Essays on Empirical Asset Pricing
ISBN (please fill in the number)

Cover design: Crasborn Graphic Designers bno, Valkenburg a.d. Geul

This book is no. 582 of the Tinbergen Institute Research Series, established through cooperation between Thela Thesis and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.
Essays on Empirical Asset Pricing

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam
op gezag van de rector magnificus
prof.dr. F.A. van der Duyn Schouten,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Economische Wetenschappen en Bedrijfskunde
op maandag 26 mei 2014 om 11.45 uur
in de aula van de universiteit,
De Boelelaan 1105

doors
Xiaoyu Shen
geboren te Anhui, China
promoter: prof.dr. A.C.F.Vorst

copromotor: dr. F.Brevik
Acknowledgements

My time in graduate school has been enriched by the help of many people. It is a great experience, now I believe, to have met and worked with various people and learned from them. This was the case with the VU finance group, the Tinbergen Insitute, and the Duisenberg School of Finance. Here I would like to take this opportunity to thank the people who have made the research in this thesis possible.

First and foremost, I would like to thank my supervisors, Prof. Ton Vorst and Dr. Frode Brevik. Without them this thesis would have never been completed. Their joint supervision was balanced and complementary. I was lucky to meet Frode at the time I was about to write my M.Phil thesis. More than that, I am thankful for his decision to let me join his research. His enthusiasm for pursuing challenging problems at the highest levels of integrity and curiosity has been a constant source of inspiration, excitement, advice, and guidance throughout our collaboration and my own research later on. I am also thankful to his openness for allowing and supporting me to independently enter into a different research area. Gaining independence in research is not an easy process. My understanding is that it means the development of cat senses for exploration of valuable research questions, and a high level of rigor as well. It was Prof. Ton Vorst who guided me through this journey. He made a great effort to instruct me to become scientifically rigor and understand the gap between theory and practice. He is also a generous supervisor in providing financial support for my research. As I look back at our mentorship, it benefits my personality.

From the VU finance group, there are many people I would like to thank. Prof Andre Lucas has always been kind to spend time with me on discussion, as he is for every graduate student in the department. Prof Philip Stork provided a long list of comments and questions on my papers, with which I prevented a few mistakes. I also thank Prof Albert Menkveld. We could often see each other at the department on weekends, and he always
had a keen interest in the progress of my research. In fact I benefited from interactions with every faculty member, as the department was the most important place for life of a graduate student. Of course, I cannot forget to mention the fellow graduate students. Everyone of them helps me to maintain a self-torturing life style.

From Tinbergen Institute, I would like to thank Prof. Wouter den Haan for his kindness of offering advices, a couple of times when I needed to make decisions for my education and career. I also thank Arianne de Jong, Judith van Kronenburg, Babs van den Berg, Ester van den Bragt, and Christina Carvalho for all of their support whenever I requested. Just as importantly, I thank my TI fellow students for the fun and wonderful moments we shared as friends. They were the great times that we spent together on parties, drinks, dinners, and the back-to-back travelings between Amsterdam and Rotterdam. I am also grateful to them for helping improve this thesis. I benefited from discussions with Istvan Barra, Falk Bruning, Yun Dai, Lerby Ergun, Xiaoyue Li, Wei Li, Jiangyu Ji, Natalya Martynova, Eriikki Silde, Patrick Tuijp, Xin Zhang, and Yueshen Zhou. I also owe a particular debt to Zhengyuan Gao, Yang Zu, and Xiye Yang, together with whom I learned a lot on high frequency econometrics.

I thank Duisenberg School of Finance for sponsoring my Ph.D study. Particularly, I enjoyed the interactions with the Dean Prof. dr. Dirk Schoenmaker and am thankful to the support from Wim Touber and Ingrid van Beek.

