{BS}\left(A_0 D, K_A(x), \tau, r, q(x), \sigma\sqrt{1-\rho^2}\right) f(x)dx \\ F_{\mathcal{T}} &= \int_{-\infty}^{x_B^*} P_{BS}\left(B_0 D, K_B(x), \tau, r, q(x), \sigma\sqrt{1-\rho^2}\right) f(x)dx \\ F_{\mathcal{A}} &= P_{BS}\left(\phi P_{A,0}, K - V, \tau, r, q_A, \sigma_A\right) \end{aligned} \tag{1.9}$$

where f is the probability density function of a standard normal random variable. The definitions for $q(x)$, x_i^*, $K_i(x)$ for $i \in \{A, B\}$ are provided in the proof of the proposition.

Corollary 2 *: For cash deals the formula for put prices in the pre-target, target, and acquired states simplify to*

$$\begin{aligned} F_{\mathcal{P}} &= P_{BS}\left(A_0 D_0, K - A_1 V, \tau, r, r - g, \sigma\right) \\ F_{\mathcal{T}} &= P_{BS}\left(B_0 D_0, K - B_1 V, \tau, r, r - g, \sigma\right) \\ F_{\mathcal{A}} &= e^{-r\tau}\left(K - V\right)^+ \end{aligned} \tag{1.10}$$

Figure 1.3 presents the volatility smile and the term structure of implied volatility for cash deals. When the deal is announced implied volatility decreases. This stems from the observation that the stock price has a smaller slope with respect to dividends in the target state and puts more weight on the cash payout which has zero volatility. Simultaneously, an

implied volatility smile emerges in which out of the money options have significantly larger values of implied volatility than at-the-money options. Although diffusive risk is low in the target state, the price may still exhibit a large move if the deal fails or succeeds. Figure 1.4 presents option prices in a stock deal. In this case the reaction of option prices when a deal is announced depends on whether the acquirer stock is more or less volatile than target dividends. When the acquirer has low volatility the reaction is similar to a cash deal. In particular, as acquirer volatility decreases the acquirer stock more closely resembles cash so that option pricing in cash and stock deals converges. The important difference occurs when the acquirer volatility is high. In this case, the level of implied volatility for the target stock is higher in both the pre-target and target state relative to the target fundamental volatility. Moreover, implied volatility increases instead of decreasing when deals with high acquirer volatility are announced. In the event of success, target shareholders will be left with the volatile acquirer stock.

1.2.3 Risk Premia

Stock and option pricing is performed under the risk neutral or equivalent martingale measure. To study risk premia including the expected returns, volatility, and the Sharpe ratio of merger arbitrage investments, I introduce market prices of risk that link the risk neutral and physical dynamics.

To price the risk embedded in target dividends and the acquirer stock I assume the capital asset pricing model (CAPM) of Sharpe (1964) and Lintner (1965b) applies. Embellishing the setup from the previous section, suppose the innovations to target dividends and the acquirer stock follow the factor structure

$$\begin{aligned} dW_t &= \rho_m dW_{m,t} + \sqrt{1-\rho_m^2}\, dW_{i,t} \\ dW_{A,t} &= \rho_m dW_{m,t} + \sqrt{1-\rho_m^2}\, dW_{Ai,t} \end{aligned} \tag{1.11}$$

where $W_{m,t}$ is interpreted as market risk that is orthogonal to the idiosyncratic errors $W_{i,t}$ and $W_{Ai,t}$.[12] I denote the relationship between market risk under the physical and risk neutral measures as

$$W_{m,t} = W^{\mathbb{P}}_{m,t} + \int_0^t \eta_s ds \tag{1.12}$$

where the process η is the price of market risk. For simplicity I consider the case where η is a constant equal to $\eta = \frac{\mu_m - r}{\sigma_m}$.[13] Then under the physical measure it follows that

$$\begin{aligned} \frac{dD_t}{D_t} &= (g + \beta_D (\mu_m - r))\, dt + \sigma \rho_m dW^{\mathbb{P}}_{m,t} + \sigma \sqrt{1 - \rho_m^2} dW_{i,t} \\ \frac{dP_{A,t}}{P_{A,t}} &= (r + \beta_A (\mu_m - r) - q_A)\, dt + \sigma_A \rho_m dW^{\mathbb{P}}_{m,t} + \sigma_A \sqrt{1 - \rho_m^2} dW_{Ai,t} \end{aligned} \tag{1.13}$$

where $\beta_D = \frac{\sigma \rho_m}{\sigma_m}$ and $\beta_A = \frac{\sigma_A \rho_m}{\sigma_m}$ are the CAPM betas of target dividends and the acquirer stock respectively.

The market prices of deal risk are defined as (κ_S, κ_F). These parameters adjust the failure and success intensity in the target state. Under the physical measure the intensities are presumed to equal[14]

$$\begin{aligned} \lambda^{\mathbb{P}}_S &= e^{-\kappa_S} \lambda_S \\ \lambda^{\mathbb{P}}_F &= e^{-\kappa_F} \lambda_F. \end{aligned} \tag{1.14}$$

The relationship between the objective probability of deal success p and risk neutral probability of deal success p_Q is given by a first order Taylor series approximation as

$$p_Q = \frac{p}{p + (1-p)\, e^{\kappa_F - \kappa_S}} \approx p - p\,(1-p)\,(\kappa_F - \kappa_S) \tag{1.15}$$

[12] In the previous section I assumed $E^{\mathbb{Q}}_t [dW_t dW_{A,t}] = \rho dt$. To be consistent with previous notation $\rho_m = \sqrt{\rho}$.

[13] That $W^{\mathbb{P}}_{m,t}$ is a Brownian motion under $\mathbb{P}$ follows from an application of Girsanov's theorem assuming $\eta \in \mathcal{L}^2$ defined as the set of adapted processes satisfying $\int_0^T \eta_s^2 ds < \infty$ a.s. This is satisfied when η is constant. For a formal statement of Girsanov's theorem see Appendix D in Duffie (2010) or Chapter 3 in Karatzas (1991).

[14] For a general version of Girsanov's Theorem with jumps see Jacod and Shiryaev (1987) and Protter (2004).

Thus, when deal risk is priced $\kappa_F - \kappa_S > 0$ it follows that $p_Q < p$. In this case either the failure intensity is larger $\kappa_F > 0$ or the success intensity is lower $\kappa_S < 0$ under the risk neutral measure relative to the objective measure. My model is agnostic about the source of deal risk. In particular, I am interested in testing model predictions for how expected returns, volatility, and the CAPM beta of merger arbitrage returns behave as a function of the deal success probability when deal risk is priced. To interpret why deal risk might be priced I present two perspective below. First, if deal outcomes covary with economic state variables merger arbitrage will receive a risk premium as in the ICAPM of Merton (1973). In support of this hypothesis, I find that deal failures are negatively correlated with the market portfolio and Mitchell and Pulvino (2001) provide evidence that merger arbitrage returns are more correlated with market returns when the market suffers a large negative realization. The implication is that deal failures may embed systematic risk. To model this in my setup one might imagine the failure intensity contains a systematic and an idiosyncratic component where only the systematic component garners a risk premium.[15] In this specification none of the qualitative results change from the simple model outlined above. An alternative view articulated in Baker and Savaşoglu (2002) is that deal risk is idiosyncratic and commands a risk premium because of limits to arbitrage. If capital constrained or risk averse arbitrageurs absorb supply shocks when deals are announced there can be a wedge between the risk neutral probability of success and the objective probability of success. In their model, the factors that determine the market price of deal risk include the size of the supply shocks relative to merger arbitrageur wealth, merger arbitrageur risk aversion, and the size of the offer premium.[16]

In my empirical analysis, I will focus on merger arbitrage returns in transactions with fixed hedge ratios. The investment strategy in this case is straightforward. For cash deals an

[15]Model deal failures as $dN_F = dN_{F,M} + dN_{F,i}$ where $dN_{F,M}$ is systematic and $dN_{F,i}$ is idiosyncratic. The intensity is $\lambda_{F,M} + \lambda_{F,i}$ under $\mathbb{Q}$ and $e^{-\kappa_{F,M}}\lambda_{F,M} + \lambda_{F,i}$ under $\mathbb{P}$. The wedge $\left(e^{-\kappa_{F,M}} - 1\right)\lambda_{F,M}$ contributes to the risk premium.

[16]The relationship between p_Q and p is given by $p_Q = p - p(1-p)\frac{XP_T}{W_A}(\gamma_A W_A)(V + \phi P_A - \frac{D}{r-g})/P_T$ in Baker and Savaşoglu (2002) where the ratio of the supply shock to merger arbitrage wealth is XP_T/W_A, merger arbitrageur risk aversion is $\gamma_A W_A$, and the size of the offer premium is $(V + \phi P_A - \frac{D}{r-g})/P_T$.

investor purchases the target stock after the deal is announced. In stock and hybrid deals an investor simultaneously purchases the target stock and sells the acquirer short in an amount equal to the exchange ratio ϕ. This motivates the following definition of instantaneous excess returns for merger arbitrage investments

$$dR^e_{MA,t} \equiv \frac{dP_t + D_t dt - \phi \left[dP_{A,t} + D_{A,t} dt - r P_{A,t} dt \right]}{P_t} - r dt \tag{1.16}$$

The definition highlights the various risks in a merger arbitrage investment. In particular, the investment increases in value if the target stock price appreciates or the target firm pays dividends. In stock and hybrid deals, the investment decreases in value if the acquirer stock price appreciates or the acquirer pays dividends. Funding occurs at the risk free rate.[17] The proposition below derives the expected return, CAPM beta, and variance for instantaneous merger arbitrage returns.

Proposition 4 ***Merger Arbitrage Risk Premia:*** *The instantaneous risk premium* $\mu_{MA,t} dt \equiv E^{\mathbb{P}}_t \left[dR^e_{MA,t} \right]$ *may be decomposed into terms that compensate investors for market risk and deal risk*

$$\mu_{MA,t} = \underbrace{\beta_{MA,t} \left(\mu_m - r \right)}_{\textit{Fundamental Risk Premium}} + \tag{1.17}$$

$$\underbrace{\left(e^{-\kappa_S} - 1 \right) \lambda_S \left[\frac{V + \phi P_{A,t}}{P_t(\mathcal{T})} - 1 \right] + \left(e^{-\kappa_F} - 1 \right) \lambda_F \left[\frac{P_t(\mathcal{P})}{P_t(\mathcal{T})} - 1 \right]}_{\textit{Deal Risk Premium}}$$

[17] The assumption investments are funded at the risk free rate is unrealistic in practice. For example, the activity of merger arbitrageurs can make it difficult for prime brokers to locate shares to short. When merger stocks are "on special" short sale proceeds can earn less than the risk free rate. There is also a risk the lender will recall the short prior to the resolution of deal uncertainty. Additionally, funding costs may exceed the risk free rate and be correlated with an investor's past performance. I abstract from these concerns and focus on how the nonlinear payoff profile in the target stock impacts the risk-return characteristics of merger arbitrage investments.

The CAPM beta for merger arbitrage investments is

$$\beta_{MA,t} = \frac{\beta_D B_0 D_t + \beta_A \left(B_2 - \phi\right) P_{A,t}}{P_t\left(\mathcal{T}\right)} \tag{1.18}$$

and the instantaneous variance $\sigma^2_{MA,t}dt \equiv var^{\mathbb{P}}_t\left(dR^e_{MA,t}\right)$ *equals*

$$\sigma^2_{MA,t} = \underbrace{\lambda^P_S\left(\frac{V+\theta P_{A,t}}{P_t(\mathcal{T})}-1\right)^2 + \lambda^P_F\left(\frac{P_t(\mathcal{P})}{P_t(\mathcal{T})}-1\right)^2}_{\textit{Deal Risk}} + \tag{1.19}$$

$$\underbrace{\frac{\left(B_0\sigma D_t\right)^2 + \left(\left(B_2-\phi\right)P_{A,t}\sigma_A\right)^2 + 2B_0\sigma D_t\left(B_2-\phi\right)\sigma_A P_{A,t}\rho}{P_t\left(\mathcal{T}\right)^2}}_{\textit{Fundamental Risk}}$$

Expected returns have two components. The first compensates investors for market risk. It equals the CAPM beta of a merger arbitrage investment multiplied by the market risk premium. The second component compensates investors for deal risk. Figure 1.5 displays these moments as a function of the probability of deal success. One prediction of the model is that volatility and CAPM beta are decreasing in the probability of success. This relates to the previous observation that the slope of the target stock price with respect to dividends decreases as the probability of success increases. Target stocks in deals with a high probability of success place low weight on volatile target dividends, which results in a small CAPM beta and low amounts of diffusive risk. The plot also highlights the contribution of deal success risk and deal failure risk to expected returns. Deal risk is hump shaped in the probability of success and is maximized for deals with an intermediate probability of success. This is intuitive. Deal risk is binary and similar to the hump shaped variance of a Bernoulli random variable which equals $p(1-p)$. Meanwhile, the Sharpe ratio is concave in the probability of success and is maximized for deals with a high probability of success that is less than one.

1.3 Data Description and Model Estimation

1.3.1 Data Description

My empirical analysis relies on three primary sources of data. Detailed information pertaining to merger activity is obtained from Thomson One Banker using the database formerly known as SDC. Deals are then matched to CRSP and OptionMetrics for stock and option data. The sample includes cash deals in addition to stock and hybrid deals with fixed exchange ratios. The sample period is 1996-2012 during which option data is available.

Deals are selected by imposing the following filters. Targets are required to be non-bankrupt public companies incorporated in the United States. Transactions that involve collars, spinoffs, asset swaps, restructurings, or recapitalizations are excluded.[18] The acquirer must seek a majority share in a deal classified as a merger, acquisition, acquisition of majority interest, acquisition of remaining interest, or acquisition of partial interest. To focus on original deal announcements I require that the original announcement date be within five days of the announcement date as reported by SDC. The original announcement date refers to the first date a target company is publicly disclosed as being a possible takeover candidate. My intent in imposing this filter is to exclude rival bids and revisions to previous offers which are outside the model developed in the previous section. Summary statistics for the deals matched to CRSP and OptionMetrics are reported in Table 1.1. The sample contains 4,364

[18]Following the literature I exclude deals with complicated or unusual features. For example, stock deals with collars embed implicit options in which the deal payoff can depend on the average acquirer stock price during a pricing period 10 to 20 days before the effective date of the merger. When the effective date is unknown ex-ante and the payoff has an Asian feature the arbitrage spread cannot be captured by statically trading European options. Instead, arbitrageurs must employ a dynamic trading strategy which involves modeling the embedded options. Collars have been studied in the literature by Officer (2004, 2006) who concludes that collars are useful in minimizing renegotiation costs, being used in nearly 20% of stock swaps between 1991-1999. Of course, excluding complicated transactions is a concern to the extent they are systematically different from the transactions studied here. For example, actively managed funds will be particularly interested in complicated deals if they represent more attractive investment opportunities. Mitchell and Pulvino (2001) provide evidence this is not the case. They report that simple deals with fixed hedge ratios are similar to the subset of complicated transactions along a number of dimensions such as the deal premium, announcement returns, failure probability, deal size, and the return for investing in the target stock from announcement to completion.

deals including 3,164 cash deals, 1,077 stock deals, and 123 hybrid deals. On average hybrid transactions are 61% financed by stock as opposed to cash. From the table it is evident that cash deals differ from stock and hybrid deals along several dimensions. Cash deals tend to be smaller, are less likely to succeed, and are shorter in duration. Another feature of the data is that acquirers are much larger than targets. This is especially true for cash deals. The subsample with option data includes 1,409 deals consisting of 921 cash deals, 417 stock deals, and 71 hybrid deals. Compared to the full sample these deals are larger and have smaller deal premiums.

The implied probability of deal success is estimated by fitting the model from the previous section to listed target stock options. The option prices used in the analysis are the mid closing price from the best bid and best offer compiled at the close by OptionMetrics.[19] To be included in the estimation I require an option price have a positive bid and offer with the bid less than or equal to the offer, positive Black-Scholes implied volatility, Black-Scholes delta less than seventy percent in absolute value, maturity between 8 and 365 calendar days, and moneyness, defined as the ratio of the strike price to the stock price, between .3 and 2. Prices are removed that violate no arbitrage bounds. I also exclude prices that represent large violations of put-call parity.[20] For each deal the model is fit to option prices in pre-announcement and post-announcement windows. The pre-announcement window is defined as the period between 25 and 5 calendar days prior to the deal announcement. Similarly, the post-announcement window includes the 1 to 10 calendar days after the announcement.[21]

[19]The model of the previous section abstracts from two features of the option data. Single stock options are American and are subject to contract adjustments when deals succeed. For cash deals the adjustment involves the acceleration of payment for in-the-money options in accordance with OCC Rule 807. I find the value of early exercise and payment acceleration is small and significantly less than typical bid-ask spreads for the short dated out of the money options used in my sample. With this in mind I estimate the model using the European option pricing formulas derived in the previous section and filter the data by removing large violations of European put-call parity.

[20]For example, option prices are excluded if $C_t^O - P_t^B + e^{-r_t(T-t)}K \cdot (1+t) \leq e^{-r_t(T-t)}F_{t,T}(1-t)$ where $F_{t,T} = S_t e^{(r_t-\delta)(T-t)}$ with $t = .025$ and $\delta = .02$. O and B denote the offer and bid. An analogous formula applies in the other direction. The transaction cost for trading the bond and forward is intended to be large. I find the no arbitrage and put call parity filters remove relatively few options. For example, in the subset of cash deals only .25% of observed prices are excluded from analysis.

[21]The results are robust to changing the definition of the pre-announcement and post-announcement windows. For example, in practice one might be interested to use a shorter post-announcement window. I

To estimate the model I require at least ten option prices be observed in both the pre-announcement and post-announcement windows.

Summary statistics for the options data and average model option pricing errors are reported in Table 1.2. From the table it is clear the model outlined in Section 1.2 fits option prices better than the Black and Scholes (1973) model, which is reassuring given the model was introduced to study the reaction of option prices to deal announcements. A more challenging competitor is an ad hoc model that allows the Black-Scholes implied volatility to be different in the pre-announcement and post-announcement periods. For cash deals the model outperforms this ad hoc competitor. In particular, the model does well relative to Black-Scholes in its ability to minimize percentage pricing errors for the post-announcement period. This suggests the model captures the implied volatility smile in the data. In contrast, the Black-Scholes model has a flat volatility surface leading to large percentage pricing errors for out of the money options which have low prices. The summary statistics also indicate that the average option prices used in the estimation are short dated and out of the money. In particular, the average maturity is four months and the average Black-Scholes delta is less than fifty percent. Additionally, the summary statistics reveal that the volume, open interest, and the percentage of put options being traded all increase after deals are announced. This is similar to the reaction in stock markets which exhibit a significant increase in the traded volume of target and acquirer stock after deals are announced.

1.3.2 Model Estimation

To estimate the model, I assume option prices are observed with measurement errors. For example, errors might stem from the bid-ask spread or asynchronous trading of particular

find a shorter post-announcement window has little impact on the substantive conclusions except that some deals with less frequently traded options are removed from the sample given my requirement that at least ten option prices be observed in both windows.

option contracts. The measurement equation is then

$$Y_{t,j} = F\left(X_{t,j}, \theta\right) + \epsilon_{t,j} \tag{1.20}$$

where $Y_{t,j}$ is the observed price of option j on date t and F is the option pricing function from Proposition 3. Each day in the pre-announcement and post-announcement window I infer the level of the fundamental D_t from the closing target stock price. The dependent variable $X_{t,j}$ collects the fundamental D_t, acquirer stock price $P_{A,t}$, state of the firm s_t, and the option characteristics including the strike, maturity, and whether the option is a put or a call. I estimate the model via non-linear least squares by minimizing the sum of the average squared pricing errors in the pre-announcement and post-announcement periods. The objective function to be minimized is,

$$Q_N(\theta) = \frac{1}{N_{pre}} \sum_{t,j \in O_{pre}} \left(Y_{t,j} - F\left(X_{t,j}, \theta\right)\right)^2 + \frac{1}{N_{post}} \sum_{t,j \in O_{post}} \left(Y_{t,j} - F\left(X_{t,j}, \theta\right)\right)^2. \tag{1.21}$$

I refer to the set of N_{pre} option prices observed before the announcement as O_{pre} and the set of N_{post} option prices observed after the announcement as O_{post}. The risk neutral parameter vector is defined as $\theta = (p_Q, \tau_Q, \lambda, \sigma)$ which consists of the probability of success p_Q, expected deal duration τ_Q, announcement intensity λ, and volatility of target dividends σ. The estimate is $\hat{\theta} = \arg\min_{\theta \in \Theta} Q_N(\theta)$. To see how this relates to the generator defined in equation 1.3 note that

$$p_Q = \frac{\lambda_S}{\lambda_S + \lambda_F} \qquad \tau_Q = \frac{1}{\lambda_S + \lambda_F} \tag{1.22}$$

The estimate of the risk neutral probability of deal success $\hat{p}_Q$ is then used to forecast deal outcomes. Below I abuse notation for convenience by denoting the estimate as p_Q and occasionally referring to the estimate as the model implied probability of deal success. Before presenting the forecasting results I illustrate a few applications to specific deals. Identification and alternative estimation techniques are discussed in Appendix 1.C.

1.3.3 Examples of Estimated Deals

Examples of cash deals are presented in Figure 1.8. First, consider Google's successful acquisition of Motorola. The $12.45 billion deal was announced on August 15, 2011 and valued Motorola at $40 per share. At the outset there was uncertainty about the outcome. The deal represented Google's largest acquisition to date and entry into the smart phone market. On the announcement date the Wall Street Journal published an article titled "Google-Motorola Deal: Signs of Regulatory Worries" which highlighted antitrust approval as a chief concern. In addition to the requirement that the deal be approved by regulators in the United States, Europe, and China, the U.S. Department of Justice had recently launched an investigation into Google's online search and advertising business resulting in a concern that the deal would be subject to heightened scrutiny. Nonetheless, stock market investors reacted positively to the announcement. Motorola's price jumped 55% to $38.12 equivalent to a share implied probability of success $\hat{p}_Q^S = .86$. Figures 1.7 and 1.8 illustrate the reaction of the volatility surface. After the announcement skewness increased and at-the-money implied volatility decreased from over 40 percent to below 20 percent. The estimated model displayed in Figure 1.8 captures the volatility smile after the announcement and relatively flat implied volatility surface before the announcement with an estimated risk neutral probability of deal success equal to $\hat{p}_Q = .81$. Ultimately, Google acquired Motorola on May 22, 2012.

For an example of a failed cash deal, consider ConAgra's bid to acquire Ralcorp announced on April 29th, 2011 for $86 per share. In this case, the stock and option market forecasts for the deal outcome diverged. To be specific, the stock market reacted positively to the announcement with a share implied probability of success equal to $\hat{p}_Q^S = .89$. During the following months Ralcorp's stock continued to trade near or above the offer. This stands in contrast to the unsupportive reception of the ConAgra bid from Ralcorp's board. According to a May 4th, 2011 New York Times article, Ralcorp adopted a poison pill and the board indicated that accepting the offer was not in the best interest of shareholders. Prospects looked particularly bleak after taking into account the fact that Ralcorp was incorporated

in Missouri, which requires two-third of shareholders to approve a deal as opposed to a majority under Delaware law, and the additional obstacle that Ralcorp had a staggered board requiring at least two annual meetings to replace a majority of directors. In contrast to the stock market reaction, option prices exhibited little response after the deal was announced. The estimated risk neutral probability of success is $\hat{p}_Q = .20$ which results in relatively flat implied volatility surfaces in the pre-target and target states. In the end, ConAgra terminated their bid on September 19th and Ralcorp subsequently executed a spin off of their subsidiary Post Cereals in January 2012.

Two examples of stock deals are included in Figure 1.9. The successful transaction between Microsoft and Visio was announced on September 15th, 1999. Microsoft offered .45 shares per Visio share in a $1.37 billion deal that valued Visio stock at $42.78. In response to the offer, the closing price of Visio stock jumped 19% from $33.50 to $39.88. At the same time, the average level of implied volatility for Visio listed options decreased from 58% to 37%. The implied volatility after the announcement was much closer to the realized volatility of 34% that Microsoft exhibited over the previous year. The model captures this reaction with an estimate of the risk neutral probability of deal success equal to $\hat{p}_Q = .73$. For an example of a failed stock deal, I also include Staples bid for Office Depot announced on September 4th, 1996. Staples offered 1.14 shares per Office Depot share valuing Office Depot at $22.23. In this case, the stock market reacted positively to the deal with Office Depot trading at a tight spread to Staples immediately after the announcement. By way of contrast, the implied volatility of Office Depot options decreased despite the fact that the realized volatility of Staples stock over the previous year was 50% which was actually higher than the implied volatility of Office Depot options prior to the announcement. The resulting estimate of the risk neutral probability of success equals $\hat{p}_Q = .30$. Ultimately the deal failed to obtain antitrust approval. As described in the July 1, 1997 New York Times article "Office Depot and Staples Merger Halted", the FTC concluded that large office supply stores occupied a specialized market in which the two companies would command significant

pricing power despite the fact that Staples and Office Depot only controlled 6 to 8 percent of the office products market.

1.4 Forecasting Deal Outcomes

Linear probability regressions are reported in Table 1.3 for cash deals and Table 1.4 for stock and hybrid deals. The estimated risk neutral probability of success from the structural model significantly forecasts deal outcomes. Moreover, the model drives out the explanatory power of the target stock and remains significant after including various control variables and fixed effects in the regressions. The results are also robust across time periods. Figure 1.10 plots the out-of-sample R^2 using different initial estimation periods. Throughout the sample, the model outperforms target stock prices and target option prices in forecasting cash deal outcomes. Taken together, the results suggest that the estimated risk neutral probability of deal success, which relies on the joint information in target stock and option prices, delivers significant and accurate forecasts for deal outcomes.

To contrast the information contained in stock prices and option prices, the regression analysis includes reduced form variables constructed from target option prices. The reduced form variables highlight the predictive content embedded in option prices and serve as additional competitors for the structural model. The motivation for the reduced form model stems from the reaction of target options to deal announcements illustrated in Figure 1.11 for cash deals and Figure 1.12 for stock deals. For cash deals, Figure 1.11 plots the percentage change in the fitted four month implied volatility of target options relative to at-the-money implied volatility before the announcement. Implied volatility decreases and a volatility smile emerges after cash deals are announced. The decrease in implied volatility is larger in deals that are ex-post successful. For stock deals, Figure 1.12 plots the percentage difference between the implied volatility of at-the-money acquirer options and the fitted four month

implied volatility of target options.[22] The structural model predicts the reaction of target options will depend on the relationship between the volatility of the target fundamental and the volatility of the acquirer stock. This is confirmed in the figure. In the subset of deals for which the target has larger risk neutral variance than the acquirer before the announcement, implied volatility of target options decreases after the announcement. Similar to cash deals, the decrease in implied volatility is larger in deals that are ex-post successful. Unlike cash deals, target options do not exhibit an increase in skewness. Moreover, implied volatility does not decrease in the subset of stock deals for which the acquirer has larger risk neutral variance than the target prior to the deal being announced. This is consistent with the model, illustrating a salient empirical difference between cash and stock deals. Motivated by these observations, the reduced form variables included in the regression analysis include the risk neutral variance and skewness of the target stock after the deal is announced and the change in the risk neutral variance and skewness before and after the announcement. These variables are constructed following Bakshi et al. (2003) who derive the risk neutral moments as functions of out-of-the money call and put prices.

Table 1.3 reports the results for cash deals. The regressions indicate that target stock prices, target option prices, and the joint information in target stock and option prices all forecast deal outcomes. For ease of interpretation, I have standardized the reduced form variables to be mean zero and standard deviation one. For example, a one standard deviation decrease in risk neutral target variance increases the probability of deal success by 13%. The analysis indicates that option prices contain more information for forecasting deal outcomes than stock prices. In particular, the reduced form model explains cash deal outcomes with an R^2 of 19%. This represents an improvement over the share implied probability which delivers an R^2 of 8%. That said, the most accurate forecasts obtained by the estimate of the risk neutral probability of deal success from the structural model which combines the

[22]The results are robust to using pre-announcement or post-announcement acquirer options. In particular, acquirer options exhibit little reaction to deal announcements and the reduced form variables constructed from acquirer option prices do not forecast deal outcomes.

information in stock and option prices. The structural model drives out the share implied probability and the reduced form variables in explaining deal outcomes. Additionally, the point estimate on the estimated risk neutral probability of success only decreases from .88 to .74 and the R^2 only increases from 34% to 38% after including the share implied probability, reduced form variables, control variables, and fixed effects. In particular, the structural model captures most of the explanatory power for forecasting deal outcomes. In contrast, the point estimate for the share implied probability decreases from .46 to -.09 when all variables are included. The results also reject the null hypothesis that the point estimate on the risk neutral probability of success is equal to one.[23] Finally, Table 1.4 reports the regression results for stock and hybrid deals. While the model remains statistically significant and continues to drive out the share implied probability of success, the results reveal that forecasting stock deal outcomes is more challenging than forecasting cash deal outcomes. After including all variables the R^2 for stock and hybrid deals only increases to 12%. This is consistent with the intuition outlined in the introduction. A challenge in stock deals is that options are not traded directly on the merger arbitrage investment. As Figure 1.2 suggests, the change in the distribution of the target stock appears to be more subtle when stock deals are announced.

The out-of-sample results in Figure 1.10 plot the R^2 for cash deals using different initial estimation periods. For each split of the full sample, I estimate the model $Y_i = \beta_0 + \beta' x_i + \epsilon_i$ in rolling monthly regressions that only use deals which have already completed or failed. Forecasts for deal outcomes Y_i are then computed for deals announced in the following month by combining $\hat{\beta}$ with the forecasting variables x_i which include the estimated risk neutral probability of success, the share implied probability of success, and the reduced form target option variables. I refer to these regressions as forecasts because the variables only depend on target stock and option data up to 10 calendar days after the deal announcement, which is

[23]Recall $p_Q \approx p - p(1-p)(\kappa_F - \kappa_S)$. The null hypothesis that deal risk is not priced, $\kappa_F = \kappa_S = 0$, implies $H_0 : \beta_0 = 0$, $\beta_1 = 1$ in the regression $Y_i = \beta_0 + \beta_1 p_{Q,i} + \epsilon_i$. Empirically, I run the regression $Y_i = \beta_0 + \beta_1 \hat{p}_{Q,i} + \epsilon_i$ where $\hat{p}_{Q,i}$ is the estimated risk neutral probability of success. When interpreting the results, it is important to note that priced deal risk and measurement error result in $\beta_1 < 1$.

significantly shorter than the average deal duration of 115 days. Figure 1.10 illustrates that the structural model outperforms target stock and option prices throughout the sample period. Additionally, the reduced form model outperforms target stock prices which is consistent with Table 1.3.

1.5 Merger Arbitrage Returns

Nonlinear risk is a key feature of merger arbitrage investments. In my sample, the median deal announcement coincides with a 26.5% jump in the target stock price and a 3.7% arbitrage spread. The resulting investment opportunity is similar to writing options or investing in high grade debt. Merger arbitrageurs have little upside. The payoffs in successful deals are small relative to the losses in failed deals. To present a new perspective on the risk-return trade-off in these investments, I study merger arbitrage in the structural model outlined in this paper and then test the model's predictions empirically in the cross section of merger arbitrage returns sorted on the share implied probability of deal success.

The model predictions are summarized in Figure 1.5 which presents the comparative statics for how the instantaneous expected returns, volatility, and Sharpe ratio of merger arbitrage investments vary with the risk neutral probability of deal success. When deal risk is priced, expected returns are hump shaped in the probability of success because deal outcomes are binary. Meanwhile, volatility decreases in the probability of success. Deals that are likely to succeed place more weight on the arbitrage spread and less weight on volatile target dividends and acquirer stock prices. The implication is that the Sharpe ratio is largest in deals with a high probability of success. Finally, note that these predictions change in the absence of deal risk. In particular, when deal risk is not priced the Sharpe ratio for merger arbitrage equals the Sharpe ratio for the market portfolio which is assumed to be a constant in my model. Empirically, I find evidence that supports these predictions which suggests that deal risk is priced. For example, a simple trading strategy that overweights deals with a high probability of success increases the monthly Sharpe ratio by nearly 50% relative to an equal

weighted portfolio, which is equivalent to nearly doubling the CAPM alpha t-statistic. The prediction that volatility is decreasing in the probability of success is also strongly supported by the data. This can be seen graphically in Figure 1.14 which plots the volatility and CAPM beta of merger arbitrage portfolios sorted on the shared implied probability of deal success. Portfolios that overweight deals with a high share implied probability of success exhibit low volatility. Further evidence is presented in Fama-MacBeth regressions. In particular, the share implied probability of success is statistically significant and obtains an R^2 of 34% in explaining the the cross section of merger arbitrage variance. Finally, I also analyze the event time properties of merger arbitrage returns. To do this, I extend my baseline model to allow the deal success intensity to be a deterministic function of event time. Calibrating the success intensity to the hump shaped pattern documented in Giglio and Shue (2014), Figure 1.6 illustrates that the model's prediction for the evolution of expected returns and volatility in event time is consistent with the data. When deal risk is priced, expected returns are hump shaped and volatility is increasing in event time.

The risk-return relationship in merger arbitrage has been addressed in a number of studies in the literature. Mitchell and Pulvino (2001) emphasize that linear asset pricing models will not properly account for the nonlinear risks in merger arbitrage investments. They advocate a contingent claims approach for analyzing whether merger arbitrage delivers excess returns based on the observation that merger arbitrage portfolios have larger CAPM betas for economic states in which the market portfolio suffers a large loss. The suggested replicating portfolio purchases a risk free bond and sells an out of the money put option on the market index. After accounting for risk and transaction costs they conclude that merger arbitrage delivered excess returns of 4% over the period from 1963-1998. Baker and Savaşoglu (2002) posit that merger stocks are priced by a group of thinly capitalized investors who absorb supply shocks when deals are announced. They find empirical support for the predictions that expected returns are increasing in idiosyncratic deal risk and deal size. Giglio and Shue (2014) present a different perspective by analyzing merger arbitrage returns and the

resolution of deal uncertainty in event time. While failure intensity is relatively constant, they find success intensity is hump shaped and highest during the four to six months after deals are announced. They also find that expected returns are hump shaped in event time. Relative to standard risk models, a trading strategy that invests during the period of high success intensity outperforms traditional buy and hold strategies.

1.5.1 Expected Returns, Volatility, and CAPM Beta

To study merger arbitrage I construct portfolio returns following the buy and hold strategy outlined in Mitchell and Pulvino (2001). Deals are added to the portfolios two business days after the original announcement date. They are held until the target firm delists from CRSP, two days pass the deal termination date, or one year elapses, whichever comes first. The termination date refers to the effective date in successful transactions and the date withdrawn in unsuccessful transactions. The portfolio is rebalanced monthly. At the start of each month, I require the offer to be at least one dollar and I exclude deals with a negative arbitrage spread. The maximum portfolio weight is 10% for an individual deal.

I construct arbitrage, cash, and high probability portfolios. The arbitrage portfolios invest in cash, stock, and hybrid deals. The cash portfolios invest exclusively in cash deals. I also introduce high probability portfolios which are new to the literature. The high probability portfolios overweight deals with a high share implied probability of success.[24] To study the cross section of merger arbitrage returns, I sort the arbitrage and cash portfolios on the share implied probability of success. I use the share implied probability, despite the fact that it doesn't forecast deal outcomes as well as the model implied probability, because it can be computed for all of the active deals in my sample each month. By way of contrast,

[24]The high probability portfolio follows the buy and hold strategy using the portfolio weights $w_{it} = \min(s_{it}/\sum s_{it},\ \bar{w})$ where $s_{it} = f(p^S_{Q,it})/f(\mu)$ and $f \sim N(\mu, \sigma)$, $\mu = .9$, $\sigma = .2$, $\bar{w} = .1$. The results are robust to alternative definitions that result in the portfolio being overweight deals with a high probability of success. For example, a cut-off strategy that only goes long deals with $p^S_{Q,it}$ greater than $\underline{p}$ with the portfolio weight $w_{it} = \min(\mathbf{1}[p^S_{Q,it} > \underline{p}]/\sum \mathbf{1}[p^S_{Q,it} > \underline{p}], \bar{w})$ has a correlation of .93 with the high probability strategy using $\underline{p} = .7$ and similar CAPM alpha of .0043 (8.68) versus .0044 (8.51) for the high probability strategy. The average number of positions for the high probability, cut-off, value, and equal weighted portfolios is (84, 54, 84, 84) and the average turnover is (.44, .46, .35, .33).

the model implied probability requires that the target firm have listed options. While value weighted portfolios are 68% invested in deals with listed options, equal weighted portfolios are only 31% invested in deals with listed options during my sample. In particular, the share implied probability allows me to construct six sorted merger arbitrage portfolios with an average of 14 deals per month. Empirically, I confirm the share implied probability tracks portfolio success rates accurately. For example, Figure 1.13 illustrates that the share implied probability explains the variation in the percentage of an equal weighted portfolio invested in successful deals with an R^2 of 58%.

Summary statistics for merger arbitrage returns are provided in Table 1.5. The constructed returns appear representative when compared to hedge fund index returns for merger arbitrage and event driven strategies. The hedge fund returns are adjusted for the fact that indices are reported net-of-fees. To obtain pre-fee returns I assume a 2% flat fee and 10% incentive fee are paid monthly as in Jurek and Stafford (2015). The hedge fund returns and constructed merger arbitrage returns are positively correlated. Additionally, the arbitrage returns are positively correlated with the market portfolio and the small-minus-big factor from Fama and French (1993). The correlation with small-minus-big follows from the observation that target firms are often small. Additionally, stock deals involve investing in the target and shorting the acquirer. Targets are typically small relative to acquirers. Additionally, the arbitrage returns are negatively correlated with momentum. Given arbitrage spreads are small it is unlikely merger stocks will be winners. However, if a deal fails or a deal failure becomes more likely a target stock may become a loser that is shorted in a momentum portfolio resulting in the negative correlation with the arbitrage returns. The returns are uncorrelated with the value factor.

Table 1.6 presents time series regressions using standard risk models. The regressions indicate that merger arbitrage loads positively on the market portfolio and small-minus-big. Additionally, merger arbitrage loads significantly on a put writing strategy shown to replicate hedge fund returns in Jurek and Stafford (2015). This is consistent with the interpretation

in Mitchell and Pulvino (2001) that merger arbitrage exhibits downside risk. The regressions also illustrate the fact that merger arbitrage earns excess returns relative to standard risk models, which is consistent with previous studies. The new observation is that the high probability portfolio, which overweights deals with a high probability of success, nearly doubles the t-statistic for the intercept relative to the equal weighted portfolio. For example, the high probability arbitrage portfolio has a CAPM alpha of 44 basis points per month with a 8.51 t-statistic. The equal weighted arbitrage portfolio has a CAPM alpha of 51 basis points per month with a 4.48 t-statistic. This result is similar for each of the risk models considered. In terms of Sharpe ratios, the high probability arbitrage portfolio earns a monthly Sharpe ratio of .55 which contrasts .36 for the equal weighted portfolio, .59 for the Hennessee merger arbitrage returns, and .43 for the Dow Jones/Credit Suisse risk arbitrage returns. This is consistent with the model's prediction that merger arbitrage obtains a higher Sharpe ratios in deals with a high probability of success. To put the size of the intercept into perspective versus other studies, Giglio and Shue (2014) report that buy and hold strategies earn 51 basis points per month during the 1991-2010 sample period and 71 basis points per month in the 1970-2010 sample period. The decrease in average merger arbitrage returns over time is consistent with the observation in Jetley and Ji (2010) that merger arbitrage spreads decreased in the early 2000s, in part due to increased assets under management dedicated to merger arbitrage.

To study the cross section of returns, I sort deals on the share implied probability of success and form six portfolios. Figure 1.14 plots the mean, volatility, Sharpe ratio, and CAPM beta for these portfolios alongside 95% confidence bands that have been block bootstrapped to account for the autocorrelation in returns. The sorted portfolios are consistent with the model's prediction that volatility and CAPM beta are decreasing in the probability of success. The Sharpe ratio also appears to be increasing in the probability of success in line with the evidence from the time series regressions. It is more challenging to determine whether average returns are hump shaped in the probability of success. In

part this stems from the high volatility of the low probability portfolios, which can be seen graphically from the wide confidence bands for the merger arbitrage portfolios with low share implied probabilities of success. Another factor contributing to the lack of precision is that my 1996-2012 sample is relatively short for the purpose of studying expected returns. To test the prediction that merger arbitrage volatility is decreasing in the probability of deal success, I run Fama-MacBeth cross-sectional regressions which are reported in Table 1.7. The coefficient on the share implied probability is negative and statistically significant, indicating that merger arbitrage variance is decreasing in the probability of deal success. As a robustness check, I also run cross sectional regressions at the level of the deal on the subset of cash deals with available options data. These regressions also support the conclusion that merger arbitrage variance is decreasing in the probability of deal success, a result which continues to hold after including the estimated risk neutral dividend volatility, deal size, acquirer attitude, and a number of other control variables in addition to calendar time and event time fixed effects. I also find that merger arbitrage volatility is increasing with the estimated risk neutral dividend volatility, which is consistent with the model.

1.5.2 Path of Expected Merger Arbitrage Returns

Finally, I apply the model to study the event time properties of merger arbitrage returns. In particular, my baseline model in Section 1.2 assumes that deal success and failure intensity are constant. This is consistent with empirical estimates for the failure intensity. However, it stands in contrast to the observation in Giglio and Shue (2014) that success intensity is hump shaped. To that end, I present an extension of the model in Appendix 1.B that allows the risk neutral deal success intensity $\lambda_S(t)$ to be a deterministic function in event time. The extension results in two primary departures from the baseline model. First, the hump shaped pattern in success intensity results in a hump shaped pattern in the probability of deal success. Second, the price continues to follow a modified Gordon Growth formula but

now the coefficients on target dividends and the deal payout are functions of event time[25]

$$P_t(\mathcal{T}) = B_0(t) D_t + B_1(t) V + B_2(t) P_{A,t}. \tag{1.23}$$

The coefficient functions $(B_0(t), B_1(t), B_2(t))$ solve linear first order differential equations with closed form solutions that are detailed in the Appendix.

Figure 1.6 illustrates the model graphically. The success and failure intensity are calibrated to Figure 2 in Giglio and Shue (2014). The success intensity starts below the failure intensity and then peaks in weeks 20 and 30 after the announcement. The plot highlights an important distinction between the success intensity and the probability of success. In particular, the probability of success takes into account the entire path of the success intensity going forward. An implication is that the probability of success is hump shaped. Additionally, the probability of success reaches its peak between weeks 10 and 20 which occurs before the peak in the success intensity between weeks 20 and 30. Intuitively, the probability of success is the average of the future success intensity, similar to the relationship between average and marginal costs. Once the success intensity has peaked its average over future dates is decreasing. Consequently, the probability of success will reach its peak before the success intensity has peaked.

The distinction between the probability of success and the success intensity is important for understanding the path of expected merger arbitrage returns. To illustrate this, I maintain the assumption that deal risk is priced and that the relationship between the risk neutral and objective intensities is given by

$$\begin{aligned} \lambda_S^{\mathbb{P}}(t) &= e^{-\kappa_S} \lambda_S(t) \\ \lambda_F^{\mathbb{P}} &= e^{-\kappa_F} \lambda_F \end{aligned} \tag{1.24}$$

[25]For notational convenience I assume the deal announcement occurs at time zero.

The market prices of deal risk (κ_S, κ_F) are constant as before. The model matches the path of expected returns in the data as illustrated in Figure 1.6. Expected returns are highest when the deal success intensity is highest and volatility is increasing in event time. Similar to Proposition 4, the contribution of deal success risk to expected merger arbitrage returns is

$$\left(e^{-\kappa_S} - 1\right) \lambda_S(t) \left[\frac{V + \phi P_{A,t}}{P_t(\mathcal{T})} - 1\right] \tag{1.25}$$

When success intensity $\lambda_S(t)$ is high the leading term above is larger. However, there could be a mitigating effect in that the price $P_t(\mathcal{T})$ is also impacted by the success intensity. A higher intensity might lead to a higher price so that expected returns would not display a hump shape in event time. Figure 1.6 illustrates this is not the case. The result follows from the observation that the probability of success, which is what matters for pricing target stocks, is the average of the success intensity going forward. Once the success intensity peaks the probability of success has already started to fall along with the target stock price. The result is a hump shaped pattern in expected returns. In summary, this analysis highlights the fact that expected returns are not necessarily proportional to volatility in nonlinear investments.

1.6 Conclusion

In conclusion, this paper contributes to the literature by providing new theoretical and empirical perspectives on mergers and acquisitions. I demonstrate that option prices embed predictive content for forecasting deal outcomes. Additionally, by estimating the risk neutral probability of success in my structural model, I use the joint information from target stock and option prices to obtain accurate out-of-sample forecasts for deal outcomes that outperform competing models that rely on the isolated information in either target stock prices or target option prices. I also employ my model to study the risk-return trade-off in merger arbitrage investments. The model predicts that merger arbitrage exhibits low

volatility and large Sharpe ratios when deals are likely to succeed. Empirically, I show that volatility is decreasing in the cross section of merger arbitrage returns sorted on the share implied probability of success. Additionally, the monthly Sharpe ratio in a portfolio that overweights deals with a high share implied probability of success is nearly 50% larger than the Sharpe ratio in an equal weighted portfolio. Finally, I demonstrate that the model is consistent with the dynamic properties of expected returns and volatility in event time. Overall, the results support the conclusion that priced deal risk is an important feature of merger arbitrage returns and that a parsimonious model for pricing merger stocks and options can be used to deliver accurate forecasts for deal outcomes, in addition to matching various features of merger arbitrage returns. Moving forward, this research can be extended in several directions. One abstraction in the analysis so far is that I have ignored rival bids and revisions to offers. Options provide a rich source of information about the distribution of the target stock price. It would be interesting to investigate whether option markets predict rival bids. Additionally, the model predicts that stock prices are positively skewed before mergers are announced. One could investigate whether positive skewness in option markets anticipates the announcement of mergers. Finally, I have observed that option prices reveal the risk neutral probability of success for corporate events that result in stock prices equaling a fixed value with a discrete probability. In addition to mergers, bankruptcy is a natural example where similar analysis may prove useful.

References

Aït-Sahalia, Y., J. Cacho-Diaz, and T. R. Hurd (2009). Portfolio choice with jumps: A closed-form solution. *Annals of Applied Probability 19*(2), 556–584.

Andrade, G., M. Mitchell, and E. Stafford (2001). New evidence and perspectives on mergers. *Journal of Economic Perspectives*, 103–120.

Baker, M., X. Pan, and J. Wurgler (2012). The effect of reference point prices on mergers and acquisitions. *Journal of Financial Economics 106*(1), 49–71.

Baker, M. and S. Savaşoglu (2002). Limited arbitrage in mergers and acquisitions. *Journal of Financial Economics 64*(1), 91–115.

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.

Barraclough, K., D. T. Robinson, T. Smith, and R. E. Whaley (2013). Using option prices to infer overpayments and synergies in m&a transactions. *Review of Financial Studies 26*(3), 695–722.

Bester, C. A., H. V. Martinez, and I. Rosu (2013). Cash mergers and the volatility smile. *SSRN Working Paper*.

Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities. *The Journal of Political Economy*, 637–654.

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

Brown, K. C. and M. V. Raymond (1986). Risk arbitrage and the prediction of successful corporate takeovers. *Financial Management*, 54–63.

Cao, C., Z. Chen, and J. Griffin (2005). Informational content of option volume prior to takeovers. *Journal of Business 78*(3), 1073–1109.

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

Duffie, D. (2010). *Dynamic Asset Pricing Theory*. Princeton University Press.

Edmans, A., I. Goldstein, and W. Jiang (2012). The real effects of financial markets: The impact of prices on takeovers. *Journal of Finance 67*(3), 933–971.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*(1), 3–56.

Giglio, S. and K. Shue (2014). No news is news: Do markets underreact to nothing? *Review of Financial Studies 27*(12), 3389–3440.

Grossman, S. J. and O. D. Hart (1980). Takeover bids, the free-rider problem, and the theory of the corporation. *Bell Journal of Economics*, 42–64.

Harford, J. (2005). What drives merger waves? *Journal of Financial Economics 77*(3), 529–560.

Jacod, J. and A. N. Shiryaev (1987). *Limit theorems for stochastic processes*, Volume 288. Springer-Verlag Berlin.

Jarrow, R. A., D. Lando, and S. M. Turnbull (1997). A markov model for the term structure of credit risk spreads. *Review of Financial Studies 10*(2), 481–523.

Jetley, G. and X. Ji (2010). The shrinking merger arbitrage spread: Reasons and implications. *Financial Analysts Journal*, 54–68.

Johannes, M. and N. Polson (2009). Mcmc methods for continuous-time financial econometrics. *Handbook of Financial Econometrics 2*, 1–72.

Jurek, J. and E. Stafford (2015). The cost of capital for alternative investments. *Journal of Finance, Forthcoming*.

Karatzas, I. (1991). *Brownian motion and stochastic calculus*, Volume 113. springer.

Lando, D. (2009). *Credit risk modeling: theory and applications*. Princeton University Press.

Lee, R. W. (2004). The moment formula for implied volatility at extreme strikes. *Mathematical Finance 14*(3), 469–480.

Leland, H. E. (1994). Corporate debt value, bond covenants, and optimal capital structure. *Journal of Finance 49*(4), 1213–1252.

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

Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica*, 867–887.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance 29*(2), 449–470.

Mitchell, M. and T. Pulvino (2001). Characteristics of risk and return in risk arbitrage. *Journal of Finance 56*(6), 2135–2175.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics 4*, 2111–2245.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Officer, M. S. (2004). Collars and renegotiation in mergers and acquisitions. *Journal of Finance 59*(6), 2719–2743.

Officer, M. S. (2006). The market pricing of implicit options in merger collars. *Journal of Business 79*(1), 115–136.

Protter, P. E. (2004). *Stochastic Integration and Differential Equations: Version 2.1*, Volume 21. Springer.

Samuelson, W. and L. Rosenthal (1986). Price movements as indicators of tender offer success. *Journal of Finance 41*(2), 481–499.

Schwert, G. W. (2000). Hostility in takeovers: in the eyes of the beholder? *Journal of Finance 55*(6), 2599–2640.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance 19*(3), 425–442.

Shleifer, A. and R. W. Vishny (2003). Stock market driven acquisitions. *Journal of Financial Economics 70*(3), 295–311.

Subramanian, A. (2004). Option pricing on stocks in mergers and acquisitions. *Journal of Finance 59*(2), 795–829.

Walkling, R. A. (1985). Predicting tender offer success: A logistic analysis. *Journal of Financial and Quantitative Analysis 20*(04), 461–478.

Tables

Deal Summary Statistics

Full Sample	Cash		Stock		Hybrid	
N	3,164		1,077		123	
Outcome	0.79	*0.41*	**0.88**	*0.33*	**0.89**	*0.32*
Duration	0.30	*0.19*	**0.38**	*0.17*	**0.42**	*0.17*
Premium	0.37	*0.27*	**0.31**	*0.25*	0.33	*0.22*
Price/Book	2.14	*1.73*	**3.01**	*2.74*	2.27	*1.49*
Price/EPS	27.31	*24.85*	28.87	*27.49*	**22.20**	*17.21*
Deal Value	0.60	*0.95*	**1.46**	*2.62*	**3.32**	*6.38*
Target Size	0.45	*0.71*	**1.15**	*2.08*	**2.68**	*5.39*
Acquirer Size	17.38	*31.54*	**5.39**	*8.89*	**7.85**	*12.88*

Option Sample	Cash		Stock		Hybrid	
N	921		417		71	
Outcome	0.79	*0.40*	0.83	*0.38*	0.90	*0.30*
Duration	0.30	*0.19*	0.37	*0.19*	0.43	*0.21*
Premium	**0.34**	*0.21*	0.28	*0.22*	0.29	*0.18*
Price/Book	**2.84**	*2.08*	**4.22**	*3.95*	**2.88**	*1.52*
Price/EPS	30.47	*28.82*	**39.14**	*40.02*	19.26	*10.56*
Deal Value	**1.93**	*2.19*	**4.26**	*6.43*	6.13	*9.38*
Target Size	**1.45**	*1.66*	**3.44**	*5.42*	4.86	*7.57*
Acquirer Size	**33.37**	*44.70*	**14.42**	*24.37*	15.44	*25.52*

Table 1.1: Deal summary statistics are reported including the sample mean alongside the standard deviation in italics. Variables are winsorized at the 5th and 95th quantiles. In the option sample bold indicates the null hypothesis of equal means between the full sample and option subsample within the same type of financing is rejected at the 1% significance level. In the full sample bold indicates the null of equal means is rejected when stock or hybrid deals are compared to cash deals. Outcome equals one for offers that are successful and zero otherwise. Duration is the time between the original announcement date and termination date of a transaction in years. Premium, Price/Book, and Price/EPS are computed using the target share price one week prior to the deal announcement. Deal Value, Target Size, and Acquirer Size are in billions. Deal Value is defined as the total consideration paid by the acquirer excluding fees and expenses including liabilities assumed if they are publicly disclosed.

Option Summary Statistics

Cash Deals			Stock and Hybrid Deals		
	Pre	Post		Pre	Post
N	260.98	61.10	N	297.98	130.92
Call	0.46	0.35	Call	0.46	0.42
T	15.25	6.01	T	15.23	6.12
τ	0.34	0.35	τ	0.32	0.33
σ	0.50	0.26	σ	0.60	0.50
Δ	0.37	0.31	Δ	0.37	0.35
Interest	0.31	0.56	Interest	0.36	0.36
Volume	0.17	0.57	Volume	0.20	0.24
	RMSE	RMSE		RMSE	RMSE
Model	0.19	0.16	Model	0.24	0.25
BS2	0.19	0.20	BS2	0.23	0.23
BS	0.72	0.52	BS	0.41	0.36
	RMSE %	RMSE %		RMSE %	RMSE %
Model	0.26	0.43	Model	0.25	0.29
BS2	0.25	0.58	BS2	0.24	0.28
BS	0.60	1.80	BS	0.34	0.40

Table 1.2: Summary statistics for option prices and average model pricing errors are reported above. The model is estimated using target stock and option prices in pre-announcement $[-25, -5)$ and post-announcement $(1, 10]$ calendar day windows surrounding deal announcements. N is the average number of options observed in each of these windows. The other variables correspond to averages across deals of the within deal average. For example, the mean of σ is the average across deals of the average Black-Scholes implied volatility before or after an announcement within a particular deal. The variables are winsorized at the 5th and 95th quantile and include the number of days from the announcement T, the fraction of call options Call, maturity τ, Black-Scholes implied volatility σ, the absolute value of the Black-Scholes delta Δ, open interest in thousands, and volume in hundreds. I also report the root mean square pricing and percentage pricing errors. The competing models include BS which is the Black-Scholes model and BS2 which is an ad hoc Black-Scholes model that uses a different implied volatility in the pre-announcement and post-announcement periods. The merger option pricing model fits option prices better than the Black-Scholes model, particularly for cash deals.

Cash Deal Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
p_Q Model	0.883***				0.739***	0.738***
	[15.80]				[11.50]	[10.12]
p_Q^S Share		0.459***		0.156*		-0.0990
		[4.77]		[2.31]		[-1.54]
Target Var			-0.130***	-0.111***		-0.00627
			[-6.47]	[-4.80]		[-0.28]
Δ Target Var			0.0532***	0.0492***		0.00595
			[8.24]	[7.33]		[0.86]
Target Skew			0.103	0.0962		0.0721
			[1.54]	[1.59]		[1.24]
Δ Target Skew			0.103*	0.0957*		0.0437
			[2.87]	[2.87]		[1.34]
Adjusted R^2	0.345	0.086	0.196	0.202	0.383	0.382
Observations	805	805	805	805	805	805
Variables					C, FE	C,FE

robust t-statistics clustered by year: * p<0.05 ** p<0.01 *** p<0.001

Table 1.3: This table reports linear probability regression results for the outcomes of cash deals. I denote the estimated risk neutral probability of deal success from the model of merger stock and option pricing as p_Q Model. To estimate the model I minimize squared option pricing errors having inferred the level of dividends from the target stock price during pre-announcement $[-25, -5)$ and post-announcement $(1, 10]$ windows that surrounding deal announcements. The structural model drives out the explanatory power of the the share implied probability of success p_Q^S computed from the target stock price and the reduced form variables constructed from target option prices including the risk neutral variance and skewness after the announcement and the change in the risk neutral variance and skewness before and after the announcement. The risk neutral moments are computed following Bakshi et al. (2003). The results are robust to the inclusion of control variables (C) in addition to time and industry fixed effects (FE). The controls include a hostile indicator equal to one if the initial reception of a deal is categorized as hostile by SDC and an indicator for whether the offer is larger than the 52-week high over the preceding $[-365, -25)$ days prior to the announcement which have been highlighted by Schwert (2000) and Baker et al. (2012) for forecasting deal outcomes.

Stock and Hybrid Deal Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
p_Q Model	0.198***				0.161**	0.127*
	[3.71]				[2.67]	[2.06]
p_Q^S Share		0.180***		0.141**		0.0554
		[3.63]		[2.34]		[0.75]
Target Var			-0.0206	-0.00497		0.0115
			[-1.12]	[-0.23]		[0.32]
Δ Target Var			0.0307***	0.0251**		0.0253*
			[3.04]	[2.34]		[2.10]
Target Skew			-0.0367*	-0.0312		-0.0317
			[-1.95]	[-1.62]		[-1.50]
Δ Target Skew			0.000569	-0.00257		-0.00508
			[0.05]	[-0.18]		[-0.32]
Adjusted R^2	0.028	0.029	0.021	0.035	0.125	0.128
Observations	480	480	480	480	480	480
Variables					C, FE	C,FE

robust t-statistics clustered by year: * p<0.10 ** p<0.05 *** p<0.01

Table 1.4: This table reports linear probability regression results for the outcomes of stock and hybrid deals. I denote the estimated risk neutral probability of deal success from the model of merger stock and option pricing as p_Q Model. To estimate the model I minimize squared option pricing errors having inferred the level of dividends from the target stock price during pre-announcement $[-25, -5)$ and post-announcement $(1, 10]$ windows that surrounding deal announcements. The structural model drives out the explanatory power of the the share implied probability of success p_Q^S computed from the target stock price and the reduced form variables computed from target option prices including the risk neutral variance and skewness after the announcement and the change in the risk neutral skewness before and after the announcement. The risk neutral moments are computed following Bakshi et al. (2003). The results are robust to the inclusion of control variables (C) in addition to time and industry fixed effects (FE). The controls include a hostile indicator equal to one if the initial reception of a deal is categorized as hostile by SDC and an indicator for whether the offer is larger than the 52-week high over the preceding $[-365, -25)$ days prior to the announcement which have been highlighted by Schwert (2000) and Baker et al. (2012) for forecasting deal outcomes.

Summary Statistics for Monthly Merger Arbitrage Returns

	Hedge Fund Indices				Merger Arbitrage Portfolios			
	DJCS ED	DJCS Arb	HEN ED	HEN Arb	Arb Equal	Cash Equal	Arb HP	Cash HP
Mean	0.0084	0.0055	0.0076	0.0066	0.0059	0.0058	0.0047	0.0050
Std. Deviation	0.0194	0.0129	0.0161	0.0111	0.0161	0.0212	0.0085	0.0107
Sharpe ratio	0.4313	0.4290	0.4711	0.5991	0.3683	0.2738	0.5589	0.4636
Min	-0.1203	-0.0641	-0.0772	-0.0496	-0.0816	-0.0834	-0.0314	-0.0452
Max	0.0487	0.0404	0.0552	0.0329	0.0779	0.0974	0.0463	0.0612
Skewness	-2.0786	-0.8719	-1.5406	-1.2290	-0.7750	-0.3059	-0.4943	-0.3563
Kurtosis	12.4903	7.0454	9.4004	6.6729	10.0699	8.4222	8.2278	9.7449
Autocorrelation	0.3542	0.2724	0.4966	0.2071	0.2560	0.3128	0.0145	0.0353
Max Drawdown	0.1653	0.0727	0.1983	0.0481	0.1834	0.2292	0.0370	0.0436
% Negative	0.2451	0.2843	0.2402	0.1961	0.2892	0.3186	0.1520	0.2108
β_{CAPM}	0.2767	0.1454	0.2408	0.1446	0.1743	0.2502	0.0746	0.1103
α_{CAPM}	0.0071	0.0049	0.0065	0.0060	0.0051	0.0047	0.0044	0.0045
t-statistic	4.7812	5.5355	5.2762	7.5461	4.4755	2.9638	8.5111	7.2228

Return Correlation Matrix

				DJCS	HEN	Arb	Cash	Arb	Cash
MKT	SMB	HML	MOM	Arb	Arb	Equal	Equal	HP	HP
1.00	0.27	-0.24	-0.33	0.53	0.61	0.51	0.55	0.41	0.48
	1.00	-0.28	0.01	0.34	0.35	0.33	0.37	0.28	0.33
		1.00	-0.06	-0.08	-0.10	-0.02	-0.05	0.01	-0.01
			1.00	-0.15	-0.17	-0.35	-0.37	-0.19	-0.19
				1.00	0.77	0.43	0.48	0.42	0.47
					1.00	0.50	0.53	0.51	0.56
						1.00	0.94	0.69	0.63
							1.00	0.64	0.66
								1.00	0.92
									1.00

Table 1.5: Hedge fund index returns are illustrated alongside the merger arbitrage returns constructed for the empirical analysis in this paper. The hedge fund indices include Dow Jones/Credit Suisse event driven (ED) and risk arbitrage (Arb) value weighted indices as well as Hennessee event driven (ED) and merger arbitrage (Arb) equal weighted indices. To adjust for the fact that indices are net-of-fees, I assume a 2% flat fee and 10% incentive fee are payable monthly following Jurek and Stafford (2015) and report pre-fee summary statistics above. The arbitrage, cash, and high probability strategies are described in Section 1.5. The Sharpe ratio and CAPM alpha of the high probability strategy more closely match the hedge fund index returns. The correlation matrix indicates that the arbitrage series are positively correlated with each other in addition to MKT and SMB. The correlation with SMB follows from the observation that merger arbitrage strategies are long target firms that are small relative to large acquirers which are sold short in stock deals.

Time Series Regressions for Merger Arbitrage Returns

Arb Equal	(1)	(2)	(3)
Intercept	0.0051***	0.0052***	0.0044***
	[4.48]	[4.31]	[3.65]
RMRF	0.1743***	0.138***	
	[4.73]	[3.89]	
SMB		0.1305***	
		[4.02]	
HML		0.0674**	
		[1.97]	
MOM		-0.0604**	
		[-2.27]	
PUT			0.2935***
			[8.35]
R^2_{adj}	0.25	0.35	0.26

Arb HP	(1)	(2)	(3)
Intercept	0.0044***	0.0043***	0.0041***
	[8.51]	[8.06]	[7.44]
RMRF	0.0746***	0.067***	
	[4.41]	[3.96]	
SMB		0.0596***	
		[2.72]	
HML		0.0418**	
		[2.1]	
MOM		-0.0092	
		[-0.6]	
PUT			0.1169***
			[4.75]
R^2_{adj}	0.17	0.21	0.15

Table 1.6: This table reports monthly excess return regressions over the period 1996-2012 ($T = 204$). The dependent variables include the buy and hold merger arbitrage returns from the arbitrage and high probability strategies. The three risk models considered include the CAPM, the Fama and French (1993) model plus momentum, and a put writing strategy outlined in Jurek and Stafford (2015) with ($Z = -1$, $L = 2$) which is found to closely match the risk profile of the HFRI broad hedge fund index. Newey and West (1987) t-statistics are reported in brackets using five lags. The analysis illustrates that the high probability strategy increases the significance of the intercept and decreases the loadings on the risk factors. The significant loading on the put writing strategy is consistent with the interpretation that merger arbitrage exhibits downside risk as discussed in Mitchell and Pulvino (2001).

Merger Arbitrage Variance Regressions

Fama-MacBeth Regressions

	(1)	(2)
β_1	-0.0021	-0.0030
	[-4.73]	[-4.40]
R^2_{adj}	0.376	0.427

Cross-Sectional Regressions

	(1)	(2)	(3)
p_Q Model	-0.0196***	-0.0177***	-0.0172***
	[-5.83]	[-4.83]	[-3.90]
σ Model		0.00364*	0.00173
		[1.95]	[0.69]
Adjusted R^2	0.044	0.054	0.084
Observations	3063	3063	3063
Variables		C	C,FE

*p<0.10 **p<0.05 ***p<0.01

Table 1.7: These tables provide evidence that merger arbitrage variance decreases in the probability of deal success. The top table reports Fama-MacBeth regressions for the cross section of merger arbitrage returns sorted on the share implied probability of success $E_t\left[Re^2_{i,t+1}\right] = \beta_{0t} + \beta_{1t} p^S_{Q,it}$ where $Re_{i,t+1}$ is the excess return and $p^S_{Q,it}$ is the average share implied probability in month t for the i-th portfolio. I report Newey-West t-statistics for the arbitrage portfolios (1) and cash portfolios (2). Both regressions indicate that merger arbitrage variance is lower in deals that are likely to succeed. Additionally, I run cross sectional regressions for cash deals with available options data $Re^2_{i,m,t+1} = \beta_0 + \beta_1 p_{Q,i} + \beta_2 \sigma_i + \beta' f_i + \nu_t + \eta_m + \epsilon_{i,m,t+1}$ and report robust t-statistics clustered by month. The variables $p_{Q,i}$ and σ_i are the estimated risk neutral probability of deal success and target dividend volatility in the i-th deal. The second specification (2) includes control variables f_i such as the size of the deal, the market return in month $t+1$, whether the acquirer is hostile, etc. The third specification (3) adds calendar time fixed effects ν_t and event time fixed effects η_m. The results are consistent with the Fama-MacBeth regressions. Merger arbitrage variance is lower in deals with a high probability of success.

Figures

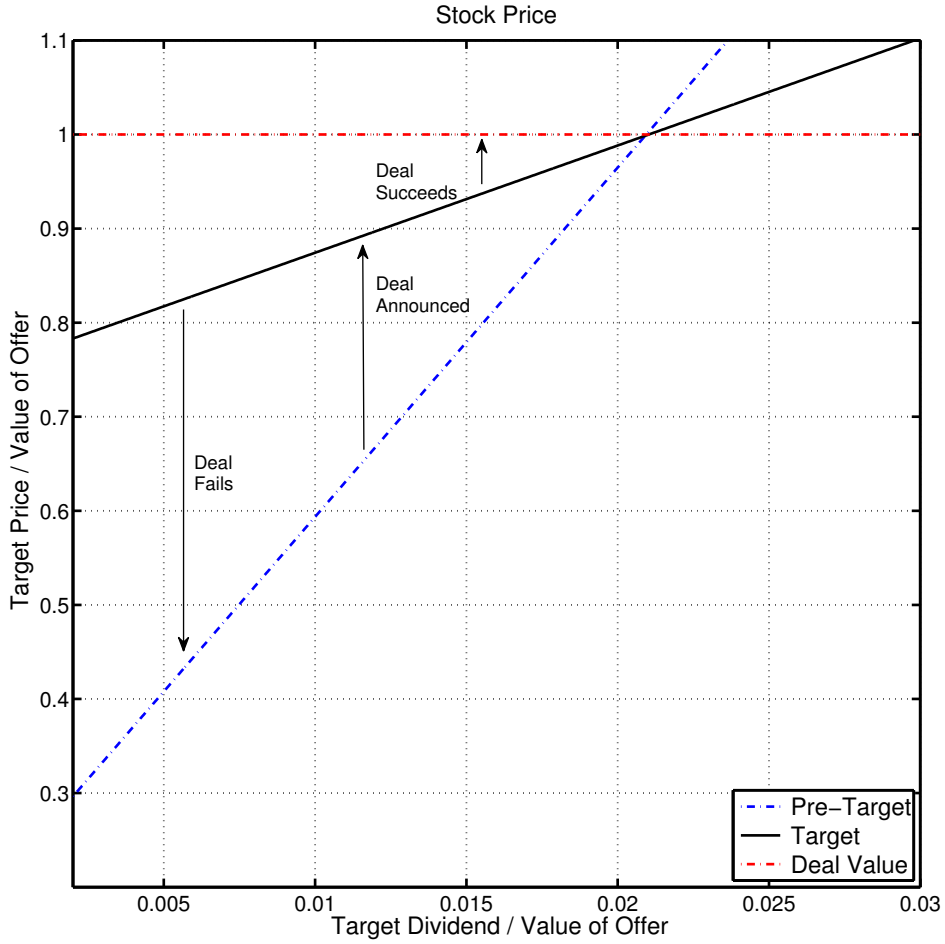


Figure 1.1: This figure plots the stock price from the model outlined in Section 1.2 for the pre-target, target, and acquired states. The offer $V + \phi P_{A,t}$ consists of fifty percent cash V and fifty percent acquirer stock $\phi P_{A,t}$. The target stock price is plotted as a function of the fundamental D_t holding the value of the offer fixed. The plot highlights two features of the model. First, transitions between firm states result in jumps in the target stock price. For example, when a deal is announced the firm transitions from the pre-target to the target state. In this case the price jumps from the blue dashed line to the black solid line depending on the prevailing level of dividends. A second feature of the model highlighted in the plot is that merger stocks are less sensitive to dividends in the target state versus the pre-target state, which can be seen graphically from the smaller slope in the target state. An implication is that stock prices have less diffusive risk in the target state which decreases the implied volatility of at-the-money options when deals are announced. The parameters are $(p_Q, \tau_Q, \lambda, r, g, q_A) = (.70, .33, .01, .03, .01, .02)$.

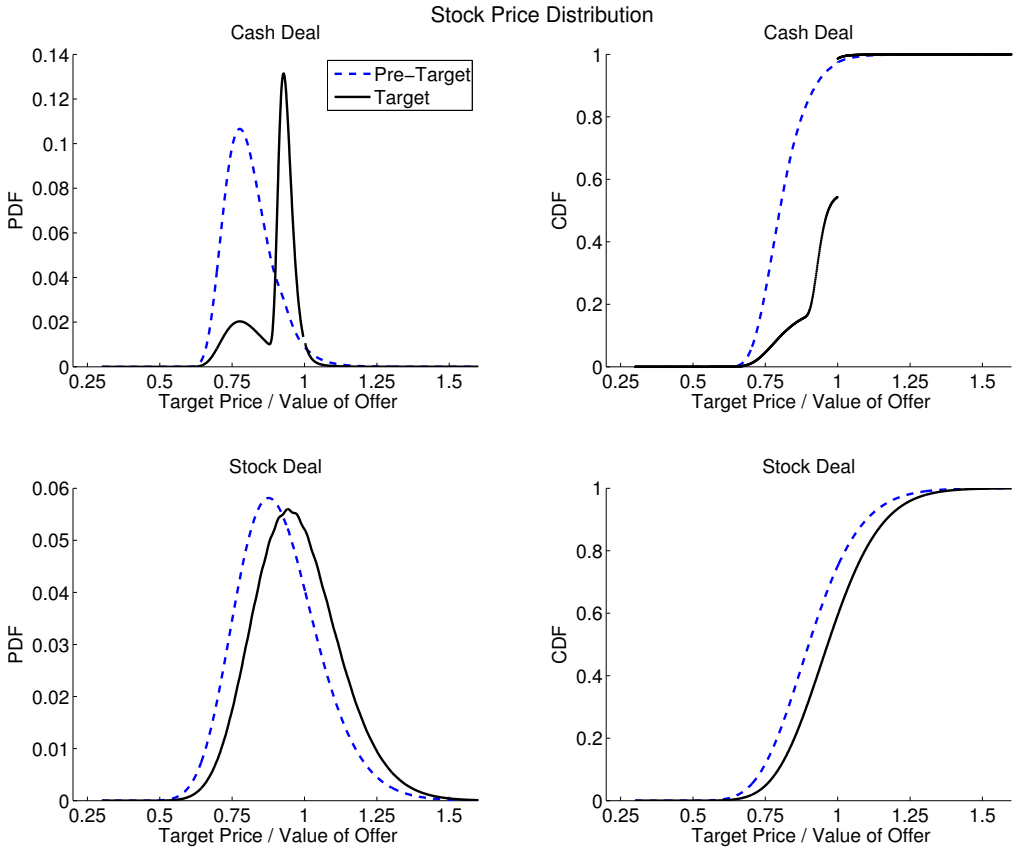


Figure 1.2: The conditional probability density function $p^{\mathbb{Q}}(P_T|D_t, s_t, P_{A,t})$ and cumulative distribution function $Pr^{\mathbb{Q}}(P_T \leq x|D_t, s_t, P_{A,t})$ are plotted at a four month horizon under the risk neutral measure. The x-axis is the stock price $P(D_t, P_{A,t}, s_t)$ divided by the offer $V+\phi P_{A,t}$. The plot includes the stock price distribution in the pre-target and target states $s_t \in \{\mathcal{P}, \mathcal{T}\}$ for deals financed by cash and stock. The deal premium relative to the fundamental is $(V+\phi P_{A,t})/(D_t/(r-g)) = 1.33$ with $D_t = .75$ and $V + \phi P_{A,t} = 50$. Cash deals have a discontinuity in the cumulative distribution function at the offer equal to the risk neutral probability of deal success prior to time T. Notably, the discontinuity is larger in the target state as the deal is already pending. The plot illustrates the change in the distribution of the target stock price when deals are announced. In both cash and stock deals, the distribution shifts to the right when deals are announced which corresponds to the jump in the target stock price. In cash deals, the target stock price has two peaks in the probability density function. The parameters are $(p_Q, \tau_Q, \lambda, r, g, q_A, \sigma, \sigma_A, \rho) = (.7, .33, .01, .03, .01, .02, .5, .25, .35)$.

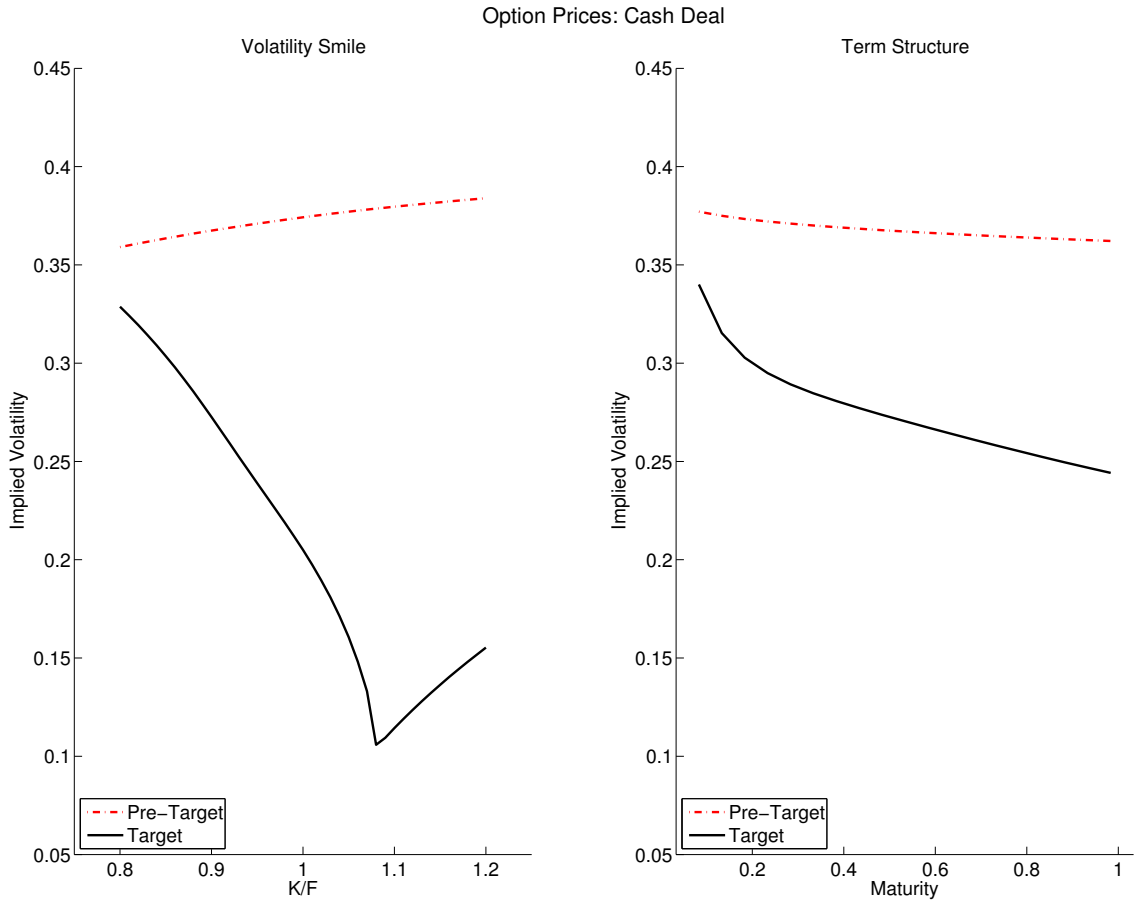


Figure 1.3: This figure plots the implied volatility of European options for a cash deal. The plot to the left illustrates the volatility smile for options with a maturity equal to half a year $\tau = 1/2$. The plot to the right illustrates the term structure of implied volatility with the strike price equal to a fixed percentage of the forward $K_\tau = .9 \times F_\tau$ where F_τ is the τ period forward price. The plot demonstrates that deal announcements coincide with a decrease in implied volatility. Additionally, option prices exhibit a volatility smile in the target state. The risk of deal failure and the corresponding negative jump in the target stock price increases the implied volatility of out-of-the money options relative to at-the-money options. Additionally, there is a kink in implied volatility when the strike price equals the offer $V/F = 1.08$. The plot to the right highlights the fact that the term structure of implied volatility is downward sloping for out of the money options. The values of the parameters are $D_t = .75$, $V = 50$, and $(p_Q, \tau_Q, \lambda, r, g, \sigma) = (.7, .33, .01, .03, .01, .5)$.

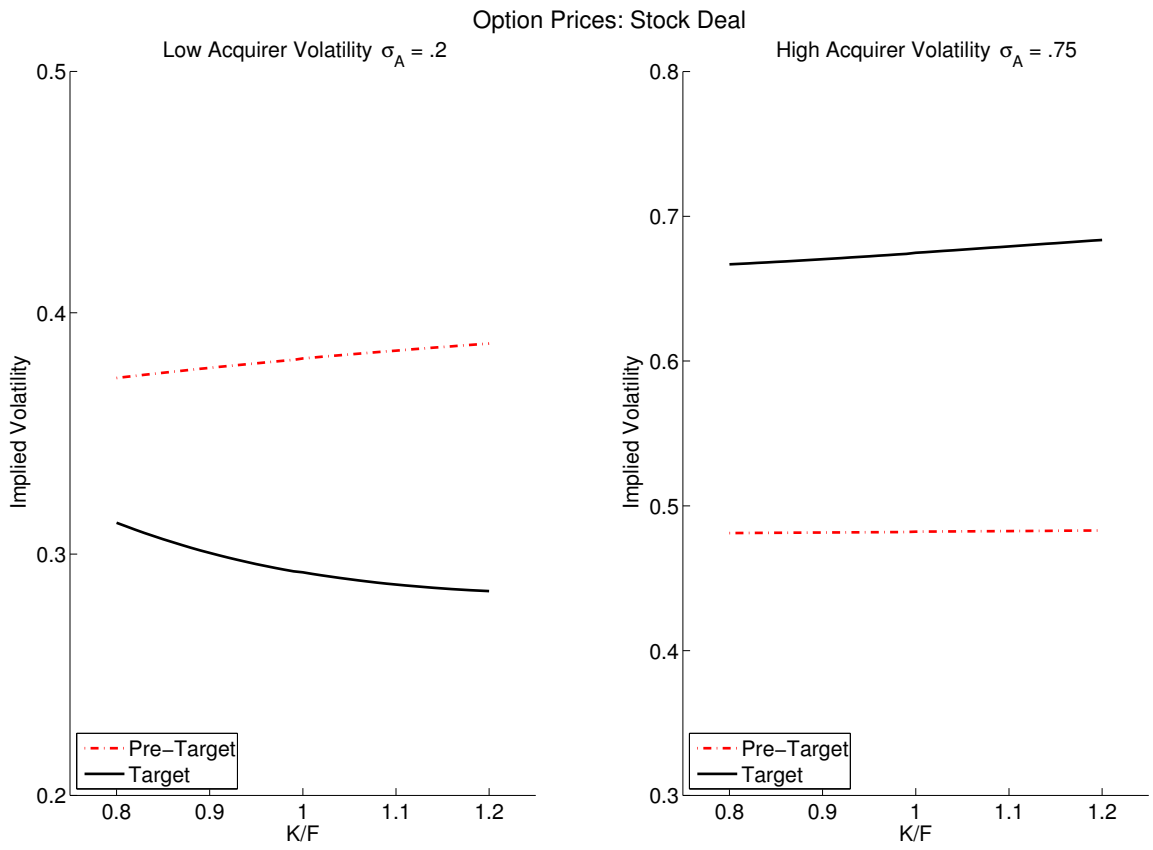


Figure 1.4: This figure plots the implied volatility smile in stock deals for which the acquirer is presumed to have either low $\sigma_A = .2$ or high $\sigma_A = .75$ volatility relative to target dividends $\sigma = .5$. The maturity of the options is half a year $\tau = .5$. When the acquirer volatility is low the reaction of option prices to deal announcements is similar to cash deals. The average level of implied volatility decreases upon the announcement of a deal and an implied volatility smile appears. As acquirer volatility decreases to zero the volatility smile will resemble cash deals. In contrast, option prices have a different reaction to deal announcements when the acquirer volatility is high relative to the volatility of target dividends. In particular, implied volatility increases after the announcement and the volatility surface is upward sloping. The values of the parameters are $D_t = .75$, $\phi P_{A,t} = 50$, and $(p_Q, \tau_Q, \lambda, r, g, \sigma, q_A, \rho) = (.7, .33, .01, .03, .01, .5, .02, .35)$.

Model Merger Arbitrage Risk Premia

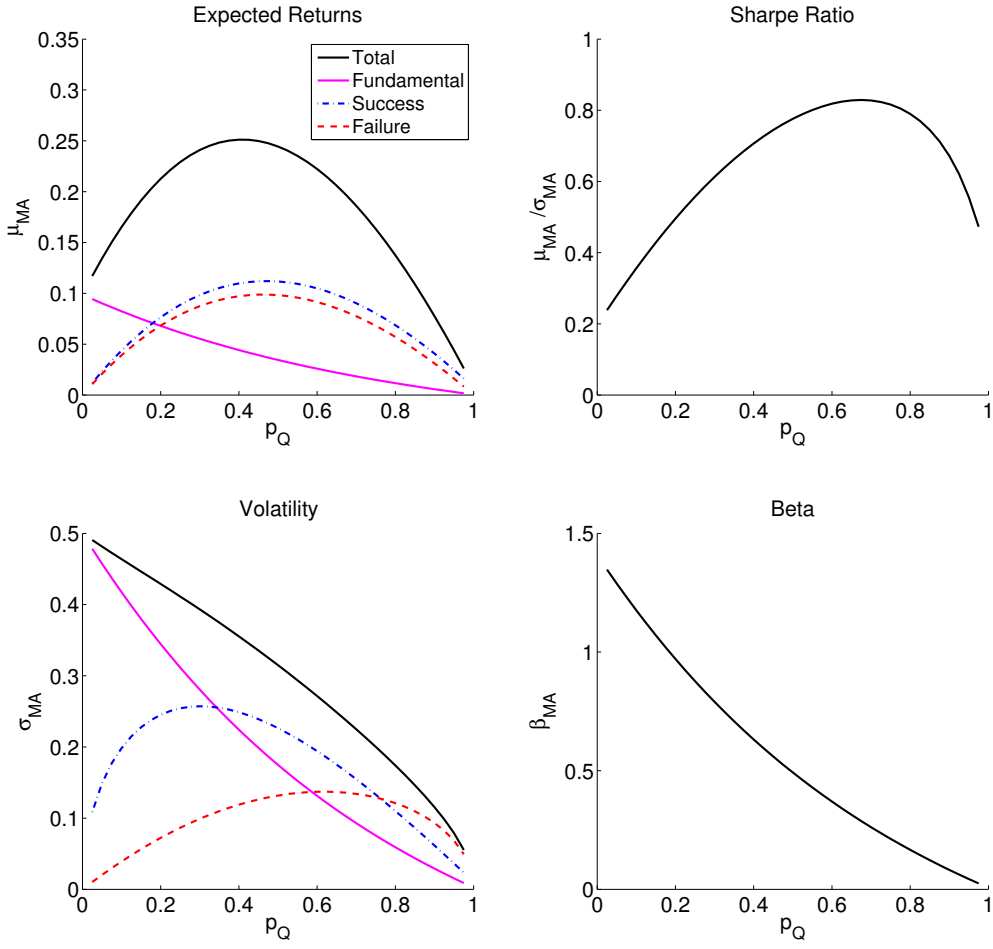


Figure 1.5: The instantaneous expected excess return, volatility, CAPM beta, and Sharpe ratio for a merger arbitrage investment in a cash deal are plotted against the risk neutral probability of deal success p_Q. The risk premium is decomposed into components that compensate investors for fundamental risk, deal success risk, and deal failure risk. The Sharpe ratio is not monotone and is maximized at a high risk neutral probability of success. Volatility and CAPM beta are decreasing functions of the probability of success. Figure 1.14 plots the empirical analog in the cross section of merger arbitrage portfolios sorted on the share implied probability of success. The parameter values are $D_t = .75$, $V = 50$, and $(\tau_Q, \lambda, r, g, \sigma) = (.33, .01, .03, .01, .5)$. The market prices of risk equal $\kappa_S = -.4$, $\kappa_F = .6$, $\eta = .35$. Similar results obtain for stock deals. The primary difference between cash and stock deals is that stock deals have a lower CAPM betas. This occurs in the model because target dividends are correlated with acquirer stock prices. Thus, being long the target and short the acquirer decreases fundamental risk

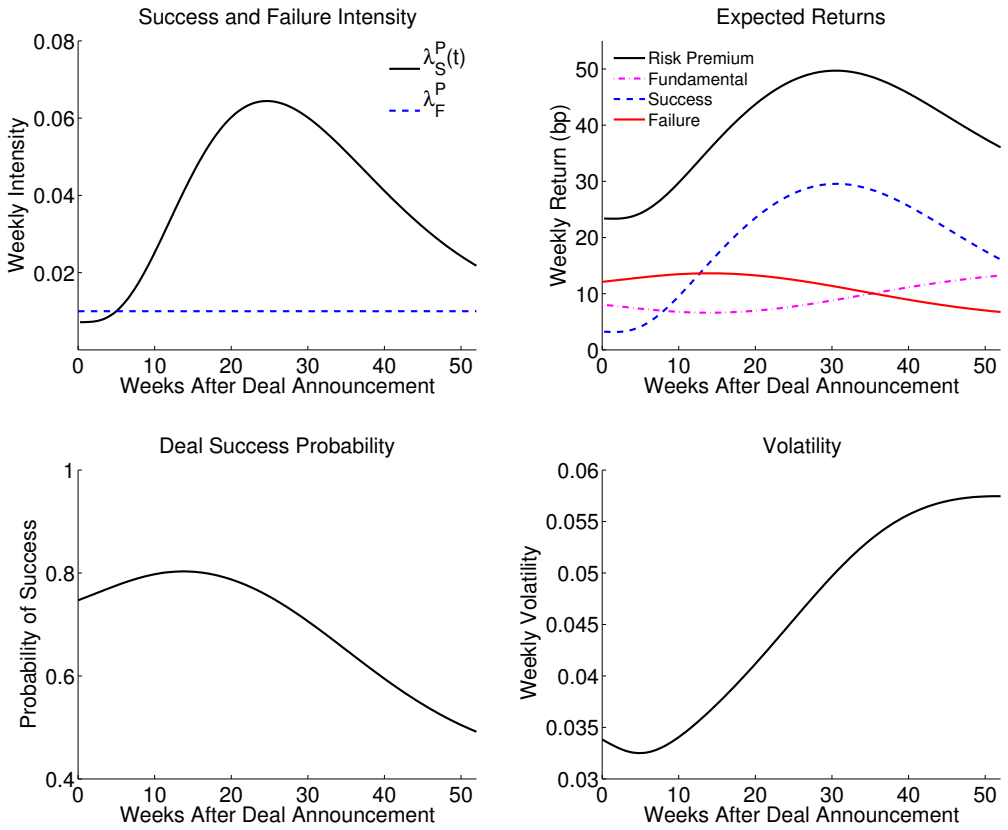


Figure 1.6: This figure illustrates expected merger arbitrage returns and volatility in event time. The analysis follows from an extension of the baseline model in Appendix 1.B that allows deal success intensity to be a deterministic function of event time $\lambda_S(t)$. To incorporate the empirical observation that the deal success intensity is hump shaped, I calibrate $\lambda_S(t) = \lambda_{S0} + (\lambda_{S1} - \lambda_{S0})\, f(t)$ to Figure 2 in Giglio and Shue (2014) where $f(\cdot)$ is the probability density function of a Gamma distribution with mean and standard deviation (μ_S, σ_S). The model predictions for expected returns and volatility are consistent with the data. Expected returns are hump shaped and volatility is increasing in event time. In the model, volatility initially declines but then begins to increase as the probability of deal success falls. Meanwhile, expected returns exhibit a hump shaped pattern because the probability of deal success reaches its peak prior to the peak in deal success intensity. When the deal success intensity reaches its peak, target stock prices are already declining which results in large expected returns in weeks 20 to 30. The parameters for the example are $(\lambda_{S0}, \lambda_{S1}, \mu_S, \sigma_S, \lambda_F, r, g, \sigma, \sigma_A, q_A, \rho, \kappa_S, \kappa_F, \eta) = (.25, 1.5, .6, .28, .85, .03, .01, .50, .25, .02, .35, -.4, .6, .35)$. Results are similar for cash and stock deals. The example above illustrates a hybrid deal with $V = 25$, $\theta = .5$, $P_{A,t} = 50$ and $D_t = .75$.

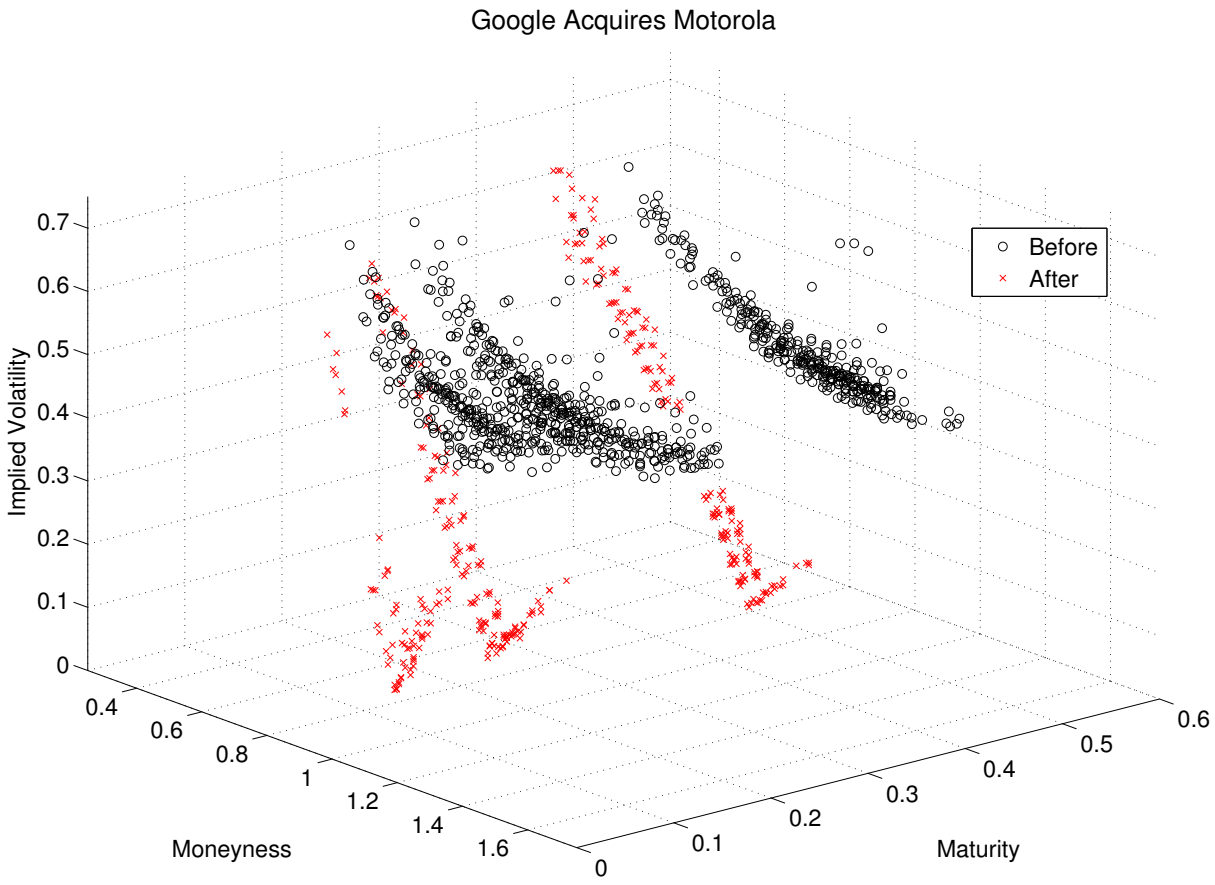


Figure 1.7: This plot illustrates the reaction of Motorola option prices to the announcement of a potential cash acquisition by Google on August 15, 2011 in a $12.45 billion deal valuing Motorola at $40 per share. The option data comes from closing prices of listed option contracts with expirations in September, October, and January. Moneyness is the strike-to-spot ratio and maturity is in years. Black-Scholes implied volatility is plotted during the pre-announcement and post-announcement periods.