Finally, my special thanks go to my parents. I am used to making my own decisions and sometimes these decisions give them feelings of uncertainty. But as time goes by, we are able to share more and more. I am happy for us.
# Contents

1 Introduction .......................................................... 1

2 Depression fears: Asset Prices with Ambiguity Averse Investors ......................................................... 11
   2.1 Introduction .................................................................. 11
   2.2 The model .................................................................. 13
      2.2.1 Worst case beliefs ................................................. 16
      2.2.2 Asset prices ........................................................ 17
      2.2.3 Other shades of robust beliefs ............................... 20
   2.3 Model estimation .......................................................... 21
      2.3.1 State specification ................................................. 21
      2.3.2 Data .................................................................. 21
      2.3.3 Preference parameters .......................................... 22
   2.4 Model predictions .......................................................... 26
      2.4.1 Term structure ..................................................... 26
      2.4.2 Asset price moments ............................................. 27
      2.4.3 Equity valuation .................................................. 28
   2.5 Conclusion .................................................................. 31

3 An Anomaly between Two Volatility Markets ......................................................................................... 35
   3.1 Introduction .................................................................. 35
   3.2 Theoretical Underpinnings ........................................... 38
      3.2.1 Option-synthesized Variance Swap Rate .................... 38
      3.2.2 The VIX and VIX Futures ....................................... 39
   3.3 Empirical Measurements ............................................... 40
      3.3.1 Data and Empirical Features ................................... 40
      3.3.2 Matching Maturities of Options and Futures .............. 42
      3.3.3 Variance of the VIX .............................................. 44
3.4 Trading On Negative Observations of the Variance of the VIX .... 46
  3.4.1 Long-Short Strategy ........................................ 46
  3.4.2 Short Strategy ............................................. 48
3.5 Robustness ..................................................... 49
  3.5.1 Number of Options for Approximating the Variance Swap Rate .. 49
  3.5.2 A benchmark for evaluating returns .......................... 50
  3.5.3 An alternative choice of matching .......................... 52
3.6 Conclusion ..................................................... 54

4 Testing Affine Stochastic Volatility Models with the VIX 67
  4.1 Introduction .................................................. 67
  4.2 AJD Models and their Implication for Expected Forward Integrated Variance 70
  4.3 VIX ........................................................... 75
  4.4 Instantaneous Variance ....................................... 76
  4.5 Data .......................................................... 77
  4.6 Test AJDs ...................................................... 79
    4.6.1 Simple OLS ................................................. 79
    4.6.2 Volatility Persistence ................................... 82
  4.7 Hedging Volatility ............................................ 85
  4.8 Conclusion .................................................... 90

5 Conclusion ....................................................... 91

Bibliography ........................................................ 98
Chapter 1

Introduction

“There are two polar approaches to this elaboration. I call them absolute pricing and relative pricing. In absolute pricing, we price each asset by reference to its exposure to fundamental sources of macroeconomic risk. ... In relative pricing, we ask a less ambitious question. We ask what we can learn about an asset value given the prices of some other assets.”


Asset price is the fundamental value to be explained by asset pricing theory. The contrast between the absolute pricing and relative pricing, reflects the positive versus normative tension present everywhere in economics. In absolute pricing, one describes investors’ preference and the economic structure, and prices an asset by reference to its exposure to the fundamental uncertainty. A key objective is to offer a plausible positive explanation of the observed historical and current prices. On the other hand, the relative pricing approach establishes an asset’s value on the basis of the values of other assets. Implicitly, it often implies a normative question, in which we have to decide whether some claims are mis-priced, and arbitrage opportunities might exist.

Almost no asset pricing study is taken purely by one of the two extremes. No exception is this thesis. It is a collection of three papers that mixes approaches of the absolute pricing and relative pricing and studies the topic “risk and uncertainty”. The first paper is a positive study in which a new robustness preference expressing “aversion to uncertainty” is brought into a consumption asset pricing model to answer the equity premium puzzle. The second and third paper are normative studies on volatility derivatives. Volatility is one of
the most popular measurements of risk in modern finance theory, and volatility derivatives are derivative products whose prices are based on the volatility of the underlying asset price. Hence prices of volatility derivatives should reflect the premium of the volatility risk. The second and third paper are devoted to comparing prices of volatility risk across different volatility markets, and studying the performances of stochastic volatility models.

The differences between uncertainty and risk, and also the associated uncertainty aversion and risk aversion are important but subtle. The recent development of economics has emphasized more and more on the distinction of uncertainty and risk, an idea dating back to Knight (1921)

“Uncertainty must be taken in a sense radically distinct from the familiar notion of Risk, from which it has never been properly separated.... The essential fact is that 'risk' means in some cases a quantity susceptible of measurement, while at other times it is something distinctly not of this character; and there are far-reaching and crucial differences in the bearings of the phenomena depending on which of the two is really present and operating.... It will appear that a measurable uncertainty, or 'risk' proper, as we shall use the term, is so far different from an unmeasurable one that it is not in effect an uncertainty at all.”