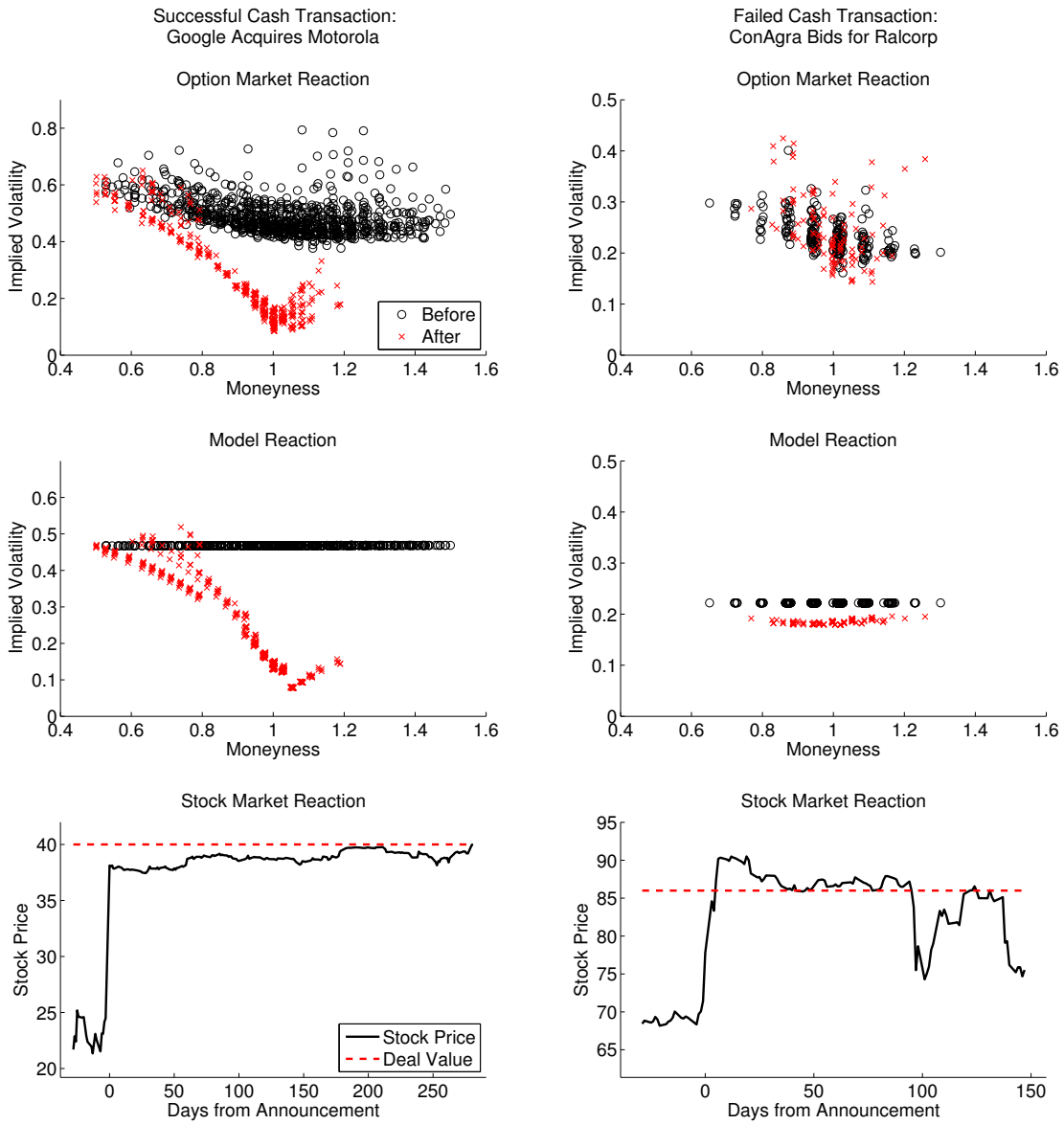


Figure 1.8: This figure plots the reaction of stock and option prices to the announcements of two cash deals. The top row plots the Black-Scholes implied volatility against moneyness for various maturities. Moneyness is defined as the strike-to-spot ratio. The middle plot illustrates the model prices at the estimated risk neutral parameters $\hat{\theta}$. For example, the risk neutral probability of success in Google's successful acquisition of Motorola is $\hat{p}_Q = .81$. This contrasts ConAgra's failed bid for Ralcorp in which $\hat{p}_Q = .20$. The last row plots the target stock price versus the offer throughout the life of the deal. While the target stock price reacts positively to both deal announcements, the lack of a reaction in target option prices for the ConAgra bid leads to the low estimate for the risk neutral probability of success

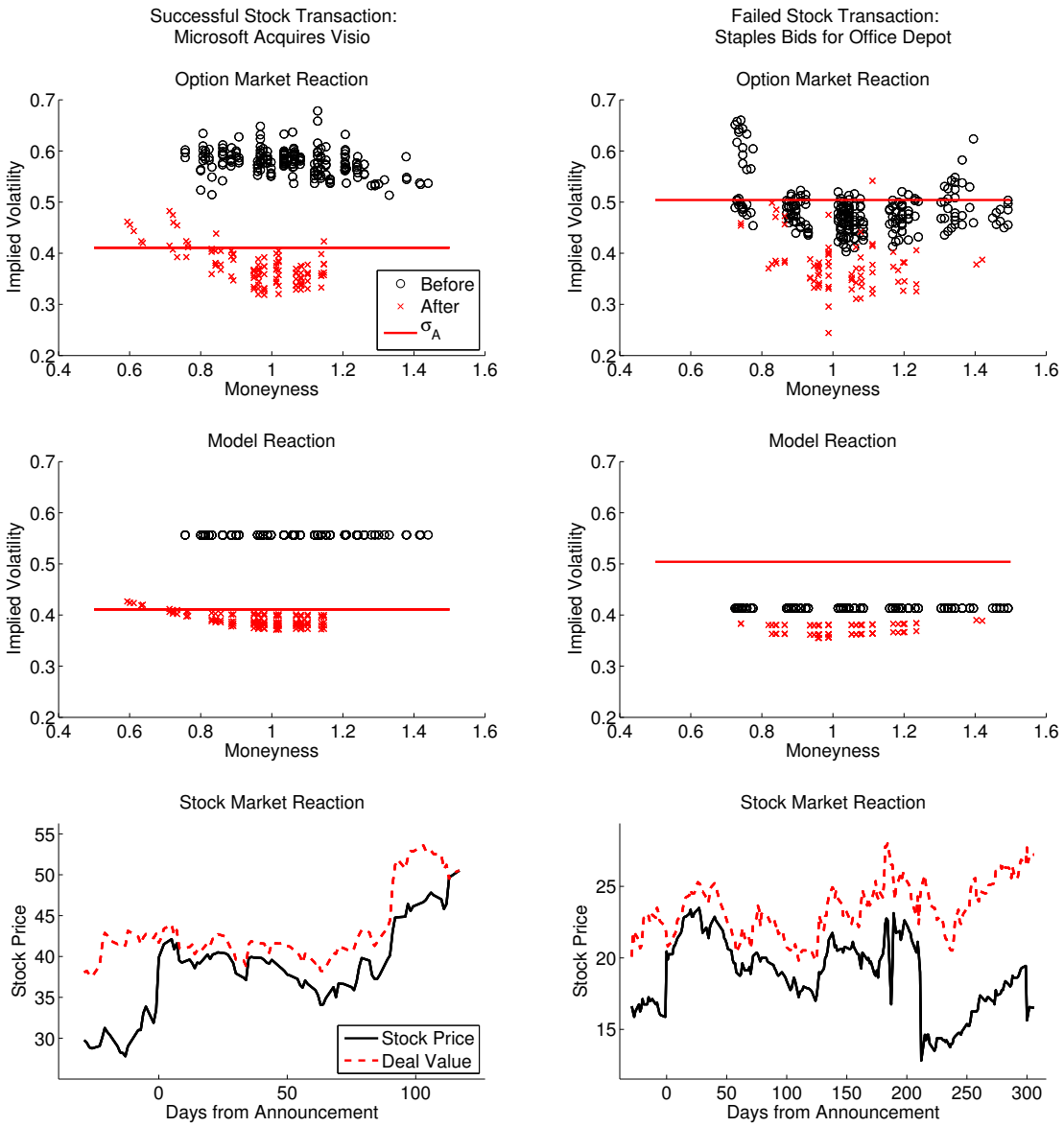


Figure 1.9: This figure plots the reaction of stock and option prices to the announcements of two stock deals. Similar to the previous plot, the top row plots the Black-Scholes implied volatility against moneyness for various maturities. Moneyness is defined as the strike-to-spot ratio. The middle plot illustrates the model prices at the estimated risk neutral parameters $\hat{\theta}$. The estimate of the risk neutral probability of success in Microsoft's successful acquisition of Visio is $\hat{p}_Q = .73$. In Staples failed bid for Office Depot $\hat{p}_Q = .30$. The last row plots the target stock price versus the offer $\phi P_{A,t}$ throughout the life of the deal. In both cases the stock market reacts positively to the announcement of the deal resulting in large share implied probabilities of success.

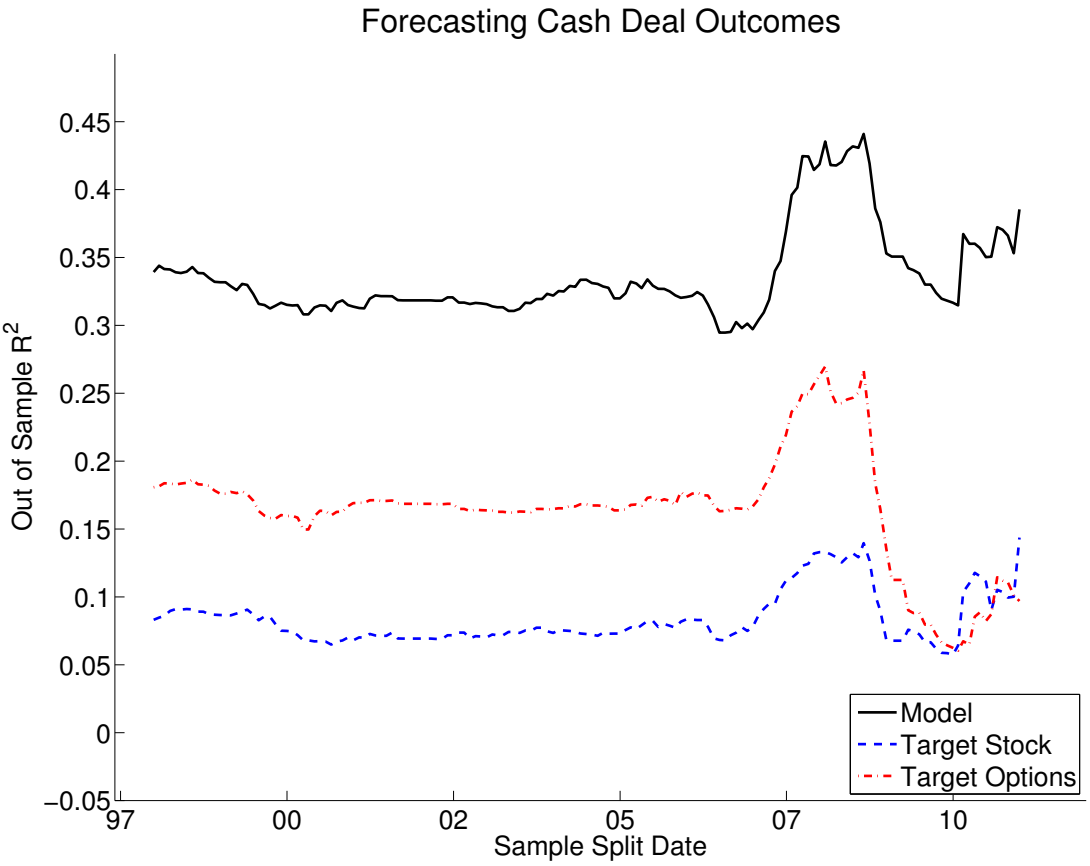


Figure 1.10: This figure plots out-of-sample R^2 for forecasting cash deal outcomes using different sample splits. The regression model is $Y_i = \beta' x_i + \epsilon_i$ where Y_i is the deal outcome and x_i represents the forecasting variables. Model refers to $x_i = p_{Q,i}$ the estimated risk neutral probability of deal success from the structural model. Target stock refers to $x_i = p^S_{Q,i}$ the share implied probability of success. Target options refers to the reduced form model constructed from target option prices. The reduced form variables include the risk neutral variance and skewness after the announcement and the change in the risk neutral variance and skewness when the deal is announced which are computed following Bakshi et al. (2003). Sample split date refers to the initial estimation period. In each month after the initial estimation period, I estimate the model and combine $\hat{\beta}$ with x_i to compute out-of-sample forecasts for the deals announced in the following month. The plot indicates the structural model, which relies on the joint information in target stock and option prices, outperforms the share implied probability and the reduced form model.

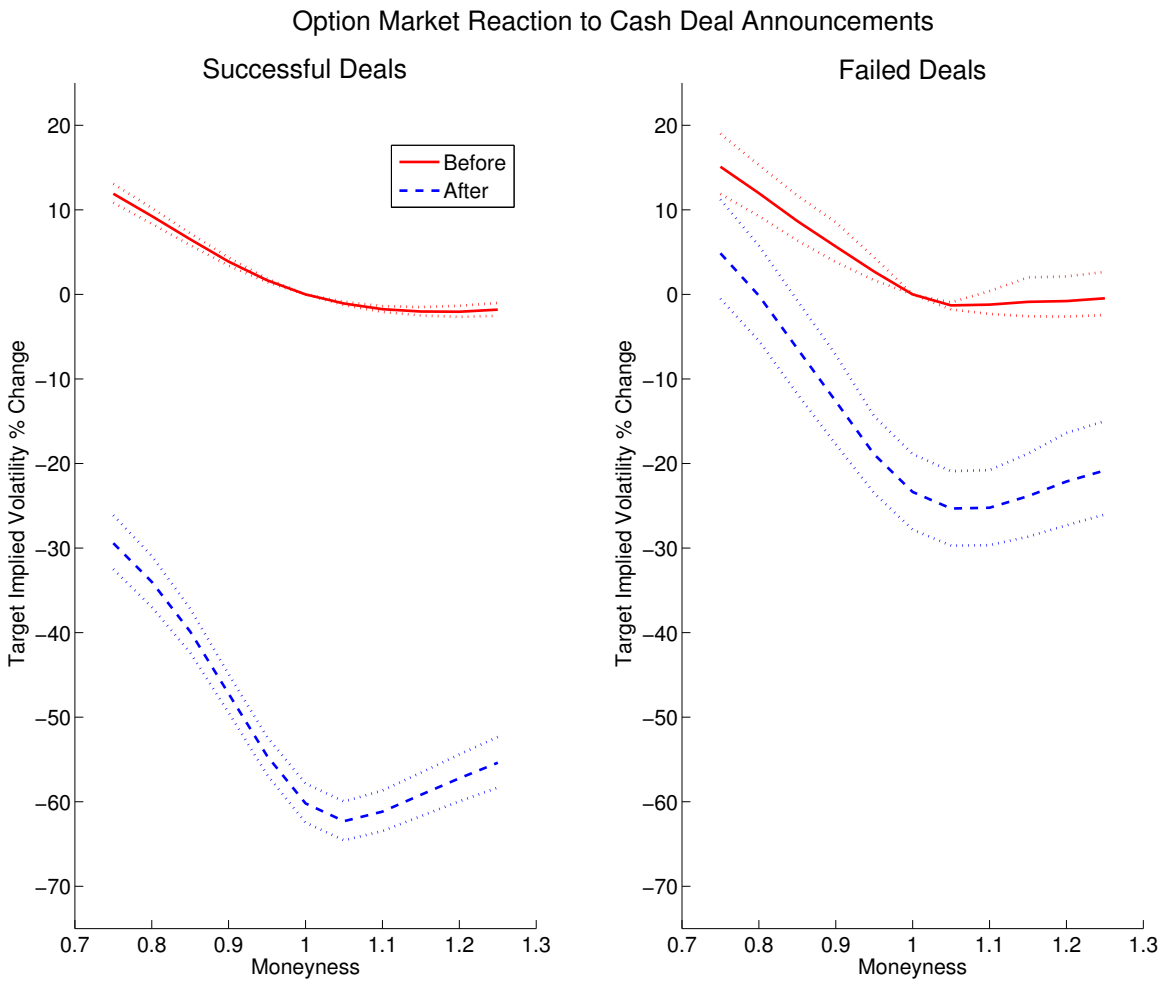


Figure 1.11: This figure plots the reaction of target option prices to cash deals. The y-axis is the percentage change in the fitted four month implied volatility relative to the pre-announcement at-the-money implied volatility. For example, the at-the-money implied volatility before the announcement has a percentage change of zero. For each deal, I obtain the fitted implied volatility by estimating the regression $\sigma^2 = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 \tau + \epsilon$ in the pre-announcement and post-announcement windows where x is log moneyness, defined as the strike to spot ratio, and τ is maturity. The figure illustrates the average reaction of target options across deals alongside bootstrapped 95% confidence intervals. The salient feature of the data is that implied volatility decreases and a more pronounced volatility smile emerges when cash deals are announced. Moreover, the decrease in implied volatility is larger in deals that are successful. The sample includes 820 cash deals between 1996-2012 of which 637 were successful and 183 failed. The larger confidence intervals for the failed subset reflects the smaller sample size. To minimize the impact of outliers or misspecification of the implied volatility function, I bounded the fitted values of implied volatility at the 5th and 95th quantiles of observed implied volatility in each regression.

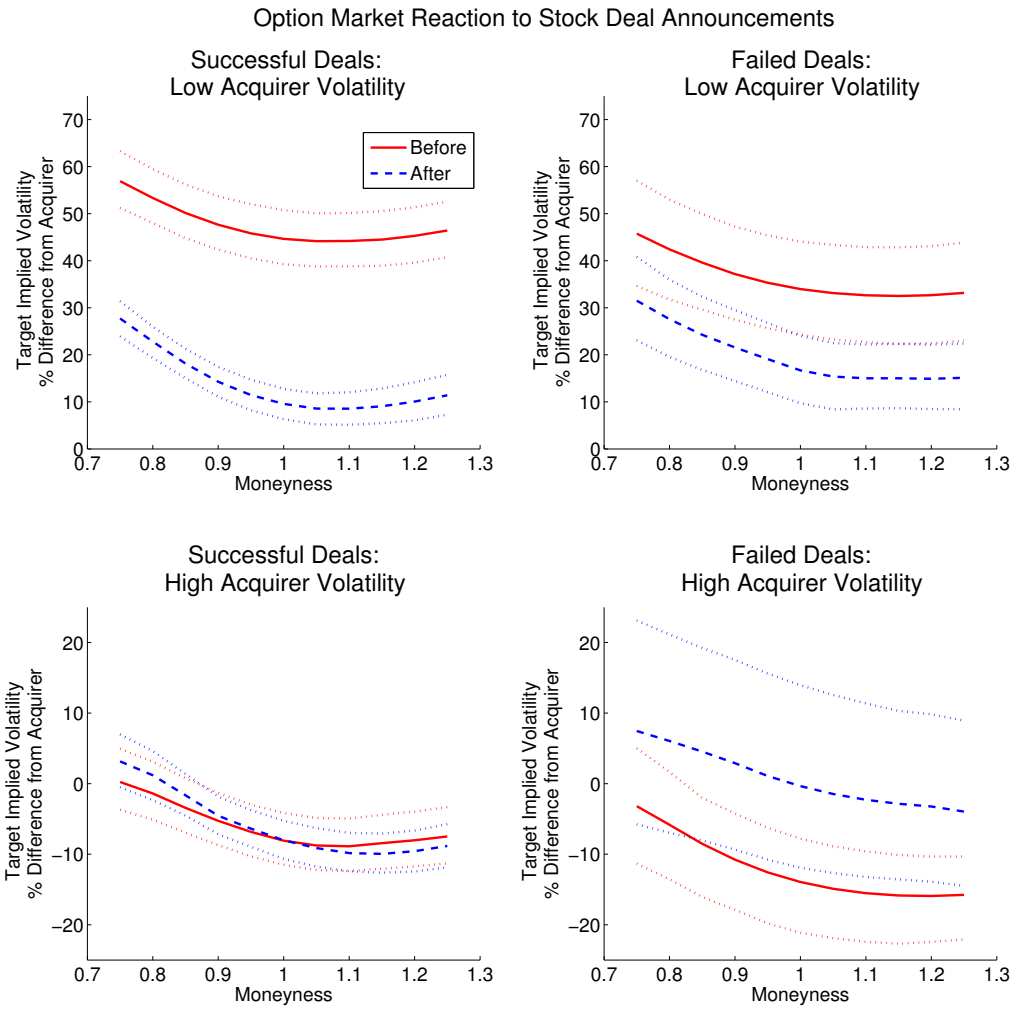


Figure 1.12: This figure plots the reaction of target option prices to stock deals. The y-axis is the percentage difference between the fitted four month implied volatility of target options and the pre-announcement at-the-money implied volatility of acquirer options. I subset the deals depending on whether the pre-announcement risk neutral variance of the target is larger or smaller than the risk neutral variance of the acquirer. The risk neutral moments are computed by following Bakshi et al. (2003). For each deal, I obtain the fitted implied volatility by estimating the regression $\sigma^2 = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 \tau + \epsilon$ in the pre-announcement and post-announcement windows where x is log moneyness, defined as the strike to spot ratio, and τ is maturity. The figure illustrates the average reaction of target options across deals alongside bootstrapped 95% confidence intervals. The plots highlight the different reactions when the acquirer has low volatility or high volatility relative to the target, which is consistent with the model as illustrated in Figure 1.4. The sample period is 1996-2012 including 272 low acquirer volatility deals of which 232 were successful and 40 failed in addition to 114 high acquirer volatility deals of which 95 were successful and 19 failed.

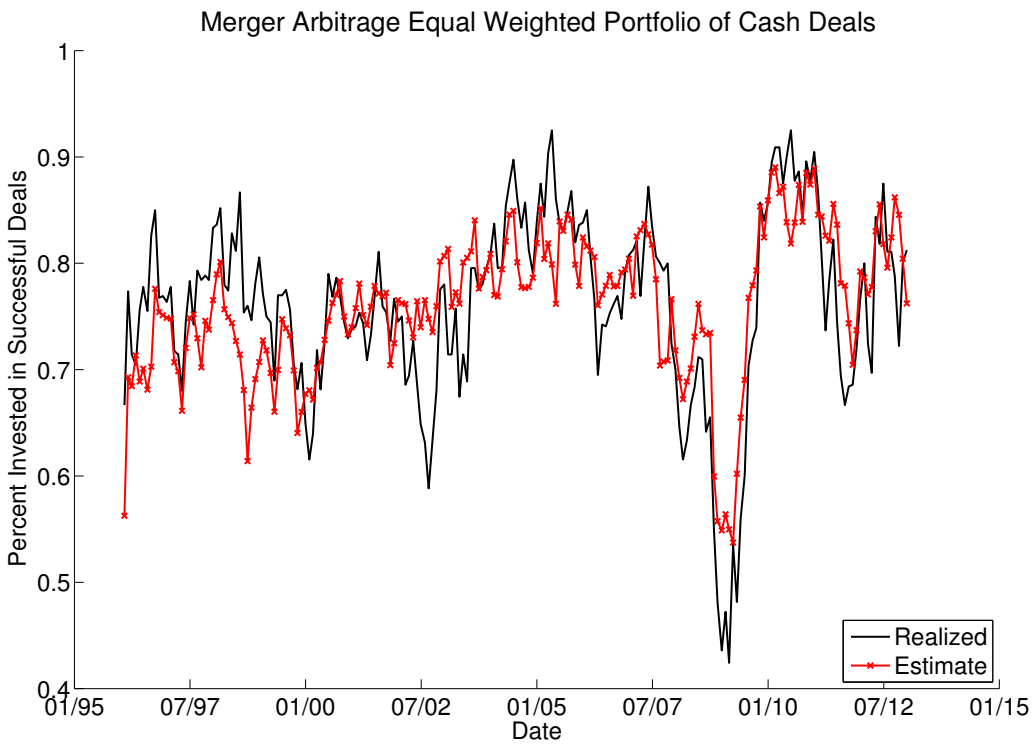


Figure 1.13: The monthly percentage of an equal weighted merger arbitrage portfolio invested in successful deals is displayed alongside the fitted value from the regression $Y_t = \beta_0 + \beta_1 \bar{p}^S_{Q,t} + \epsilon_t$ where Y_t is the percentage of the portfolio invested in successful deals and $\bar{p}^S_{Q,t}$ is the average share implied probability of success each month for all of the deals in the portfolio. The correlation of the time series is 76% and the parameter estimates are $\hat{\beta}_0 = .29$ (4.41), $\hat{\beta}_1 = .66$ (7.39), $R^2_{adj} = .58$. The t-statistics are computed with Newey and West (1987) standard errors with five lags. For each deal I compute the share implied probability of success each month $p^S_{Q,it}$ from the target stock price $P_{T,it}$, the offer $V_i + \phi_i P_{A,it}$, and the fall back price $P_{F,i}$ defined as the average target stock price over the window $[-25, -5)$ days prior to the deal announcement following Brown and Raymond (1986). The target stock price and offer are observed the last trading day of the previous month for deals already in the portfolio. In the month deals are added to the portfolio the target stock price and offer are observed two days after the deal announcement. The average number of deals in the portfolio each month is 56.

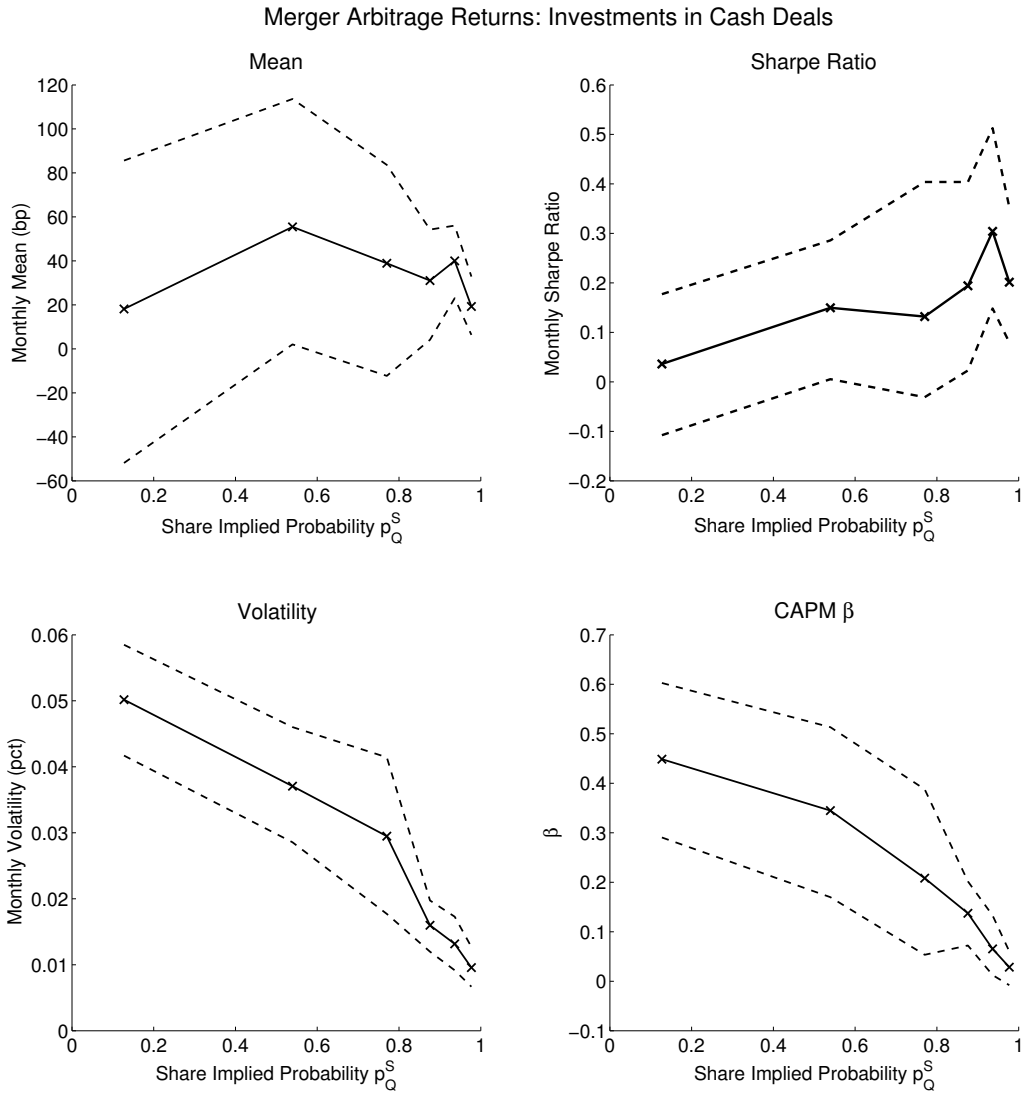


Figure 1.14: This figure plots the mean, volatility, Sharpe ratio, and CAPM beta of monthly merger arbitrage returns for portfolios that are invested in cash deals and sorted on the share implied probability of success p_Q^S. Figure 1.5 presents the analogous moments in the model. The x-axis denotes the average share implied probability of success over the sample period from 1996-2012 ($T = 204$). Dashed lines provide 95% confidence intervals that are block bootstrapped to account for autocorrelation in returns. Each portfolio has an average of 9.5 deals per month. The plot is consistent with the model predictions that volatility and beta are decreasing functions of success probability. Additionally, the plot does not reject the predictions that merger arbitrage Sharpe ratios and expected returns are hump shaped in the success probability.

APPENDIX

1.A Proofs and Derivations

Proposition 1:

As dividends are specified under an equivalent martingale measure $\mathbb{Q}$ prices may be computed by solving the risk neutral pricing partial differential equation. A helpful guess is that prices follow a modified Gordon growth formula

$$P(D, P_A, s) = \begin{cases} A_0 D + A_1 V + A_2 P_A & s_t \in \mathcal{P} \\ B_0 D + B_1 V + B_2 P_A & s_t \in \mathcal{T} \end{cases} \tag{1.26}$$

The pricing pdes in the pre-target and target states are then

$$\begin{aligned} D + gA_0 D + (r - q_A) A_2 P_A + \lambda \left[P(\mathcal{T}) - P(\mathcal{P}) \right] &= rP(\mathcal{P}) \\ D + gB_0 D + (r - q_A) B_2 P_A + \lambda_S \left[P(\mathcal{A}) - P(\mathcal{T}) \right] + \lambda_F \left[P(\mathcal{P}) - P(\mathcal{T}) \right] &= rP(\mathcal{T}) \end{aligned} \tag{1.27}$$

where $P(\mathcal{A}) = V + \phi P_A$. Matching coefficients on (D, V, P_A) results in six equations in six unknowns whose solution is

$$\begin{aligned} A_0 &= \tfrac{1+\lambda B_0}{r-g+\lambda} & B_0 &= \tfrac{1+\lambda_F/(r-g+\lambda)}{\lambda_S+\lambda_F+r-g-\frac{\lambda\lambda_F}{r-g+\lambda}} \\ A_1 &= \tfrac{\lambda B_1}{r+\lambda} & B_1 &= \tfrac{\lambda_S}{\lambda_S+\lambda_F+r-\frac{\lambda\lambda_F}{r+\lambda}} \\ A_2 &= \tfrac{\lambda B_2}{q_A+\lambda} & B_2 &= \tfrac{\lambda_S \phi}{\lambda_S+\lambda_F+q_A-\frac{\lambda\lambda_F}{q_A+\lambda}} \end{aligned} \tag{1.28}$$

Note A_0 and B_0 are non-negative if $r > g$ and the price in the acquired state satisfies $P(\mathcal{A}) = V + \phi P_A$. ■

Corollary 1:

From the proof of Proposition 1 it is immediate prices converge to the stated amounts as the deal announcement intensity decreases to zero. To see this also equals the discounted expected payoff in the target state when $\lambda = 0$ let $\tau = \min\{N_S, N_F\} \sim \exp(\lambda_S + \lambda_F)$ be the time of failure or success and denote the probability density function of τ as f_τ. Conditioned on τ the probability of success is $\lambda_S/(\lambda_S + \lambda_F)$. Define X_τ as the payoff when the deal fails or succeeds. By Fubini's theorem it follows that

$$\begin{aligned} P_0 &= E_0^{\mathbb{Q}}\left[\int_0^\tau e^{-ru} D_u du + e^{-r\tau} X_\tau\right] \\ &= \int_0^\infty \Big[\int_0^t e^{-(r-g)u} D_0 du + \\ &\qquad \tfrac{\lambda_S}{\lambda_S+\lambda_F}\left[V e^{-rt} + \phi P_{A0} e^{-q_A t}\right] + \tfrac{\lambda_F}{\lambda_S+\lambda_F}\tfrac{D_0}{r-g} e^{-(r-g)t}\Big] f_\tau(t) dt \\ &= \underbrace{\frac{D_0}{\lambda_S + \lambda_F + r - g}}_{NPV\ of\ Dividends\ Before\ \tau} + \\ &\qquad \underbrace{\frac{\lambda_S}{\lambda_S+\lambda_F+r} V + \frac{\lambda_S}{\lambda_S+\lambda_F+q_A}\phi P_{A0}}_{NPV\ of\ Success\ Payout} + \underbrace{\frac{\lambda_F}{\lambda_S+\lambda_F+r-g}\frac{D_0}{r-g}}_{NPV\ of\ Failure\ Payout} \end{aligned} \tag{1.29}$$

Rearranging one can verify the result matches the price given in the corollary. ■

Proposition 2:

For a continuous time-homogeneous Markov chain the transition probabilities equal $Q(t) = \exp(\Lambda t)$ which solves the Chapman-Kolmogorov forward and backward equations as discussed in Appendix C of Lando (2009). Alternatively, this can be derived by taking the limit of the discrete approximation. For horizon t discretize $[0, t]$ into n equal periods. The discrete approximation is $P_n = \left[I + \Lambda t \frac{1}{n}\right]^n$ which in the limit equals $\lim_{n\to\infty} P_n = \exp(\Lambda t) = \sum_{k=0}^\infty (\Lambda t)^k / k!$ as discussed in Jarrow et al. (1997). To compute the transition probabilities in closed form diagonalize the generator matrix into $\Lambda = PVP^{-1}$ and it follows that

$$Q(t) = I + \sum_{k=1}^{\infty} \frac{\Lambda^k t^k}{k!} = P \exp(Vt) P^{-1} \tag{1.30}$$

The eigenvalues of Λ are (v_1, v_2) reported in the proposition and $v_3 = 0$. The eigenvectors are

$$P = \begin{bmatrix} 1 & 1 & 1 \\ \frac{v_1}{\lambda} + 1 & \frac{v_2}{\lambda} + 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \tag{1.31}$$

Computing the inverse of P and simplifying leads to the result in the proposition. ■

Proposition 3:

At time maturity the firm is in either the pre-target, target, or acquired state. Employing the assumption firm transitions are independent of $(W_T, W_{A,T})$ the price of an option can be broken into three terms equal to the risk neutral probability of being in a particular state at maturity multiplied by the expected discounted payoff conditioned on the state. Let $t = 0$ and denote the maturity as T. For example, in the target state

$$E_0^{\mathbb{Q}}\left[e^{-rT}\left(P_T - K\right)^+ | s_0 \in \mathcal{T}\right] = q_{21}(T)F_{\mathcal{P}} + q_{22}(T)F_{\mathcal{T}} + q_{23}(T)F_{\mathcal{A}} \tag{1.32}$$

where

$$\begin{aligned} F_{\mathcal{P}} &= E_0^{\mathbb{Q}}\left[e^{-rT}\left(K - A_0 D_T - A_1 V - A_2 P_{A,T}\right)^+\right] \\ F_{\mathcal{T}} &= E_0^{\mathbb{Q}}\left[e^{-rT}\left(K - B_0 D_T - B_1 V - B_2 P_{A,T}\right)^+\right] \\ F_{\mathcal{A}} &= E_0^{\mathbb{Q}}\left[e^{-rT}\left(K - V - \theta P_{A,T}\right)^+\right] \end{aligned} \tag{1.33}$$

The result follows from rewriting the expectations above using the Black-Scholes formula. Define

$$\begin{aligned} q(x) &\equiv r - g - \tfrac{\sigma\rho x}{\sqrt{T}} + \tfrac{1}{2}\sigma^2\rho^2 \\ K_A(x) &\equiv K - A_1 V - A_2 P_{A0} e^{(r - q_A - \frac{1}{2}\sigma_A^2)T + \sigma_A\sqrt{T}x} \\ x_A^* &\equiv \frac{\ln\left(\frac{K - A_1 V}{A_2 P_A}\right) - \left(r - q_A - \frac{1}{2}\sigma_A^2\right)T}{\sigma_A\sqrt{T}} \end{aligned} \tag{1.34}$$

The definition for $K_B(x)$, x_B^* simply replace the coefficients (A_0, A_1) with (B_0, B_1) . To illustrate the idea I will derive $F_{\mathcal{P}}$ below. Note that $y|x \sim N(\rho x, (1 - \rho^2))$ and $x \sim N(0, 1)$. The change of

variables in the third line is $z = (y - \rho x)/\sqrt{1-\rho^2}$.

$$\begin{aligned} F_{\mathcal{P}} &= E_0^{\mathbb{Q}}\left[e^{-rT}\left(K - A_0 D_T - A_1 V - A_2 P_{A,T}\right)^+\right] \\ &= \int_{-\infty}^{x_A^*}\int e^{-rT}\left[K_A(x) - A_0 D e^{\left(g-\frac{1}{2}\sigma^2\right)T+\sigma\sqrt{T}y}\right]^+ f(y|x)f(x)dydx \\ &= \int_{-\infty}^{x_A^*}\int e^{-rT}\left[K_A(x) - A_0 D e^{\left(g-\frac{1}{2}\sigma^2\right)T+\sigma\sqrt{T}\left(\rho x+\sqrt{1-\rho^2}z\right)}\right]^+ f(z)dzf(x)dx \\ &= \int_{-\infty}^{x_A^*} P_{BS}\left(A_0 D, K_A(x), T, r, q(x), \sigma\sqrt{1-\rho^2}\right) f(x)dx \end{aligned} \tag{1.35}$$

The method for computing $F_{\mathcal{T}}$ and $F_{\mathcal{A}}$ is analogous. Call option prices may be computed directly as above or by put-call parity. ■

Proposition 4

To compute the instantaneous merger arbitrage risk premium apply Ito's lemma

$$\begin{aligned} \mu_t &= E_t^{\mathbb{P}}\left[\frac{dP_t + D_t dt - \phi\left[dP_{A,t} + D_{A,t}dt - rP_{A,t}dt\right]}{P_t}\frac{1}{dt} - r\right] \\ &= \frac{\left[\beta_D B_0 D + \beta_A(B_2-\phi)P_A\right](\mu_m - r) + \left(\lambda_S^{\mathbb{P}} - \lambda_S\right)\left(V + \phi P_{A,t} - P_t(\mathcal{T})\right) + \left(\lambda_F^{\mathbb{P}} - \lambda_F\right)\left(P_t(\mathcal{P}) - P_t(\mathcal{T})\right)}{P_t(\mathcal{T})} \\ &\quad + \frac{D_t + B_0 g D_t + B_2(r - q_A)P_{A,t} + \lambda_S\left(V + \phi P_{A,t} - P_t(\mathcal{T})\right) + \lambda_F\left(P_t(\mathcal{P}) - P_t(\mathcal{T})\right)}{P_t(\mathcal{T})} - r \\ &= \frac{\beta_D B_0 D + \beta_A(B_2-\phi)P_A}{P_t(\mathcal{T})}(\mu_m - r) \\ &\quad + \left(e^{-\kappa_F} - 1\right)\lambda_F\left(\frac{P_t(\mathcal{P})}{P_t(\mathcal{T})} - 1\right) \\ &\quad + \left(e^{-\kappa_S} - 1\right)\lambda_S\left(\frac{V + \phi P_{A,t}}{P_t(\mathcal{T})} - 1\right) \end{aligned} \tag{1.36}$$

The simplification from the second to the third line follows from recognizing the pricing pde in the target state and setting it equal to zero. The market return is assumed to follow $\frac{dP_m}{P_m} = (\mu_m - r)\,dt + \sigma_m dW_{m,t}^{\mathbb{P}}$. Hence,

$$\begin{aligned} \beta_{T,t} &= \frac{cov_t\left(dR_{MA}^e, \frac{dP_m}{P_m}\right)}{cov_t\left(\frac{dP_m}{P_m}, \frac{dP_m}{P_m}\right)} \\ &= \frac{\frac{\sigma\rho_m B_0 D_t + \sigma_A \rho_m (B_2-\phi)P_{A,t}}{P_t(\mathcal{T})}}{\sigma_m^2}\sigma_m \\ &= \frac{\frac{\sigma\rho_m}{\sigma_m}B_0 D_t + \frac{\sigma_A \rho_m}{\sigma_m}(B_2-\phi)P_{A,t}}{P_t(\mathcal{T})} \\ &= \frac{\beta_D B_0 D_t + \beta_A(B_2-\phi)P_{A,t}}{P_t(\mathcal{T})} \end{aligned} \tag{1.37}$$

The derivation of $\sigma_{MA,t}^2$ follows via a similar application of Ito's lemma. ■

Stock Price Distribution

The cumulative distribution function of the target stock price at time t for horizon T may be computed by conditioning on the state of the firm at time T. In particular,

$$\begin{aligned} Pr^{\mathbb{Q}}\left(P_T \leq u|\mathcal{F}_t\right) &= q_{21}\left(\tau\right) Pr^{\mathbb{Q}}\left(P_T \leq u|D_t, P_{A,t}, s_T \in \mathcal{P}\right) \\ &+ q_{22}\left(\tau\right) Pr^{\mathbb{Q}}\left(P_T \leq u|D_t, P_{A,t}, s_T \in \mathcal{T}\right) \\ &+ q_{23}\left(\tau\right) Pr^{\mathbb{Q}}\left(P_T \leq u|D_t, P_{A,t}, s_T \in \mathcal{A}\right) \end{aligned} \tag{1.38}$$

which equals

$$\begin{aligned} Pr^{\mathbb{Q}}(P_T \leq u|\mathcal{F}_t) &= q_{21}(\tau) \cdot \int \Phi\left(\frac{\tilde{u}_A(x)-\rho x}{\sqrt{1-\rho^2}}\right)\phi(x)dx \\ &+ q_{22}(\tau) \cdot \int \Phi\left(\frac{\tilde{u}_B(x)-\rho x}{\sqrt{1-\rho^2}}\right)\phi(x)dx \\ &+ q_{23}(\tau) \cdot [\mathbf{1}[\phi = 0] \cdot \mathbf{1}[V \leq u] + \\ &\quad \mathbf{1}[\phi > 0] \cdot \Phi\left(\frac{\ln\left(\frac{u-V}{\phi P_A}\right)-\left(r-q_A-\frac{1}{2}\sigma_A^2\right)\tau}{\sigma_A\sqrt{\tau}}\right)\Bigg] \end{aligned} \tag{1.39}$$

where I define

$$\tilde{u}_A(x) \equiv \frac{\ln\left(\frac{u-A_1V-A_2P_{A,t}e^{(r-q_A-\frac{1}{2}\sigma_A^2)\tau+\sigma_A\sqrt{\tau}x}}{A_0D_t}\right)-\left(g-\frac{1}{2}\sigma^2\right)\tau}{\sigma\sqrt{\tau}} \tag{1.40}$$

The definition of $\tilde{u}_B(x)$ is analogous replacing the coefficients (A_0, A_1, A_2) with (B_0, B_1, B_2). The distribution in the pre-target state is obtained by replacing the transition probabilities $(q_{21}(\tau), q_{22}(\tau), q_{23}(\tau))$ with $(q_{11}(\tau), q_{12}(\tau), q_{13}(\tau))$. For cash deals there is a discontinuity in the cumulative distribution function at the offer V. This results in a kink in option prices and the Black-Scholes implied volatility as the strike price approaches V. Bester et al. (2013) obtain a similar result analytically employing the assumption in their model that the fall back price is lognormal so Black-Scholes option pricing applies in the event of deal failure. The contribution here is to emphasize the fact that kink in option prices and implied volatility is model free. It occurs in option prices and Black-Scholes implied volatility because there is a discrete probability of deal success in which target shareholders receive a fixed amount. To see this define the implied volatility function as $\sigma(K)$ which solves $C(K) = C_{BS}(S, K, \tau, r, q, \sigma(K))$. Differentiating both sides of the equation with

respect to the strike and rearranging yields

$$\frac{\partial \sigma}{\partial K} = \frac{\frac{\partial C}{\partial K} - \frac{\partial C_{BS}}{\partial K}}{\frac{\partial C_{BS}}{\partial \sigma}} \tag{1.41}$$

From the formula $C(K) = \int_K^\infty e^{-r\tau}\,(P-K)\,p_t^{\mathbb{Q}}(P)dP$ it follows that

$$\frac{\partial C}{\partial K} = -e^{-r\tau}\left(1 - Pr_t^{\mathbb{Q}}\left(P_T \leq K\right)\right) \tag{1.42}$$

Thus, if there is a discontinuity in the cumulative distribution function at the point x there will be a kink in implied volatility whose magnitude is proportional to $g(x) \equiv \lim_{K\downarrow x} Pr_t^{\mathbb{Q}}(P_T \leq K) - \lim_{K\uparrow x} Pr_t^{\mathbb{Q}}(P_T \leq K)$. For cash deals in the target state my model results in the kink equal to

$$e^{r\tau}\left(\frac{\partial \sigma}{\partial K}|_{K\downarrow V} - \frac{\partial \sigma}{\partial K}|_{K\uparrow V}\right) = \frac{g(V)}{\frac{\partial C_{BS}}{\partial \sigma}} = \frac{q_{23}(\tau)}{\frac{\partial C_{BS}}{\partial \sigma}} \tag{1.43}$$

1.B Model Extensions and Discussion

Below I present extensions of the model that allow for merger waves, time varying deal success intensity, and deal failures that are correlated with the target dividend process.

Deterministic Success Intensity

I allow for deterministic success intensity to incorporate the observation in Giglio and Shue (2014) that success intensity is hump shaped in event time. Prices continue to follow a modified Gordon growth formula as in the baseline model. The key difference is that event time becomes a state variable. To be specific, let the generator equal

$$\Lambda(t) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ \lambda_N & \lambda_2 & \lambda & 0 \\ \lambda_F & 0 & \lambda_3\left(\tilde{t}\right) & \lambda_S\left(\tilde{t}\right) \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{1.44}$$

where $\lambda_S(\cdot)$ is a deterministic function of time and $\tilde{t} \equiv (t-\tau)^+$ where τ is the transition time from the pre-target to the target state. This generator introduces two differences from the baseline model outlined in equation 1.3. First, if a transition to the target state occurs the success intensity evolves deterministically according to $\lambda_S(t-\tau)$. This complicates analysis as the success intensity is no longer a constant. At the same time, the simplifying assumption has been made that a firm in the target state may not return to the pre-target state. This facilitates the computation of transition probabilities and allows us to obtain prices via backward induction. For example, the risk neutral probability of remaining in the target state after a length of time τ equals $q_{\mathcal{T},\tau} = e^{-\lambda_F \tau - \int_0^\tau \lambda_S(u)du}$. Prices in this model are outlined in the proposition below.

Proposition 5 ***Target Stock Price with Deterministic Success Intensity:*** *Consider a firm whose dividend process follows equation 1.1 that transitions across deal states according to an inhomogeneous Markov chain with generator $\Lambda(t)$ given by equation 1.44 where λ_S is a continuous deterministic function of time satisfying $\lim_{t\to\infty} \lambda_S(t) = \lambda_{S,\infty}$. If $r-g>0$ the price is given below. The coefficients (A_0, A_1, A_2) and $(B_0(\tilde{t}), B_1(\tilde{t}), B_2(\tilde{t}))$ are provided in the proof.*

$$P(D_t, P_{A,t}, s_t, t) = \begin{cases} A_0 D_t + A_1 V + A_2 P_{A,t} & s_t \in \mathcal{P} \\ B_0(\tilde{t}) D_t + B_1(\tilde{t}) V + B_2(\tilde{t}) P_{A,t} & s_t \in \mathcal{T} \end{cases} \tag{1.45}$$

Proof of Proposition 5:

For notational convenience suppose $\tau = 0$. In the target state the value of the firm in the event of success equals $V + \theta P_{A,t}$ while the value of the firm in the event of failure equals $\frac{D_t}{r-g}$. As a result, the pricing pde is

$$\begin{aligned} &D + [B_0'(t)D + B_1'(t)V + B_2'(t)P_A] + B_0(t)gD \\ &\quad + B_2(t)(r-q_A)P_A + \lambda_S(t)[V + \phi P_A - P(\mathcal{T})] + \lambda_F\left[\tfrac{D}{r-g} - P(\mathcal{T})\right] = rP(\mathcal{T}). \end{aligned} \tag{1.46}$$

Matching coefficients on (D, V, P_A) as in Proposition 1 leads to the following ordinary differential equations

$$\begin{aligned} 1 + \tfrac{\lambda_F}{r-g} + B_0'(t) &= B_0(t)\left[\lambda_S(t) + \lambda_F + r - g\right] \\ \lambda_S(t) + B_1'(t) &= B_1(t)\left[\lambda_S(t) + \lambda_F + r\right] \\ \lambda_S(t)\phi + B_2'(t) &= B_2(t)\left[\lambda_S(t) + \lambda_F + q_A\right] \end{aligned} \tag{1.47}$$

These are first order linear differential equations whose general solution $f' + pf = g$ is known to be $f = \frac{\int e^{\int p} g + c}{e^{\int p}}$ given our assumption $\lambda_S(t)$ is continuous. The solution is

$$B_0(t) = \frac{\left(1 + \frac{\lambda_F}{r-g}\right)\left[\int_0^\infty e^{-\int_0^u \lambda_S(v)dv - (\lambda_F + r - g)u} du - \int_0^t e^{-\int_0^s \lambda_S(u)du - (\lambda_F + r - g)u} du\right]}{e^{-\int_0^t \lambda_S(u)du - (\lambda_F + r - g)t}} \tag{1.48}$$

which satisfies the boundary condition $\lim_{t\to\infty} B_0(t) = \frac{1 + \lambda_F/(r-g)}{\lambda_{S,\infty} + \lambda_F + r - g}$. Numerical solutions can be obtained by Runge Kutta. In the pre-target state prices satisfy

$$D + A_0 g D + A_2(r - q_A)P_A + \lambda\left[P(\mathcal{T}, 0) - P(\mathcal{P})\right] + \lambda_N\left[\frac{D}{r-g} - P(\mathcal{P})\right] = rP(\mathcal{P}) \tag{1.49}$$

where $P(\mathcal{T}, 0) = B_0(0)D + B_1(0)V + B_2(0)P_A$. Solving I find

$$\begin{aligned} A_0 &= \tfrac{1 + \lambda B_0(0) + \lambda_N/(r-g)}{\lambda + \lambda_N + r - g} \\ A_1 &= \tfrac{\lambda B_1(0)}{\lambda + \lambda_N + r} \\ A_2 &= \tfrac{\lambda B_2(0)}{\lambda + \lambda_N + r} \end{aligned} \tag{1.50}$$

■

Time Varying Announcement Intensity

The observation that mergers arrive in waves has been documented in Shleifer and Vishny (2003) and Harford (2005). Taking this fact as given one approach to incorporate merger waves in the model is to assume the announcement intensity contains a common and time varying component

across firms. In particular, suppose λ_t follows a Cox-Ingersoll-Ross square root process[26]

$$d\lambda_t = \kappa(\bar{\lambda} - \lambda_t)dt + \sigma_\lambda \sqrt{\lambda_t} dW_{\lambda,t} \tag{1.51}$$

where $E_t^{\mathbb{Q}}\left[dW_{\lambda,t}dW_t\right] = \rho_\lambda dt$ and $E_t^{\mathbb{Q}}\left[dW_{\lambda,t}dW_{A,t}\right] = \rho_\lambda dt$. Firm transitions evolve according to a time-inhomogeneous Markov chain whose generator under the risk neutral measure is

$$\Lambda_t = \begin{bmatrix} 0 & 0 & 0 & 0 \\ \lambda_N & \lambda_{2,t} & \lambda_T + \lambda_t & 0 \\ \lambda_F & 0 & \lambda_3 & \lambda_S \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{1.52}$$

The generator embeds two distinctions from the baseline model. Deal announcement intensity is no longer a constant. It equals a firm specific component λ_T plus the common and time varying component λ_t. Additionally, the target firm may never revert to being a pre-target. The latter assumption allows for prices to be computed by backward induction and facilitates the computation of transition probabilities, similar to the model with deterministic success intensity. For example, the risk neutral probability of staying in the pre-target state after a length of time τ may be computed as

$$\begin{aligned} q_{P,\tau} &= E_0^{\mathbb{Q}}\left[\int_\tau^\infty \int_\tau^\infty \lambda_N e^{-\lambda_N u}\left(\lambda_T + \lambda_v\right) e^{-\int_0^v (\lambda_T+\lambda_s)ds} du dv\right] \\ &= e^{-\lambda_N \tau} E_0^{\mathbb{Q}}\left[-e^{-\int_0^v (\lambda_T+\lambda_s)ds}|_\tau^\infty\right] \\ &= e^{-\lambda_N \tau} E_0^{\mathbb{Q}}\left[e^{-\int_0^\tau (\lambda_T+\lambda_s)ds}\right] \\ &= e^{-(\lambda_N+\lambda_T)\tau - A(\tau) - B(\tau)\lambda_0} \end{aligned} \tag{1.53}$$

where $A(\tau)$ and $B(\tau)$ are the coefficient functions for zero coupon bond prices in the CIR model with maturity τ. The remaining transition probabilities may be computed in a similar fashion. This calculation would not apply if firms could transition back and forth between the pre-target and target states, although it would provide a useful approximation for short lengths of time. The share price is derived below. By combining the transition probabilities with the share price for the case when $\rho_\lambda = 0$ option prices can be computed as in the baseline model.

[26] I assume the Feller condition $2\kappa\bar{\lambda} \geq \sigma_\lambda^2$ is satisfied so the intensity remains positive.

Proposition 6 ***Target Stock Price Incorporating Merger Waves:*** *Consider a firm whose dividend process follows equation 1.1 that transitions across deal states according to a Markov chain with generator $\Lambda(t)$ given by equation 1.52. If $r-g>0$ the price in the target state $s\in\mathcal{T}$ is given by*

$$P=\frac{D}{r-g}\frac{\lambda_F+r-g}{\lambda_S+\lambda_F+r-g}+V\frac{\lambda_S}{\lambda_S+\lambda_F+r}+\phi P_A\frac{\lambda_S}{\lambda_S+\lambda_F+q_A} \tag{1.54}$$

The price in the pre-target state $s\in\mathcal{P}$ is

$$\begin{aligned}P &= \tfrac{D}{r-g}\tfrac{\lambda_F+r-g}{\lambda_S+\lambda_F+r-g}\left[1+\lambda_S\tfrac{\lambda_N+r-g}{\lambda_F+r-g}\int_0^\infty e^{-A(u)-B(u)\lambda-(\lambda_T+\lambda_N+r-g)u}du\right]\\ &\quad +V\tfrac{\lambda_S}{\lambda_S+\lambda_F+r}\left[1-(\lambda_N+r)\int_0^\infty e^{-A(u)-B(u)\lambda-(\lambda_T+\lambda_N+r)u}du\right]\\ &\quad +\phi P_A\tfrac{\lambda_S}{\lambda_S+\lambda_F+q_A}\left[1-(\lambda_N+q_A)\int_0^\infty e^{-A(u)-B(u)\lambda-(\lambda_T+\lambda_N+q_A)u}du\right]\end{aligned} \tag{1.55}$$

when $\rho_\lambda=0$. In the event $\rho_\lambda\neq 0$ the price may be computed by solving the ordinary differential equations reported in the proposition.

Proof of Proposition 6:

Prices may be computed via backward induction. Consider the target state. In the event the current deal fails the firm is never the target of a subsequent takeover and the share is worth the discounted value of dividends under the risk neutral measure which is of course the familiar Gordon growth formula. If the deal succeeds the target share is exchanged for $V+\phi P_A$. Let $x=\log(D)$ and $x_A=\log(P_A)$ and define the operator

$$\begin{aligned}\mathcal{A}f=&\left[g-\tfrac{1}{2}\sigma^2\right]\tfrac{\partial f}{\partial x}+\tfrac{1}{2}\sigma^2\tfrac{\partial^2 f}{\partial x^2}+\left[r-q_A-\tfrac{1}{2}\sigma_A^2\right]\tfrac{\partial f}{\partial x_A}+\tfrac{1}{2}\sigma_A^2\tfrac{\partial^2 f}{\partial x_A^2}+\sigma\sigma_A\rho\tfrac{\partial^2 f}{\partial x\partial x_A}\\ &+\kappa\left[\bar{\lambda}-\lambda\right]\tfrac{\partial f}{\partial\lambda}+\tfrac{1}{2}\sigma_\lambda^2\lambda\tfrac{\partial^2 f}{\partial\lambda^2}+\sigma\sigma_\lambda\sqrt{\lambda}\rho_\lambda\tfrac{\partial^2 f}{\partial x\partial\lambda}+\sigma_A\sigma_\lambda\sqrt{\lambda}\rho_\lambda\tfrac{\partial^2 f}{\partial x_A\partial\lambda}\end{aligned} \tag{1.56}$$

By the Feynman-Kac theorem the price in the target state must satisfy

$$e^x+\mathcal{A}P(\mathcal{T})+\lambda_S\left[V+\phi e^{x_A}-P(\mathcal{T})\right]+\lambda_F\left[\frac{e^x}{r-g}-P(\mathcal{T})\right]=rP(\mathcal{T}) \tag{1.57}$$

One can verify the equation in the proposition satisfies the partial differential equation above and the transversality condition $\lim_{T\to\infty}E\left[e^{-rT}P\right]=0$ as long as $r>g$. Next consider the state when

the firm is a pre-target. In this case the relevant pricing partial differential equation is

$$e^x + \mathcal{A}P + (\lambda_T + \lambda)\left[b_0 e^x + b_1 V + b_2 e^{x_A} - P\right] + \lambda_N \left[\frac{e^x}{r-g} - P\right] = rP \tag{1.58}$$

where $b_0,\ b_1,\ b_2$ correspond to the coefficients in the target state just analyzed. With the guess $P = a_0(\lambda)e^x + a_1(\lambda)V + a_2(\lambda)e^{x_A}$ observe the partial differential equation separates into three ordinary differential equations,

$$\begin{aligned} 0 &= -\left(\lambda_T + \lambda + \lambda_N + r - g\right) a_0(\lambda) + \kappa(\bar{\lambda} - \lambda) a_0'(\lambda) \\ &\quad + \sigma\sigma_\lambda \sqrt{\lambda}\rho_\lambda a_0'(\lambda) + \tfrac{1}{2}\sigma_\lambda^2 \lambda a_0''(\lambda) + (\lambda_T + \lambda)\, b_0 + 1 + \tfrac{\lambda_N}{r-g} \end{aligned} \tag{1.59}$$

$$\begin{aligned} 0 &= -(\lambda_T + \lambda + \lambda_N + r) a_1(\lambda) + \kappa(\bar{\lambda} - \lambda) a_1'(\lambda) \\ &\quad + \tfrac{1}{2}\sigma_\lambda^2 \lambda a_1''(\lambda) + (\lambda_T + \lambda)\, b_1 \\ 0 &= -\left(\lambda_T + \lambda + \lambda_N + q_A\right) a_2(\lambda) + \kappa(\bar{\lambda} - \lambda) a_2'(\lambda) \\ &\quad + \sigma_A\sigma_\lambda \sqrt{\lambda}\rho_\lambda a_2'(\lambda) + \tfrac{1}{2}\sigma_\lambda^2 \lambda a_2''(\lambda) + (\lambda_T + \lambda)\, b_2 \end{aligned} \tag{1.59}$$

Based on the similarity to bond pricing in the CIR model I can guess and verify a solution for the case when $\rho_\lambda = 0$. To illustrate the idea let $A(u)$ and $B(u)$ be coefficient functions for zero coupon bond pricing in the CIR model

$$\begin{aligned} \gamma &= \tfrac{1}{2}\sqrt{\kappa^2 + 2\sigma_\lambda^2} \\ A(u) &= -\tfrac{2\kappa\bar{\lambda}}{\sigma^2} \log\left[\frac{\gamma e^{\frac{1}{2}\kappa u}}{\gamma\cosh(\gamma u) + \frac{1}{2}\kappa\sinh(\gamma u)}\right] \\ B(u) &= \frac{\sinh(\gamma u)}{\gamma\cosh(\gamma u) + \frac{1}{2}\kappa\sinh(\gamma u)} \end{aligned} \tag{1.60}$$

It is well known

$$-B(u)\kappa(\bar{\lambda} - \lambda) + \frac{1}{2}B(u)^2\sigma_\lambda^2\lambda - \lambda = -A'(u) - B'(u)\lambda \tag{1.61}$$

To solve the odes guess

$$a_0(\lambda) = \kappa_0\left[1 - \kappa_1 \int_0^\infty e^{-A(u) - B(u)\lambda - (\lambda_T + \lambda_N + r - g)u} du\right] \tag{1.62}$$

Observe

$$\int_0^\infty \sup_{\lambda\in\mathbb{R}} |B(u)\cdot e^{-A(u)-B(u)\lambda-(\lambda_T+\lambda_N+r-g)u}|du \leq \frac{1}{\gamma+\frac{1}{2}\kappa}\int_0^\infty e^{-(\lambda_T+\lambda_N+r-g)u}du < \infty \qquad (1.63)$$

which follows from $e^{-A(u)-B(u)\lambda} \leq 1$ for all $(u,\lambda) \geq 0$ and $\sup_{u\in\mathbb{R}} B(u) = \lim_{u\to\infty} B(u) = \frac{1}{\gamma+\frac{1}{2}\kappa}$. As a result, Lemma 3.6 from Newey and McFadden (1994) applies which proves we can differentiate under the integral sign by an application of the Dominated Convergence Theorem. To identify the constant terms plug into the ode

$$\begin{aligned}
0 &= -(\lambda_T+\lambda+\lambda_N+r-g)\,a_0(\lambda) + \kappa(\bar{\lambda}-\lambda)a_0'(\lambda) + \tfrac{1}{2}\sigma_\lambda^2\lambda a_0''(\lambda) + 1 + (\lambda_T+\lambda)\,b_0 + \tfrac{\lambda_N}{r-g} \\
&= -b_0\kappa_1\int_0^\infty \left[-(\lambda_T+\lambda_N+r-g) - \lambda - \kappa(\bar{\lambda}-\lambda)B(u) + \tfrac{1}{2}\sigma_\lambda^2\lambda B(u)^2\right]\cdot \\
&\quad e^{-A(u)-B(u)\lambda-(\lambda_T+\lambda_N+r-g)u}du - b_0\,(\lambda_N+r-g) + 1 + \tfrac{\lambda_N}{r-g} \\
&= -b_0\kappa_1\int_0^\infty \left[-(\lambda_T+\lambda_N+r-g) - A'(u) - B'(u)\lambda\right] e^{-A(u)-B(u)\lambda-(\lambda_T+\lambda_N+r-g)u}du \\
&\quad -b_0\,(\lambda_N+r-g) + 1 + \tfrac{\lambda_N}{r-g} \\
&= -b_0\kappa_1\left[e^{-A(u)-B(u)\lambda-(\lambda_T+\lambda_N+r-g)u}|_0^\infty\right] - b_0\,(\lambda_N+r-g) + 1 + \tfrac{\lambda_N}{r-g} \\
&= b_0\kappa_1 - b_0(\lambda_N+r-g) + 1 + \tfrac{\lambda_N}{r-g}
\end{aligned} \qquad (1.64)$$

In the second line I set $\kappa_0 = b_0$. After solving for κ_1 and simplifying I find

$$a_0(\lambda) = \frac{1}{r-g}\frac{\lambda_F+r-g}{\lambda_S+\lambda_F+r-g}\left[1+\lambda_S\frac{\lambda_N+r-g}{\lambda_F+r-g}\int_0^\infty e^{-A(u)-B(u)\lambda-(\lambda_T+\lambda_N+r-g)u}du\right] \qquad (1.65)$$

Observe that $\lim_{\lambda\to\infty} a_0(\lambda) = \frac{1}{r-g}\frac{\lambda_F+r-g}{\lambda_S+\lambda_F+r-g} = b_0$. Intuitively as λ increases it becomes more likely a jump to the target state will occur immediately and the coefficient converges to a_0 which is the dividend coefficient in the target state. The other odes are solved in a similar fashion and the transversality condition is easily verified leading to the pre-target share price given in the proposition. ■

1.C Model Estimation

In what follows I discuss identification, alternative methods for estimation, and the computation of risk neutral moments for the reduced form model. Recall my model is estimated by nonlinear

least squares as discussed in Section 1.3. Following the terminology in Newey and McFadden (1994) identification requires the nonlinear least squares conditional mean identification condition $F(X_{t,j}, \theta_0) \neq F(X_{t,j}, \theta)$. This follows from the observation that mean squared option pricing errors are minimized by the merger option pricing function F from Proposition 3 evaluated at the true parameter θ_0 underlying the data generating process. First, I discuss the economic intuition behind identification and then I prove my model is identified. For intuition, note that the option pricing function is differentiable so it suffices to show that $\nabla_\theta F(X_{t,j}, \theta_0) \neq 0$. Comparative statics for a cash deal are illustrated in Figure 1.15. The parameter $\theta = (p_Q, \tau_Q, \lambda, \sigma)$ includes the risk neutral probability of deal success, expected deal duration, announcement intensity, and volatility of the target fundamental. Changing the value of each of these parameters impacts option prices. For example, increasing p_Q leads to an increase in the merger stock price for the typical case in which the arbitrage spread is positive by increasing the coefficient on the deal payout and decreasing the coefficient on the target fundamental. This lowers diffusive risk which has the effect of decreasing option prices. Simultaneously, the increased price in the target state results in a larger loss in the event of deal failure is reflected in the increase in the skewness of implied volatility. When τ_Q increases average deal resolution takes a longer amount of time. As a result, the firm spends more time in the target state during which diffusive risk is low. This decreases the skewness of implied volatility because deal failures and deal successes are less imminent. Changing τ_Q also has an impact on the stock price via discounting, but this effect is relatively minor. The announcement intensity has an impact on option prices as increases in λ result in faster transitions from the pre-target to the target state. This puts more weight on cash in the merger stock price and results in the firm spending more time in the target state, similar to the combined effect of increasing p_Q and τ_Q as discussed above. The result is lower option prices in both the pre-target and target state. Finally, increasing the fundamental volatility σ increases diffusive risk in both states which increases option prices.

It is straightforward to show target volatility is identified because volatility does not enter in the transition probabilities or the coefficients in the stock price. As such, it is immediate that $\partial F/\partial \sigma > 0$ because Black-Scholes vega is positive. A direct proof for the other parameters is more involved. For example, in the target state the partial derivative with respect to the probability of

success leads to the computation

$$\frac{\partial F\left(\mathcal{T}\right)}{\partial p_Q} = \frac{\partial q_{21}}{\partial p_Q} F_{\mathcal{P}} + q_{21} \frac{\partial F_{\mathcal{P}}}{\partial p_Q} + \frac{\partial q_{22}}{\partial p_Q} F_{\mathcal{T}} + q_{22} \frac{\partial F_{\mathcal{T}}}{\partial p_Q} + \frac{\partial q_{23}}{\partial p_Q} F_{\mathcal{A}} \tag{1.66}$$

This approach is tedious as the probability of success enters in $(F_{\mathcal{P}}, F_{\mathcal{T}})$ in the Black-Scholes arguments for the stock price and the strike price as well as the upper bound of integration. Additionally, a number of the terms in the expression above have different signs. Instead, it is more productive to proceed by contradiction. Suppose there exists $\theta_1 \in \Theta$ such that $F(X_{t,j}, \theta_0) = F(X_{t,j}, \theta_1)$ but $\theta_0 \neq \theta_1$. As I have already discussed it must be the case that $\sigma_0 = \sigma_1$. The argument to show the other parameters are also the same follows by considering the properties of short dated options. Suppose the firm is in the pre-target state. As the time to maturity converges to zero it follows that

$$\begin{aligned} \lim_{\tau \to 0} F_{\mathcal{P}}(\theta_0) &= \left(K - A_{00} D - A_{10} V - A_{20} P_A\right)^+ \\ &= \left(K - A_{01} D - A_{11} V - A_{21} P_A\right)^+ \\ &= \lim_{\tau \to 0} F_{\mathcal{P}}(\theta_1) \end{aligned} \tag{1.67}$$

Where (A_{00}, A_{10}, A_{20}) are the coefficients under θ_0 and (A_{01}, A_{11}, A_{21}) are the coefficients under θ_1. Thus, it follows from short dated in the money put options in the pre-target state that the A coefficients are the same under θ_0 and θ_1. An analogous argument applies in the target state to show the B coefficients are the same. Then, the contradiction that $\theta_0 = \theta_1$ follows from the observation that (p_Q, τ_Q, λ) can be recovered from (A_1, B_0, B_1) in cash deals and (A_2, B_0, B_2) in stock deals. For example, in cash deals $\lambda = rA_1/(B_1 - A_1)$ and (λ_S, λ_F) can be obtained from the linear equations below

$$\begin{bmatrix} B_0 & \frac{B_0(r-g)-1}{r-g+\lambda} \\ B_1 - 1 & B_1 \frac{r}{r+\lambda} \end{bmatrix} \begin{bmatrix} \lambda_S \\ \lambda_F \end{bmatrix} = \begin{bmatrix} 1 - B_0(r-g) \\ -B_1 r \end{bmatrix} \tag{1.68}$$

This requires the two by two matrix be invertible which is satisfied in the examples I considered.

In my estimation strategy the stock price is used to infer the value of the fundamental. Option prices provide the additional restrictions which identify the parameters. The implicit assumption is that observation errors in option prices are large relative to observation errors in stock prices.

An alternative approach is to assume stock and option prices are both observed with error. In this case the target fundamental is a latent variable and the model can be written as

$$\begin{aligned} S_{t_i} &= P(X_{t_i,j},\theta) + \eta_{t_i} \\ Y_{t_i,j} &= F(X_{t_i,j},\theta) + \epsilon_{t_i,j} \\ d_{t_i} &= d_{t_{i-1}} + \left(g - \tfrac{1}{2}\sigma^2\right)\Delta t_i + \sigma\sqrt{\Delta t_i}u_{t_i} \end{aligned} \tag{1.69}$$

I denote the observation errors for the stock price as η_{t_i}, the amount of time between observations as $\Delta t_{i+1} = t_{i+1} - t_i$, and the log of target dividends as $d \equiv \log(D)$. Under the assumption the errors are normally distributed the model can be estimated by maximum likelihood or by following a Bayesian approach as discussed in Johannes and Polson (2009). Empirically, I find the estimate of the risk neutral probability of deal success is similar for nonlinear least squares and maximum likelihood.

In contrast to the structural model, the reduced form model relies on the computation of risk neutral moments of the target stock. Bakshi et al. (2003) show the risk neutral variance and skewness of log returns can be replicated with a portfolio of out-of-the-money call and put options. To implement their result I run the parametric regression $\sigma_{BS} = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 \tau + \epsilon$ where σ_{BS} is the Black-Scholes implied volatility of target options, $x = \log\left(\frac{K}{S}\right)$ is log-moneyness, and τ is maturity for each deal in the pre-announcement and post-announcement periods. I then use the fitted value of implied volatility truncated at the 5th and 95th quantile of observed implied volatility to estimate the risk neutral moments at a four month horizon. The truncation is performed so the risk neutral moments remain finite. In particular, the formula for the risk neutral moments requires the prices of deep out-of-the money call and put options which are obtained by extrapolation. The assumption that tails are bounded is stronger than the bound derived in Lee (2004) who shows, in the absence of arbitrage, that the tail of Black-Scholes implied volatility is bounded by twice the square root of the absolute value of log-moneyness divided by the maturity.

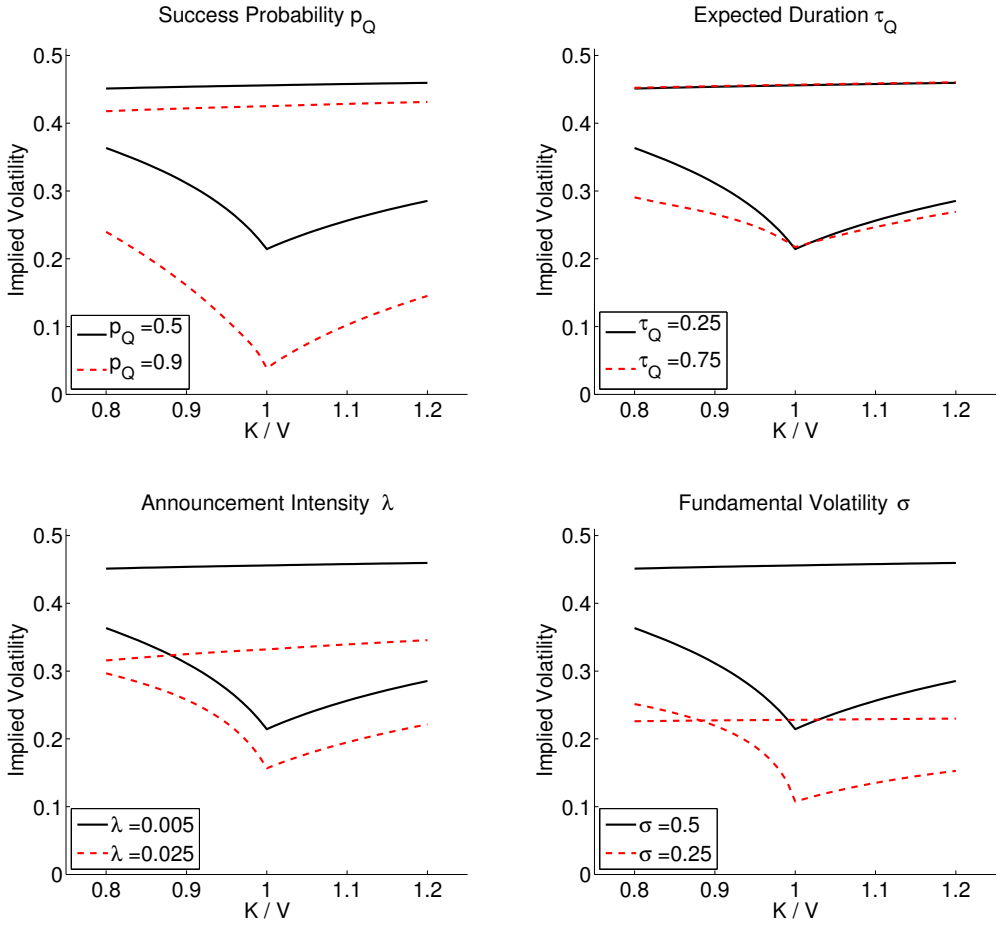


Figure 1.15: The figure above presents comparative statics for a cash deal. I plot the implied volatility function in black for the baseline parameters $\theta = (.5, .25, .005, .5)$. The top left plot increases the risk neutral probability of deal success to $p_Q = .9$ which results in a decrease in implied volatility and increase in skewness. The top right plot increases the risk neutral expected deal duration to $\tau_Q = .75$. This decreases the skewness of the implied volatility function. Intuitively, average deal resolution takes a longer amount of time so the firm will be a target for a longer period during which the slope of the target stock price with respect to dividends is low. Increasing the announcement intensity to $\lambda = .025$ decreases the implied volatility in the pre-target state and the target state. This contrasts the increase in the risk neutral probability of success which had little impact on the pre-target state above because the baseline announcement intensity is low. Finally, decreasing fundamental volatility to $\sigma = .25$ decreases implied volatility in both states. Note implied volatility in the figure is kinked as the strike approaches V consistent with the analytical result discussed earlier in the appendix.

CHAPTER 2

HIGH FREQUENCY MACHINE-READABLE NEWS AND ASSET PRICES

Machine-readable news allows high frequency traders to quantify the qualitative content in the news on a millisecond time scale. The resulting data can serve as an input for quantitative trading strategies and risk management. In this paper, which incorporates material from a joint paper with Yacine Aït-Sahalia and Jiangmin Xu, I examine how machine-readable news from Thomson Reuters relates to stock prices for the members of the Dow Jones Industrial Average from 2008 to 2013. I find that positive news sentiment forecasts high returns and low volatility, and that large quantities of news forecast high volatility and high volumes. In an application of these observations, I show that news sentiment can be used to increase the returns and Sharpe ratios of contrarian trading strategies. The outperformance is particularly significant at the open when overnight news increases the risk of adverse selection and informed trading. Finally, I demonstrate that intraday patterns in the arrival of news are contemporaneous with patterns in realized volatility and volume.

2.1 Introduction

The news plays a central role in financial markets. It serves as a primary source of information about expected cash flows, discount rates, and investor sentiment. In recent years, high frequency trading has motivated the development of new technologies that allow traders to rapidly analyze the news. Machine-readable news, which refers to the use of natural language processing algorithms to provide real-time analysis of the news, is a leading example of these technologies that highlights the importance of speed in equities and futures trading.[1] By subscribing to news analytics services and performing in-house analysis, sophisticated traders may now obtain a high frequency perspective on the daily flow of information that serves as a new signal for quantitative trading strategies and risk management.

In this paper, I investigate how machine-readable news from Thomson Reuters relates to intraday stock market activity. This builds on the work in Tetlock (2007) and Tetlock et al. (2008) who use textual analysis to quantify the qualitative content in the news. They find that negative sentiment in the media predicts aggregate and individual stock returns at a daily frequency. I extend this analysis to an intraday frequency by constructing news sentiment and news quantity variables that forecast returns, volatility, and volume at a five-minute frequency for the stocks in the Dow Jones Industrial Average from 2008 to 2013. In standard predictability regressions, positive news sentiment forecasts higher returns and lower volatility, while large quantities of news forecast higher volatility and higher volumes. As a robustness check, I confirm these results continue to hold across different periods of time and for different intraday sampling frequencies.

Then, to illustrate the economic significance of the return predictability, I show that intraday news sentiment can be used to improve the performance of contrarian trading

[1]For theoretical discussions of the importance of speed in high frequency trading see Aït-Sahalia and Saglam (2014) and Foucault et al. (2015). Aït-Sahalia and Saglam (2014) studies the impact of speed on market making strategies while Foucault et al. (2015) investigates how speed affects directional trading strategies based on the news. For a survey of empirical and theoretical research on high frequency trading see Jones (2013).

strategies. In particular, I modify the baseline strategy outlined in Khandani and Lo (2011) that buys the losers and sells the winners every five-minutes within the trading day. To be specific, my news adjusted strategy buys the losers with non-negative news sentiment and sells the winners with non-positive news sentiment.[2] In comparison to the baseline strategy, the news adjusted strategy earns a higher Sharpe ratio and higher average returns. Moreover, I show that a similar improvement applies to a second contrarian trading strategy that is based on trading against jumps in prices. In both cases, the news adjustment results in an improvement. Additionally, I find the largest improvements occur during the financial crisis and during the first hour of the trading day. This suggests that news sentiment is most valuable for forecasting returns when there is a heightened risk of asymmetric information and informed trading.

In addition to forecasting, I also document intraday patterns in the arrival of news that coincide with patterns in realized volatility and volume.[3] For example, most of the news each day is released before the open or after the close. Additionally, within the trading day, news arrives most frequently in the morning, the first minute within the hour, and the first seconds within the minute. Furthermore, I illustrate that the intraday periods with high news arrival rates are contemporaneous with high levels of realized volatility and large trading volumes. By highlighting this pattern at different frequencies, the results suggest that information in the news is, at least partially, responsible for driving the intraday patterns in stock market activity.[4] In particular, while there is a concern about reverse causality between realized volatility and the news at lower frequencies, it seems less plausible that high levels of

[2]In my context, winners and losers refer to the stocks in the Dow whose returns were in the upper and lower tercile of returns over the previous five-minutes. A detailed description of this strategy and a related contrarian trading strategy are included in Section 2.6. The news sentiment and news quantity variables are defined in Section 2.5.

[3]In addition to intraday patterns, there are also seasonal patterns in news arrivals. For example, Tetlock et al. (2008) observes that news is more likely to arrive during quarterly earnings seasons. I confirm this observation in Figure 2.1.

[4]These observations contribute to the literature that studies intraday patterns in stock market activity. For example, see Wood et al. (1985); Admati and Pfleiderer (1988); Andersen and Bollerslev (1998), and Stroud and Johannes (2014).

volatility in the first few seconds within a minute are causing a large number of news items to be released at that instant or during the first few seconds within the subsequent minute.

The layout for the remainder of the paper is as follows. Section 2.2 provides a literature review of the existing research that studies asset pricing and the news. Section 2.3 describes the machine-readable news data. Section 2.4 investigates the stock market activity around intraday news including the average response to the news and a number of specific examples that illustrate market reactions to anticipated and unanticipated news. Section 2.5 defines the news sentiment and news quantity variables and illustrates how these variables forecast intraday returns, volatility, and volume. In an application of these results, Section 2.6 uses intraday news sentiment to improve the performance of contrarian trading strategies. Section 2.7 concludes with a brief discussion of the results and suggestions for future research.

2.2 Literature Review

This paper contributes to the literature by investigating the relationship between asset prices and high frequency machine-readable news. The news data is from Thomson Reuters News Analytics, which uses natural language processing algorithms to rapidly determine the relevance, sentiment, and novelty of firm-specific news. The focus on machine-readable news contrasts a large body of work in the literature that has studied the impact of scheduled news. To be specific, I use the term scheduled news to refer to news that arrives at pre-determined times in a format that is structured around widely followed numerical content. Often, scheduled news contains information that is relevant for asset pricing. For example, earnings announcements and macroeconomic announcements are types of scheduled news which reveal information about the level of firm earnings, unemployment, inflation, and interest rates to the market. By comparing the realization of these announcements to analyst forecasts, it is possible to investigate how prices respond to estimated shocks in scheduled news.[5] In contrast

[5]The surprise or unexpected shock typically refers to $(X_{i,t} - E_{t-1}[X_{i,t}])/S_{i,t-1}$, the standardized analyst forecast error, where $X_{i,t}$ is the variable of interest such as earnings per share or non-farm payrolls, $E_{t-1}[X_{i,t}]$ is the mean or median analyst forecast before the announcement, and $S_{i,t-1}$ is a scaling variable such as

to scheduled news, unscheduled news is difficult to analyze because it arrives asynchronously and includes the unstructured text in articles and headlines. Moreover, the sheer volume of unscheduled news and its coverage of qualitative information pose challenges for traditional analysis. To circumvent these issues, I follow the recent literature that employs natural language processing algorithms to analyze scheduled and unscheduled news simultaneously.

With respect to scheduled news, there is long-standing evidence that earnings announcements impact asset prices.[6] For example, stock prices tend to exhibit increased average returns, realized volatility, and CAPM betas on earnings announcement dates. In option markets, the term-structure of implied volatility is also affected. In particular, there is typically an increase in the implied volatility of short-dated at-the-money options before earnings are announced and a decrease after the announcement. As a result, the cross-sectional and time-series behavior of option prices provide ex-ante and ex-post indications of the extent to which earnings impact stock return volatility. Finally, it is well-known that stock prices exhibit pre-announcement and post-announcement earnings drifts. The post-announcement drift can be illustrated by sorting stocks on earnings shocks. Historically, a portfolio that is long stocks with positive earnings shocks and short stocks with negative earnings shocks has earned a large abnormal return.

Beyond earnings, the response of various assets classes including stocks, bonds, and currencies to macroeconomic announcements has been extensively studied in the literature.[7] In part, this research is motivated by early work that highlighted the difficulty of explaining

the sample standard deviation of lagged forecast errors or lagged prices. For earnings, Tetlock et al. (2008) reports that the standardized analyst forecast errors are similar to the statistical errors from a seasonal random walk with drift which is used in Bernard and Thomas (1989).

[6]For a discussion of pre-announcement and post-announcement earnings drifts see Ball and Brown (1968) and Bernard and Thomas (1989). The earnings announcement premium was first documented in the accounting literature by Beaver (1968) and has subsequently been studied in Chari et al. (1988), Frazzini and Lamont (2007) and Savor and Wilson (2015). With respect to volatility, Patton and Verardo (2012) find an increase in realized betas estimated from high frequency returns on earnings announcement dates while Dubinsky and Johannes (2006) study option pricing around earnings announcements.

[7]For example, see Ederington and Lee (1993), Berry and Howe (1994), Jones et al. (1998), Fleming and Remolona (1999), Andersen and Bollerslev (1998), Balduzzi et al. (2001), Flannery and Protopapadakis (2002), Andersen et al. (2003), Green (2004), Bernanke and Kuttner (2005), Boyd et al. (2005), Andersen et al. (2007), Savor and Wilson (2013), Savor and Wilson (2014), and Lucca and Moench (2015).

asset price movements.[8] Since then, numerous studies have documented how the market reaction to macroeconomic announcements is consistent with rational expectations. In particular, realized volatility and volume tend to increase around scheduled announcements, prices jump when there are large unexpected shocks in macroeconomic variables like unemployment or interest rates, and returns exhibit low autocorrelation after announcements are released. Additionally, on announcement dates, asset prices tend to earn higher returns, exhibit larger amounts of realized volatility, and are more consistent with economic theory in the respect that market beta is highly correlated with the cross section of returns. In summary, the empirical findings suggest that macroeconomic announcements convey important information to the market which is quickly impounded into prices.

In addition to the studies focused on scheduled news, several recent papers have employed textual analysis to investigate the combined information in scheduled and unscheduled news. Tetlock et al. (2008) is a leading example. They use the fraction of negative words in Wall Street Journal and Dow Jones stories to predict the returns and earnings of S&P 500 firms from 1980 to 2004. Their approach highlights the additional information in firm-specific stories that is not subsumed by quantified measures of firm performance. In particular, negative words are found to significantly forecast daily returns and quarterly earnings after controlling for lagged values of returns, earnings, analyst forecasts, and accounting variables. This complements Tetlock (2007) who finds that negative sentiment in the "Abreast of the Market" column from the Wall Street Journal forecasts daily returns for the Dow Jones Industrial Average and trading volume on the New York Stock Exchange.[9]

[8]Seminal early papers include Shiller (1981), French and Roll (1986), Black (1986), Roll (1988), and Cutler et al. (1989). As a specific example, Cutler et al. (1989) find that innovations in macroeconomic and financial variables can only explain 15% to 20% of aggregate stock market returns at a monthly frequency from 1926-1986. Moreover, even when ex-post information is used to explain contemporaneous returns, the explanatory power of their unrestricted vector autoregressions does not increase above 40%. To further illustrate the point, they also examine the fifty largest post-war aggregate daily price movements and find that, although some days coincide with important information releases, for most days it is difficult to link major market moves to economic releases or other pertinent information.

[9]In particular, the largest eigenvector in the covariance matrix of General Inquirer word count categories for the "Abreast of the Market Column" loads heavily on the negative, weak, fail, and fall word categories. Tetlock (2007) interprets this eigenvector as a media pessimism factor and finds that high values of pessimism forecast negative daily returns whereas high absolute values of pessimism forecast high trading volumes.

Beyond improved forecasts, textual analysis has also been used to examine the hypothesis that investors overreact to the news. Specifically, Tetlock (2011) investigates the market reaction to stale news. In his context, staleness is defined as the average textual similarity of a news story to the previous ten stories about the same firm in the Dow Jones newswires. In line with the conjecture that stale stories contain less new information that is relevant for asset pricing, volatility and volume are significantly lower on days with high average staleness. At the same time, there is also evidence that weekly return reversals are larger after stale news, and that individual investors trade more aggressively on stale news. This suggests that individual investors may overreact to information in stale news that has already been incorporated into prices. In a related study, Von Beschwitz et al. (2015) investigate the market reaction to high frequency machine-readable news by exploiting a unique feature of RavenPack data. In particular, they identify a number of stories that were originally released as having high relevance that were later reclassified as having low relevance in an updated version of the RavenPack data.[10] In reaction to these false positives, they find a short-term impact on returns that begins to mean revert after 30 seconds and completely reverses after 2 minutes. This is similar to the return reversals following stale news in Tetlock (2011). Both studies indicate that the media has an impact on financial markets that is distinct from the information conveyed in the news, which builds on the findings in Engelberg and Parsons (2011) and Peress (2014) who study the causal impact of the media.

Additionally, there are a number of other financial applications where textual analysis is used to examine unscheduled news. For example, Antweiler and Frank (2004) analyze internet stock message boards. They find that disagreement on message boards is contemporaneous with increased trading, and that message posting helps to predict volatility, but they do not find economically significant return predictability in internet chat rooms. Additionally, Loughran and McDonald (2011) develop a negative word list for financial texts

[10]RavenPack is a competitor of Thomson Reuters News Analytics that provides machine-read news for articles in the Dow Jones Newswire. The interpretation in Von Beschwitz et al. (2015) relies on the assumption that the updated versions of RavenPack have improved natural language processing algorithms that do a better job of classifying relevant versus not relevant news.

that can be used as a substitute for the negative word list in the Harvard Psychosociological Dictionary. They document a significant relationship between excess returns and the fraction of negative words in annual 10-K reports using their dictionary whereas no significant relationship holds for the Harvard Dictionary.[11] Jegadeesh and Wu (2013) extend this work. They propose a new method for determining document sentiment that uses the market reaction to 10-K filings to estimate the strength of different words.[12] They find a significant relationship between filing date returns and their sentiment scores obtained from positive, negative, and combined positive and negative word lists. This contrasts the previous literature which emphasizes the importance of negative words. Finally, Groß-Klußmann and Hautsch (2011) examine high frequency machine-readable news from Thomson Reuters. They document significant increases in volatility and volume when relevant news is released both unconditionally and in vector autoregressions, but they do not find evidence of intraday return predictability.

2.3 Data Description

I use two primary sources of data in my analysis. First, machine-readable news is obtained from Thomson Reuters News Analytics. The machine-readable news dataset includes Thomson Reuters stories and headlines that are analyzed with natural language processing algorithms to determine the relevance, sentiment, and novelty of the news for individual companies. To study how the news interacts with asset prices, I also obtain intraday trade and quote data from the TAQ database for the constituents of the Dow Jones Industrial

[11] In addition to using the fraction of negative words, Loughran and McDonald (2011) also show that a term weighting scheme, which downplays the importance of frequently used words, results in a significant relationship between both of the dictionaries and the excess returns around the file date, which are defined as the 4-day return from the filing date to 3 days after the filing date minus the CRSP value-weighted market index return.

[12] The innovation in Jegadeesh and Wu (2013) is to determine the term weights based on the 10-K filing returns. Much of the previous literature uses a proportional weighting scheme, based on the fraction of negative or positive words relative to the total number of words, to determine sentiment. In contrast to a proportional weighting scheme, Jegadeesh and Wu (2013) illustrate that their term weighting method is empirically robust to the dictionary used to compile the positive or negative word lists and continues to produce significant results even if half of the words are randomly discarded.

Average. Sections 2.3.1 and 2.3.2 provide overviews of the Thomson Reuters and TAQ datasets.

2.3.1 Thomson Reuters Machine-Readable News

I use Thomson Reuters News Analytics (TRNA) to investigate the relationship between machine-readable news and asset prices. TRNA employs natural language processing algorithms to analyze the contents of Thomson Reuters articles and headlines. The results are published in real-time to subscribing clients which include high frequency hedge funds and algorithmic traders. The dataset consists of a number of variables such as the sentiment, relevance, and novelty of the news along with millisecond time-stamps indicating when the data was transmitted to Reuters clients.

In this paper, I use the sentiment, relevance, and novelty variables. The sentiment variable indicates whether a news item, which can be an article or headline, is positive, neutral, or negative for a specific company. The relevance variable is a number between zero and one indicating the relevance of a news item for a specific company. It is determined by comparing the number of times a company is mentioned within a news item in contrast to other companies or assets. Higher values indicate the news is more relevant for a particular company. For example, relevance is equal to one for companies that are mentioned in the headline of a news item.[13] Finally, novelty is obtained by tracking the take or sequence number of a story. The take is equal to one for the first-take and is updated by one for subsequent takes or corrections.

Table 2.1 reports summary statistics for the news data. It contrasts how the data varies for the Dow versus a subset of 4,864 firms matched to CRSP by ticker symbol. The sample period is 2008 to 2013. The table includes the time series averages for each of the variables alongside the median and the 25th and 75th quantiles. For example, Firms Per Day indicates

[13]I consider articles to be relevant in the remainder of the paper if the relevance variable is greater than or equal to .9. This seems a natural cutoff after inspecting the histogram of the relevance variable which is reported in the Appendix along with an example of the TRNA data. For a comprehensive description of the TRNA dataset see Mitra and Mitra (2011).

News Data Summary Statistics

	Dow				CRSP			
	Mean	25th	50th	75th	Mean	25th	50th	75th
Firms Per Day	27.8	27.0	28.0	29.0	669.6	576.0	660.0	754.0
News Per Firm	12.9	9.5	12.2	15.6	4.4	3.8	4.3	4.9
Intraday News Per Firm	6.0	4.2	5.6	7.4	1.7	1.3	1.6	1.9
% Relevant	44.5	39.6	44.6	50.0	63.4	58.9	63.2	68.1
% Relevant Positive	17.9	13.8	17.8	21.7	25.1	21.3	25.7	29.1
% Relevant Negative	13.1	9.5	12.5	16.0	15.5	13.4	15.3	17.2
% Articles (first-take)	72.4	66.2	72.2	78.9	64.2	58.5	64.4	70.4
% Alerts	20.0	15.0	19.3	24.8	31.7	25.8	31.2	37.5
News Take	1.7	1.3	1.5	1.8	1.9	1.5	1.8	2.1
Number of Companies	4.7	3.6	4.2	5.5	4.6	2.8	3.7	6.3
Number of Words	461.5	395.6	443.3	513.3	417.0	329.7	376.8	499.1
Number of Sentiment Words	177.1	153.9	172.9	195.4	167.5	152.0	167.6	184.0

Table 2.1: This table reports summary statistics for the Thomson Reuters News Analytics data on trading days from 2008 to 2013. The time series average and median are included alongside the 25th and 75th quantiles for the Dow and a subset of 4,864 firms matched to CRSP. In contrast to the CRSP firms, the Dow has more news per firm, a higher percentage of firms with news, a lower percentage of relevant news, and a higher percentage of articles versus alerts. Beyond these differences, the Dow and CRSP subsets have similar novelty of the news as measured by the news item take variable. There are also similar numbers of companies mentioned, words, and sentiment words in a typical news item across the two subsets. The percent of relevant news is the fraction of news items with relevance greater than or equal to .9. The percent of relevant positive and negative news is the fraction of news items that are relevant and have sentiment equal to 1 or -1. The news take is equal to 1 for the first-take of a story and is updated by 1 for subsequent alerts or takes. In particular, the percent of first-take articles plus alerts is less than 100% because there can be corrections and subsequent takes for the same story. Finally, sentiment words refers to the number of words in a news item that are deemed relevant for a particular firm and intraday news refers to news items arriving between 9:30am and 4:00pm when US equity markets are open.

that 28 firms in the Dow have a news item on an average trading day versus 670 firms in the CRSP subset. Other variables include the number of news items per firm, the number of intraday news items per firm, and the percentages of relevant news, relevant positive news, relevant negative news, first-take articles, and alerts. Additionally, I include summary statistics for the news take, number of words, number of sentiment words, and number of companies mentioned in a news item. The table indicates that the Dow receives more media attention than the firms in the CRSP subset. This is evident from the Dow's higher percentage of firms with news items, higher number of news items per firm, and lower percentage of relevant stories. In particular, one interpretation of the low percentage of relevant stories for the Dow is that financial media writers prefer to focus on the Dow when there is no relevant news to report because it is less costly to write articles about large firms that are widely followed. In addition to the differences, there are also a number of similarities between the Dow and CRSP subsets. For example, there is similar news novelty and a similar number of words, sentiment words, and companies mentioned in a typical news item for both subsets. In particular, the average news item is a first-take article that references 4 to 5 companies with more than 100 words of text.