To make a clarification not rigorous but sufficient for this thesis, uncertainty in this thesis means the ambiguity of a probabilistic distribution, and risk means the randomness of a probability distribution.

The first paper adopts the absolute pricing approach and studies a preference-based consumption asset pricing model. A new robustness preference is proposed and applied to study the link between the valuation of the aggregate U.S stock market and the business cycles. Though a large literature in finance is written directly on models of returns, the choice of a reduced form model becomes not as promising as for returns of individual assets when it comes to explain the returns of the aggregate market. The true risk factors which the market is exposed to are macroeconomics in nature. Therefore the explanation has to be about the systemic risk factors. On this point the aggregate consumption uncertainty is widely believed to be the risk source, as a tradition inherited by modern finance from macroeconomics. In these models, the stream of the uncertain consumption flow is discounted by the marginal rate of substitution that is decided by the representative
agent’s utility function. However, the classic consumption-based models that focus on the representative agent formulation of the model with time-separable power utility fail the empirical examinations. The failures are reflected on the several inconsistencies between the model predictions and the empirical phenomena of the prices of the fundamental assets in the US market: the aggregate stock market, the short-term treasury bills, and the long-term treasury bonds. Usually economies where investors have time-separable preferences feature counterfactually high risk-free rates and an equity premium that is too low. (Mehra and Prescott, 1985; Weil, 1989; Hansen and Jagannathan, 1991). They also feature term structures of real interest rates that are on average slightly downward sloping (Backus et al. (1989)), while empirically long term bonds earn positive excess returns. The same is also true for Epstein-Zin preferences (Piazzesi and Schneider, 2006; Beeler and Campbell, 2009).

In the attempt to overhaul the canonical consumption-based paradigm, a new generation of models take account of new preferences, or new restrictions on the dynamics of the fundamental cash flows, or market inefficiency such as incompleteness and limited participation. The shared view is that one needs a model in which the market price of risk is time varying and higher than in the standard consumption-based model. The first paper in this thesis refines the rational expectations models to attribute model specification doubts to the agents in their making of investment decisions. As it turns out, the robustness against the rational expectation belief of the fundamental state of the economy makes the consumption uncertainty riskier in the eyes of robust investors and pushes down the equity prices.

In our model, robustness kicks in as the artificial representative investor fears about the preciseness of his belief of the state of the economy in the business cycles. The cycles are modeled in the form of regime-switching patterns of aggregate consumption growth. The endowed regimes and the switch are aimed to capture the empirical feature of the aggregate consumptions, which like many macroeconomic and financial series tends to exhibit different behaviors during economic downturns and upturns. Therefore to value the claim to the aggregate consumption, the agent needs to take account of the variation of the good-or-bad state. His expectation of the states in the future is formed on the basis of the transition dynamics of the states that are in the model governed by a finite state Hidden Markov chain (HMM). In a simple Markov chain, the state is visible. In HMM, the state is not directly visible, but signals dependent on the state, such as consumption
and GDP, are visible. In this way, the outputs of the economy are considered at the higher layer on top of the underlying state. However, the agent cannot know the exact state through a top-to-bottom inference, as the output signals are also subject to other shocks. In the model economy, the shocks are Gaussian noise whose volatility is uniquely decided by the hidden regime.

As the HMM implies, the agent’s probabilistic belief of the future states is conditional on his estimation of the conditional distribution of the current state of the economy. For that the agent needs to form some kind of ”best estimate” using all the past observations of signals. Here the difference between a rational expectations agent and a robust agent emerges. Given the Markovian structure and the Gaussian specification of the noise, a rational expectations agent solves this optimal filtering problem in a sequential updating procedure, that is widely known to economists as ”Hamilton filter”. However, a robust agent distrusts the rational expectation estimation and assigns more weights in his belief to the worst state according to the formulation of his robustness.