Figure 2.1 illustrates intraday and quarterly patterns in relevant news arrivals. Specifically, it plots the frequency of intraday news arrivals, the time series of the daily quantity of news, and the spectrum of the daily quantity of news. The top plot shows the frequency of intraday relevant news arrivals for the stocks in the Dow. It highlights how firm-specific news clusters before the open and after the close. In fact, this pattern has only become more stark over time. During recent years the amount of news released right after the close has increased whereas the amount of news released right before the close has decreased. To the extent that firms influence the timing of their news releases, for example, by scheduling earnings or corporate announcements, the results suggest that firms may prefer to report relevant news when the market is closed and that this preference has become more significant over time. In addition to the intraday patterns, the middle plot shows the time series of

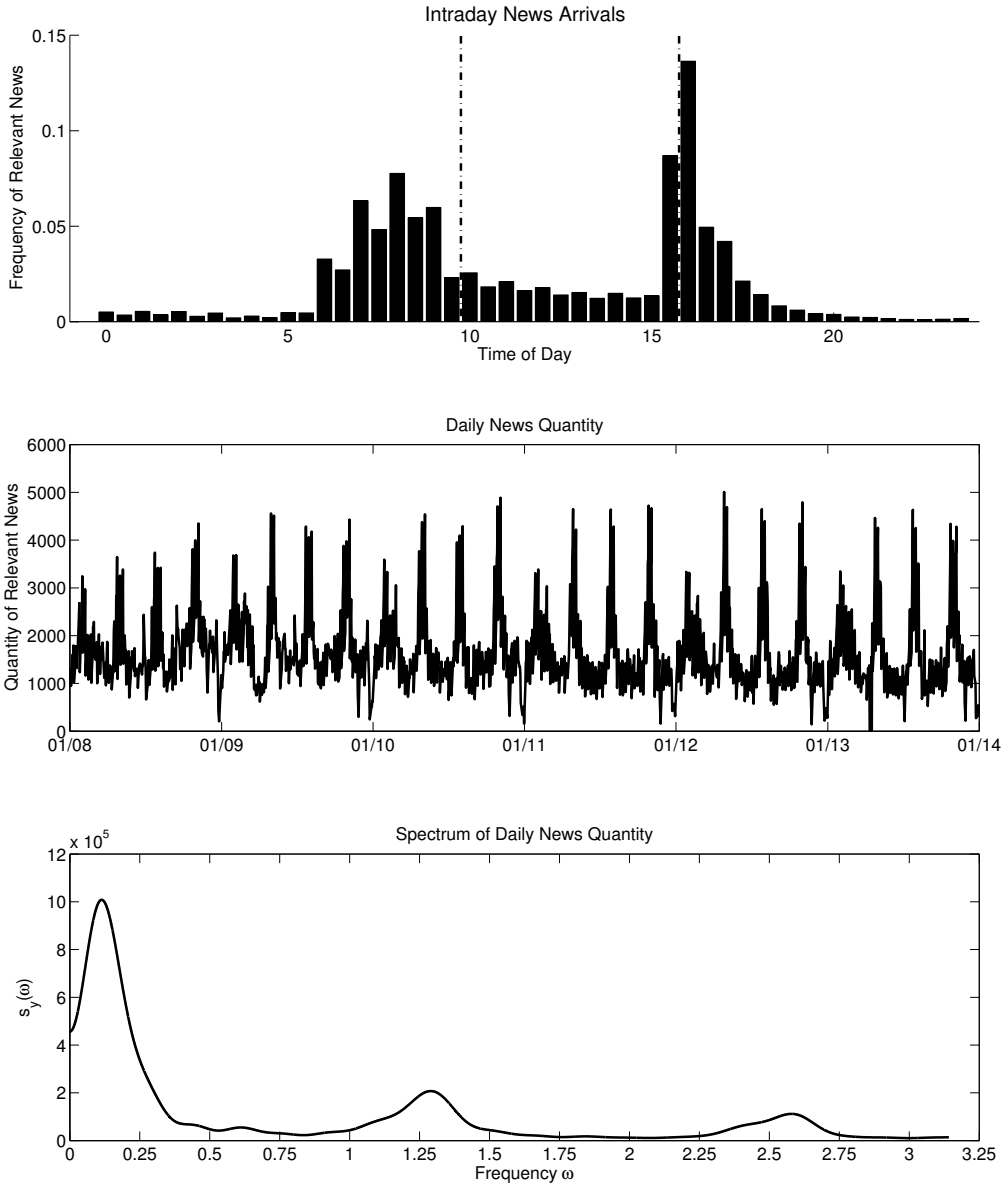


Figure 2.1: This figure illustrates seasonal patterns in the arrival of relevant news on business days from 2008 to 2013. The top plot shows the intraday arrival of news. Many of the stories arrive outside of trading hours. In particular, the arrival of news items clusters before the open and around the close of trading which are denoted with dashed lines. The middle plot highlights a different feature of the data by plotting the time series of the daily quantity of unique relevant headlines. It demonstrates how the quantity of news increases during quarterly earnings announcement seasons. Finally, the bottom plot reports the Bartlett estimate of the spectrum for the daily quantity of news following Hamilton (1994) Section 6.3. Integrating the spectrum over different frequencies decomposes the contribution of low versus high frequencies components in the news quantity variance. In particular, I find that low frequencies corresponding to periods greater than 20 business days account for 58% of the news quantity variance whereas high frequencies corresponding to periods less than 20 business days account for 42% of the news quantity variance.

the daily quantity of unique relevant headlines for the firms in the Dow. This plot reveals a significant increase in the quantity of news during quarterly earnings announcement seasons, a pattern that is also documented in Tetlock et al. (2008). Finally, the last plot shows an estimate of the spectrum of the daily quantity of news to decompose the high and low frequency components in the variance of the quantity of news. I find that both high and low frequency components contribute to the news quantity variance. In particular, low frequencies corresponding to periods greater than 20 business days account for 58% of the variance whereas high frequencies corresponding to periods less than 20 business days account for 42% of the variance.

2.3.2 Trade and Quote Data

Intraday trade and quote data are obtained from the Trade and Quote (TAQ) database for the constituents of the Dow Jones Industrial Average (Dow) over the sample period from 2008 to 2013. I use the midpoint of the National Best Bid and Offer (NBBO) derived by WRDS to obtain a time series for intraday stock prices at a one second frequency.[14] In concept, the NBBO refers to the highest bid and lowest offer among all the available quotes across all the national exchanges at each point in time. The SEC defined the NBBO as part of Regulation NMS, which was approved in 2005 and implemented in 2007, to mandate that brokers execute customer trades at or within the best prevailing market prices. To compute the NBBO, WRDS compares the last quote every second from the consolidated tape at each of the national exchanges to find the highest bid and the lowest offer in addition to the

[14]While TAQ contains a substantial amount of information about intraday trades and quotes, it is also worth noting what is not available in the data. For example, the consolidated quotes reported in TAQ cannot be used to reconstruct the limit order book for individual exchanges. Also, the trade data does not indicate trader identities, whether trades are buyer or seller originated, nor who the makers and takers of liquidity are, with the latter distinction being an important source of rebates on some exchanges. Lastly, TAQ does not include odd lots, which are trades for less than 100 shares, as these transactions are not reported to the consolidated tape. O'Hara et al. (2014) find that odd lots comprised about 25% of the trades and 6% of the volume in a sample of 120 NASDAQ stocks at the end of 2011. They further argue that informed trading, including as much as 39% of price discovery, is now increasingly happening in odd lot transactions as high frequency and algorithmic traders break up large orders into small pieces.

largest size for each of these quotes. The result is an intraday time series of the last NBBO each second along with its depth for all of the stocks with available quote data.[15]

I also use the TAQ trade data to compute the volume traded and the volume weighted average price each second. Then, I combine the one second time series obtained from the raw trade and quote data to construct the variables used in my empirical analysis which include returns, realized volatility, volume, spread, depth, illiquidity, depth imbalance, and order imbalance for each stock in the Dow. Detailed definitions for these variables are provided in the appendix. Finally, to proxy for the market portfolio, I use returns from the S&P 500 exchange traded fund whose ticker symbol is SPY. In contrast to computing market returns directly from the index constituents, the advantage of my approach is that abnormal and excess returns may be interpreted as long-short positions which involve only two trades, being long a members of the Dow and short the SPY exchange traded fund.

2.4 Market Activity Around Intraday News

I investigate stock market activity around intraday news in the discussion below. The section is structured as follows. Section 2.4.1 presents three examples illustrating how prices interact with the news. Section 2.4.2 builds on the specific examples by documenting the average response of signed returns, realized variance, and volume to the news. In summary, I find that relevant news with less than fifty words, which I call short relevant news, is particularly significant for market activity. After short relevant news, signed returns exhibit a statistically and economically significant drift while realized variance and volume increase over 40%. Finally, I find that intraday periods with high news arrival rates are contemporaneous with larger trading volumes and higher values of realized variance at various frequencies.

[15]I would like to thank WRDS for making their NBBO algorithm and its output available. To clean the data, the NBBO algorithm excludes quotes if the bid is greater than or equal to offer (crossed and locked quotes), the bid or bid size is zero, the offer or offer size is zero, or the quote condition indicates a trading halt, cancelled quote, or another reason for exclusion (as indicated by the TAQ MODE variable). Similarly, when using the trade data I require that trades have a positive price, non-zero size, and correction code equal to 0 or 1 indicating a "good trade" as defined in the TAQ user manual.

2.4.1 Examples

The SEC Charges Goldman Sachs with Fraud

On April 16th, 2010 the SEC charged Goldman Sachs (GS) with securities fraud for its actions related to the structuring and marketing of a synthetic CDO tied to subprime mortgages known as Abacus 2007-AC1. Thomson Reuters first reported the charges at 10:39:50 am. The news released at the time of the announcement including the headlines, relevance, sentiment, and novelty are detailed in the table below.[16]

Time	Relevance	Sentiment	Take	Headline
10:39:50.50	1	-1	1	SEC CHARGES GOLDMAN SACHS GROUP INC <GS.N> WITH FRAUD IN STRUCT
10:41:03.03	1	-1	2	SEC ALLEGES GOLDMAN, ONE VICE PRESIDENT DEFRAUDED INVESTORS BY
10:41:31.31	1	-1	3	SEC SAYS THE VICE PRESIDENT IS FABRICE TOURRE
10:42:14.14	1	0	4	SEC ALLEGES THAT PAULSON & CO HEDGE FUND PAID GOLDMAN TO STRUCT
10:43:07.07	1	-1	5	SEC ALLEGES THAT GOLDMAN STRUCTURED A SYNTHETIC CDO THAT HINGED
10:44:13.13	1	0	6	SEC ALLEGES THAT PAULSON & CO PAID GOLDMAN $15 MLN FOR STRUCTURI
10:44:39.39	1	-1	7	SEC SAYS INVESTORS IN LIABILITIES OF ABACUS ALLEGED TO HAVE LOST
10:45:23.23	1	-1	8	SEC SAYS FILES COMPLAINT AGAINST GOLDMAN SACHS, TOURRE IN U.S. DIS
10:46:40.40	1	-1	9	SEC SEEKS FOR GOLDMAN, TOURRE TO DISGORGE ILLEGAL PROFITS, PAY CIV
10:47:45.45	1	-1	1	BRIEF-SEC charges Goldman Sachs with fraud over subprime-related CDO
10:48:35.35	1	-1	1	GOLDMAN SACHS <GS.N> SHARES DROP MORE THAN 7 PCT AFTER SEC CHAR
10:50:11.11	1	-1	1	U.S. STOCKS ADD LOSSES, S&P 500 <.SPX> HEADS TOWARD SESSION LOWS AFT
10:52:10.10	1	-1	2	GOLDMAN SACHS <GS.N> SHARES EXTEND LOSSES, DROP 10 PCT
10:53:55.55	1	-1	2	U.S. STOCKS EXTEND LOSSES, S&P 500 <.SPX> FALLS 1 PCT AS BANKS WEIGH
10:57:22.22	1	-1	1	STOCKS NEWS US-Goldman shares fall after SEC charges

Prior to the SEC's announcement, April 16th had begun as a relatively quiet Friday. The biggest story of the morning was that Bank of America had reported first quarter earnings of $.28 per share beating analyst expectations of $.10 per share. Despite this positive news for financial stocks, GS opened .74% lower and continued to sell-off an additional 1% during

[16]The full headline is included in the TRNA dataset. I only report the first line of the headline to fit each news item on one line.

the first hour of trading. Then, the situation began to rapidly deteriorate with GS stock declining over 8% in the minutes prior to the fraud accusation being publicly announced.

Figure 2.2 illustrates the stock market reaction to the SEC charges. Specifically, the top left plot shows the cumulative log return for GS stock from the open to the close. The top right plot zooms in on the returns at the announcement with dashed black lines indicating the arrival of negative relevant news. Finally, the bottom right plot shows the dollar volume traded at the announcement relative to the average dollar volume traded throughout the day. Together these plots indicate that the sell-off accelerated alongside a sharp increase in the dollar volume traded shortly after 10:37 am, pre-dating the first Thomson Reuters story at 10:39:50 am. This suggests that market participants either anticipated the charges or that another news vendor broke the story first. After the announcement, GS shares continued to decline losing as much as 16.5% at the intraday low.

Beyond the stock market reaction, Figure 2.2 also illustrates the intraday arrival of news. The bottom left plot shows the time when news arrived against the cumulative signed sentiment words. I define the latter quantity as the running sum of sentiment multiplied by sentiment words updated throughout the day for each story attributed to GS. In particular, sentiment is -1,0, or 1 for negative, neutral, and positive stories and sentiment words refers to the number of relevant words in a news item for a particular stock. The plot indicates that a large amount of negative news followed the announcement of the SEC charges, as seen by the increasing number of negative sentiment words that were released throughout the trading day. In total, 123 relevant news items, 107 negative relevant news items, and -29,555 signed sentiment words were attributed to GS between the open and the close. The average duration between these news items was 2 minutes and 15 seconds with a maximum duration of 13 minutes and 7 seconds. To put these numbers in perspective, this was the largest quantity of news, relevant news, negative relevant news, and negative signed sentiment words during any trading session from 2008 to 2013 for GS. Over the same period of time the average amount

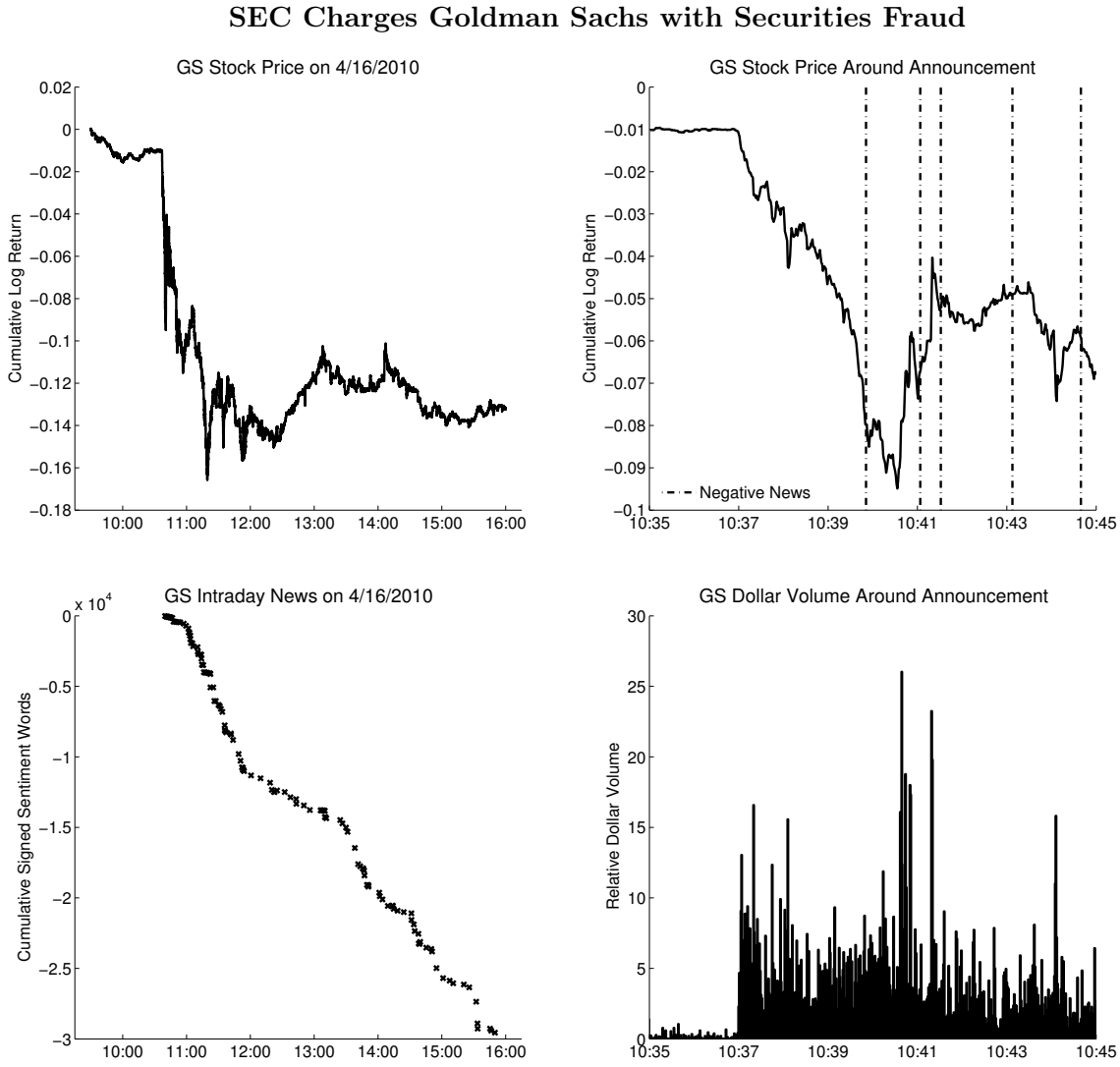


Figure 2.2: On April 16th, 2010 the SEC charged Goldman Sachs with securities fraud for its actions in structuring and marketing a synthetic CDO tied to subprime mortgages known as Abacus 2007-AC1. This figure illustrates the stock market reaction to the SEC charges and the arrival of news throughout the trading day. In particular, the top left plot shows the cumulative return for GS stock from the open to the close. The top right plot zooms in on the returns at the announcement with dashed black lines indicating the arrival of negative relevant news. The bottom right plot shows the dollar volume traded at the announcement relative to the average dollar volume traded throughout the day. Finally, the bottom left plot shows the cumulative signed sentiment words for GS. The plots illustrate that the sell-off accelerated shortly after 10:37 am alongside a sharp increase in the dollar volume traded. By the time Thomson Reuters first announced the charges at 10:39:50 am GS had depreciated over 8%. This suggests that the charges were anticipated by market participants or that another news vendor had broken the story first. After the announcement more negative news was released and GS continued to trade lower.

of news was 11.6 items per day consisting of 4.4 relevant stories, 2.3 negative relevant stories, and -864 signed sentiment words.

IBM Announces A Share Buyback Program

IBM's board of directors authorized a $15 billion share buyback program on February 26th, 2008 that was reported by Thomson Reuters at 10:52:02 am. Concurrently, IBM raised its earnings per share forecast for the 2008 fiscal year. Both of these announcements were positive and significant for shareholders. In particular, the share buyback program represented 10% of IBM's $151 billion market capitalization. The news released at the time of the announcement including the headlines, relevance, sentiment, and novelty are detailed in the table below.

Time	Relevance	Sentiment	Take	Headline
10:52:02.02	1	1	1	IBM Board of Directors Authorizes $15 Billion for Stock Repurchase <IBM.N>
10:52:05.05	1	1	1	IBM BOARD OF DIRECTORS AUTHORIZES $15 BILLION FOR STOCK REPURC
10:52:11.11	1	1	2	IBM <IBM.N> SAYS RAISES FULL-YEAR EPS OUTLOOK
10:52:18.18	1	0	3	IBM <IBM.N> SAYS 2008 FULL-YEAR EARNING PER SHARE OF $8.20 TO $8.30
10:53:20.20	1	-1	1	CORRECTED-IBM <IBM.N> SEES 2008 FULL-YEAR EARNING PER SHARE OF
10:53:20.20	1	1	2	RPT-IBM BOARD OF DIRECTORS AUTHORIZES $15 BILLION FOR STOCK REP
10:53:21.21	1	1	3	RPT-IBM <IBM.N> SAYS RAISES FULL-YEAR EPS OUTLOOK
10:53:36.36	1	0	4	IBM <IBM.N> REUTERS ESTIMATES FY 2008 SHR VIEW $8.22
10:54:05.05	1	1	1	DOW <.DJI> TURNS POSITIVE ON IBM BUYBACK, RAISED OUTLOOK
10:54:14.14	1	1	5	IBM <IBM.N> SAYS AT THIS TIME DOES NOT SEE REQUESTING BOARD APP
10:55:02.02	1	1	1	REG-Intnl Bus. Mach IBM Board of Directors Authorizes $15 Billion for Stock Repurch
10:57:29.29	1	0	2	S&P 500 <.SPX>, NASDAQ <.IXIC> ALSO GAIN ON IBM NEWS
11:02:26.26	1	1	1	IBM says board approves $15 bln share buyback
11:04:58.58	1	1	1	US STOCKS-Wall St briefly turns up on IBM news
11:05:32.32	1	1	1	US STOCKS-Wall St briefly turns up on IBM news

Figure 2.3 illustrates the stock market reaction and the intraday arrival of news on February 26th, 2008. The top left plot shows the cumulative log return for IBM from the open to the close. The top right plot zooms in on the returns at the announcement with

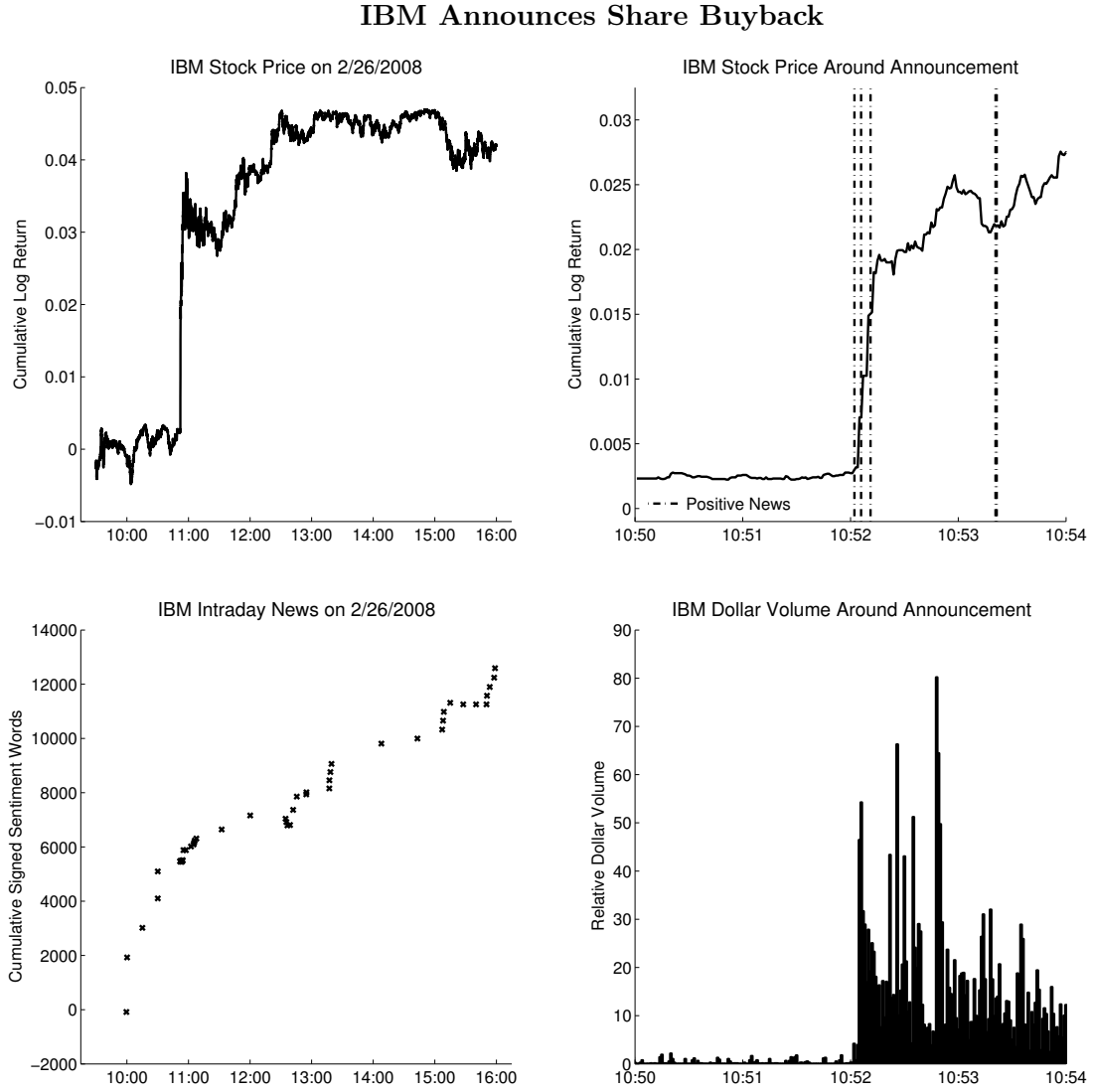


Figure 2.3: On February 26th, 2008 the IBM board of directors approved a $15 billion share buyback program and raised its earnings per share forecast. These announcements were positive and significant for shareholders with the buyback program representing 10% of IBM's $151 billion market capitalization. This figure illustrates the stock market reaction to the announcement and the arrival of news throughout the trading day. The top left plot shows the cumulative return for IBM stock from the open to the close. The top right plot zooms in on the returns at the announcement with dashed black lines indicating the arrival of positive news. After the announcement IBM increased 1.5% within 10 seconds. It then continued to rally closing 4.2% higher than the open. The stark reaction to the news suggests that market participants did not anticipate the buyback program. Further evidence can be found in the bottom right plot which shows a jump in the dollar volume traded at the time of the announcement relative to the average dollar volume traded throughout the day. Finally, the bottom left plot shows the arrival of news against the cumulative amount of signed sentiment words. The positive announcement of the share buyback program was followed by more positive news throughout the day.

dashed black lines indicating the arrival of positive news. These plots suggest that IBM's share buyback program was not anticipated by market participants. In particular, IBM's stock price was flat before the announcement. After the announcement, IBM shares jumped 1.5% in 10 seconds. This suggests the share buyback program was not anticipated by the market, which contrasts the previous example in which GS depreciated significantly before the SEC charges were reported by Thomson Reuters. Beyond this difference, there are also a number of similarities between the GS and IBM examples. In particular, there is a sharp increase in the dollar volume traded at the time of the announcement. Additionally, more positive news followed IBM's original share buyback announcement including 44 news items, 39 relevant stories, 30 positive relevant stories, and 7,132 signed sentiment words before the close of trading. Finally, IBM's share price continued to drift up in the post-announcement period similar to the decline in GS's stock price after the SEC charges. By the end of the day IBM closed 4.2% higher.

Verizon Wireless Acquires Alltel

On June 6th, 2008 Verizon Wireless agreed to buy Alltel for $28.1 billion from Atlantis Holdings, a special purpose vehicle created by TPG Capital and Goldman Sachs. Thomson Reuters reported the deal at 9:50:41 am. The news released at the time of the announcement including the headlines, relevance, sentiment, and novelty are detailed in the table below.

The morning of the merger announcement Vodafone indicated at 2:27:30 am that Verizon was in talks to buy Alltel.[17] Overnight Verizon was up .46% on this news. Figure 2.4 plots the intraday stock market reaction and arrival of news. Specifically, the top left plot shows that Verizon's stock increased over 6% in the first hour of trading. The top right plot shows

[17]According to the New York Times, the deal between Verizon and Alltel was "one of the quickest flips in corporate history". It was announced less than nine months after Atlantis had acquired Alltel and was motivated by the banks who, after financing the original leveraged buyout, had taken losses on debt they had been unable to sell. Beyond the pressure from the banks, there had also been longstanding rumors of a potential deal between Verizon and Alltel that dated back to 2005. Originally, Vodafone was concerned that a deal with Alltel would dilute their 45% stake in Verizon. The deal was ultimately financed with debt to avoid dilution. It established Verizon as the nation's largest cellular telephone provider, jumping ahead of AT&T which fell to the number two spot.

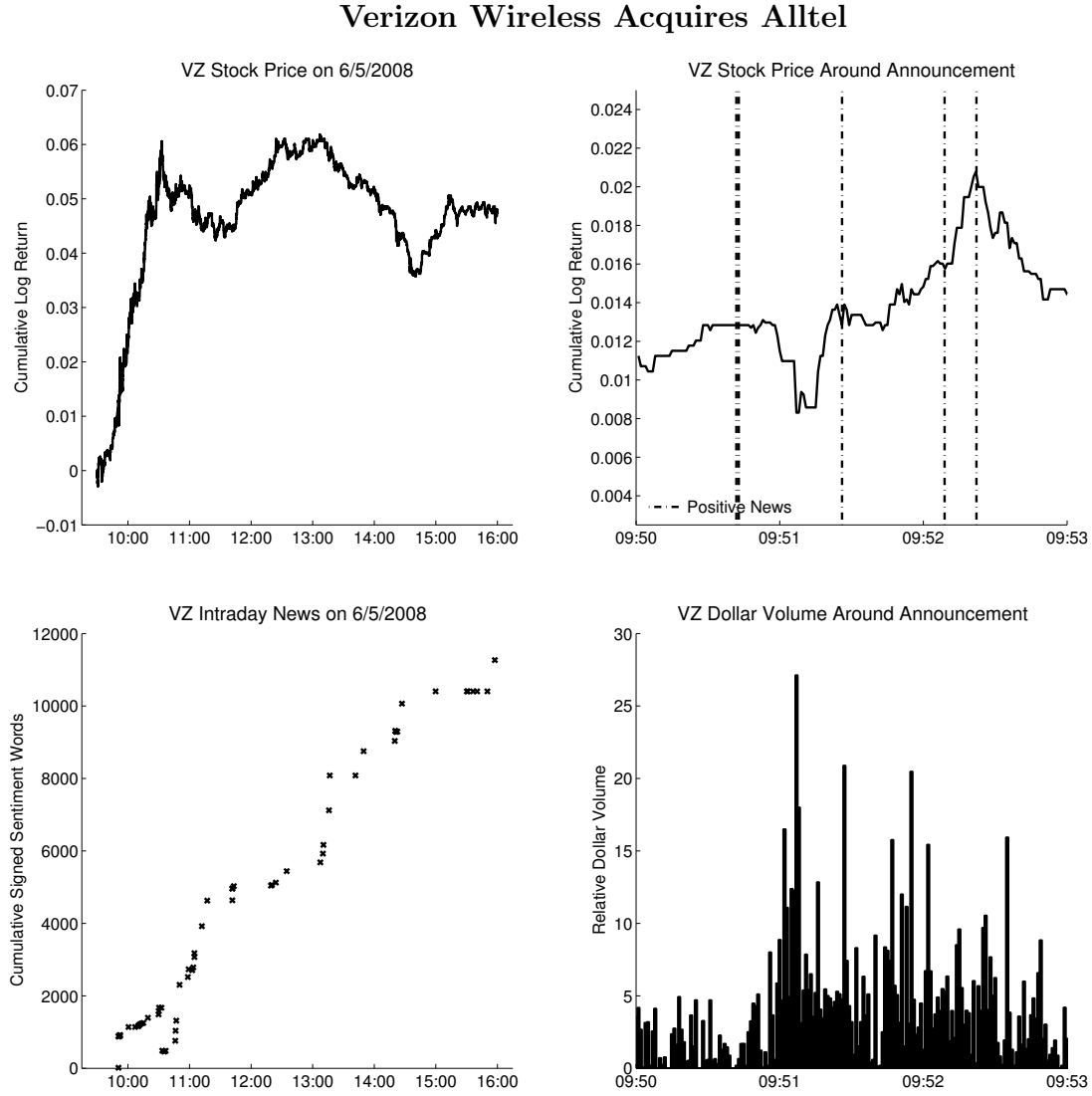


Figure 2.4: On June 6th, 2008 Verizon Wireless agreed to buy Alltel for $28.1 billion from Atlantis Holdings, a special purpose vehicle created by TPG Capital and Goldman Sachs. The deal established Verizon as the nation's largest cellular telephone provider allowing it to jump ahead of AT&T Wireless which fell to the number two spot. This figure illustrates the stock market reaction to the announcement and the arrival of news throughout the trading day. The top left plot shows the cumulative return for VZ stock from the open to the close. The top right plot zooms in on the returns at the announcement with dashed black lines indicating the arrival of positive news. The bottom right plot shows the dollar volume traded at the announcement relative to the average dollar volume traded throughout the day. Finally, the bottom left plot shows the cumulative signed sentiment words for VZ throughout the trading day. The deal announcement at 9:50:41 am was followed by more positive news throughout the day.

Time	Relevance	Sentiment	Take	Headline
09:50:41.41	1	1	1	VERIZON WIRELESS TO ACQUIRE ALLTEL; WILL EXPAND NATION'S MOST
09:50:42.42	1	1	1	Verizon Wireless To Acquire Alltel; Will Expand Nation's Most Reliable Wireless Net
09:51:05.05	1	0	2	VERIZON WIRELESS SAYS WILL ACQUIRE THE EQUITY OF ALLTEL FOR
09:51:26.26	1	1	3	VERIZON WIRELESS SAYS BASED ON ALLTEL'S PROJECTED NET DEBT AT
09:52:02.02	1	-1	4	VERIZON WIRELESS SEES SYNERGIES WITH NET PRESENT VALUE, AFTER
09:52:08.08	1	1	5	VERIZON WIRELESS SAYS ENTERED INTO AN AGREEMENT WITH ALLTEL
09:52:22.22	1	1	6	VERIZON WIRELESS SEES SYNERGIES TO GENERATE INCREMENTAL COST
10:00:32.32	1	1	1	Verizon Wireless to buy Alltel in $28.1 bln deal
10:06:54.54	1	0	1	Hartford, Vermont Residents to Benefit From Verizon Wireless Network Expansion
10:09:43.43	1	1	1	VERIZON COMMUNICATIONS INC <VZ.N> SAYS AGREED TO PREPURCHA
10:10:21.21	1	1	2	VERIZON COMMUNICATIONS INC <VZ.N> SEES DEAL CASH FLOW POSITI
10:10:38.38	1	1	3	VERIZON COMMUNICATIONS INC <VZ.N> SEES RUN RATE SYNERGIES FUL
10:11:43.43	1	1	4	VERIZON COMMUNICATIONS INC <VZ.N> SEES INTEGRATION COSTS FOR
10:13:11.11	1	1	5	VERIZON COMMUNICATIONS INC <VZ.N> CFO DOES NOT SEE VERIZON
10:14:57.57	1	1	6	VERIZON COMMUNICATIONS INC <VZ.N> CEO SAYS HE APPROACHED PRI

that Verizon declined slightly after the deal was announced at 9:50:41 am but then continued to rally. One interpretation of the slow response is that traders might have needed more time or information to determine the value of the deal. This might also explain why the large increase in Verizon's share price did not occur overnight despite an indication from Vodafone that Verizon was in talks to buy Alltel. Similar to the previous examples, positive news at the time of the announcement was followed by more positive news throughout the trading day. In total, Verizon had 64 news items, 54 relevant items, 42 positive relevant items, and 11,269 signed sentiment words which compares to an average of 8.1 news items, 5.2 relevant news items 3.1 positive relevant items, and 1,559 signed sentiment words during the 2008 to 2013 sample period.

2.4.2 Returns, Volatility, and Volume Around the News

Event Time Response After Intraday News

Figure 2.5 plots the average signed returns, realized variance, and volume from five minutes before to five minutes after news arrivals between 9:40am and 3:50pm for the members of the Dow from 2008 to 2013. To focus on intraday stories, I exclude news that arrives during the first and last ten minutes of the trading day. The plot reveals that stock prices react the most to relevant first-take news with fewer than 51 words, which I call short relevant news.

In the plot, the solid black line, dashed black line, and marked black line correspond to the average response for short relevant news, relevant news, and not relevant news respectively. The shaded grey area is a pointwise 95% confidence interval for short relevant news. The top figure illustrates the average cumulative signed log return. The signed return is equal to the cumulative return multiplied by 1 for positive news and -1 for negative news. I exclude neutral news whose sentiment is 0. The lower figures plot the realized variance and dollar volume traded relative to the average realized variance and dollar volume traded over the previous day. I focus on the relative realized variance and volume for ease of interpretation and to take into account the fact that high frequency returns exhibit persistent conditional heteroskedasticity. In total, the sample includes 9,710 short relevant news items, 48,065 relevant news items, and 74,817 not relevant news items.

As seen in the plot, the market reaction is strongest after short relevant news. The top figure indicates that signed returns exhibit an upward drift and earn a statistically and economically significant average return of 4 basis points in the five minutes that follow short relevant news.[18] To provide a sense for the economic magnitude of these returns, note that an annualized return of 5% corresponds to an average return of 2 basis points per day. By way of contrast, the stocks in the Dow average 6.7 short relevant news items per day which earn an average of 28 basis points in less than 35 minutes. These returns are large in an

[18] As a robustness check, I confirm that similar results hold for excess and abnormal returns in addition to returns that have been winsorized at the 1% and 99% quantiles to mitigate the impact of outliers.

Stock Market Activity Around Intraday News

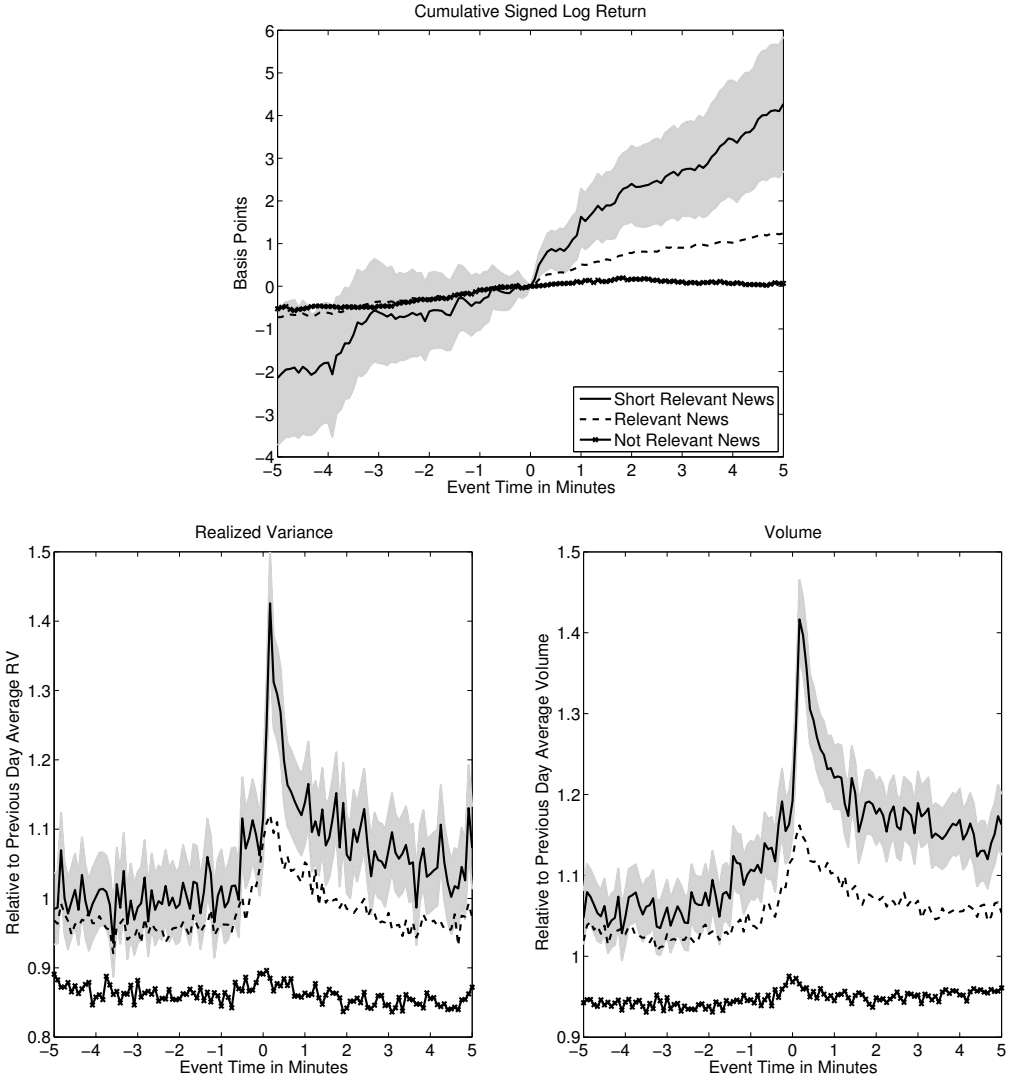


Figure 2.5: This figure plots the average cumulative signed returns, realized variance, and volume around intraday news from 9:40am to 3:40pm for the stocks in the Dow from 2008 to 2013. Signed returns are equal to returns multiplied by sentiment. Realized variance is the squared five-second return relative to the average squared return over the previous day. Similarly, volume is the dollar volume traded relative to the average dollar volume traded over a five-second interval on the previous day. The solid black line corresponds to the average cumulative signed return, realized variance, and volume for relevant first-take news with fewer than 51 words, which I call short relevant news. The dashed black line and the marked black line correspond to relevant first-take news and not relevant first-take news. The grey shaded area is the pointwise 95% confidence interval around short relevant news. I exclude neutral news. The sample includes 9,710 short relevant news items, 48,065 relevant news items, and 74,817 not relevant news items. Similar results are obtained when using abnormal returns and returns that are winsorized at the 1% and 99% quantiles to remove outliers.

absolute sense. Additionally, on a relative basis, the returns after short relevant news are larger than the returns after relevant and not relevant news. For example, relevant news is followed by a statistically significant cumulative signed return of 1 basis point in five minutes whereas not relevant news is followed by insignificant returns that are essentially equal to zero throughout the entire event window. Finally, the lower figure provides another indication of the importance of short relevant news. In particular, the realized variance and volume increase over 40% after short relevant news. This contrasts an increase of about 15% after relevant news and little reaction after not relevant news.

To illustrate the robustness of these unconditional results, I estimate the following regression

$$Y_{i,t} = \sum_{k\in\{P,N\}} \sum_{h=-10}^{10} \delta_h^k I_{i,t+h}^k + \sum_{j=1}^{6} \beta_{ij} D_{j,t} + \sum_{j=2}^{74} \phi_j H_{j,t} + \beta_C Controls_{i,t-1} + \epsilon_{i,t} \tag{2.1}$$

where the dependent variable $Y_{i,t}$ is equal to returns, realized variance, or volume and the dummy variable $I_{i,t+h}^k$ indicates the arrival of positive or negative $k \in \{P, N\}$ short relevant news at time $t+h$ for stock i.[19] This specification is adapted from Patton and Verardo (2012). Specifically, this regressions allows me to control for firm-year fixed effects β_{ij} where $D_{j,t}$ is a dummy variable for year j and intraday fixed effects ϕ_j where $H_{j,t}$ is a dummy variable for the intraday time period j. I also control for lagged values of abnormal returns, market returns, realized variance, volume, bid-offer spread, depth, illiquidity, depth imbalance, and order imbalance, which are defined in the Appendix. The estimated coefficients for the first four leads and lags δ_h^k are reported in Table 2.2 along with robust t-statistics that are clustered by time. The results are consistent with Figure 2.5. For example, when positive short relevant news is released, returns exhibit a statistically significant upward drift during the five-minute period that is contemporaneous with the news and in the subsequent

[19] In this context, the dependent variables are five minute log returns, the squared five minute log return relative to the average squared return over the previous day, and the dollar volume traded over a five minute interval relative to the average dollar volume traded over the previous day, which is analogous to Figure 2.5.

Stock Market Activity Around Intraday News

	Returns		Realized Variance		Volume	
Model	(1)	(2)	(1)	(2)	(1)	(2)
δ^P_{-4}	-0.67	-0.67	0.04	0.01	0.04***	-0.00
	[-0.93]	[-0.94]	[1.42]	[0.44]	[3.46]	[-0.31]
δ^P_{-3}	0.37	0.38	0.04	-0.00	0.08***	0.03***
	[0.53]	[0.55]	[1.32]	[-0.07]	[6.74]	[3.54]
δ^P_{-2}	-2.06***	-2.05***	0.13***	0.06**	0.10***	0.04***
	[-2.82]	[-2.80]	[4.19]	[2.12]	[8.22]	[4.08]
δ^P_{-1}	0.90	0.97	0.18***	0.09***	0.14***	0.07***
	[0.79]	[0.86]	[5.35]	[2.76]	[11.42]	[6.83]
δ^P_0	2.44***	2.49***	0.37***	0.27***	0.19***	0.13***
	[2.66]	[2.72]	[9.73]	[7.41]	[14.66]	[11.96]
δ^P_1	2.07**	2.12***	0.18***	0.10***	0.09***	0.04***
	[2.52]	[2.58]	[5.61]	[3.19]	[8.14]	[4.13]
δ^P_2	0.92	0.96	0.13***	0.03	0.09***	0.03***
	[1.21]	[1.27]	[4.22]	[1.07]	[8.04]	[3.05]
δ^P_3	-0.48	-0.43	0.14***	0.04	0.10***	0.03***
	[-0.56]	[-0.50]	[4.42]	[1.26]	[8.37]	[3.14]
δ^P_4	0.54	0.57	0.18***	0.05	0.09***	0.03***
	[0.70]	[0.74]	[5.44]	[1.62]	[7.87]	[2.80]
δ^N_{-4}	0.11	0.14	0.04	0.02	0.05***	0.02**
	[0.12]	[0.15]	[1.48]	[0.70]	[4.75]	[2.05]
δ^N_{-3}	-0.23	-0.19	0.09***	0.06**	0.07***	0.03***
	[-0.25]	[-0.20]	[2.99]	[2.09]	[6.41]	[3.79]
δ^N_{-2}	1.79*	1.78*	0.06*	0.01	0.08***	0.03***
	[1.88]	[1.87]	[1.95]	[0.22]	[7.41]	[3.37]
δ^N_{-1}	-2.72***	-2.73***	0.21***	0.15***	0.14***	0.07***
	[-2.92]	[-2.94]	[6.48]	[4.63]	[11.65]	[7.97]
δ^N_0	-2.64***	-2.61***	0.29***	0.22***	0.18***	0.13***
	[-2.65]	[-2.63]	[8.33]	[6.54]	[13.83]	[12.00]
δ^N_1	-1.60	-1.58	0.18***	0.11***	0.10***	0.04***
	[-1.63]	[-1.61]	[5.50]	[3.63]	[8.37]	[4.82]
δ^N_2	-1.59	-1.54	0.19***	0.13***	0.10***	0.05***
	[-1.57]	[-1.53]	[5.83]	[4.11]	[8.37]	[5.08]
δ^N_3	-0.26	-0.19	0.12***	0.05*	0.08***	0.03***
	[-0.31]	[-0.23]	[3.90]	[1.67]	[7.30]	[3.56]
δ^N_4	-1.16	-1.09	0.11***	0.02	0.08***	0.02**
	[-1.31]	[-1.24]	[3.60]	[0.73]	[6.95]	[2.33]

robust t-statistics clustered by time: * p<0.10, ** p<0.05, *** p<0.01

Table 2.2: This table reports the coefficients on the first four leads and lags for short relevant news for the regression in equation 2.1 estimated for the stocks in the Dow from 2008 to 2013 between 9:40am and 3:50pm ($T = 3,261,071,\ N = 39$). The results indicate the robustness of the unconditional reactions in Figure 2.5. The second specification (2) accounts for fixed effects and control variables that are described in Section 2.4.2.

five-minute period. Adding fixed effects and control variables has a limited impact on these results, which can be seen by the small difference from the first and second specifications for the return regressions in Table 2.2. Additionally, when negative short relevant news is released, returns exhibit a statistically significant negative drift during the five-minute period that is contemporaneous with the news and in the five-minute period that precedes the news. This illustrates a difference between positive and negative news that is not highlighted in Figure 2.5. In particular, the market appears to anticipate negative news resulting in a pre-announcement drift whereas positive news is not anticipated resulting in a post-announcement drift. Interestingly, this is consistent with the GS and IBM examples from the previous discussion. Finally, there are also large increases in realized variance and volume before and after the arrival of short relevant news. For example, the impact on realized variance is 27% for positive news and 22% for negative news after accounting for fixed effects and control variables. Additionally, there is also a larger increase in the realized variance between the first lag and the period that is contemporaneous with the news for positive news, which is consistent with the interpretation that positive news is less anticipated than negative news.