To illustrate the robustness estimation, we consider a simplified version of the model. Suppose at each time there are a finite number of states possible for the economy $I = \{i = 1, 2, ..., I\}$, and from time 0 to $t$ a sequence of signals, for example, the aggregate consumption, $C(0), C(1)...C(t)$ is observed. Conditional on the observations of historical consumption, the rational expectation estimation of the state likelihoods at time $t$ assigns positive probabilities to each state, with $\pi_i > 0$ and $\sum_{i=1}^{I} \pi_i = 1$. Provided that the current economy was known almost surely to be at the state $i$, the utility $V_i$ that corresponds to the coming consumption stream would be the same for both the rational expectations agent and the robust agent. The value function of the rational expectations agent is the result of averaging $V_i$ by his belief of the state probabilities, that is, $\sum_{i=1}^{I} \pi_i V_i$. Therefore it is recognized that the rational expectations agent we consider is Bayesian. In contrast to the rational expectations agent, the robust agent dislikes the fact the rational expectation belief is probabilistic in nature and is concerned about this uncertainty of the estimated state distribution.

Robustness in economics is recently actively advocated by Hansen and Sargent (see, e.g., Hansen and Sargent (2001), Hansen and Sargent (2007b), and Hansen (2007)). The justifications come in from both philosophical and pragmatic reasoning. Perhaps the most obvious reason why we think robustness is attractive is that it can help explain the aggre-
gate equity premium that the rational expectations cannot. In comparison to the rational expectations agent, the robust agent is more cautious in using the Bayesian estimates of the state distribution as he knows it is only an approximation to the true world in the model. Out of the aversion to this ambiguity, he seeks robust decision rules in case that the rational expectations estimation turns out to be too optimistic. In this spirit he picks an alternative state distribution that assigns more weights to the worst state. Hence the consumption uncertainty becomes riskier for him.

Without any constraint the robust agent would assign probability one to the worst state. Though we let the agent put doubts of the state estimation on the table, wanting robustness should be plausible. It would be an obvious contradiction to the agent’s knowledge that the underlying regime switches according to a Markov chain, if he always believed that the economy is in the worst case. Therefore we restrict the space within which the robust agent can choose the alternative state distribution. This space is defined through the concept of relative entropy, which is a non-symmetric measure of the difference between two probability distributions P and Q.

In a general sense, if P and Q are probability measures over a set X, and P is absolutely continuous with respect to Q, then the relative entropy of P with respect to Q is defined as

$$\mathcal{R}(P, Q) = \int_X \log \left( \frac{dP}{dQ} \right) dP dQ$$

where \( \frac{dP}{dQ} \) is the Radon-Nikodym derivative of P with respect to Q, and provided the integral exits.

The relative entropy we consider is the one of the distribution of future consumption growth rates induced by the robust belief of the current state distribution with respect to that induced by the rational expectations belief. An entropy ball consists of all the alternative distributions that are mutually continuous to each other and whose relative entropy with respect to the distribution induced by the rational expectation estimation is no greater than the maximum relative entropy. We let the robust agent pick the most pessimistic belief of the state distribution subject to the constraint of this entropy ball. In consequence, we find uncertainty aversion can explain the empirically low risk-free rate, the bond premium, the equity premium, and equity returns.
The second and third paper turn to a different subject. They are normative studies of the volatility index (VIX) and volatility derivatives. Volatility is widely used as a measurement of risk, and measures how much variation in asset returns exists from the average return. This seemingly clear definition, however, can be thought-provoking on account of that an inherent part of it has to do with the sampling frequency of returns. In the data, returns at daily, weekly, and monthly frequency present different empirical features and so do the measured return volatility. Finding a volatility model that can consistently describe these stylized features is far from trivial. With an econometrics or derivatives pricing model, one soon realizes that volatility is often a latent variable in a parametrized setting. Think of an ARCH or GARCH model, only return is the visible variable and volatility is treated as latent. From this point of view, it seems that volatility is difficult to trade.

On the other hand, volatility trading is appealing in its own characteristics. Typically volatility rises when uncertainty increases, and goes up and tends to stay up when returns are low. In fact, buying and selling volatility through options has long been common to traders. The search for possibility of trading volatility as an independent asset class calls for a volatility benchmark. Ideally this benchmark should be a standard as objective as possible.