The importance of short relevant news can be motivated by the examples from the previous section. In particular, the SEC charges, share buyback, and merger were originally announced as short relevant news items. After these announcements, more relevant news about the same event was released throughout the trading day. Often the follow-up stories had a similar sentiment and longer word count in comparison to the original announcement. This suggests that pertinent intraday information may arrive first in the form of short headlines or articles that are followed by a longer sequence of stories about the same event, or at least this appears to be the case in the Thomson Reuters dataset. To the extent that traders react the most to the first stories containing new information that is relevant for stock prices, one would expect a larger reaction in the returns, realized variance, and volume after

short relevant news in comparison to relevant news, assuming short relevant news arrives first. Figure 2.5 provides support for this hypothesis.

To relate these findings to the literature, this is the first paper to my knowledge that emphasizes the importance of short relevant news. For example, Von Beschwitz et al. (2015) find that relevant news obtains an average signed return of 2 basis points after two minutes for all the stocks in TAQ from April 2009 to September 2012 using RavenPack data. To reconcile their result with mine, the larger returns they find for relevant news might be attributed to the differences in the news data, the sample period, or the inclusion of small stocks in their sample. As another example, Groß-Klußmann and Hautsch (2011) use Thomson Reuters data and find that relevant news is not followed by statistically significant returns for a sample of 39 large stocks on the London Stock Exchange between January 2007 and June 2008. Tetlock (2011) is perhaps most closely related to my results. He finds that stock prices respond less to stale news where staleness is defined as the textual similarity of a news item to the previous ten stories about the same firm. Importantly, his study does not include news items with less than 50 words due to the difficulty in determining the textual similarity between articles with a small number of words. As such, my results are complementary to his finding that stock prices react more to news that is not stale.

News vs. No News Returns

To further investigate the impact of the news on intraday stock returns, Table 2.3 displays summary statistics for five-minute signed returns following the news and Figure 2.6 plots the log empirical density of these returns. As was noted before, short relevant news and relevant news are followed by statistically significant signed returns of 4bp and 1bp respectively. When the data is winsorized to control for outliers the average signed returns following the news decrease to 2.8bp and .7bp but their statistical significance increases from t-statistics of 5.4 and 4.8 to 6.6 and 5.5. Additionally, the table highlights the high levels of volatility and excess kurtosis after the news. In particular, I include the returns from non-overlapping

five minute windows that had no news as a point of comparison to the signed returns after the news. The table indicates that the no news returns had a standard deviation of 20bp which is significantly lower than the 45bp, 57bp, and 79bp standard deviations following not relevant, relevant, and short relevant news. This result is also illustrated visually in Figure 2.6 which plots the log empirical density of the returns along with a normal distribution whose mean and standard deviation are matched to the sample moments of the no news returns. While all the high frequency returns exhibit excess kurtosis and fatter tails than a normal distribution, the plot indicates that volatility and kurtosis increase even more for returns that follow the news.

Intraday Patterns in News Arrivals, Realized Variance, and Volume

Finally, I also investigate how intraday patterns in the arrival of news are related to realized variance and volume. The results are reported in Figure 2.7. In particular, the left panel plots the frequency of relevant news arrivals every 15 minutes within the day, minute within the hour, and second within the minute from 2008 to 2013 for the stocks in the Dow between 10:00am and 3:30pm. The right panel plots the realized variance and volume over the analogous intraday periods relative to the average level of realized variance and volume each day. To focus on intraday patterns, I exclude the first and last thirty minutes of trading each day.[20]

The plots indicate that news arrivals exhibit intraday seasonality at different frequencies. News arrives most frequently in the morning, the first minute within an hour, and the first seconds within a minute. The periods with high news arrival rates are also contemporaneous with higher levels of realized variance and volume. By highlighting this pattern at different frequencies, the results suggest that news arrivals are driving the increased levels of realized variance and volume. For example, while there is a concern about reverse causality at lower frequencies, it seems less plausible that high levels of volatility in the first few seconds within

[20]The results are similar but slightly noisier when the first and last thirty minutes of trading are included in the analysis.

Stock Returns After the News: Summary Statistics and Log Density Plot

News Type	SR	R	NR	NN	SRW	RW	NRW	NNW
Mean (bp)	4.3	1.2	0.1	0.0	2.8	0.7	0.1	-0.0
Std. Dev. (bp)	79.1	57.0	44.9	20.0	41.9	29.5	26.5	16.0
t-Statistic	5.4	4.8	0.5	0.1	6.6	5.5	0.8	-1.4
Skew	9.1	6.4	-0.2	0.7	1.1	0.5	-0.0	-0.0
Kurtosis	212.9	310.0	322.7	191.7	11.5	9.9	8.6	6.0
Max (%)	19.8	19.8	17.0	21.6	2.1	1.3	1.0	0.6
Min (%)	-10.2	-15.8	-19.1	-17.3	-1.5	-1.1	-1.1	-0.6
Sample Size	9,710	48,065	74,817	3.11e6	9,710	48,065	74,817	3.11e6

Table 2.3: The table above reports summary statistics for five-minute signed returns after short relevant, relevant news, not relevant news, and periods with no news from 9:40am to 3:40pm for the stocks in the Dow from 2008 to 2013. Consistent with Figure 2.5, short relevant news and relevant news are followed by positive signed returns that are statistically significant. Returns are also more volatile after the news. To illustrate the robustness of these results, the second set of column winsorizes returns at the 1% and 99% quantiles.

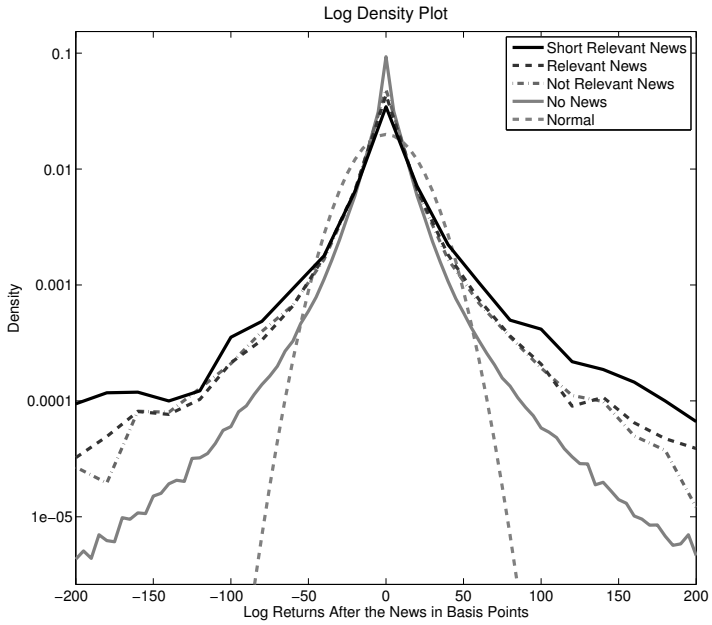


Figure 2.6: The figure above plots the log kernel density estimate of five-minute returns following short relevant news, relevant news, not relevant news, and periods with no news during the analogous period to the table above to illustrate the increase in volatility after the news. Additionally, I include the log pdf of a normal distribution whose mean and standard deviation match the no news returns to highlight the fat tails and excess kurtosis of the high frequency returns.

Intraday Patterns in News Arrivals, Realized Variance, and Volume

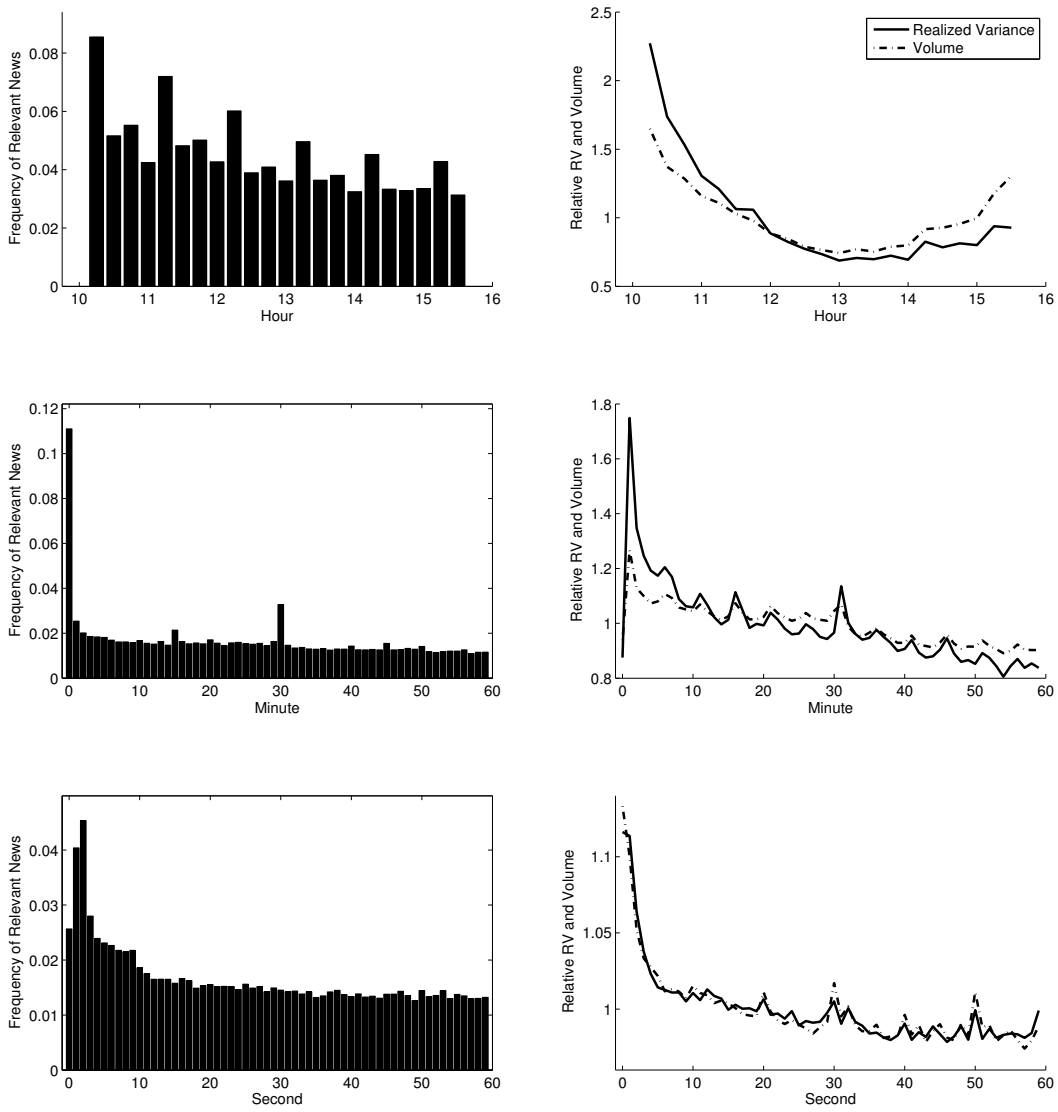


Figure 2.7: The figure above illustrates intraday patterns in the arrival of news alongside the average realized variance and volume for the stocks in the Dow from 2008 to 2013. In particular, the plots on the left show the frequency of relevant news arrivals every 15 minutes within the day, minute within the hour, and second within the minute between 10:00am and 3:30pm. The plots on the right show the average realized variance and volume over the analogous intraday periods. To account for the fact that high frequency returns exhibit highly persistent conditional heteroskedasticity, I compute the realized variance during intraday periods relative to the average realized variance throughout the day. I call this the relative realized variance. The relative volume is computed in a similar fashion. The plots indicate that news arrives more frequently in the first and thirtieth minute within an hour and during the first few seconds within a minute which are contemporaneous with higher volumes and volatility. The top plots contrast the well known U-shaped pattern in realized variance and volume at a lower frequency against the intraday arrival of news.

a minute are increasing the number of news items being released at the start of the next minute. The analysis cannot, however, rule out other microstructure explanations that might explain these patterns. For example, order execution algorithms might be designed to trade at the start of a minute or an hour which could contribute to the observed patterns in realized variance and volume while being unrelated to the news.

2.5 Forecasting Intraday Returns, Volatility, and Volume

The results from the previous section suggest that machine-readable news forecasts high frequency returns, volatility, and volume. To investigate this hypothesis in a dynamic setting, I construct news sentiment and news quantity variables throughout the trading day and examine the forecasting content of these variables in cross-sectional regressions which are specified as

$$Y_{i,t+1} = \beta_i + \beta_S NS_{i,t} + \beta_Q NQ_{i,t} + \beta_C Controls_{i,t} + \epsilon_{i,t+1}, \tag{2.2}$$

where the dependent variable $Y_{i,t+1}$ is either abnormal returns $R^a_{i,t+1}$, volatility $|R^a_{i,t+1}|$, or volume $VLM_{i,t+1}$, and the explanatory variables are news sentiment $NS_{i,t}$, news quantity $NQ_{i,t}$, and a vector of control variables $Controls_{i,t}$ that can include abnormal returns $R^a_{i,t}$, realized variance $RV_{i,t}$, volume $VLM_{i,t}$, bid-offer spread $SPD_{i,t}$, depth $DPH_{i,t}$, illiquidity $ILQ_{i,t}$, depth imbalance $DIB_{i,t}$, and order imbalance $OIB_{i,t}$. Each of these variables is defined in the Appendix. I estimate the model for individual stocks, as a panel, and in Fama and MacBeth (1973) regressions. The results indicate that positive news sentiment forecasts high returns and low volatility, and that large quantities of news forecast high realized volatility and high volumes.[21]

[21]The results are robust to whether raw returns or abnormal returns are used in the analysis. Abnormal returns are defined as $R^a_{i,t} = R_{i,t} - \hat{\beta}_{i,t} R_{m,t}$ where the realized beta is estimated from twenty days of lagged five minute returns using a subsample estimator following Zhang et al. (2005). I report the regression results for abnormal returns and for convenience will interchangeably use the terms returns and abnormal returns in this section.

The news sentiment and news quantity variables are motivated by Figure 2.5. In particular, the event time analysis from the previous section indicates that signed returns exhibit an upward drift after relevant news and that realized variance and volume increase after relevant news regardless of the sign of the news. Together, these results motivate the definition of two variables which I call news sentiment and news quantity. The first variable, news sentiment, tracks the sentiment of the news to forecast returns. The second variable, news quantity, tracks the quantity of news to forecast realized volatility and volume.

I define the news sentiment $NS_{i,t}$ as the average sentiment from the ten most recent news items that are relevant first-takes for stock i at time t.[22] By construction, the news sentiment is between -1 and 1. It averages the sentiment scores from individual news items which equal -1, 0, or 1 for negative, neutral, or positive stories as determined by Thomson Reuters' algorithms. In this manner, news sentiment distinguishes between positive and negative news with higher values corresponding to higher fractions of positive news in recent stories.

I define the news quantity $NQ_{i,t}$ as the standardized number of relevant first-takes for stock i at time t. To compute this variable, I obtain the number of relevant first-takes for each stock between the previous day's close and the intraday observation times. Then, for each intraday time, I standardize the quantity of news with the mean and standard deviation of the lagged quantity of news at the same intraday time over the previous sixty business days. The resulting variable has a mean and standard deviation that are approximately equal to zero and one. It is also skewed to the right and is uncorrelated with news sentiment. To be specific, the pointwise correlation between the news sentiment and the news quantity is -.05.

Table 2.4 reports panel regressions for forecasting returns, volatility, and volume with the news variables at a five-minute frequency from 9:30am to 10:30am for the stocks in the Dow

[22]The results are similar if the average is taken over the most recent five or twenty news items. Additionally, I find similar results if the news sentiment is defined as the average sentiment of relevant first-takes over a fixed amount of time, such as five days, as opposed to a fixed amount of news items.

Forecasting Returns, Volatility, and Volume: Panel Regressions

	Returns			Volatility			Volume		
Model	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
NS	0.21***	0.20***	0.20***	-1.98***	-0.81***	-0.86***	-0.02***	-0.01**	-0.01*
	[4.05]	[4.26]	[4.37]	[-4.58]	[-4.53]	[-4.61]	[-2.65]	[-2.17]	[-1.78]
NQ	-0.10	-0.11*	-0.12*	1.53***	0.23***	0.36***	0.30***	0.14***	0.14***
	[-1.24]	[-1.80]	[-1.95]	[9.98]	[3.57]	[5.69]	[13.01]	[14.54]	[14.68]
Ra		-0.90***	-0.90***		-0.12*	-0.08		0.01***	0.01***
		[-5.51]	[-5.61]		[-1.90]	[-1.17]		[2.90]	[3.03]
RV		-0.25	-0.24		9.91***	9.80***		0.03***	0.03***
		[-1.43]	[-1.39]		[17.49]	[17.65]		[5.08]	[5.74]
VLM		0.07	0.10		2.67***	2.52***		0.54***	0.56***
		[0.87]	[1.12]		[15.67]	[12.93]		[26.33]	[28.55]
SPD			-0.15**			1.44***			-0.08***
			[-2.33]			[11.20]			[-11.01]
DPH			0.05			-1.20***			0.03***
			[0.89]			[-7.16]			[7.31]
ILQ			-0.01			0.60***			0.02***
			[-0.20]			[8.27]			[9.65]
DIB			0.03			-0.07**			-0.00
			[0.66]			[-2.06]			[-0.35]
OIB			-0.02			-0.07*			-0.01**
			[-0.23]			[-1.72]			[-2.46]
R^2_{adj}	0.000	0.001	0.001	0.112	0.282	0.291	0.114	0.421	0.430

robust t-statistics clustered by time and firm: * p<0.10, ** p<0.05, *** p<0.01

Table 2.4: This table reports panel regression estimates for forecasting abnormal returns, volatility, and volume at a five-minute frequency for the stocks in the Dow between 9:30am and 10:30am from 2008 to 2013 ($T = 528,984$, $N = 39$). The regression model is $Y_{i,t+1} = \beta_i + \beta_S NS_{i,t} + \beta_Q NQ_{i,t} + \beta_C Controls_{i,t} + \epsilon_{i,t+1}$. The dependent variable $Y_{i,t+1}$ is either abnormal returns $R^a_{i,t+1}$, volatility $|R^a_{i,t+1}|$, or volume $VLM_{i,t+1}$, and the explanatory variables are lagged values of news sentiment $NS_{i,t}$, news quantity $NQ_{i,t}$, abnormal returns $Ra_{i,t}$, realized variance $RV_{i,t}$, volume $VLM_{i,t}$, bid-offer spread $SPD_{i,t}$, depth $DPH_{i,t}$, illiquidity $ILQ_{i,t}$, depth imbalance $DIB_{i,t}$, and order imbalance $OIB_{i,t}$. For ease of interpretation, I have standardized the explanatory variables to be mean zero and standard deviation one. For example, a one standard deviation increase in news sentiment $NS_{i,t}$ forecasts a .2 basis point increase in abnormal returns over a five-minute horizon. The results indicate that high values of news sentiment forecasts higher returns and that large news quantities forecast increases in volatility and volume. I compute t-statistics following Petersen (2009) that are double clustered by time and firm.

Forecasting Returns for Individual Stocks

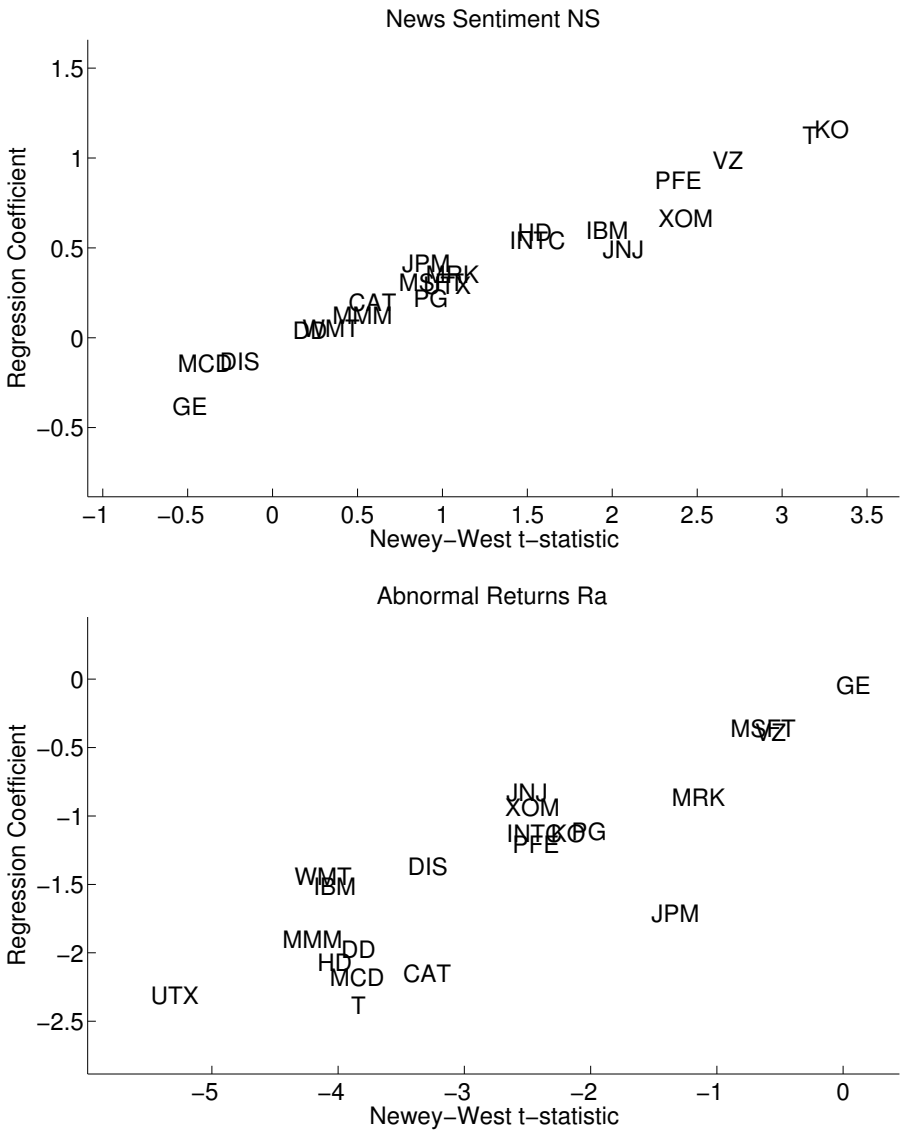


Figure 2.8: This figure plots the regression coefficients and Newey and West (1987) t-statistics for the news sentiment $\hat{\beta}_S$ and abnormal returns $\hat{\beta}_{C,1}$ in the model $\sum_{j=1}^{h} R^a_{i,t+j} = \beta_i + \beta_S NS_{i,t} + \beta_Q NQ_{i,t} + \beta_C[R^a_{i,t}\, RV_{i,t}\, VLM_{i,t}] + \epsilon^h_{i,t+1}$ for forecasting fifteen-minute returns ($h = 3$). I estimate the model for the individual stocks in the Dow between 9:30am to 10:30am. The explanatory variables are standardized to be mean zero and standard deviation one. The results are reported for the subset of stocks in the Dow that were members during the entire 2008 to 2013 sample period ($N = 21$) to ensure the regressions are run over the same period of time for each stock. The plot highlights the cross-sectional differences in the strength of news sentiment and abnormal returns in forecasting returns. Similar to the panel regression results, most stocks have positive coefficients on news sentiment and negative coefficients on abnormal returns, many of which are statistically significant.

from 2008 to 2013. The results indicate that positive news sentiment forecasts high returns, low volatility, and low volumes, and that large news quantities forecast high volatility and high volumes. In the regressions, I focus on the first hour of trading because this is the intraday period when traders incorporate overnight information into prices. In particular, one might hypothesize that the news variables would deliver more accurate forecasts at the open because overnight news increases the risk of informed trading and adverse selection. Empirically, I confirm this conjecture.[23] While the qualitative results in the panel regressions are similar throughout the trading day, the most significant forecasts are obtained in the morning.

Additionally, the results indicate that the news variables remain significant when lagged values of the controls are included in the regressions. This is notable as several of the non-news variables are known to forecast returns, volatility, or volume. For example, the negative coefficient on lagged abnormal returns in the return forecasting regression provides evidence of intraday mean reversion while the positive coefficients on realized variance and volume in the volatility and volume regressions are consistent with the fact that high frequency volatility and volume are highly persistent. After controlling for these relationships and standardizing the forecasting variables, I find that a one standard deviation increase in news sentiment forecasts a .20bp increase in abnormal returns and .86bp decrease in absolute returns at a five-minute horizon. Similarly, a one standard deviation increase in news quantity forecasts a .36bp increase in absolute returns and a 14% increase in the dollar volume traded. Finally, I confirm as a robustness check that the news variables remain significant and deliver similar forecasts at longer horizons.[24]

In addition to the panel regressions, Figure 2.8 illustrates the regression coefficients and t-statistics for news sentiment and abnormal returns when forecasting returns at the level of the firm. This contrasts the panel regressions which implicitly assume the stocks in the Dow have the same exposure to the forecasting variables. In particular, Figure 2.8 highlights

[23]Panel regressions for the full trading day are reported in Table 2.9 in the Appendix.

[24]The results for the panel regressions over longer horizons are available upon request.

the cross-sectional differences in the regression coefficients for news sentiment and abnormal returns across firms. While there is some heterogeneity, the results also indicate that most stocks have positive coefficients on news sentiment and negative coefficients on abnormal returns which is qualitatively similar to the panel regressions.

To present another perspective, Table 2.5 reports Fama and MacBeth (1973) regressions for forecasting returns at different horizons. Similar to the previous results, the Fama-MacBeth regressions indicate that positive news sentiment forecasts higher returns. Moreover, the Fama-MacBeth regressions indicate that the magnitude of the coefficient on news sentiment increases with horizon. For example, in the third specification which includes all of the forecasting variables in the regression, a one standard deviation increase in news sentiment forecasts a .23bp increase in abnormal returns at a ten-minute horizon versus a .57bp increase in abnormal returns at a thirty-minute horizon. As before, there is also a negative coefficient on lagged abnormal returns which provides evidence of intraday mean reversion. Similar to news sentiment, the magnitude of the coefficient on lagged returns increases with horizon.

Additionally, the Fama and MacBeth (1973) regressions provide insight into the time series behavior of the return predictability. Figure 2.9 plots three-month rolling averages of the coefficients on news sentiment and lagged abnormal returns in addition to the R^2_{adj} from the cross-sectional regressions. The results indicate that the greatest return predictability coincided with the financial crisis from the end of 2008 through early 2009. This is the period with the largest R^2_{adj} and the largest magnitudes of the regression coefficients on news sentiment and lagged returns. The plot also indicates that predictability has decreased in recent years. While this might be attributed to the proliferation of news analytics services or the rise of high frequency trading, there is also evidence that the decrease in return predictability is related to the low levels of market volatility that have followed the crisis. For example, I find that the correlation between the realized volatility of the SPY exchange-

Forecasting Returns: Fama-MacBeth Regressions

Horizon	5 mn	10 mn	15 mn	30 mn	60 mn
Model (1)					
NS	0.19***	0.32***	0.43***	0.74***	1.04***
	[4.33]	[4.11]	[3.95]	[3.97]	[3.8]
NQ	-0.08	-0.15	-0.18	-0.2	-0.23
	[-1.55]	[-1.58]	[-1.31]	[-0.9]	[-0.7]
Model (2)					
NS	0.12***	0.2***	0.28***	0.52***	0.66***
	[3.17]	[3.12]	[3.18]	[3.64]	[3.1]
NQ	-0.05	-0.07	-0.08	-0.1	-0.13
	[-1.08]	[-0.79]	[-0.74]	[-0.55]	[-0.45]
Ra	-0.66***	-1.09***	-1.06***	-1.07***	-1.33***
	[-10.04]	[-11.9]	[-10.33]	[-9.09]	[-8.82]
RV	-0.22***	-0.32**	-0.41*	-0.76**	-1.56***
	[-2.58]	[-2.2]	[-1.94]	[-2.07]	[-2.68]
VLM	0.15***	0.19**	0.34***	0.58***	0.91***
	[2.59]	[1.97]	[2.75]	[3.04]	[3.07]
Model (3)					
NS	0.11***	0.23***	0.32***	0.57***	0.71***
	[2.77]	[3.43]	[3.54]	[3.9]	[3.32]
NQ	-0.03	-0.04	-0.07	-0.09	-0.19
	[-0.54]	[-0.49]	[-0.63]	[-0.47]	[-0.68]
Ra	-0.46***	-0.9***	-0.9***	-0.76***	-1.16***
	[-5.57]	[-7.99]	[-7.13]	[-4.79]	[-5.6]
RV	-0.24***	-0.34**	-0.43**	-0.79**	-1.56***
	[-2.65]	[-2.22]	[-2.04]	[-2.17]	[-2.69]
VLM	0.12*	0.16	0.37***	0.61***	1.02***
	[1.74]	[1.42]	[2.67]	[2.88]	[3.06]
SPD	0.05	0.07	0.13	0.24	0.21
	[1.14]	[0.95]	[1.34]	[1.64]	[0.95]
DPH	-0.06	-0.04	-0.07	-0.07	-0.08
	[-1.43]	[-0.62]	[-0.75]	[-0.53]	[-0.44]
ILQ	-0.08	-0.11	-0.11	-0.09	-0.02
	[-1.47]	[-1.49]	[-1.25]	[-0.75]	[-0.13]
DIB	0.06	0.1*	0.06	0.15	0.37***
	[1.39]	[1.7]	[0.79]	[1.48]	[2.72]
OIB	-0.34***	-0.48***	-0.49***	-0.61***	-0.52***
	[-7.76]	[-7.76]	[-6.67]	[-5.89]	[-3.84]

Newey-West t-statistics: * p<0.10 ** p<0.05 *** p<0.01

Table 2.5: This table reports Fama and MacBeth (1973) regressions for forecasting abnormal returns in the model $\sum_{j=1}^{h} R^a_{i,t+j} = \beta_i + \beta_S NS_{i,t} + \beta_Q NQ_{i,t} + \beta_C Controls_{i,t} + \epsilon^h_{i,t+1}$ at different horizons $h \in \{1, 2, 3, 6, 12\}$ for the stocks in the Dow between 9:30am and 10:30am from 2008 to 2013. The coefficients are the time series averages from the cross sectional regressions with Newey and West (1987) t-statistics reported to account for overlapping returns. The results indicate that news sentiment significantly forecasts returns with a magnitude that increases with the forecasting horizon.

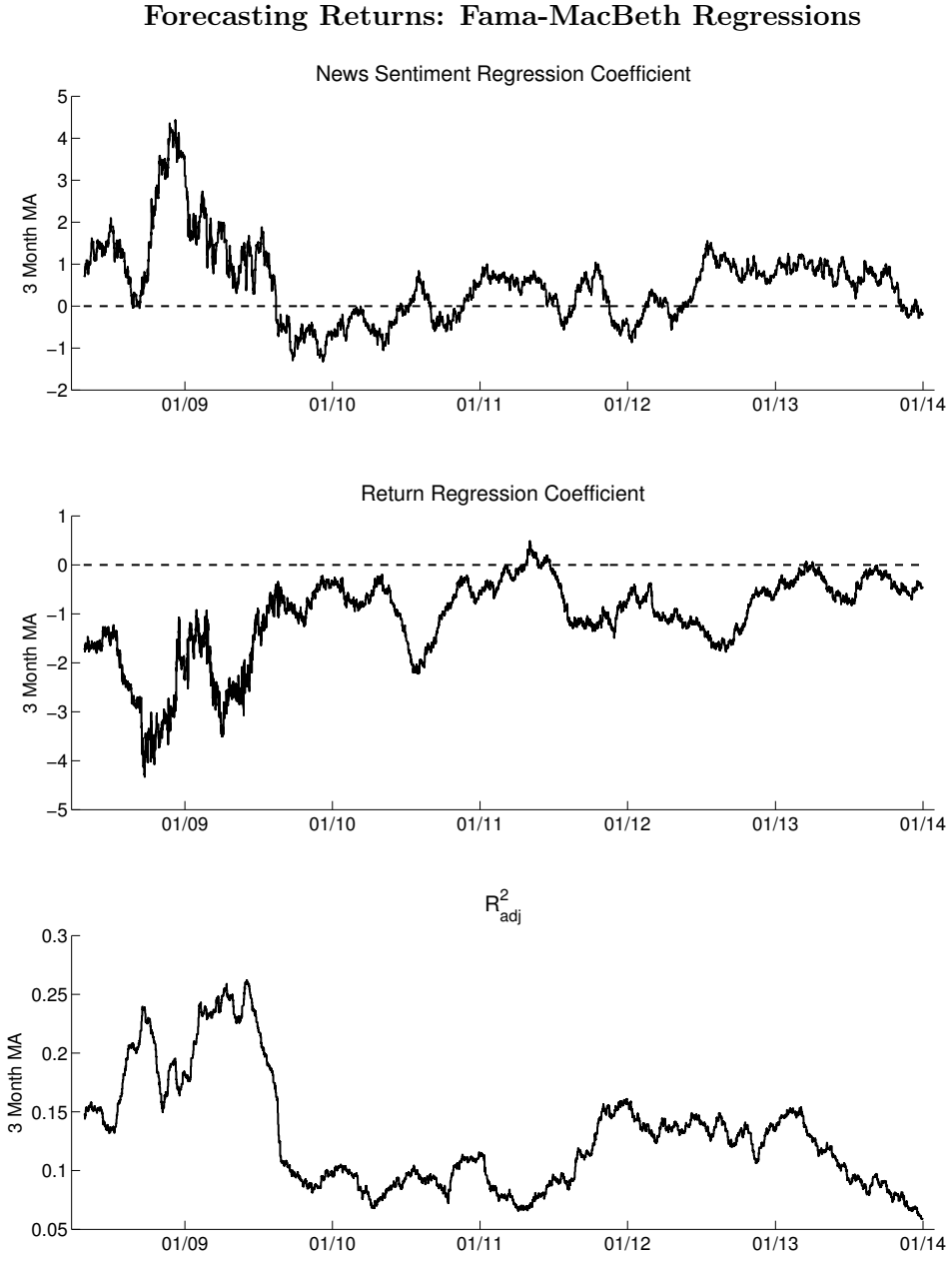


Figure 2.9: This figure plots the three-month moving average of the explanatory power R^2_{adj} and the regression coefficients for news sentiment $NS_{i,t}$ and abnormal returns in Model (2) from the Fama and MacBeth (1973) cross-sectional regressions in Table 2.5 at a thirty minute horizon ($h = 6$). The plot suggests that return predictability has varied over time. In particular, the period with the strongest predictability coincides with the financial crisis from late 2008 through early 2009. Since then, predictability has decreased which can be seen by the lower values of the regression coefficients and the smaller R^2_{adj} at the end of the sample period.

traded fund and the three-month moving average of the cross-sectional R^2_{adj} is 47% during my sample period.

In summary, the regression analysis in this section provides evidence that news sentiment forecasts returns and that news quantity forecasts volatility and volume. To investigate the economic significance of these findings, I explore whether the news data can be used to enhance contrarian trading strategies in the next section.

2.6 News Adjusted Contrarian Trading Strategies

In this section, I construct the returns from news adjusted contrarian trading strategies to illustrate the news sentiment's economic significance in forecasting returns. Specifically, I modify a market-making strategy outlined in Khandani and Lo (2011) and a related liquidity providing strategy to only take long (short) positions when news sentiment is non-negative (non-positive). In comparison to the original contrarian strategies, the news adjusted strategies produce higher returns and Sharpe ratios with a substantial amount of the outperformance occurring from the end of 2008 through early 2009. This is consistent with the results from the previous section and is the same period when news sentiment obtains the greatest return predictability in the Fama and MacBeth (1973) regressions. As before, I focus on the first hour of trading from 9:30am to 10:30am. In particular, I find the news adjustment is most valuable early in the day when there is a heightened risk of adverse selection as a result of overnight and pre-open news.

Contrarian strategies provide liquidity in response to price movements. They are theoretically motivated by Grossman and Miller (1988) and Kyle (1985). In these frameworks, the provision of immediacy and the risk of being asymmetrically selected result in mean-reverting prices. To motivate the news adjusted strategies, Campbell et al. (1993) observe that changes in prices are driven by supply shocks, which create mean-reversion, and new information about cash flows and discount rates, which are permanent. I interpret the news sentiment as a signal about new information and modify the original strategies to avoid trading when

changes in prices are consistent with the news. To be specific, the first strategy, which I call the contrarian strategy, follows Khandani and Lo (2011). Every five minutes from 9:35am to 10:25am I buy (sell) the stocks in the Dow whose lagged five-minute abnormal returns are in the lowest (highest) tercile. This can be interpreted as buying the losers and selling the winners, which is the opposite of a momentum strategy. Portfolio returns are computed as

$$R^p_{i,t} = \frac{\sum_{i=1}^{N} \omega_{i,t} Ra_{i,t}}{\frac{1}{2}\sum_{i=1}^{N} |\omega_{i,t}|} \tag{2.3}$$

with $\omega_{i,t} \in \{-1, 0, 1\}$ for short, neutral, and long positions. This assumes Regulation-T margin is satisfied with invested capital equal to fifty percent of the market value of the net long and short positions. I then modify the baseline contrarian strategy to only go long (short) the stocks in the Dow with non-negative (non-positive) news sentiment whose lagged five-minute abnormal returns are in the lowest (highest) tercile, which I call the contrarian news strategy. Compared to the baseline strategy, the contrarian news strategy takes into account the sentiment of the news when formulating new positions. Because the contrarian news strategy only trades when news sentiment is not in the same direction as lagged returns, it trades less frequently and has about half the number of positions and half the portfolio turnover when compared to the baseline contrarian strategy.

I also consider a strategy that provides liquidity for stocks that exhibit large price movements relative to realized volatility. This differs from the contrarian trading strategy described above which only takes into account the absolute size of returns. In particular, I estimate the spot variance from the truncated realized variance of lagged returns over the previous day at each intraday point in time. To mitigate the impact of microstructure noise I use a sparse five-minute sampling frequency and average the truncated realized variance across subsamples. This avoids discarding data and reduces the estimator's variability following Zhang et al. (2005).[25] I then go long (short) stocks whose lagged five-minute

[25]I use lagged returns when computing the realized variance to avoid look ahead bias in the forecasting regressions and trading strategies. For further details about realized variance estimation see Aït-Sahalia and Jacod (2014) Chapters 7, 8, and 10. In computing the truncated realized variance, I truncate returns that

Summary Statistics for Contrarian Returns
In the First Hour of Trading

Strategy	C	CN	J	JN	MKT
Mean (bp)	0.88	1.24	1.45	2.01	0.02
Std. Deviation (bp)	14.60	18.78	44.20	41.66	16.45
Sharpe Ratio	0.06	0.07	0.03	0.05	0.00
t-Statistic	7.66	8.44	4.20	6.15	0.13
Min	-0.02	-0.03	-0.17	-0.12	-0.02
Max	0.03	0.02	0.11	0.14	0.04
Skewness	0.79	0.42	-3.41	1.29	0.74
Kurtosis	25.13	27.40	265.08	187.43	27.56
Autocorrelation	0.02	0.03	-0.03	-0.01	-0.00
Negative	0.47	0.47	0.42	0.38	0.48
Number Positions	19.91	10.80	5.44	2.92	
Turnover	0.87	0.47	0.32	0.17	

Return Correlation Matrix

C	CN	J	JN	MKT
1.00	0.73	0.49	0.38	0.01
	1.00	0.31	0.49	-0.05
		1.00	0.61	0.01
			1.00	-0.04
				1.00

Table 2.6: The tables above contrast the returns from high frequency contrarian trading strategies between 9:30am and 10:30am. The portfolios are rebalanced every five minutes and are hedged with the SPY ETF using realized betas as hedge ratios. In particular, the contrarian strategy (C) goes long (short) the stocks in the Dow whose lagged five-minute abnormal returns are in the lower (upper) tercile. The first trade occurs at 9:35am and no overnight positions are allowed. The jump strategy (J) goes long (short) the stocks in the Dow whose lagged abnormal returns exceed a negative (positive) two standard deviation move relative to the estimated spot variance. The contrarian news (CN) and jump news (JN) strategies modify the original strategies by only going long (short) when the news sentiment NS is non-negative (non-positive). The table indicates that the news adjustment increases the average returns and the Sharpe ratios for both strategies. Additionally, the news adjustment lowers the turnover and the number of positions relative to the baseline strategies. Finally, to provide a benchmark, I include the returns from the SPY ETF during the first hour of trading which are denoted as MKT.

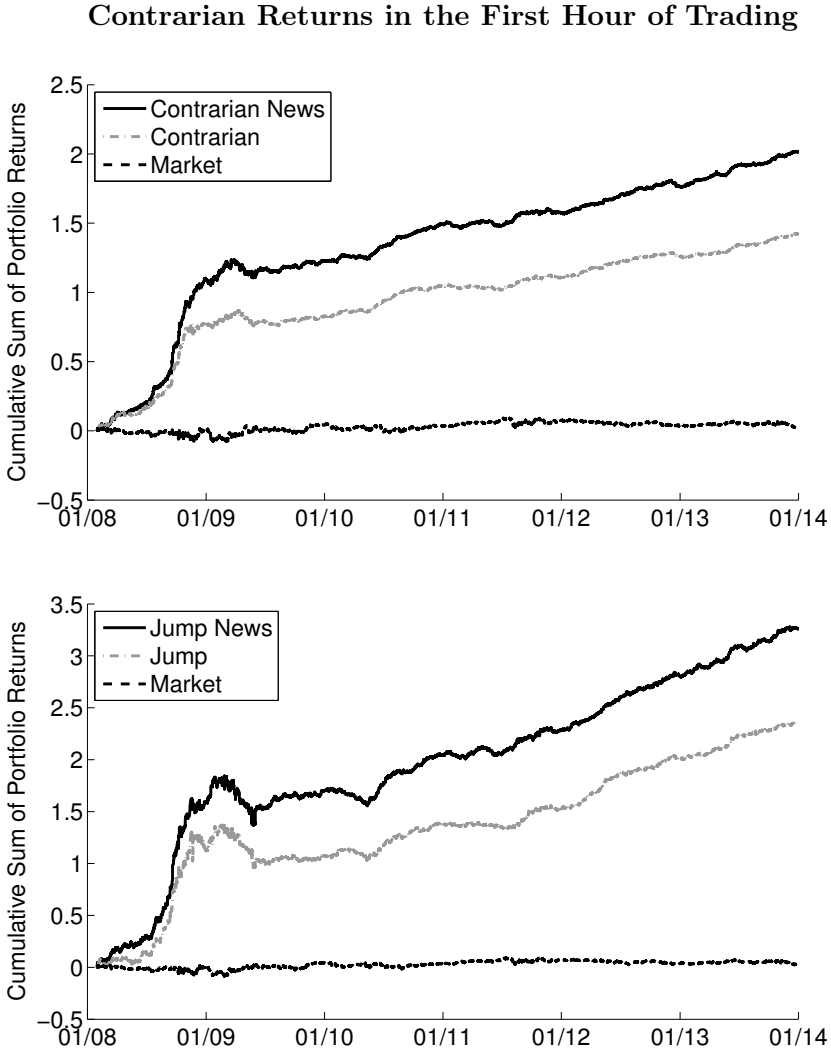


Figure 2.10: This figure plots the cumulative sum of returns (i.e., without compounding) for the high frequency contrarian trading strategies between 9:30am to 10:30am. The portfolios are rebalanced every five minutes and are hedged with the SPY ETF using realized betas as hedge ratios. The top plot shows the contrarian and contrarian news strategies. The contrarian strategy goes long (short) the stocks in the Dow whose abnormal returns are in the lower (upper) tercile of lagged five-minute returns with the first trade occurring at 9:35am. The bottom plot shows the jump and jump news strategies. The jump strategy goes long (short) the stocks in the Dow whose lagged abnormal returns exceed a negative (positive) two standard deviation move relative to the estimated spot variance. The news strategies follow the original strategies but only going long (short) when the news sentiment NS is non-negative (non-positive). The plot indicates that the news adjustment improved the performance of the contrarian strategies substantially at the end of 2008 and early 2009. Additionally, under the idealized assumption of no transaction costs, the plot indicates that the contrarian strategies produced significantly higher returns than the market.

abnormal returns exceed a negative (positive) two standard deviation move relative to the estimated spot variance. I call this the jump strategy based on the interpretation of large price movements as jump discontinuities as in the jump detection test of Lee and Mykland (2008). Then, I modify the jump strategy to only go long (short) stocks with non-negative (non-positive) news sentiment whose lagged five-minute abnormal returns exceed a negative (positive) two standard deviation move relative to the estimated spot variance. I refer to the modified strategy as the jump news strategy. As before, the news strategy has about half the number of positions and turnover in comparison to the baseline strategy.

I report summary statistics from the five-minute returns of the various strategies in Table 2.6. The results indicate that the news adjustment increases the average returns and the Sharpe ratios of the baseline strategies. Specifically, the news adjustment increases the average returns from .88bp to 1.24bp for the contrarian strategy and from 1.45bp to 2.01bp for the jump strategy. The news adjustment also increases the standard deviation from 15bp to 19bp for the contrarian strategy, which is partly driven by a diversification effect as the average number of positions decreases from 20 to 11. For the jump strategy, the news adjustment actually decreases the standard deviation from 44bp to 42bp despite the smaller number of positions. It also increases the skewness from -3.41 to 1.29. In both cases, the increase in average returns more than compensates for the change in volatility leading to higher Sharpe ratios. Additionally, I find that the returns from the contrarian strategies are significantly higher than the corresponding returns for the market. To be specific, the SPY exchange traded fund earned an insignificant .02bp during the first hour of trading in my sample period. Of course, the caveat in comparing the market returns with the contrarian returns is that I do not account for transaction costs. As the table indicates, the contrarian strategies require a large number of trades and a high portfolio turnover rate. Finally, I plot the cumulative returns of the various strategies in Figure 2.10 to highlight their performance

are more than a four standard deviation move relative to a first stage estimate of the realized variance. In this context, large price movements may be interpreted as jumps as in the jump detection test of Lee and Mykland (2008).

over time. While all of the contrarian strategies performed well during the financial crisis, Figure 2.10 indicates that the news adjustment was particularly valuable from the end of 2008 through early 2009 and that the outperformance of the news strategies persisted throughout the sample period.

2.7 Conclusion

In summary, the analysis in this paper highlights the forecasting content of machine-readable news. During the trading day, realized variance and volume increase over 40% after short relevant news is released, and signed returns exhibit a statistically and economically significant drift of 4 basis points in the five minutes that follow short relevant news. In addition to these unconditional reactions, I show that similar results continue to hold after accounting for firm-year fixed effects, intraday fixed effects, and a number of control variables. Moreover, my regression analysis highlights the difference between positive and negative news. In contrast to positive news which has a longer post-announcement drift, negative news is preceded by a pre-announcement drift which suggests the market anticipated negative but not positive news for the stocks in the Dow during my sample period.

The market reaction to intraday news motivates my subsequent analysis of the forecasting content of machine-readable news. In particular, I show that positive news sentiment forecasts higher returns and lower volatility, and that large quantities of news forecasts higher volatility and larger trading volumes. I illustrate the robustness of these results in panel and Fama-MacBeth cross-sectional regressions and in multiple regression analysis for individual stocks. Then, to demonstrate the economic significance of the return predictability, I show that the news sentiment can be used to improve the performance of contrarian trading strategies. In particular, news adjusted contrarian strategies that only go long when news sentiment is non-negative and short when news sentiment is non-positive earn higher Sharpe ratios and higher average returns in comparison to baseline contrarian strategies that ignore the news and only trade on intraday mean reversion. I also find the most significant

return predictability and outperformance of the news adjusted strategies during the financial crisis, which suggests that news sentiment is valuable during periods with increased risks of asymmetric information and informed trading. Finally, I document intraday patterns in the arrival of news within the hour and within the minute and demonstrate that increases in news arrival rates are contemporaneous with increases in realized variance and volume.

These results contribute to the literature that studies the impact of unscheduled news on financial markets by documenting a number of new empirical facts related to the interaction between asset prices and machine-readable news within the trading day. Moving forward, this research motivates further investigation along a number of lines. For example, more could be done to study the causal relationship between volatility, volume, and intraday news. While my results suggest that intraday news causes increases in realized volatility and trading volumes at various frequencies, I do not rule out other microstructure explanations that could be driving the patterns such as order execution algorithms. Additionally, it would be useful to differentiate the impact of overnight and intraday news on prices. The challenge, of course, is that markets are closed when overnight news is released. Nonetheless, overnight news is clearly important given the fact that it accounts for a majority of firm-specific news releases. Two possible approaches to study overnight news include an analysis of after hours trading around earnings announcements or a study of the impact of pre-open news on opening auctions. Additionally, it would be interesting to know whether the results in this paper would be enhanced by using the combined data from multiple news analytics service providers. In this study, I have relied on intraday news data from Thomson Reuters. There are now several news analytics services including Thomson Reuters, RavenPack, and Bloomberg which provide machine-readable news by covering different financial publishers. To the extent that these companies compete over stories, using data from one provider introduces measurement error. Finally, it would be useful to know which of the news analytics services provides the most relevant information to the market both in terms of the speed

that stories are conveyed to subscribers and in terms of the accuracy of the natural language processing algorithms that are used to analyze the news.

References

Admati, A. R. and P. Pfleiderer (1988). A theory of intraday patterns: Volume and price variability. *Review of Financial Studies 1*(1), 3–40.

Aït-Sahalia, Y. and J. Jacod (2014). *High-Frequency Financial Econometrics*. Princeton University Press.

Aït-Sahalia, Y. and M. Saglam (2014). High frequency traders: Taking advantage of speed. *SSRN Working Paper*.

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

Andersen, T. G. and T. Bollerslev (1998). Deutsche mark-dollar volatility: Intraday activity patterns, macroeconomic announcements, and longer run dependencies. *Journal of Finance 53*(1), 219–265.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and C. Vega (2003). Micro effects of macro announcements: Real-time price discovery in foreign exchange. *American Economic Review 93*(1), 38–62.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and C. Vega (2007). Real-time price discovery in global stock, bond and foreign exchange markets. *Journal of International Economics 73*, 251–277.

Antweiler, W. and M. Z. Frank (2004). Is all that talk just noise? The information content of internet stock message boards. *Journal of Finance 59*(3), 1259–1294.

Balduzzi, P., E. J. Elton, and T. C. Green (2001). Economic news and bond prices: Evidence from the u.s. treasury market. *Journal of Financial and Quantitative Analysis 36*(4), 523–543.

Ball, R. and P. Brown (1968). An empirical evaluation of accounting income numbers. *Journal of Accounting Research*, 159–178.

Beaver, W. H. (1968). The information content of annual earnings announcements. *Journal of Accounting Research*, 67–92.

Bernanke, B. S. and K. N. Kuttner (2005). What explains the stock market's reaction to federal reserve policy? *Journal of Finance 60*(3), 1221–1257.

Bernard, V. L. and J. K. Thomas (1989). Post-earnings-announcement drift: Delayed price response or risk premium? *Journal of Accounting Research 27*, 1–36.

Berry, T. D. and K. M. Howe (1994). Public information arrival. *Journal of Finance 49*(4), 1331–1346.

Black, F. (1986). Noise. *Journal of Finance 41*(3), 529–543.

Boyd, J. H., J. Hu, and R. Jagannathan (2005). The stock market's reaction to unemployment news: Why bad news is usually good for stocks. *Journal of Finance 60*(2), 649–672.

Campbell, J. Y., S. J. Grossman, and J. Wang (1993). Trading volume and serial correlation in stock returns. *Quarterly Journal of Economics 108*(4), 905–939.

Chari, V. V., R. Jagannathan, and A. R. Ofer (1988). Seasonalities in security returns: The case of earnings announcements. *Journal of Financial Economics 21*(1), 101–121.

Cutler, D. M., J. M. Poterba, and L. H. Summers (1989). What moves stock prices? *Journal of Portfolio Management 15*(3), 4–12.

Dubinsky, A. and M. Johannes (2006). Fundamental uncertainty, earnings announcements and equity prices. *Columbia University Working Paper*.

Ederington, L. H. and J. H. Lee (1993). How markets process information: News releases and volatility. *Journal of Finance 48*(4), 1161–1191.

Engelberg, J. E. and C. A. Parsons (2011). The causal impact of media in financial markets. *Journal of Finance 66*(1), 67–97.

Fama, E. F. and J. D. MacBeth (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy 81*, 607–636.

Flannery, M. J. and A. A. Protopapadakis (2002). Macroeconomic factors do influence aggregate stock returns. *Review of Financial Studies 15*(3), 751–782.

Fleming, M. J. and E. M. Remolona (1999). Price formation and liquidity in the u.s. treasury market: The response to public information. *Journal of Finance 54*(5), 1901–1915.

Foucault, T., J. Hombert, and I. Roşu (2015). News trading and speed. *Journal of Finance, Forthcoming*.

Frazzini, A. and O. A. Lamont (2007). The earnings announcement premium and trading volume. *NBER Working paper no. 13090*.

French, K. R. and R. Roll (1986). Stock return variances: The arrival of information and the reaction of traders. *Journal of Financial Economics 17*(1), 5–26.

Green, T. C. (2004). Economic news and the impact of trading on bond prices. *Journal of Finance 59*(3), 1201–1234.

Groß-Klußmann, A. and N. Hautsch (2011). When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions. *Journal of Empirical Finance 18*(2), 321–340.

Grossman, S. J. and M. H. Miller (1988). Liquidity and market structure. *Journal of Finance 43*(3), 617–633.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press Princeton.

Jegadeesh, N. and D. Wu (2013). Word power: A new approach for content analysis. *Journal of Financial Economics 110*, 712–729.

Jones, C. M. (2013). What do we know about high-frequency trading? *Columbia Business School Research Paper* (13-11).

Jones, C. M., O. Lamont, and R. L. Lumsdaine (1998). Macroeconomic news and bond market volatility. *Journal of Financial Economics 47*(3), 315–337.

Khandani, A. E. and A. W. Lo (2011). What happened to the quants in august 2007? Evidence from factors and transactions data. *Journal of Financial Markets 14*(1), 1–46.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica 53*(6), 1315–1335.

Lee, C. and M. J. Ready (1991). Inferring trade direction from intraday data. *Journal of Finance 46*(2), 733–746.

Lee, S. S. and P. A. Mykland (2008). Jumps in financial markets: A new nonparametric test and jump dynamics. *Review of Financial Studies 21*(6), 2535–2563.

Loughran, T. and B. McDonald (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-ks. *Journal of Finance 66*(1), 35–65.

Lucca, D. O. and E. Moench (2015). The pre-fomc announcement drift. *Journal of Finance 70*(1), 329–371.

Mitra, G. and L. Mitra (Eds.) (2011). *The Handbook of News Analytics in Finance*. John Wiley & Sons.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

O'Hara, M., C. Yao, and M. Ye (2014). What's not there: Odd lots and market data. *Journal of Finance 69*(5), 2199–2236.

Patton, A. J. and M. Verardo (2012). Does beta move with news? Firm-specific information flows and learning about profitability. *Review of Financial Studies 25*(9), 2789–2839.

Peress, J. (2014). The media and the diffusion of information in financial markets: Evidence from newspaper strikes. *Journal of Finance 69*(5), 2007–2043.

Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies 22*(1), 435–480.

Roll, R. (1988). R2. *Journal of Finance 43*(3), 541–566.

Savor, P. and M. Wilson (2013). How much do investors care about macroeconomic risk? Evidence from scheduled economic announcements. *Journal of Financial and Quantitative Analysis 48*(2), 343–375.

Savor, P. and M. Wilson (2014). Asset pricing: A tale of two days. *Journal of Financial Economics 113*(2), 171–201.

Savor, P. G. and M. I. Wilson (2015). Earnings announcements and systematic risk. *Journal of Finance, Forthcoming*.

Shiller, R. (1981). Do stock prices move too much to be justified by subsequent dividends. *American Economic Review 71*(3), 421–436.

Stroud, J. R. and M. S. Johannes (2014). Bayesian modeling and forecasting of 24-hour high-frequency volatility. *Journal of the American Statistical Association 109*(508), 1368–1384.

Tetlock, P. (2011). All the news that's fit to reprint: Do investors react to stale information? *Review of Financial Studies 24*(5), 1481–1512.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance 62*(3), 1139–1168.

Tetlock, P. C., M. Saar-Tsechansky, and S. Macskassy (2008). More than words: Quantifying language to measure firms' fundamentals. *Journal of Finance 63*(3), 1437–1467.

Von Beschwitz, B., D. B. Keim, and M. Massa (2015). First to read the news: News analytics and institutional trading. *SSRN Working Paper*.

Wood, R. A., T. H. McInish, and J. K. Ord (1985). An investigation of transactions data for nyse stocks. *Journal of Finance 40*(3), 723–739.

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association 100*(472), 1394–1411.

APPENDIX

2.A Variable Definitions

The tables below report definitions and summary statistics for the variables used in the empirical analysis. I compute the variables every five minutes from 9:30am to 4:00pm for the stocks in the Dow on trading days from 2008 to 2013 excluding half-sessions and the flash crash on May 6th, 2010.[26] In total, the sample includes 1,477 days and 3.44 million observations per variable.

The news variables are derived from relevant first-take news in the Thomson Reuters database. I define the news sentiment $NS_{i,t}$ as the average sentiment from the ten most recent news items that are categorized as relevant first-takes for stock i at time t. By construction, news sentiment is between -1 and 1. It averages the sentiment scores from individual news items which are equal to -1, 0, or 1 for stories that are negative, neutral, or positive according to Thomson Reuters' algorithms. The summary statistics indicate that news sentiment has been positive on average, with a mean equal to .18, and roughly symmetric during my sample period. I also compute the quantity of news. Specifically, the quantity of news $NQ_{i,t}$ is the number of standardized relevant first-takes that have been released between the previous day's close and the intraday time t for stock i.[27] In contrast to sentiment, news quantity

[26] To mitigate the influence of outliers on the regression analysis in Section 2.5, I truncate the variables at the upper end of the range indicated in the summary statistics table. The results are robust to whether the variables are truncated or not. In particular, the truncation is above the 99th quantile for each of the variables. Returns are not truncated.

[27] I standardize the news quantity every five minutes from the open to the close with the mean and standard deviation of the lagged news quantity over the previous sixty business days at the analogous intraday time.

is skewed to the right with a mean and standard deviation that are approximately equal to zero and one. The pointwise correlation between news sentiment and news quantity is -.05 which is close to zero.

Intraday log returns are obtained from NBBO mid-prices. I denote these returns as $R_{i,t}$, $R^a_{i,t}$, and $R^e_{i,t}$ where abnormal and excess returns are defined as $R^a_{i,t} \equiv R_{i,t} - \hat{\beta}_{i,t} R_{m,t}$ and $R^e_{i,t} \equiv R_{i,t} - R_{m,t}$. I use returns from the SPY exchange traded fund to proxy for the market $R_{m,t}$. The realized beta $\hat{\beta}_{i,t}$ is estimated from twenty days of lagged five-minute returns with a subsample estimator following Zhang et al. (2005). The summary statistics table reports the mean and standard deviation of returns in basis points. In particular, I find that the standard deviation of five-minute log returns is 23 basis points which is larger than the standard deviations of abnormal returns and market returns which are equal to 18.5 basis points and 13 basis points respectively. The table also indicates that returns exhibit negative skewness and excess kurtosis, which are well known features of high frequency stock returns. To estimate the realized variance $RV_{i,t}$, I compute the subsample estimate from one day of lagged five-minute returns. The summary statistics table reports the mean and standard deviation of the annualized realized variance assuming there are 250 trading days per year with 78 five-minute intervals per day. The table indicates that realized variance is skewed to the right. It is also a highly persistent time series.

In addition to returns and realized variance, I also compute the volume, spread, depth, depth imbalance, order balance, and illiquidity over five minute intraday intervals.[28] In particular, volume $VLM_{i,t}$ refers to the dollar volume traded during an intraday interval relative to the average volume over the previous twenty days. Spread $SPD_{i,t}$ is the average spread during an intraday interval relative to the average spread over the previous twenty

[28]My implementation of the Lee and Ready (1991) algorithm proceeds as follows. The dollar volume traded each second is classified as buyer (seller) initiated if the average traded price weighted by volume is closer to the last best offer (best bid) from the previous second. If the average traded price equals the midpoint of the NBBO at the end of the previous second a "tick test" classifies the dollar volume as buyer (seller) initiated when the average traded price increased (decreased) over the last second. I leave the volume unsigned if the average traded price equals the NBBO midpoint and average traded price from the previous second.

days. Depth $DPH_{i,t}$ is the dollar bid size plus dollar offer size during an intraday interval relative to the average depth over the previous twenty days. Illiquidity $ILQ_{i,t}$ is the absolute return divided by the dollar volume traded during an intraday interval relative to the average illiquidity over the previous twenty days, which is defined following Amihud (2002). The mean of volume, spread, depth, and illiquidity is close to one as each of these variables is defined relative to its lagged twenty day moving average. By way of contrast, the mean of depth imbalance and order imbalance are close to zero. Specifically, depth imbalance $DIB_{i,t}$ is the average bid minus offer size divided by the average bid plus offer size within an intraday interval. Order imbalance $OIB_{i,t}$ is the difference between buyer minus seller initiated dollar volume divided by the total dollar volume traded during an intraday interval. I classify trades as buyer or seller initiated with the Lee and Ready (1991) algorithm.

2.B Tables

Variable Definitions

Variable	Description	Definition
$NS_{i,t}$	News Sentiment	Average sentiment of last 10 news items
$NQ_{i,t}$	News Quantity	Standardized quantity of news
$R_{i,t}$	Return	Log return of stock i
$R_{m,t}$	Market Return	Log return of SPY
$Ra_{i,t}$	Abnormal Return	$R_{i,t} - \hat{\beta}_{i,t} R_{m,t}$
$Re_{i,t}$	Excess Return	$R_{i,t} - R_{m,t}$
$RV_{i,t}$	Realized Volatility	Estimated from one day of lagged returns
$VLM_{i,t}$	Volume	$\frac{\text{Dollar volume}}{\text{Avg. dollar volume previous 20 days}}$
$SPD_{i,t}$	Spread	$\frac{\text{Bid-offer spread}}{\text{Avg. bid-offer spread previous 20 days}}$
$DPH_{i,t}$	Depth	$\frac{\text{Dollar bid plus dollar offer size}}{\text{Avg. dollar bid plus dollar offer size previous 20 days}}$
$ILQ_{i,t}$	Illiquidity	$\frac{\lvert R_{i,t}\rvert \text{ divided by dollar volume}}{\text{Avg. } \lvert R_{i,t}\rvert \text{ divided by dollar volume previous 20 days}}$
$DIB_{i,t}$	Depth Imbalance	$\frac{\text{Bid minus offer size}}{\text{Bid plus offer size}}$
$OIB_{i,t}$	Order Imbalance	$\frac{\text{Buyer minus seller dollar volume}}{\text{Dollar Volume}}$

Variable Summary Statistics

Variable	Mean	Std. Dev.	Skew	Kurtosis	1st	25th	50th	75th	99th	Range
$NS_{i,t}$	0.18	0.43	-0.27	2.62	-0.80	-0.10	0.20	0.50	1.00	$[-1, 1]$
$NQ_{i,t}$	-0.01	0.93	3.01	14.21	-1.00	-0.48	-0.31	0.13	4.91	$[-5, 5]$
$R_{i,t}$	-0.04	23.21	-2.08	772.89	-62.24	-7.14	0.00	7.17	61.76	
$R_{m,t}$	0.02	13.16	0.40	27.95	-38.49	-4.87	0.00	5.00	37.92	
$R^a_{i,t}$	-0.06	18.56	-4.36	1759.54	-46.37	-5.58	0.00	5.54	45.93	
$R^e_{i,t}$	-0.06	19.17	-3.93	1581.03	-47.97	-5.86	0.00	5.82	47.56	
$RV_{i,t}$	0.08	0.20	6.35	50.91	0.01	0.02	0.03	0.06	1.13	$[0, 2]$
$VLM_{i,t}$	1.00	0.92	4.05	29.02	0.17	0.49	0.75	1.18	4.81	$[0, 10]$
$SPD_{i,t}$	1.00	0.29	7.24	79.15	0.60	0.92	0.97	1.02	2.19	$[0, 5]$
$DPH_{i,t}$	1.00	0.60	3.09	17.09	0.27	0.66	0.86	1.14	3.74	$[0, 5]$
$ILQ_{i,t}$	1.01	0.97	1.79	8.37	0.00	0.31	0.76	1.44	4.32	$[0, 10]$
$DIB_{i,t}$	0.00	0.15	0.01	4.29	-0.39	-0.08	0.00	0.09	0.40	$[-1, 1]$
$OIB_{i,t}$	-0.01	0.29	0.04	3.29	-0.71	-0.19	-0.01	0.17	0.70	$[-1, 1]$

Table 2.7: The tables above report definitions and summary statistics for the variables used in my empirical analysis. The sample period is 2008 to 2013 during which I compute the variables every five minutes from 9:30am to 4:00pm for the stocks in the Dow. To mitigate the impact of outliers on the forecasting regressions in Section 2.5, I truncate the variables at the upper end of the range. The results are robust to whether the truncation is performed or not. Returns are not truncated.

Example of Thomson Reuters News Analytics Data on 11/7/13

IDN_TIME	STOCK_RIC	RELEVANCE	SENTIMENT	TAKE_SEQNO	SENT_POS	SENT_NEUT	SENT_NEG	SENT_WORDS	TOT_WORDS	TOT_SENTS	NUM_CO	ORGID
20:26:00.00	MW.N	1.00	-1	1	0.14	0.18	0.68	224	239	11	2	54616
20:27:20.20	ATEA.O	1.00	1	1	0.84	0.12	0.04	319	418	22	1	82093
20:27:36.36	AAPL.O	0.13	0	1	0.28	0.71	0.01	47	710	31	3	19775
20:27:36.36	CSCO.O	0.13	0	1	0.28	0.71	0.01	47	710	31	3	21628
20:27:36.36	GOOG.O	0.13	0	1	0.28	0.71	0.01	47	710	31	3	100365604
20:28:02.02	S.N	1.00	1	1	0.68	0.17	0.15	828	828	31	1	112758474
20:29:23.23	ACCO.N	0.48	0	1	0.20	0.78	0.01	302	799	36	1	38056
20:29:44.44	ARE.TO	1.00	-1	1	0.08	0.14	0.78	144	222	18	1	19872
20:30:09.09	CTIP3.SA	1.00	1	1	0.77	0.18	0.05	23	23	1	1	106185112
20:30:45.45	FNP.N	1.00	1	1	0.85	0.12	0.03	21	21	2	1	26862
20:31:10.10	BIG.N	1.00	0	1	0.33	0.38	0.29	29	29	2	1	22119
20:31:11.11	PCG.N	1.00	-1	1	0.06	0.13	0.81	27	215	7	2	28906
20:31:28.28	TUC.N	1.00	-1	1	0.36	0.15	0.49	342	451	19	1	119862
20:31:30.30	DIS.N	0.05	-1	1	0.24	0.13	0.63	1010	1010	98	14	12289
20:31:30.30	FB.O	0.05	-1	1	0.24	0.13	0.63	1010	1010	98	14	102739130
20:31:30.30	NVDA.O	0.05	-1	1	0.24	0.13	0.63	1010	1010	98	14	133446
20:31:30.30	PCLN.O	0.05	-1	1	0.24	0.13	0.63	1010	1010	98	14	134648
20:31:30.30	QCOM.O	0.10	-1	1	0.06	0.13	0.82	106	1010	98	14	48662
20:31:30.30	TSLA.O	0.05	-1	1	0.06	0.13	0.82	90	1010	98	14	103324829

Table 2.8: The table above reports an example of the Thomson Reuters News Analytics (TRNA) data on November 7th, 2013. In this paper, I rely on the sentiment variable to sign the news as negative, neutral, or positive stories and on the relevance and take_seqno variables to identify relevant first-take news items. I also use the tot_words variable to find short relevant news. Beyond the variables included above, the TRNA dataset also includes the headline text, genre, and topic of the news as well as the number of linguistically linked stories over various horizons. Additionally, there are also indicator variables for sell-side analyst rating changes, market commentary, and alerts. For a comprehensive description of the TRNA dataset see Mitra and Mitra (2011).

Forecasting Returns, Volatility, and Volume from 9:30am to 4:00pm: Panel Regressions

	Returns			Volatility			Volume		
Model	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
NS	0.06***	0.05***	0.05***	-1.43***	-0.46***	-0.48***	-0.02***	-0.00*	-0.00
	[3.80]	[3.02]	[2.93]	[-4.43]	[-3.91]	[-3.95]	[-2.90]	[-1.76]	[-0.55]
NQ	-0.05**	-0.04**	-0.04**	0.81***	-0.26***	-0.24***	0.14***	0.05***	0.05***
	[-2.11]	[-2.35]	[-2.47]	[6.72]	[-5.68]	[-5.37]	[13.06]	[12.34]	[12.77]
Ra		-0.76***	-0.77***		-0.07**	-0.05		0.00***	0.00***
		[-6.70]	[-6.31]		[-2.13]	[-1.56]		[3.36]	[3.20]
RV		-0.14**	-0.14**		8.00***	7.82***		0.02***	0.03***
		[-2.23]	[-2.19]		[17.72]	[17.67]		[8.05]	[8.33]
VLM		0.01	0.04		3.33***	3.44***		0.58***	0.55***
		[0.17]	[0.83]		[19.25]	[16.28]		[58.04]	[48.19]
SPD			-0.07*			1.11***			-0.03***
			[-1.83]			[12.29]			[-7.44]
DPH			-0.03			-0.43***			0.10***
			[-0.81]			[-3.38]			[19.43]
ILQ			0.04			0.57***			0.00
			[1.63]			[9.94]			[1.35]
DIB			0.08***			-0.02			-0.00**
			[4.79]			[-0.95]			[-2.49]
OIB			0.07			-0.03			-0.00
			[0.97]			[-1.57]			[-1.04]
R^2_{adj}	0.000	0.002	0.002	0.084	0.292	0.300	0.024	0.407	0.420

robust t-statistics clustered by time and firm: * p<0.10, ** p<0.05, *** p<0.01

Table 2.9: This table reports panel regressions $Y_{i,t+1} = \beta_i + \beta_S NS_{i,t} + \beta_Q NQ_{i,t} + \beta_C Controls_{i,t} + \epsilon_{i,t+1}$ for forecasting returns, volatility, and volume at a five-minute frequency for the stocks in the Dow between 9:30am and 4:00pm from 2008 to 2013 ($T = 3,440,970$, $N = 39$). It differs from Table 2.4 with respect to the intraday time period. Instead of focusing on the first hour of trading, this table reports panel regressions for the full trading day from 9:30am to 4:00pm. While the qualitative results from the earlier regressions remain, the magnitude of the coefficients are lower. For example, in the regressions above a one standard deviation increase in news sentiment forecasts an increase in five minute returns of .04 basis points. This is significantly smaller than the .21 basis point increase during the first hour of trading. The coefficients are also smaller in magnitude for news quantity in the volume regressions and news sentiment in the volatility regressions. Additionally, the regressions above indicate that the sign on news quantity changes after including the additional variables in the volatility regressions, which is not the case during the first hour of trading.

2.C Figures

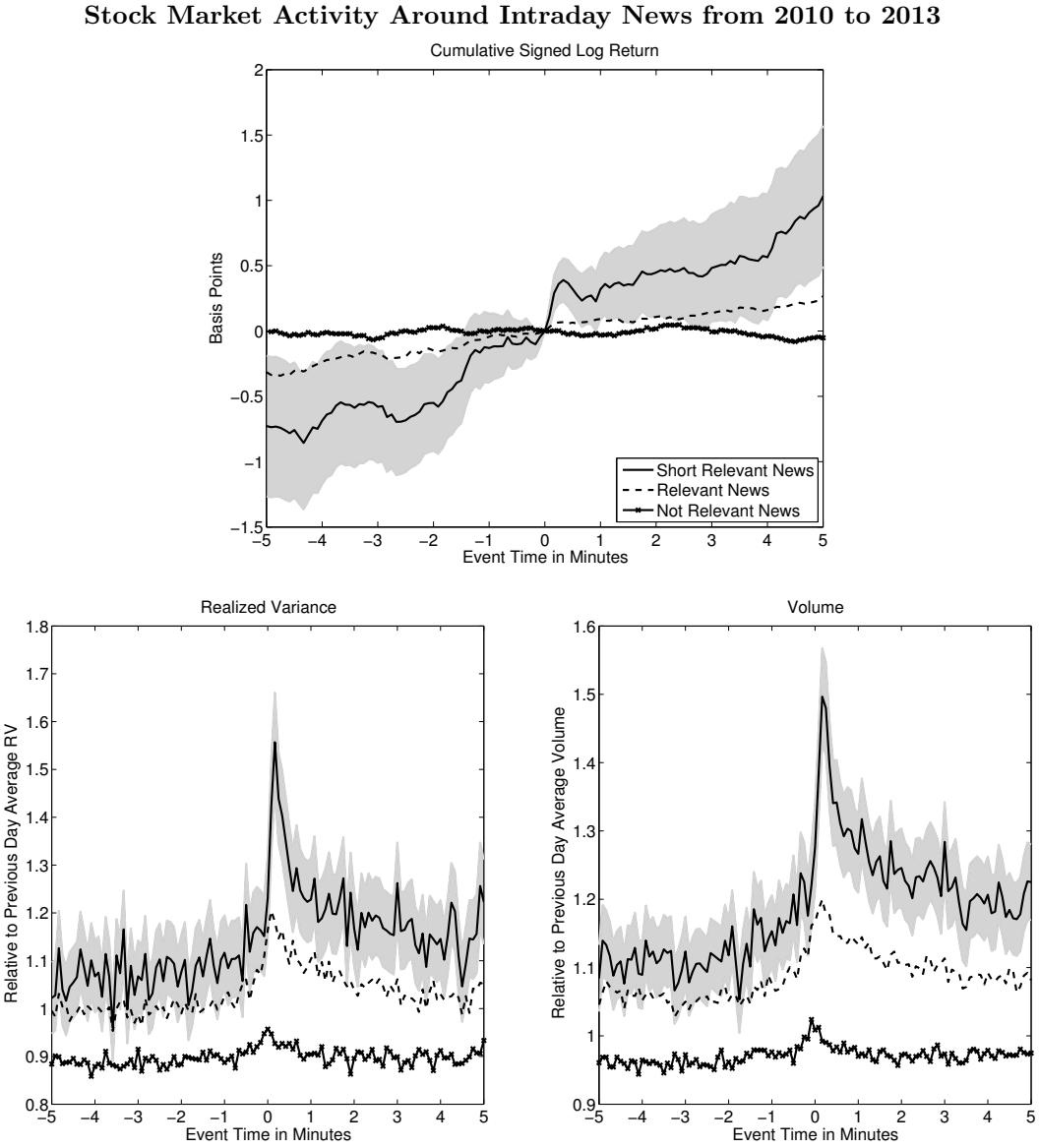


Figure 2.11: This figure plots the average cumulative signed abnormal returns, realized variance, and volume around intraday news for the Dow from 2010 to 2013. It is analogous to Figure 2.5 but focuses on the last four years of the sample which exclude the financial crisis. The 2010-2013 sample includes 5,374 short relevant news items, 29,813 relevant news items, and 47,388 not relevant news items. In comparison to Figure 2.5, the realized variance and volume increase about 50% after short relevant news which is a similar magnitude to the full sample. Meanwhile, signed abnormal returns exhibit a statistically and economically significant drift of 1 basis point five minutes after short relevant news arrives, which is a similar qualitative response but a smaller quantitative result in comparison to the 4 basis point when the financial crisis is included.

Histograms of Relevance, Sentiment, and Novelty News Variables

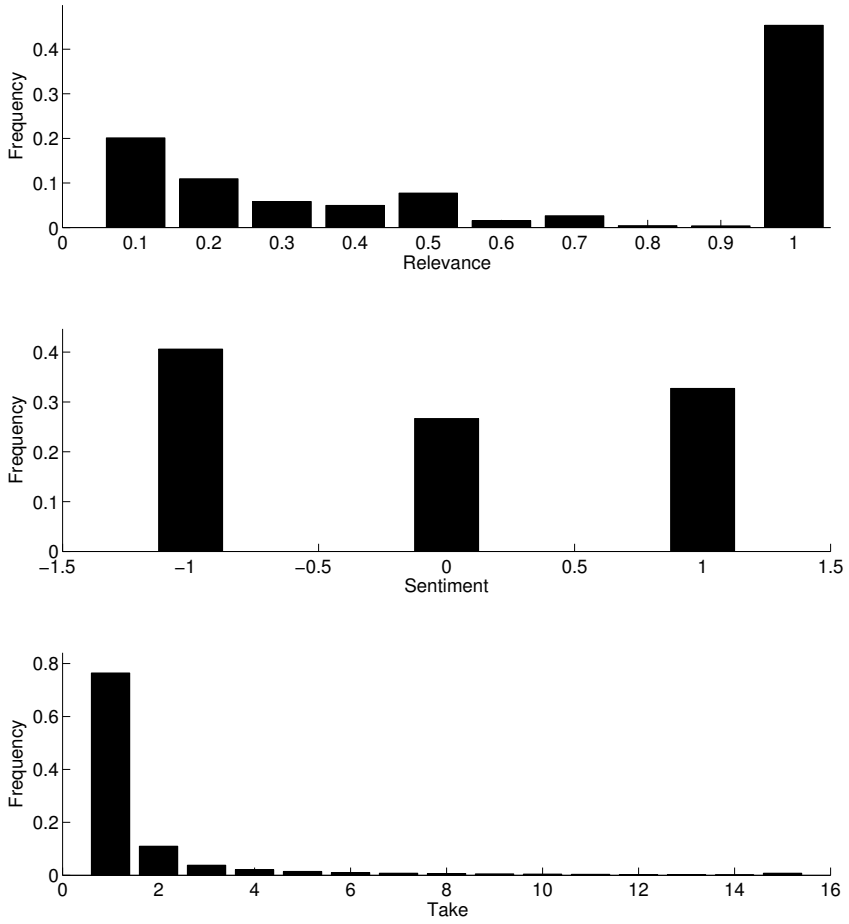


Figure 2.12: This figure reports histograms for the relevance, sentiment, and novelty variables in the Thomson Reuters News Analytics dataset for the stocks in the Dow during the 2008 to 2013 sample period. Relevance is a real valued number between zero and one indicating how relevant a news item is for a particular company. Higher values for relevance indicate the news is more relevant for a particular company. Sentiment indicates the tone of the author. It equals -1, 0, or 1 for negative, neutral, and positive stories as classified by Thomson Reuters' algorithms. Novelty is obtained by tracking the take or sequence number of a story. In this paper I often focus on relevant first-take news. This refers to news items for which the take is equal to one and the relevance is greater than or equal to .9.

CHAPTER 3

HIGH FREQUENCY TESTS OF CONDITIONAL LINEAR FACTOR MODELS

In this paper, I propose a nonparametric test of dynamic asset pricing models whose expected returns satisfy a conditional beta relationship. To test the conditional moment restriction on expected returns, I use high frequency data to estimate time varying betas with a local realized volatility estimator from the financial econometrics literature. Empirically, my test rejects the hypothesis that the conditional CAPM holds for the stocks in the Dow from 2008 to 2013. The impact of market microstructure noise and asynchronous data on my approach is also discussed.

3.1 Introduction

High frequency data can be used to test models with time varying coefficients. This is a subject of interest in empirical asset pricing as many dynamic asset pricing models predict that expected excess returns are a linear combination of time varying betas multiplied by conditional risk premiums on economic state variables. A leading example is the intertemporal capital asset pricing model (ICAPM) of Merton (1973) for which the conditional capital asset pricing model (CAPM) is a special case. These models are often called conditional linear factor models. The latter description stems from their implication that the stochastic discount factor can be expressed as a linear combination of economic state variables and time varying coefficients. From either the stochastic discount factor or beta representation, the insight remains the same. High frequency data permits the estimation of time varying coefficients which allows for a direct test of the conditional moment restriction on expected returns in dynamic asset pricing models.[1]

Factor models have their origin in the seminal work pioneered by Sharpe (1964) and Lintner (1965b). They show that the expected excess return for any asset is equal to beta multiplied by the excess return on the wealth portfolio in a static equilibrium with mean-variance investors. Along a different line of reasoning, Ross (1976a) shows the absence of arbitrage implies the risk premium for an asset is approximately equal to the beta weighted sum of factor risk premiums. Merton (1973) and Breeden (1979) extend these static results to a dynamic setting with time varying investment opportunities in which heterogeneous agents maximize the expected utility of lifetime consumption. Their crucial insight in the theory of portfolio choice is that optimal asset allocation requires more than rebalancing a market portfolio and risk free asset. Investors also demand hedging portfolios. In particular,

[1]My approach differs from the use of managed portfolio which are discussed in Chapter 8 of Cochrane (2005). For example, it follows that $E[z_t p_t] = E[m_{t+1} x_{t+1} z_t]$ for $z_t \in \mathcal{F}_t$ from $p_t = E_t[m_{t+1} x_{t+1}]$. Similarly, the parameterization $m_{t+1} = a(z_t) + b(z_t) f_{t+1}$ leads to an unconditional moment condition. In either of these cases, GMM can be used to test the asset pricing model. By way of contrast, my approach is to use high frequency data to estimate β_t so the conditional moment restriction $E_t[R_{t+1}] = \beta_t \lambda_t$ can be tested directly.

hedging portfolios are the portfolios of assets most highly correlated with economic state variables. They are used to hedge time varying investment opportunities that are correlated with shocks to wealth and consumption. In this setup, aggregating demands and imposing market clearing implies that excess expected returns equal time varying betas multiplied by the market's excess expected return and the excess expected return of the hedging portfolios.

Empirically, the static CAPM is rejected by the data. For example, Jensen et al. (1972) provide evidence that low beta assets earn higher returns than the CAPM predicts or equivalently that the empirical security market line is too flat. Additionally, the well known value, size, and momentum effects (Fama and French (1993), Jegadeesh and Titman (1993), Asness (1994)) are not explained by the CAPM despite having earned excess returns over long periods of time across different markets. One might attribute these mispricings to the CAPM's static construction. To correct for this, there are several approaches to account for time varying investment opportunities. In particular, one might use a multi-factor model where the additional factors serve as a proxy for the hedging portfolios as in the ICAPM. Of course, this approach is subject to the critique of Fama (1991) who famously dubbed the ICAPM a fishing license for practitioners. The concern is that researchers will find factors that explain the cross section of returns in-sample that have a dubious or unclear relation to the real economy. This can easily lead to spurious forecasts for out-of-sample returns. Campbell and Vuolteenaho (2004) provide a partial response to this criticism by estimating a two beta model with a bad cash flow beta and a good discount rate beta motivated by the approximate present value decomposition of Campbell and Shiller (1988). They conclude that their model adequately explains the value effect and is less easily criticized for its choice of hedging portfolios.

As an alternative to multiple factor models, another line of research investigates the possibility that the CAPM holds conditionally. As Hansen and Richard (1987) put it, a portfolio can be conditionally mean-variance efficient every period while not being mean-variance efficient unconditionally. This introduces the statistical problem of deciding how to

estimate a time varying beta. One approach followed by Lettau and Ludvigson (2001) is to estimate a conditional consumption CAPM. The theoretical justification for their approach is closely tied to the work of Breeden (1979). Breeden (1979) questions how the ICAPM can make salient predictions about expected returns given its silence on what constitutes the vector of economic state variables. To mitigate this problem he demonstrates in a dynamic setting analogous to that of the ICAPM that a single beta relationship holds where beta is measured as the covariance of an asset with instantaneous changes in aggregate real consumption. This representation stems from the observation that aggregate consumption is perfectly negatively correlated with the marginal utility of wealth. Lettau and Ludvigson (2001) set out to test a conditional C-CAPM by replacing a time varying coefficient in the stochastic discount factor with an instrument or conditioning variable that is multiplied by a constant coefficient. Their instrument is the consumption to wealth ratio which they interpret as the dividend paid from aggregate wealth. Using a linearization argument similar to Campbell and Shiller (1988), they show that the consumption to wealth ratio will be high when either expected returns to the wealth portfolio are high or expected consumption growth rates are high. In this sense the consumption to wealth ratio proxies for time varying investment opportunities. Lettau and Ludvigson (2001) conclude that a conditional C-CAPM can help to explain the value effect.

Lewellen and Nagel (2006) respond critically to this claim. They argue on theoretical grounds that the beta from a conditional CAPM would have to exhibit unrealistically high levels of volatility to explain the pricing errors observed from the value and momentum effects. Empirically they analyze a conditional version of the Sharpe-Lintner CAPM and a consumption CAPM following the approach of Breeden et al. (1989), where consumption is replaced with a portfolio of assets that can be measured at higher frequencies and is most highly correlated with consumption at low frequencies. Instead of proxying for time varying coefficients with an instrument and a constant coefficient Lewellen and Nagel (2006) employ a direct approach. They estimate conditional betas with daily observations over quarters

assuming the value of beta is relatively constant within each quarter. They find betas exhibit significant variation over time. Additionally they observe an average of the beta time series is close to an unconditional beta computed from a static CAPM, and that betas seem to be persistent although not as persistent as typical business cycle macroeconomic variables. Lastly they find a statistically significant relationship between betas and the treasury bill rate, dividend yields and the slope of the term structure. Interestingly they do not find that conditional betas are related to the consumption to wealth ratio variable used by Lettau and Ludvigson (2001). Despite the time variation in beta Lewellen and Nagel (2006) ultimately conclude a conditional CAPM is not sufficient to explain the value and momentum effects.

I propose a test of conditional linear factor models that makes two contributions to this literature. First, I extend the setting from that of testing the conditional CAPM to testing a conditional multi-factor model. Second, I take seriously the notion that betas vary over time and must be estimated at each time t in my sample $[0, T]$. This contrasts Lewellen and Nagel (2006) which makes the inconsistent assumption that betas exhibit significant time variation across quarters but are constant within quarters. My strategy is possible due to the availability of high frequency data. In particular, it is now possible to obtain intraday data for increasingly long time horizons for a large cross section of assets. A novel feature of my approach is that I can in principle exploit all of the available data. To be specific, I employ a local volatility estimator whose properties are described in Jacod and Protter (2011) to estimate time varying betas. My estimator is consistent as the time step between observed asset prices decreases to zero which requires the use of intraday data. To test linear factor models I then focus on the moment condition for expected returns. Under the null hypothesis that the model holds my proposed statistic converges to a multivariate normal distribution. By applying my test to longer time series and more assets in the cross section, the power of my inference increases.

This paper builds on earlier work that has used high frequency data and rolling regressions to estimate conditional variances and covariances (Foster and Nelson (1996)). One challenge

for this approach is that high frequency data has limited precision as a result of market microstructure noise. I study the impact that noise has on my approach in a Monte Carlo experiment in the Appendix. I find noisy and asynchronous data pose severe problems for my baseline statistic. With these observations in mind, I employ recently developed volatility estimators that are robust to certain types of noise and asynchronous trading in my empirical application.

The remainder of the paper proceeds as follows. Section 3.2 introduces the principal idea behind the test. First, I analyze an unfeasible test in which the econometrician knows the instantaneous values of beta. Then, I make my procedure feasible by using a consistent estimator for beta. One of the key results is that the difference between the feasible and infeasible tests uniformly converges in probability to zero. Section 3.3 illustrates an empirical application in which I test the conditional CAPM for the stocks in the Dow from 2008 to 2013. I find that the conditional CAPM is rejected by the data. Section 3.4 concludes with a brief discussion of the results and suggestions for future research.

3.2 Asset Pricing Tests

3.2.1 Infeasible Test

I assume there exists a probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ satisfying the usual conditions on which there is an adapted process $X_t \in \mathcal{F}_t$ that models log asset prices as,

Assumption 1 *X_t is a d-dimensional continuous Itô semimartingale*

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s \tag{3.1}$$

where volatility is a positive and locally bounded Itô semimartingale with $c = \sigma\sigma'$ taking values in the set of symmetric positive definite $d \times d$ matrices. The drift is a locally bounded and progressively measurable process.

The beta representation of a conditional linear factor model implies the moment condition,

$$\mu_t^p - r_t = \beta_t \left(\mu_t^f - r_t \right) \tag{3.2}$$

where r_t is the risk free rate, μ_t^p is the expected return on asset p, μ_t^f is the vector of expected returns of the factor portfolios, and β_t is the multiple regression coefficient of asset p onto the factor portfolios. An example of a model with this feature is the ICAPM of Merton (1973). Another example is the consumption CAPM of Breeden (1979) where the single factor is aggregate consumption.

To test this asset pricing model, suppose I could sample the unobservable process

$$dX^{UO} \equiv dX^P - r \cdot dt - \beta \left[dX^f - r \cdot dt \right] \tag{3.3}$$

where dX^P is the instantaneous return on a portfolio of assets. By construction the drift of dX^{UO} is equal to zero under the null hypothesis that the conditional factor model holds. Now consider a discretized increment of dX^{UO} or

$$\begin{aligned} \Delta_i^n X^{UO} &= X^{UO}_{T(n,i)} - X^{UO}_{T(n,i-1)} \\ &= \int_{(i-1)\Delta_n}^{i\Delta_n} \mu_s^{UO} ds + \int_{(i-1)\Delta_n}^{i\Delta_n} \sigma_s^{UO} dW_s \end{aligned} \tag{3.4}$$

Here I use the notation $T(n,i) = i\Delta_n$ for $i = \{1, \ldots, m_n\}$ over $[0,T]$ where $m_n = [T/\Delta_n]$ and $\Delta_n \to 0$ as $n \to \infty$. Under the null hypothesis it is immediate that $\int_{(i-1)\Delta_n}^{i\Delta_n} \mu_s^{UO} ds = 0$ for all i, n. This forms the basis for my test. Given the process X is continuous I have the discretized version $X_t^n - X_0 \equiv \sum_{i=1}^{[t/\Delta_n]} \Delta_i^n X^{UO}$ satisfies

$$X^n - X_0 \overset{ucp}{\to} X - X_0 \tag{3.5}$$

where *ucp* denotes uniform convergence in probability or $\sup_{t\leq T}\|X_t^n - X_t\| \xrightarrow{\mathbb{P}} 0 \ \forall \ T$. Additionally, the realized variance satisfies the following relationship,

$$\hat{C}_T^{UO} = \sum_{i=1}^{[T/\Delta_n]} \Delta_i^n X^{UO} (\Delta_i^n X^{UO})' \overset{ucp}{\to} C_T^{UO} = \int_0^T c_s^{UO} ds. \tag{3.6}$$

The consistency of this estimator depends crucially on the fact that X is continuous and that the process is observed without noise. I will return to these concerns when discussing empirical implementation. For the time being I restrain myself to the idealized setting. Combining the two facts above it is natural to introduce the normalized statistic:

$$S_{T_i}^n := \left[\sum_{i=[T_{i-1}/\Delta_n]}^{[T_i/\Delta_n]} \Delta_i^n X^{UO} (\Delta_i^n X^{UO})' \right]^{-1/2} \left(\sum_{i=[T_{i-1}/\Delta_n]}^{[T_i/\Delta_n]} \Delta_i^n X^{UO} \right) \tag{3.7}$$

where $(T_1, T_2, \ldots, T_N)$ is a non-random partition of $[0, T]$ for some $N > 0$. With my current setup I establish the following result:

Proposition 1 ***Infeasible Test:*** *Under Assumption 1 and supposing $\int_{T_{i-1}}^{T_i} c_s^{UO} ds$ is invertible for all i w.p.1 it follows that*

$$(S_{T_1}^n, S_{T_2}^n, \ldots, S_{T_N}^n) \ \xrightarrow{\mathbb{P}} \ S \ \sim MN(\mu, I_{N\cdot d}) \tag{3.8}$$

where μ is a random vector with components

$$\mu_i \equiv \left(\int_{T_{i-1}}^{T_i} c_s^{UO} ds \right)^{-1/2} \left[\int_{T_{i-1}}^{T_i} \mu_s^{UO} ds \right] \tag{3.9}$$

Under the null hypothesis the conditional linear factor model holds $\mu = 0$.

From this point I am interested in the test $H_0 : \ \mu = 0$ against $H_1 : \ \mu \neq 0$. In principal there are a number of ways to handle this small sample testing problem. An important hurdle is that the distribution of μ is unknown. Contrast this with the classical case where μ is unknown but nonrandom. In this simplified case, the Neyman-Pearson lemma suggests

I reject the null hypothesis for large values of $\mu' S$. So even in the case where S is normally distributed with unknown fixed mean μ there is no uniformly most powerful test. If one restricts attention to tests invariant to the orthogonal group of transformations then it turns out the optimal test rejects for large values of $S'S$ as discussed in Müller (2011). Invariance is perhaps a desirable property if I intend to be agnostic about the timing and sign of normalized pricing errors, hopefully making my approach appropriate for a variety of assets under different market conditions. Note also that $S'S \sim \chi^2_{N \cdot d}$ under H_0 making it easy to compute critical values.

A natural concern then is whether my test leads to powerful inference. Figure 3.1 plots the power as a function of the pricing error and the number of observations. In particular, suppose μ^{UO} and σ^{UO} are constants and $d = 1$. Then, the mean of the statistic S is an information ratio multiplied by the square root of time $\mu = \mu^{UO}/\sigma^{UO}\sqrt{dt}$. For example, if $\mu^{UO} = .05$, $\sigma^{UO} = .20$, and $dt = 1$ then $\mu = .25$. In this case, Figure 3.1 suggests that I would need 20-50 observations for a powerful procedure. Intuitively, the power increases when I test more assets for longer periods of time.

3.2.2 Feasible Test

The challenge in practice is that I do not observe β and I cannot compute $\Delta_i^n X^{UO}$. To circumvent this problem I use a local volatility estimator to estimate β and then construct the observable increment $\Delta_i^n X^O$ defined as,

$$\Delta_i^n X^O \equiv \Delta_i^n X^p - r_{(i-1)\Delta} \Delta_i^n - \hat{\beta}_{(i-1)\Delta} \left[\Delta_i^n X^f - r_{(i-1)\Delta} \Delta_i^n \right] \tag{3.10}$$

To estimate β I can compute $\hat{\beta}_{(i-1)\Delta} = \left[\hat{c}^{pf}_{i-1} \right] \left[\hat{c}^{f}_{i-1} \right]^{-1}$ where

$$\hat{c}^{jk}_{i-1} \equiv \hat{c}^{jk}_{i-k_n}(k_n, t-) = \frac{1}{k_n \Delta_n} \sum_{m=0}^{k_n - 1} \Delta_{i+m}^n X^j \Delta_{i+m}^n X^k. \tag{3.11}$$

The local volatility estimator $\hat{c}(k_n, t-) \xrightarrow{\mathbb{P}} c_{t-}$ is consistent for each time t. Consistency coupled with the following assumption drive my results for the next proposition.