The early literature that writes about developing volatility indexes and listing related volatility derivatives is small but prescient. A key breakthrough that later shaped the landscape of volatility trading was the discovery of the link between the price of a contract that pays the natural log of an underlying asset price at its expiry, and the limit of the sum of squared returns. The sum of squared returns is named “realized variance”. Realized variance in the limit of continuous sampling is a consistent estimator of the corresponding quadratic variation, if we think the asset price evolves in the form of a continuous time semimartingale. Under a general assumption that the underlying asset price follows a diffusion process driven by a Brownian motion, the price of the log contract is equal to half of the expectation in the risk neutral measure of the realized variance up to the expiry of the log contract (see Neuberger (1990) and Dupire (1993)). Therefore pricing realized variance is equivalent to pricing a log contract.

One of the direct applications of this research outcome is found in the trading in variance swaps. A variance swap is a financial derivatives instrument that allows two parties
to change the predetermined variance swap rate for the realized variance between the
inception and expiry of the swap contract. Therefore the payoff to the long party is the
difference between the realized variance and the fixed swap rate. The first documented
trading of a variance swap is a deal in 1993 by Michael Weber who was then at the Union
Bank of Switzerland (UBS). According to Weber, UBS bought one million pounds per
volatility point on the FTSE 100. The initial swap rate was quoted at the at-the-money-
forward implied volatility less one volatility point for safety. Later the contract was valued
by the method of Neuberger (1990). According to Carr and Lee (2009), variance swaps
on stock indexes began to take off in 1998 due to the historically high implied volatilities
then. It was such an environment in which the hedge funds sold variance swap rates that
exceeded their econometric forecasts of realized variance in the future, and banks on the
other side were happy to buy them as long as it was possible to earn even higher premiums
from shorting a strip of options and delta-hedging them.

The reason that banks sold options is due to the theoretical justification that under
the same assumption mentioned earlier, the log contract and hence the variance swap can
be replicated by a static position in European options and a dynamic trading in futures,
on the same underlying asset and at the same maturity date. Since the theoretical ex-
ante trading cost of futures is zero, the variance contract can be hedged by the coterminal
European options. This insight later led to the birth of the modern Chicago Board of Op-
tions Exchange (CBOE) volatility index (VIX) in 2003. The VIX is a volatility benchmark
derived from SPX options to reflect the market expectation of the volatility of SP500 over
the next 30 day period. It is an enhancement of the old volatility index (VXO), but now
inferring the variance swap rate from the option prices becoming the theoretical under-
pinning of the new methodology. So VIX is the square root of the next 30 day variance
swap rate for SP500.

The creation of the volatility index opens the door to derivatives contracts written on
it. Shortly after, the futures contracts written on VIX were introduced for trading. Over
time the VIX futures attract more and more liquidity. Hence it becomes interesting to
know the efficiency of price discovery in the VIX futures market. Given the price parity
between the log contract and the variance swap, we expect a relation between the futures
prices and the log contract price as well. The price of the log contract can be inferred
from SPX option prices. Actually there is an upper bound of the futures price that can be
represented as a function of option prices. The validity of this bounding relation becomes
the question that the second paper in this thesis aims to answer.

This empirical exercise tries to form a normative statement about the efficiency of price discovery between the VIX futures market and SPX options market. Recall the case of the boom of variance swap trading in 1998. At that time banks engaged in a cross-market arbitrage in which they bought variance swap rates at low prices and sold the option-implied variance rate at high prices. Carr and Lee (2009) attribute the opportunity for banks of making such a profit to lack of infrastructure for hedge funds to delta-hedge options back then. It is hard to imagine that this limitation is still around today. On the other hand, the possibility of a cross-market mispricing cannot be excluded ex ante, if market frictions exist.