Assumption 2 *The eigenvalues of the covariance matrix for the factor portfolios are bounded*

$$\sup_{s\leq t} \|(c_s^f)^{-1}\| = \sup_{s\leq t} \|P_s \Lambda_s^{-1} P_s'\| \leq \sup_{s\leq t} \|\Lambda_s^{-1}\| \leq K \tag{3.12}$$

The proposition below shows that the error between the feasible and infeasible observations converges uniformly in probability to zero.

Proposition 2 ***Feasible Test:*** *Under Assumption 1 and Assumption 2, as* $\Delta_n \to 0$ *and* $k_n \Delta_n \to 0$ *it follows that*

$$\sum_{i=k_n+1}^{[\cdot/\Delta_n]} \Delta_i^n X^O \xrightarrow{ucp} X^{UO} - X_0 \tag{3.13}$$

Moreover,

$$\sum_{i=k_n+1}^{[t/\Delta_n]} \Delta_i^n X^O (\Delta_i^n X^O)' \xrightarrow{ucp} C_t^{UO} \tag{3.14}$$

Corollary 1 ***Feasible Test:*** *Define*

$$\hat{S}_{T_i}^n = \left[\sum_{i=[T_{i-1}/\Delta_n]}^{[T_i/\Delta_n]} \Delta_i^n X^O (\Delta_i^n X^O)' \right]^{-1/2} \left[\sum_{i=[T_{i-1}/\Delta_n]}^{[T_i/\Delta_n]} \Delta_i^n X^O \right] \tag{3.15}$$

Then the feasible test statistic satisfies

$$\hat{S}^n = (\hat{S}_{T_1}^n, \ldots, \hat{S}_{T_N}^n) \xrightarrow{\mathbb{P}} S \sim MN(\mu, I_{N \cdot D}) \tag{3.16}$$

where μ *has the same definition as in Proposition 1 and under* $H_0 : \mu = 0$.

3.3 Empirical Application

To illustrate an empirical application of my test I compute the standardized pricing errors for the conditional CAPM using high frequency data from the Trade and Quote (TAQ) database which is described in the second chapter of this dissertation. For convenience, I focus on the firms that were members of the Dow during the entire 2008 to 2013 sample period. This leaves me with 21 firms. I then compute empirical analog to equation 3.10 at a five-minute sampling frequency $\Delta_n = 5$ minutes with the SPY exchange traded fund serving as my proxy for the market portfolio.[2]

The results are reported in Table 3.1 and Figure 3.2. The pricing errors for the conditional CAPM are large and statistically significant. This is true across time and across firms. For example, in the cross section, the only stocks for which the conditional CAPM is not rejected are United Technologies (UTX) and Merck (MRK). The model is rejected for all of the remaining firms with the largest pricing errors occurring for International Business Machines (IBM), 3M (MMM), and McDonald's (MCD). In the time dimension, the smallest pricing errors actually occurred during 2008 and 2009 while larger errors occur later in the sample. In part this reflects the fact that realized volatility, whose inverse is multiplied by the pricing errors, was high during the financial crisis. In unreported results, I also check to see whether the conditional CAPM holds during different intraday periods, days of the week, or months. I find larger standardized pricing errors on Monday, Tuesday, and Wednesday and smaller errors on Thursday and Friday. The smallest intraday standardized errors occur from 10:30am-11:30am and 1:30pm-2:30pm but the largest error occurs during lunch from 11:30am-12:30pm. Finally, the largest standardized errors occur during January and July while the lowest errors occur during September and October. As before, the pricing

[2]This corresponds to the abnormal return $Ra_{i,t}$ defined in the second chapter. As before, I estimate the realized beta from twenty days of lagged five-minute returns with a subsample estimator following Zhang et al. (2005) to account for market microstructure noise.

errors during all of these periods are large and statistically significant, indicating that the conditional CAPM is rejected by the data.

3.4 Conclusion

This paper illustrates a new way to test conditional linear factors models using high frequency data. A novel feature of my approach is that I can exploit the information in intraday prices for multiple assets over long horizons. In an empirical application of my test, I find that the conditional CAPM is rejected by the stocks in the Dow from 2008 to 2013. This is consistent with the rejection of the conditional CAPM in Lewellen and Nagel (2006). To extend this work, it would be valuable to develop a similar test that is robust to jumps and microstructure noise. In the appendix I show that microstructure noise and asynchronous data pose serious challenges for my test, and to account for this in my empirical application I use sparse sampling and a subsample estimator to estimate realized betas for the stocks in the Dow. Additionally, it would be interesting to explore whether the time variation in high frequency realized betas is consistent with the assumption in Lewellen and Nagel (2006) that betas are relatively constant over quarters. Finally, I could extend my empirical analysis to intraday multi-factor models. For example, the second chapter of this dissertation noted that contrarian trading strategies earn abnormal returns within the trading day. This test, in principal, would allow a researcher to test whether a market factor and a liquidity or contrarian trading factor could explain the cross section of high frequency returns.

References

Aït-Sahalia, Y., J. Fan, and D. Xiu (2010). High-frequency covariance estimates with noisy and asynchronous financial data. *Journal of the American Statistical Association 105*(492), 1504–1517.

Asness, C. (1994). Variables that explain stock returns. PhD Dissertation, University of Chicago.

Barndorff-Nielsen, O., P. Hansen, A. Lunde, and N. Shephard (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica 76*(6), 1481–1536.

Breeden, D. (1979). An intertemporal asset pricing model with stochastic consumption and investment opportunities. *Journal of Financial Economics 7*(3), 265–296.

Breeden, D., M. Gibbons, and R. Litzenberger (1989). Empirical tests of the consumption capm. *Journal of Finance 44*, 231–262.

Campbell, J. and R. Shiller (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies 1*(3), 195–228.

Campbell, J. and T. Vuolteenaho (2004). Bad beta, good beta. *American Economic Review 94*(5), 1249–1275.

Christensen, K., S. Kinnebrock, and M. Podolskij (2010). Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *Journal of Econometrics 159*(1), 116–133.

Cochrane, J. (2005). *Asset Pricing*. Princeton University Press.

Epps, T. (1979). Comovements in stock prices in the very short run. *Journal of the American Statistical Association 74*(366a), 291–298.

Fama, E. (1991). Efficient capital markets: Ii. *Journal of Finance 46*(5), 1575–1617.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*(1), 3–56.

Foster, P. and D. Nelson (1996). Continuous record asymptotics for rolling sample variance estimators. *Econometrica 64*(1), 139–174.

Hansen, L. and S. Richard (1987). The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. *Econometrica*, 587–613.

Jacod, J. (2007). Statistics and high frequency data. *Unpublished paper*.

Jacod, J. and P. Protter (2011). *Discretization of processes*, Volume 67. Springer Verlag.

Jegadeesh, N. and S. Titman (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance 48*(1), 65–91.

Jensen, M. C., F. Black, and M. S. Scholes (1972). The capital asset pricing model: Some empirical tests. In M. C. Jensen (Ed.), *Studies in the Theory of Capital Markets*. Praeger.

Lettau, M. and S. Ludvigson (2001). Resurrecting the (c) capm: A cross-sectional test when risk premia are time-varying. *Journal of Political Economy 109*(6), 1238–1287.

Lewellen, J. and S. Nagel (2006). The conditional capm does not explain asset-pricing anomalies. *Journal of Financial Economics 82*(2), 289–314.

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

Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica*, 867–887.

Müller, U. (2011). Time series lecture notes: Invariance.

Protter, P. E. (2004). *Stochastic Integration and Differential Equations: Version 2.1*, Volume 21. Springer.

Ross, S. (1976a). The arbitrage theory of capital asset pricing. *Journal of Economic Theory 13*, 341–360.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance 19*(3), 425–442.

Xiu, D. (2010). Quasi-maximum likelihood estimation of volatility with high frequency data. *Journal of Econometrics 159*(1), 235–250.

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association 100*(472), 1394–1411.

Tables

Conditional CAPM Pricing Errors

Ticker	2008	2009	2010	2011	2012	2013	$S'S$
CAT	-0.53	-2.38	4.36	-3.87	-4.47	-3.71	73.76
DD	-2.61	0.70	1.06	-1.12	-2.68	2.89	25.30
DIS	1.40	2.26	4.70	1.53	5.98	4.22	84.98
GE	-1.76	-5.27	-0.64	-1.76	2.93	-6.78	88.96
HD	0.35	-0.06	0.19	3.56	8.69	-1.90	91.93
IBM	1.45	11.69	16.54	8.62	2.75	-1.91	497.89
INTC	0.69	-1.62	-4.33	6.81	-1.13	0.11	69.58
JNJ	0.55	-2.54	-4.96	1.79	2.34	5.77	73.26
JPM	0.72	-0.73	-1.68	-2.10	3.58	1.47	23.24
KO	-0.69	2.38	4.65	-1.05	-6.73	-5.13	100.46
MCD	2.46	-3.14	2.56	2.91	-12.06	-0.69	176.75
MMM	0.71	3.88	-5.22	-0.96	8.06	18.77	460.88
MRK	-0.18	1.25	-1.26	-2.76	-0.20	-0.86	11.55
MSFT	-1.52	1.69	-4.26	-0.63	-2.45	5.09	55.61
PFE	-1.46	-2.22	-3.16	5.19	6.98	3.07	102.13
PG	1.23	-3.93	7.91	1.31	3.73	4.26	113.26
T	0.29	-3.76	-3.65	-5.63	4.09	-3.95	91.57
UTX	0.25	1.98	-1.83	0.38	0.66	0.35	8.03
VZ	-0.67	-3.11	-3.64	2.62	-1.64	-1.55	35.31
WMT	3.16	-3.67	-2.16	3.00	5.12	3.06	72.71
XOM	3.59	-1.94	-1.38	9.22	6.63	0.68	148.06
$S'S$	52.19	285.78	546.53	342.61	588.34	589.78	2405.22

Table 3.1: This table reports the standardized pricing errors from equation 3.15 for the conditional CAPM computed each year for the members of the Dow from 2008 to 2013. The 1% critical values for a χ^2_N distribution are 38.93 ($N = 21$) and 16.81 ($N = 6$). As the table suggests, the conditional CAPM is strongly rejected. It is interesting to observe that the smallest pricing errors occurred during 2008 and 2009 which coincides with the financial crisis.

Figures

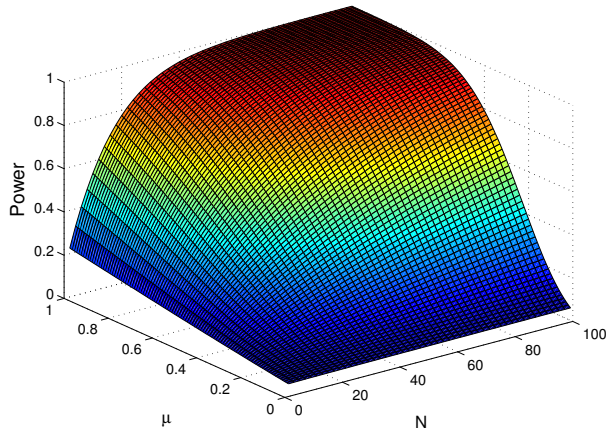


Figure 3.1: The power of rejecting the null hypothesis $H_0 : S \sim N(0, I_N)$ versus the alternative $H_1 : S \sim N(\mu \mathbf{1}, I_N)$ for $\mu \in \{.01, \ldots, 1\}$ with $N \in \{2, 4, \ldots, 100\}$ observations using $S'S$ as the test statistic.

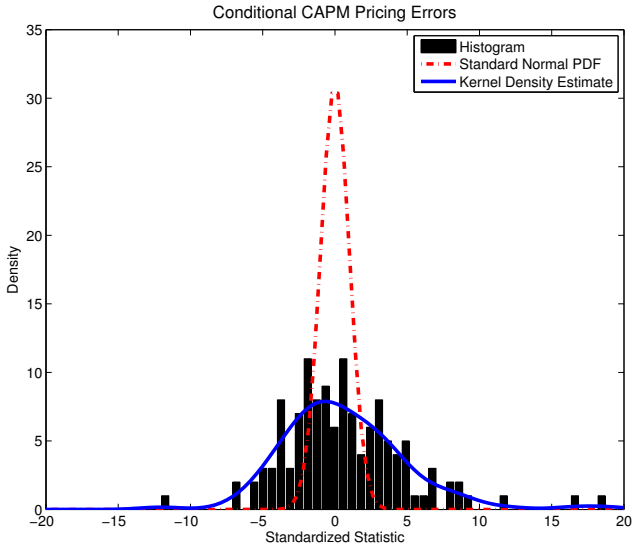


Figure 3.2: This plot illustrates the standardized pricing errors from equation 3.15 for the conditional CAPM computed each year for the members of the Dow from 2008 to 2013. As the plot suggests, the conditional CAPM is rejected by the data.

APPENDIX

3.A Proofs and Derivations

Proposition 1:

To show Proposition 1 note that $X^n - X_0 \overset{ucp}{\to} X - X_0$ and $\hat{C}^{UO} \overset{ucp}{\to} C^{UO}$, for example, by Theorem 3.3.1 in Jacod and Protter (2011). The uniform convergence then implies the joint finite dimensional convergence of

$$\begin{aligned}(X^n_{T_1} - X_0, \ldots, X^n_{T_N} - X_0, \hat{C}^{UO}_{T_1}, \ldots, \hat{C}^{UO}_{T_N}) \\ \overset{\mathbb{P}}{\to} (X_{T_1} - X_0, \ldots, X_{T_N} - X_0, C_{T_1}, \ldots, C_{T_N})\end{aligned} \tag{3.17}$$

To see this let $\hat{Y}_t = X^n_t - X_0$ for notational simplicity and observe that

$$\begin{aligned}Pr(\sup_{t\leq T} \|(\hat{Y}_t, \hat{C}_t) - (Y_t, C_t)\| > \epsilon) &\leq Pr(\sup_{t\leq T} \|\hat{Y}_t - Y_t\| + \sup_{t\leq T} \|\hat{C}_t - C_t\| > \epsilon) \\ &\leq Pr(\sup_{t\leq T} \|\hat{Y}_t - Y_t\| > \tfrac{\epsilon}{2}) + Pr(\sup_{t\leq T} \|\hat{C}_t - C_t\| > \tfrac{\epsilon}{2}) \\ &\to 0\end{aligned} \tag{3.18}$$

Now introduce $h = (h_1, \ldots, h_N)$ where

$$h_i(x_1, \ldots, x_{T_N}, y_1, \ldots, y_{T_N}) = (y_i - y_{i-1})^{-1/2} (x_i - x_{i-1}). \tag{3.19}$$

This is a continuous mapping with my assumption that $C_{T_i} - C_{T_{i-1}}$ is invertible with probability one so I conclude

$$\begin{aligned} S^n &= h(\hat{Y}_{T_1}, \ldots, \hat{C}^{UO}_{T_N}) \\ &\xrightarrow{\mathbb{P}} S \\ &= h(Y_{T_1}, \ldots, C_{T_N}) \\ &= \left(\left(\int_0^{T_1} c_s^{UO} ds \right)^{-1/2} \left[\int_0^{T_1} \mu_s^{UO} ds + \int_0^{T_1} \sigma_s^{UO} dW_s \right], \ldots, \right. \\ &\quad \left. \left(\int_{T_{N-1}}^{T_N} c_s^{UO} ds \right)^{-1/2} \left[\int_{T_{N-1}}^{T_N} \mu_s^{UO} ds + \int_{T_{N-1}}^{T_N} \sigma_s^{UO} dW_s \right] \right) \end{aligned} \tag{3.20}$$

To show the next claim that $S \mid \mu \sim N(\mu, I_N)$ where $\mu_i \equiv \left(\int_{T_{i-1}}^{T_i} c_s^{UO} ds \right)^{-1/2} \int_{T_{i-1}}^{T_i} \mu_s^{UO} ds$ repeated applications of the law of iterated expectations shows that

$$\begin{aligned} E[e^{iu'S} \mid \mu] &= E[e^{i(u_1'S_1 + \cdots + u_N'S_N)} \mid \mu] \\ &= E[e^{i(u_1'S_1 + \cdots + u_{N-1}'S_{N-1})} E_{N-1}[e^{iu_N'S_N}] \mid \mu] \\ &= E[e^{i(u_1'S_1 + \cdots + u_{N-1}'S_{N-1})} E_{N-1, \{\sigma_s^{UO}\}_{T_{N-1}}^{T_N}}[e^{iu_N'S_N}] \mid \mu] \\ &= E[e^{i(u_1S_1 + \cdots + u_{N-1}S_{N-1})} e^{iu_N'\mu_N - \frac{1}{2}u_N'u_N} \mid \mu] \\ &= \cdots \\ &= e^{iu'\mu - \frac{1}{2}u'Iu}. \end{aligned} \tag{3.21}$$

The fourth line follows from the observation that $\int_{T_{N-1}}^{T_N} \mu_s^{UO} ds + \int_{T_{N-1}}^{T_N} \sigma_s^{UO} dW_s \mid (X_N, Y_N) \sim N(X_N,\ Y_N)$ where $(X_N, Y_N) \equiv (\int_{T_{N-1}}^{T_N} \mu_s^{UO} ds, \int_{T_{N-1}}^{T_N} c_s^{UO} ds)$. Under the null hypothesis that the asset pricing model holds $\mu_t^{UO} = 0$ for every t so it follows that $\mu = 0$. ■

Proposition 2:

The proof for Proposition 2 follows the discussion in Jacod and Protter (2011). I can break up $\Delta_i^n X^O$ into

$$\begin{aligned} \Delta_i^n X^O &= \int_{(i-1)\Delta}^{i\Delta} \mu_s^p - r_{(i-1)\Delta} - \hat{\beta}_{(i-1)\Delta} \left[\mu_s^f - r_{(i-1)\Delta} \right] ds \\ &\quad + \int_{(i-1)\Delta}^{i\Delta} \sigma_s^p - \hat{\beta}_{(i-1)\Delta} \sigma_s^f dW_s \\ &= \int_{(i-1)\Delta}^{i\Delta} \mu_s^p - r_s - \beta_s \left[\mu_s^f - r_s \right] ds \\ &\quad + \int_{(i-1)\Delta}^{i\Delta} \sigma_s^p - \beta_{(i-1)\Delta} \sigma_s^f dW_s + \xi_i^n \end{aligned} \tag{3.22}$$

where

$$\xi_i^n \equiv \int_{(i-1)\Delta}^{i\Delta} r_s + \beta_s \left[\mu_s^f - r_s\right] - r_{(i-1)\Delta} \tag{3.23}$$
$$-\hat{\beta}_{(i-1)\Delta}\left[\mu_s^f - r_{(i-1)\Delta}\right] ds + \int_{(i-1)\Delta}^{i\Delta} \left[\beta_{(i-1)\Delta} - \hat{\beta}_{(i-1)\Delta}\right] \sigma_s^f dW_s$$

Through a series of small steps I will argue

$$R_t = \sum_{i=k_n+1}^{[t/\Delta_n]} \xi_i^n \overset{ucp}{\to} 0. \tag{3.24}$$

Key to this result will be repeated use of the Doob and Davis-Burkholder-Gundy inequality which implies the bounds 2.1.44 in Jacod and Protter (2011). These inequalities depend on boundedness of (μ, σ, r, β). This is no problem by an application of the localization procedure outlined in 4.4.1 of Jacod and Protter (2011) which allows me to replace my locally bounded coefficients from Assumption 1 with bounded coefficients. Relatedly, I abuse notation slightly by letting K denote a generic constant whose value may change from line to line. First, I argue that

$$\sum_{i=k_n+1}^{[t/\Delta_n]} \int_{(i-1)\Delta}^{i\Delta} r_s + \beta_s \left[\mu_s^f - r_s\right] - r_{(i-1)\Delta} - \beta_{(i-1)\Delta}\left[\mu_s^f - r_{(i-1)\Delta}\right] ds \overset{ucp}{\to} 0 \tag{3.25}$$

I can observe that in the case where I consider a $D \times 1$ vector of assets it suffices to show each of the components uniformly converges in probability to zero. Since my procedure below applies to any of the components I restrict attention to the case where $D = 1$. In this setting it is sufficient to argue

$$\frac{1}{\Delta} \sup_i E| \int_{(i-1)\Delta}^{i\Delta} r_s + \beta_s \left[\mu_s^f - r_s\right] - r_{(i-1)\Delta} - \beta_{(i-1)\Delta}\left[\mu_s^f - r_{(i-1)\Delta}\right] ds| \to 0 \tag{3.26}$$

by Markov's inequality. To start I can observe that an application of Itô's lemma implies β is a continuous Ito semimartingale as β is a C^2 function of c. As such,

$$\begin{aligned}
&E|\int_{(i-1)\Delta}^{i\Delta}(\beta_s-\beta_{(i-1)\Delta})\mu_s^f ds| \\
&\leq \quad E\int_{(i-1)\Delta}^{i\Delta}|(\beta_s-\beta_{(i-1)\Delta})\mu_s^f|ds \\
&\leq \quad E\left(\int_{(i-1)\Delta}^{i\Delta}\|\beta_s-\beta_{(i-1)\Delta}\|^2 ds\right)^{1/2}\left(\int_{(i-1)\Delta}^{i\Delta}\|\mu_s^f\|^2 ds\right)^{1/2} \\
&\leq E\left(\int_{(i-1)\Delta}^{i\Delta}\|\beta_s-\beta_{(i-1)\Delta}\|^2 ds\right)^{1/2}(K\Delta)^{1/2} \\
&\leq \left(E\int_{(i-1)\Delta}^{i\Delta}\|\beta_s-\beta_{(i-1)\Delta}\|^2 ds\right)^{1/2}K\Delta^{1/2} \\
&\leq \left(\int_{(i-1)\Delta}^{i\Delta}E\|\beta_s-\beta_{(i-1)\Delta}\|^2 ds\right)^{1/2}K\Delta^{1/2} \\
&\leq \left(\int_{(i-1)\Delta}^{i\Delta}K(s-(i-1)\Delta)ds\right)^{1/2}K\Delta^{1/2} \\
&\leq \left(K\tfrac{\Delta^2}{2}\right)^{1/2}K\Delta^{1/2} \\
&\leq K\Delta^{3/2}.
\end{aligned} \tag{3.27}$$

where the first line is immediate, the second uses Cauchy Schwarz, the third uses the localization assumption $\|\mu_t(\omega)\| \leq K$, the fourth uses Jensen's inequality, the fifth uses Fubini's theorem, and the sixth uses the inequality 2.1.44. By a similar logic,

$$\begin{aligned}
E|\int_{(i-1)\Delta}^{i\Delta}r_s - r_{(i-1)\Delta}ds| &\leq \int_{(i-1)\Delta}^{i\Delta}E|r_s-r_{(i-1)\Delta}|ds \\
&\leq K\int_{(i-1)\Delta}^{i\Delta}(s-(i-1)\Delta)^{1/2}ds \\
&= K\Delta^{3/2}
\end{aligned} \tag{3.28}$$

Lastly, I can note βr is also an Ito process. The same logic then applies,

$$E|\int_{(i-1)\Delta}^{i\Delta}\beta_{(i-1)\Delta}r_{(i-1)\Delta}-\beta_s r_s| \leq K\Delta^{3/2} \tag{3.29}$$

Putting these pieces together with the triangle inequality yields the desired result. Next let's show

$$\sum_{i=k_n+1}^{[t/\Delta_n]}\int_{(i-1)\Delta}^{i\Delta}\left[\hat{\beta}_{(i-1)\Delta}-\beta_{(i-1)\Delta}\right]\left[\mu_s^f - r_{(i-1)\Delta}\right]ds \overset{ucp}{\to} 0 \tag{3.30}$$

I can note from Cauchy Schwarz and the boundedness assumption on the coefficients

$$\int_{(i-1)\Delta}^{i\Delta} | \left[\hat{\beta}_{(i-1)\Delta} - \beta_{(i-1)\Delta}\right] \left[\mu_s^f - r_{(i-1)\Delta}\right] | ds \quad \leq \quad K\Delta \|\hat{\beta}_{(i-1)\Delta} - \beta_{(i-1)\Delta}\| \tag{3.31}$$

hence

$$\begin{aligned} &| \sum_{i=k_n+1}^{[t/\Delta_n]} \int_{(i-1)\Delta}^{i\Delta} \left[\hat{\beta}_{(i-1)\Delta} - \beta_{(i-1)\Delta}\right] \left[\mu_s^f - r_{(i-1)\Delta}\right] ds| \\ &\leq K\Delta \sum_{i=k_n+1}^{[t/\Delta_n]} \|\hat{\beta}_{(i-1)\Delta} - \beta_{(i-1)\Delta}\| \\ &= K \int_{k_n\Delta_n}^{[t/\Delta_n]\Delta_n} \|\hat{\beta}_s - \beta_s\| ds. \end{aligned} \tag{3.32}$$

Here I define for $t \geq 0$ that $\beta_t = \beta_{(i-1)\Delta} \cdot \mathbf{1}[t > k_n\Delta_n]$ whenever $(i-1)\Delta \leq t < i\Delta$ and analogously for $\hat{\beta}$. The argument from here follows closely the proof of Theorem 9.4.1 in Jacod and Protter (2011). Namely, suppose I knew that $\|\hat{\beta}_s - \beta_s\| < B$. Then I could apply the Dominated Convergence Theorem to conclude

$$\begin{aligned} \lim_{n\to\infty} E| \int_0^{[t/\Delta]\Delta} \|\hat{\beta}_s - \beta_s\| ds| &= \lim_{n\to\infty} \int_{k\Delta}^{[t/\Delta]\Delta} E\|\hat{\beta}_s - \beta_s\| ds \\ &\leq \int_0^t \lim_{n\to\infty} E\|\hat{\beta}_s - \beta_s\| ds \\ &= \int_0^t E[\lim_{n\to\infty} \|\hat{\beta}_s - \beta_s\|] ds \\ &= 0 \end{aligned} \tag{3.33}$$

where $\lim_{n\to\infty} \|\hat{\beta}_s - \beta_s\| \xrightarrow{p} 0$ by the continuous mapping theorem applied by the pointwise convergence of $\hat{c}(k_n, t-) \xrightarrow{\mathbb{P}} c_{t-} = c_t$ from Theorem 9.3.2 of Jacod and Protter (2011). Relaxing the boundedness assumption above is not trivial. In Theorem 9.4.1 Jacod and Protter use a Lipschitz condition and I pursue a similar strategy here by making a strong assumption. Clearly,

$$\beta(c) = c^{pf} \left[c^f\right]^{-1} \tag{3.34}$$

is not Lipschitz. But by diagonalization I have

$$\|\beta(c)\| \leq \|c^{pf}\| \cdot \|P\| \|\Lambda^{-1}\| \|P'\| \leq \|c^{pf}\| \|\Lambda^{-1}\|. \tag{3.35}$$

Assumption 2 implies $\sup_{s\leq t} \|\Lambda_s^{-1}\| < K$. Thus if I assume β behaves in a Lipschitz manner by bounding the lowest eigenvalue in the covariance matrix above for all times $s \in [0, t]$ uniformly the

rest of the argument in 9.4.1 can be applied in order to apply the DCT. Lastly, I will show

$$\sum \left(\hat{\beta}_{(i-1)\Delta} - \beta_{(i-1)\Delta}\right) \int_{(i-1)\Delta}^{i\Delta} \sigma_s^f dW_s \overset{ucp}{\to} 0 \tag{3.36}$$

Here I follow Lemma 4.1 from Jacod (2007) which indicates $\sum_{i=k_n+1}^{[t/\Delta_n]} \chi_i^n \overset{ucp}{\to} 0$ as long as the following conditions are met:

$$\sum_{i=k_n+1}^{[t/\Delta_n]} E_{i-1}^n \chi_i^n \overset{ucp}{\to} 0 \quad \sum_{i=k_n+1}^{[t/\Delta_n]} E_{i-1}^n \|\chi_i^n\|^2 \overset{\mathbb{P}}{\to} 0 \, \forall t > 0. \tag{3.37}$$

Define

$$\chi_i^n := \left(\hat{\beta}_{(i-1)\Delta} - \beta_{(i-1)\Delta}\right) \int_{(i-1)\Delta}^{i\Delta} \sigma_s^f dW_s. \tag{3.38}$$

This is a martingale difference sequence so the first condition is met trivially. Next I have

$$\begin{aligned} E_{i-1}^n \|\chi_i^n\|^2 &= \textstyle\sum_{j=1}^d (\hat{\beta}_{(i-1)\Delta,j} - \beta_{(i-1)\Delta,j})^2 E_{i-1}^n \left(\int_{(i-1)\Delta}^{i\Delta} \sigma_{s,j}^f dW_s\right)^2 \\ &\leq K\Delta \textstyle\sum_{j=1}^d (\hat{\beta}_{(i-1)\Delta,j} - \beta_{(i-1)\Delta,j})^2 \\ &= K\Delta \|\hat{\beta}_{(i-1)\Delta} - \beta_{(i-1)\Delta}\|^2. \end{aligned} \tag{3.39}$$

Then I find

$$\begin{aligned} \textstyle\sum_{i=k_n}^{[t/\Delta_n]} E_{i-1}^n \|\chi_i^n\|^2 &\leq K\Delta \textstyle\sum_{i=k_n+1}^{[t/\Delta_n]} \|\hat{\beta}_{(i-1)\Delta} - \beta_{(i-1)\Delta}\|^2 \\ &= K \textstyle\int_{k_n\Delta_n}^{[t/\Delta_n]\Delta_n} \|\hat{\beta}_s - \beta_s\|^2 ds. \end{aligned} \tag{3.40}$$

Applying the DCT by the same argument as in the previous lemma I get the desired result. This establishes the result $R_t \overset{ucp}{\to} 0$. my next step is to demonstrate

$$\sum_{i=k_n+1}^{[t/\Delta_n]} \Delta_i^n X^O \left(\Delta_i^n X^O\right)' \overset{ucp}{\to} C_t^{UO} \tag{3.41}$$

From my representation

$$\begin{aligned}\Delta_i^n X^O &= \int_{(i-1)\Delta}^{i\Delta}\left(\mu_s^p - r_s - \beta_s\left[\mu_s^f - r_s\right]\right)ds \\ &\quad + \int_{(i-1)\Delta}^{i\Delta}\left(\sigma_s^p - \beta_{(i-1)\Delta}\sigma_s^f\right)dW_s + \xi_i^n \\ &= \Delta_i^n X^{UO} + \int_{(i-1)\Delta}^{i\Delta}(\beta_s - \beta_{(i-1)\Delta})\sigma_s^f dW_s + \xi_i^n \\ &:= \Delta_i^n X^{UO} + \Delta_i^n N + \xi_i^n.\end{aligned} \tag{3.42}$$

Now I have a book keeping exercise. Namely, I must show the entries of $\Delta_i^n N(\Delta_i^n N)'$ and $\xi_i^n(\xi_i^n)'$ as well as all the cross terms uniformly converge in probability to zero. In what follows my arguments apply to arbitrary entries of the matrices above, hence it suffices to discuss the case where $\Delta_i^n X^O$ is a scalar. For the $(\xi_i^n)^2$ terms I find

$$\begin{aligned}&E|\sup_{s\leq t}\textstyle\sum_{i=k_n}^{[s/\Delta_n]}(\xi_i^n)^2| \\ &= \textstyle\sum_{i=k_n}^{[t/\Delta_n]} E(\xi_i^n)^2 \\ &\leq K\cdot\textstyle\sum_{i=k_n}^{[t/\Delta_n]} E\int_{(i-1)\Delta}^{i\Delta}\|\hat{\beta}_{(i-1)\Delta} - \beta_{(i-1)\Delta}\|^2 ds + o(1) \\ &= K\Delta\textstyle\sum_{i=k_n}^{[t/\Delta_n]} E\int_{(i-1)\Delta}^{i\Delta}\|\hat{\beta}_{(i-1)\Delta} - \beta_{(i-1)\Delta}\|^2 ds + o(1) \\ &\leq \int_0^t E\|\hat{\beta}_s - \beta_s\|^2 ds + o(1) \\ &\to 0\end{aligned} \tag{3.43}$$

by an application of the DCT. Similarly,

$$\begin{aligned}&E|\sup_{s\leq t}\textstyle\sum_{i=k_n}^{[s/\Delta_n]}\xi_i^n\Delta_i^n X^{UO}| \\ &\leq \textstyle\sum_{i=k_n}^{[t/\Delta_n]} E|\xi_i^n\Delta_i^n X^{UO}| \\ &\leq \textstyle\sum K\Delta^{1/2}\left(E(\xi_i^n)^2\right)^{1/2} \\ &\leq \textstyle\sum K\Delta^{1/2}\left(\Delta E\|\hat{\beta}_{(i-1)\Delta} - \beta_{(i-1)\Delta}\|^2 + \Delta^{3/2} + \Delta^2\right)^{1/2} \\ &\leq K\Delta\textstyle\sum E\|\hat{\beta}_{(i-1)\Delta} - \beta_{(i-1)\Delta}\|^2 + o(1) \\ &\to 0\end{aligned} \tag{3.44}$$

An analogous argument holds for $\xi_i^n \Delta_i^n N$. For the $(\Delta_i^n N)^2$ term I find by an application of Ito isometry

$$\begin{aligned} \sup_i E(\Delta_i^n N)^2 &\leq \int_{(i-1)\Delta}^{i\Delta} E\|(\beta_s - \beta_{i-1})\tilde{\sigma}_s\|^2 ds \\ &\leq K \int_{(i-1)\Delta}^{i\Delta} (s - (i-1)\Delta) ds \\ &\leq K\Delta^2 \end{aligned} \tag{3.45}$$

Finally an application of Thm 29 from Protter (2004) implies

$$\begin{aligned} \sup_i E|\Delta_i^n X^{UO} \Delta_i^n N| &\leq K\Delta^{1/2} \left(E\left(\Delta_i^n N\right)^2 \right)^{1/2} \\ &= K\Delta^{1/2} \left(\int_{(i-1)\Delta}^{i\Delta} E\|(\beta_s - \beta_{(i-1)\Delta})\sigma_s^f\|^2 ds \right)^{1/2} \\ &\leq K\Delta^{1/2} \left(\int_{(i-1)\Delta}^{i\Delta} (s - (i-1)\Delta) ds \right)^{1/2} \\ &\leq K\Delta^{3/2} \end{aligned} \tag{3.46}$$

As such I conclude

$$\sum_{i=k_n+1}^{[t/\Delta_n]} (\Delta_i^n X^O)^2 = \sum_{i=k_n+1}^{[t/\Delta_n]} (\Delta_i^n X^{UO})^2 + B_t \overset{ucp}{\to} [X^{UO}, X^{UO}]_t = C_t^{UO} \tag{3.47}$$

by $B_t \overset{ucp}{\to} 0$ from the arguments above. The last equality holds as X^{UO} is a continuous Itô semimartingale. The corollary is then immediate as $\sum_{i=k_n+1}^{[t/\Delta_n]} \Delta_i^n X^O \overset{ucp}{\to} \int_0^t \mu_s^{UO} ds + \int_0^t \sigma_s^{UO} dW_s$ and $\sum_{i=k_n+1}^{[t/\Delta_n]} \Delta_i^n X^O (\Delta_i^n X^O)' \overset{ucp}{\to} C_t^{UO}$. ■

3.B Noise and Asynchronous Data

Market microstructure noise and asynchronous data pose challenges when working with high frequency data. I investigate the impact of noise on my test in the Monte Carlo simulations below. Following the literature, I model noise in the observed price process as $\tilde{X}_t = X_t + U_t$ where U_t is i.i.d and independent of X. For a construction of the noise process see Jacod and Protter (2011) Chapter 16. In this setting realized variance estimators have been shown by Zhang et al. (2005) to diverge linearly in the number of increments, or when scaled by the number of increments estimate twice the variance of the microstructure noise. Additionally, Epps (1979) made the observation, known as the Epps effect, that asynchronous data biases correlation estimates toward zero as the

sampling frequency Δ_n decreases. The combination of these effects would tend to bias my estimates of β leading to potentially spurious inference.

To study the impact of noisy data I contrast using the realized variance estimators suggested in Section 3.2 with a multivariate pre-average estimator discussed in Christensen et al. (2010).[3] The pre-average estimator is defined as,

$$[X,\hat{X}]_T \equiv \frac{n}{n-k_n+2}\frac{12}{k_n}\sum_{i=0}^{n-k_n+1}\bar{X}_i^n(\bar{X}_i^n)' \tag{3.48}$$

$$\bar{X}_i^n \equiv \frac{1}{k_n}\left(\sum_{j=k_n/2}^{k_n-1}X_{i+j} - \sum_{j=0}^{k_n/2-1}X_{i+j}\right) \tag{3.49}$$

where $k_n = n^{1/2+\delta}$ for a small $\delta \in (0,1/2)$. That $\delta > 0$ ensures the estimator is consistent. Intuitively, the pre-average estimator eliminates the noise by averaging an increasing number of observations over small windows and then differences averages in nearby windows. Christensen et al. (2010) derive the asymptotic properties of this estimator in the presence of noise and demonstrates that it performs well in the presence of asynchronous data.

My Monte Carlo simulation is from the Heston model. Following Aït-Sahalia et al. (2010), I consider the data generating process

$$\begin{aligned} dX_{it} &= \mu_{it}dt + \sigma_{it}dW_{it} \\ d\sigma_{it}^2 &= \kappa_i(\bar{\sigma}_i^2 - \sigma_{it-}^2)dt + s_i\sigma_{it-}dB_{it} + \sigma_{it-}J_{it}dN_{it} \\ dr_t &= \kappa_r(\bar{r} - r_t)dt + s_r\sqrt{r_t}dB_{rt} \end{aligned} \tag{3.50}$$

and I impose an ICAPM model with one hedging portfolio

$$\mu_{pt} = r_t + \beta_t \begin{bmatrix} \mu_m - r_t \\ \mu_h - r_t \end{bmatrix}. \tag{3.51}$$

The parameters are detailed in Figure 3.3. The initial condition for volatility is sampled from the stationary distribution of the CIR process, $\sigma_{i0}^2 \sim \Gamma(2\kappa_i\bar{\sigma}_i^2/s_i^2, s_i^2/2\kappa_i)$. Jumps are of size $J_t = \exp(z)$

[3]Alternatively, one might use a subsample estimator Zhang et al. (2005), quasi-maximum likelihood estimator Xiu (2010) and Aït-Sahalia et al. (2010), or a realized kernel estimator Barndorff-Nielsen et al. (2008).

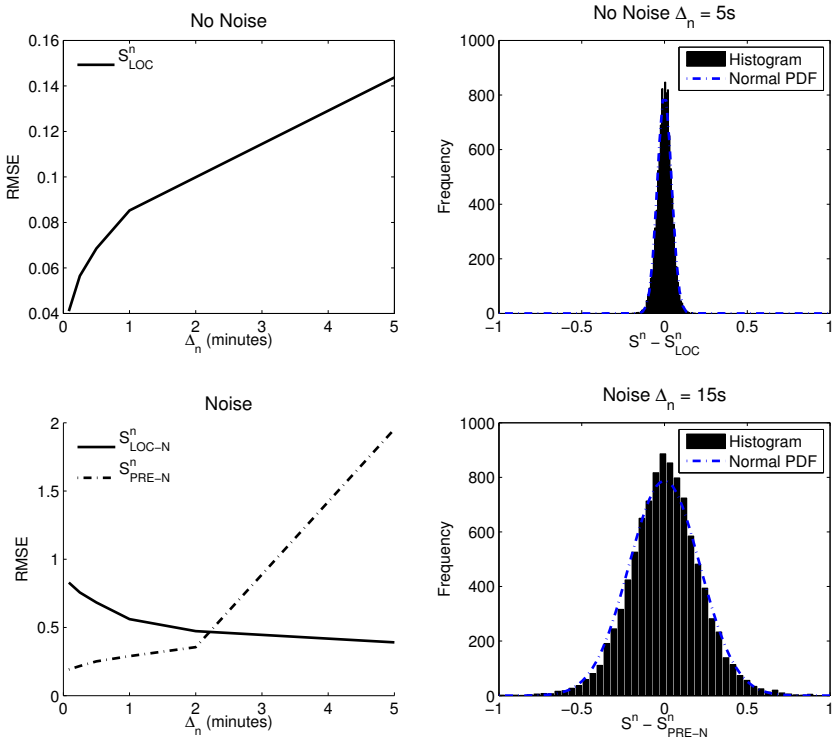
where $z \sim N(\theta_i, \alpha_i)$ and N_{it} is a Poisson Process with intensity λ_i. The correlation of the diffusive components are $E[dW_{it}dW_{jt}] = \rho \cdot dt$, $E[dW_{it}dB_{it}] = \rho_i \cdot dt$, and $E[dW_{jt}dB_{it}] = 0 \cdot dt$ when $i \neq j$.

My study contrasts how the feasible and infeasible statistics perform when the data is generated in the model above at a $\Delta_n = 1s$ frequency for a 6.5 hour trading day. Under the case with no noise or data desynchronization I see from the plots in Figure 3.3 that the root mean square error decreases as the sampling frequency increases and that $S^n - \hat{S}^n_{LOC}$ appears to be normally distributed. Recall that S^n is the infeasible statistic in equation 3.7. I use the notation $\hat{S}^n_{LOC}$ to denote the feasible statistic in equation 3.15.

To generate the noisy data, I desynchronize the observations by generating random intervals from an exponential distribution and random times within each interval when the various assets are observed. Then, I contaminate the data with iid noise that is distributed $N(0, a_i^2)$ and resync the data with the Refresh Time scheme suggested in Aït-Sahalia et al. (2010). Finally, I compute the feasible estimators $\hat{S}^n_{LOC-N}$ and $\hat{S}^n_{PRE-N}$ which use the local volatility estimator and pre-average estimator respectively to estimate beta and standardize the statistics. Figure 3.3 indicates that the root mean square error for both of these statistics is significantly larger after the data is contaminated with noise. Moreover, as the sampling frequency increases the root mean square error for $\hat{S}^n_{LOC-N}$ actually increases, in line with the observation in Zhang et al. (2005) that the realized variance diverges in the presence of noise. In contrast the root mean square error of $\hat{S}^n_{PRE-N}$ decreases but at a slow rate and is still higher at a 5 second sampling frequency than the error at a 5 minute sampling frequency in the case with no noise. The analysis suggests that noisy data significantly hinders the performance of my test.

Finally, Figure 3.4 provides a visual illustration for how noise impacts my estimation of beta. The true value of beta is plotted in blue. I see in the presence of no noise the realized estimator from Section 2 tracks the true value quite closely using a one day rolling window. After adding noise and desynchronizing the prices, the realized estimator is heavily biased toward 0 illustrating the Epps effect. The pre-average estimator does a better job at estimating the true value of beta but seems to exhibit the typical bias versus variance trade-off familiar to non-parametric statistics. In particular, the figure suggests that the pre-average estimator is quite volatile. I could increase the lag window to lower volatility but this would increase the bias.

Monte Carlo Study: The Impact of Noise and Asynchronous Data



Parameter Values for Monte Carlo Study

Asset	X_{i0}	μ_i	κ_i	s_i	$\bar{\sigma}_i^2$	ρ_i	λ_i	θ_i	α_i	ρ	a_i
$i = p$	$\log(100)$	-	6	.5	.10	-.60	36	-6	1.2	.4	.001
$i = m$	$\log(200)$	.05	5	.3	.05	-.75	12	-5	.8	.4	.001
$i = h$	$\log(50)$	.04	4	.4	.05	-.75	24	-5	.8	.4	.001
$i = r$	$\Gamma(2\kappa_r\bar{\sigma}_r^2/s_r^2, s_r^2/2\kappa_r)$	-	5	.5	.02	-	-	-	-	-	-

Figure 3.3: This plot illustrates the impact of noise and asynchronous data. The Monte Carlo simulation is described in Appendix 3.B. The top plots consider the case with no noise. In particular, the top left plot shows that the root mean square error between S^n and the feasible $\hat{S}^n$ decreases as the sampling frequency Δ_n decreases in 1,000 draws. The top right plot shows the difference between S^n and $\hat{S}^n_{LOC}$ appears normally distributed at a $\Delta_n = 5$ second sampling frequency in 10,000 draws. The bottom plots perform the analogous computations for the case with noisy data. In this case, the realized beta computed from the local volatility estimator performs poorly. The root mean square error of $S^n - \hat{S}^n_{LOC-N}$ actually increases as the sampling frequency decreases, which contrasts the case with no noise. While the root mean square error between $S^n - \hat{S}^n_{PRE-N}$ does decrease when using the pre-average estimator, the histogram in the bottom right plot reveals that the differences between the feasible and infeasible statistics are much larger in the presence of noise.

Estimating Beta: The Impact of Noise and Asynchronous Data

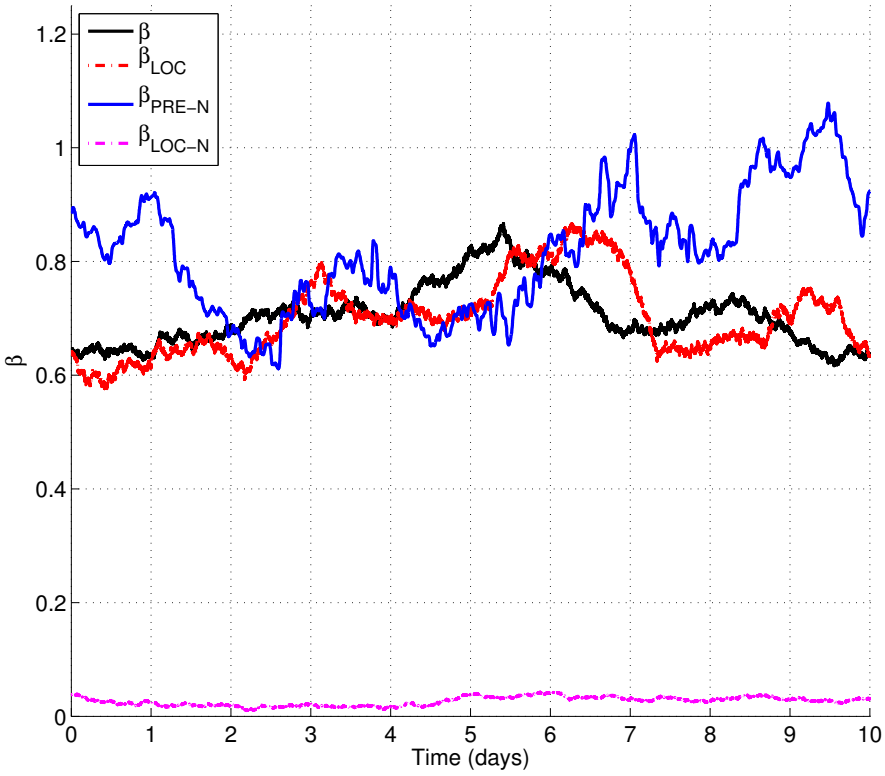


Figure 3.4: This plot illustrates the difficulties noisy data pose for estimating β. The black line plots the true beta from the Monte Carlo simulation described in Appendix 3.B. Before contaminating the data, I estimate the realized beta β_{LOC} using the local volatility estimator from Section 3.2 from $\Delta_n = 5$ second observations with a rolling window of approximately 1 day. The realized beta β_{LOC} is plotted in red and tracks the true beta quite closely. I then desynchronize the data and add noise to the observations. The realized beta from the local volatility estimator using the noisy data is denoted as β_{LOC-N} and is plotted as the dashed magenta line. As is evident from the plot, β_{LOC-N} is heavily biased toward 0 which illustrates the Epps effect. The pre-average estimator is also included in blue β_{PRE-N} using a 4.5 day window. It tracks the true beta more closely than β_{LOC-N} but is quite volatile.