The idea is to calculate the cost of hedging the variance swap by options and compare it to the price of the VIX futures, given that the variance swap and VIX futures are priced on the variance and volatility over about the same period. This exercise discovers a recurrent inconsistency that the expected bounding relation between the VIX futures price and the variance swap price implied by options is broken. In fact, combining S&P500 index option prices at different maturities and strike prices provides a model free measurement of expected return variances up to the option maturities, while market prices of futures on the volatility index VIX measure the expected return volatilities over the thirty days after the futures maturities. If options and futures markets are integrated, the same pricing kernel should apply to both, and the option-implied variance and futures-implied volatility can be combined to give the risk-neutral variance of the VIX. But taking them together often implies a negative measurement of the variance of the VIX, a mathematical impossibility. The significance of the found inconsistency is further supported by the high returns of the trading strategies that are designed to exploit the inconsistency. Two types of trading strategies are developed. A negative observation of implied risk-neutral variance indicates that the relative price of the option implied variance is too low relative to the volatility implied from VIX futures. A natural way to trade this anomaly is to short the VIX futures and long the option-anesthetized variance swap. In doing so we find the trading profits are large. But this result is only based on not taking account of the bid-ask spreads, and the transaction cost of the portfolio of a large number of options can actually be significant enough to bias the results. Therefore a shorting strategy which only shorts VIX futures using market prices is implemented and found sharp positive returns as well. The results clearly send a signal to further reflect on the volatility market.
The precise valuation of VIX futures and other volatility derivatives depends on a parametric volatility model and among all sorts of models the continuous time models are often preferred due to their relative computational tractability. Meanwhile, an empirical examination of volatility models can shed light on the validity of the assumptions that lie behind the hedging of variance swap by coterminal options. The assumptions require the path of the underlying asset price is continuous and the uncertainty in prices is induced by stochastic integrals with respect to a Brownian motion. However, using different volatility models can lead to different conclusions. For example, in a framework where pricing variance swap is done on a general exponential Levy process which is further stochastically time-changed by a continuous integrable clock, the model assumptions are different and in consequence the arguments in support of the VIX index are not valid any more.

What particularly makes the exercise of examining volatility models interesting is the availability of high frequency price data in recent time. There is a natural connection between the realized measure of variance and the semi-martingale stochastic process that the efficient market price is believed to follow, as the former reflects the quadratic variation of the latter. As a motivation, if the efficient log-price follows a pure diffusion process,

\[ dS(t) = u_t dt + \sigma_t dW_t \]

and we have equally spaced observations \( \Delta S_{t(i)}, i = 0, 1...n \) over \([0, T]\):

\[ t(i) = i \Delta t, \Delta t = T/n \]

then we will have

\[ \lim_{\Delta t \to 0} \sum (\Delta S_{t(i)})^2 = \int_0^t \sigma_u^2 du \] (1.1)

Even there is a possibility that the relation above is violated by the convolutions with jumps in returns and microstructure noise, solutions have been found to filter out the convoluted factors and it is possible to develop non-parametric statistics to estimate the integrated and spot variance. Naturally these nonparametric volatility estimators serve as touchstones for evaluating parametric volatility models.

Using a nonparametric estimation of volatility, the third paper provides an evaluation of the affine volatility models. As these models are consistent with the theory of the VIX, this exercise can be seen as a first step to explore the assumptions behind the VIX.
The class of the affine volatility models includes a wide scope, not only the pure diffusion process driven by Brownian motion, but also stochastic volatility models with Possion jumps. Presence of Possion jumps causes approximation errors in the VIX and Jiang and Tian (2005) and Carr and Wu (2009a) argue that they are small. A key feature of the affine volatility models is that there is a one-to-one linear mapping between the instantaneous variance and the VIX square. The instantaneous variance can be estimated by high frequency returns, and therefore this feature can be empirically examined. We find that indeed jumps do not have an big impact on VIX, and that the linear relationship is ruined by volatility persistence. Though volatility persistence is a well known feature, unfortunately so far it has not been taken into account by most derivatives models due to computational reasons. We also investigate the hedging of the variance swap rate under the affine models, and find a systematic error.

The thesis consists of five chapters in total. The three papers are presented in order from chapter 2 to 4. The final chapter concludes. There I first summarize the most important findings of this dissertation. Then I shortly discuss what these results imply for any further exploration.
Chapter 2

Depression fears: Asset Prices with Ambiguity Averse Investors

This chapter is co-authored with Frode Brevik.

2.1 Introduction

Since the second world war, the US has not experienced a period of economic distress of the magnitude of the great depression, but the possibility of such events should be considered by investors holding risky securities. We include a depression state in the set of fundamental states in a Hidden-Markov model. In the economy we consider, observations of the underlying growth state is blurred by Gaussian noise, and investors can never fully exclude the possibility that the economy is actually in the depression state. Investors treat the underlying state of the economy as ambiguous and are wary of trusting fully the filtered state probabilities implied by historical data when making investment choices. Instead they seek to make robust decisions that work well also if their assessment turned out too optimistic. This has first order effects on asset prices.

The laboratory endowment economy is estimated using only macroeconomic time series. Both trend growth rates and volatilities are subject to regime switching. Trend growth regimes capture business cycle states, while consumption volatility regimes capture low frequency fluctuations in macroeconomic risk. In particular, the great moderation is captured as a low volatility regime. In addition to these states which we take as being the fundamental states that actually occurred in the post-war period, we give agents knowledge about another possible state that the economy can enter: a state calibrated to match
the experience during the great depression. Investors know the underlying structure of the economy, but only see realized growth rates.

Departing from a benchmark rational expectations specification (the reference model), we let the representative investor seek decision rules that are robust to state uncertainty. This robust representative investor prices financial assets different from a bayesian investor. A bayesian investor would reduce the problem of pricing with uncertain states to weighing the prices of the claims that would have obtained if the state was observed with the posterior probability vector that comes from the filtering. (See, e.g. Detemple (1986), David (1997), Veronesi (2000) and Ai (2010). Our robust representative investor prices asset with worst case probabilities that are endogenously shifted more towards states of the economy where his continuation utility is low. In particular, under the measure he uses for pricing he endogenously attributes a much greater probability to the depression state than what a bayesian investor would do. He fears depressions.

The key role of time-varying growth rates and volatilities is common with the model of Bansal and Yaron (2004). But we share the skepticism of Hansen and Sargent (2009) that these should be directly observable to investors. Lettau et al. (2008) provides an account of the relation between macroeconomic volatility and valuation that is close in spirit to ours. We also build on a series of papers studying the effects of robustness and learning on asset prices, including Cagetti et al. (2002), Maenhout (2004), Hansen and Sargent (2007a, 2011), Hansen (2007), and Leippold et al. (2008). We differ from all these papers by incorporating a depression state and in the way we incorporate ambiguity aversion.

We follow Hansen and Sargent (2007a) in considering forward-looking concerns for misspecification, but because the set of worst case models the agent chooses from share the model hyper-parameters of the benchmark model, we do not need to resort to discounted entropy. We argue that predicted relative entropy is what investors should care about and show that the model is better able to match asset prices when investors use this statistic instead of other commonly used measures of model divergence.

Together with the proposed version of robustness, standard time-separable log-utility preference can take account of many features in the asset price data. Usually economies where investors have time-separable preferences feature counterfactually high risk-free rates and an equity premium that is too low. (Mehra and Prescott, 1985; Weil, 1989; Hansen and Jagannathan, 1991). They also feature term structures of real interest rates that are on average slightly downward sloping (Backus et al. (1989)), while empirically long term bonds earn positive excess returns. The same is also true for Epstein-Zin preferences (Piazzesi and Schneider, 2006; Beeler and Campbell, 2009). In contrast, in
economies where the representative investor are sufficiently averse to uncertainty about the state of the economy, real-yield curves are on average upward sloping (Brevik, 2009). We use this feature to pick preference parameters that match the empirical term structure of interest rates. The performance of the model is judged on how it matches other asset pricing dimensions. At this calibration, the model generates a plausible equity premium of 6.12%, but the implied Sharpe ratio of equity is low by historic standards because the model returns are—at the parameters needed to match the empirical interest rates—too volatile. Especially noteworthy is that valuation ratios in the model economy soar during spells of low volatility. Hence the model can replicate the run-up of equity valuations in the period of Great Moderation.

Our paper is also related to a growing body of literature that builds on Rietz (1988)’s idea of rare economic events recently reinvigorated by Robert Barro (2006; 2009) and others (e.g. Gabaix, 2008). While in this literature the event is observable, the investors in our model do not know whether it has taken place or not. A depression is hidden by noise. The model investors fear that even normal looking growth rates are dressed up realizations from a depression regime.

The structure of the rest of the paper is the following: Section 2.2 describes the stochastic processes driving the economy, the evolution of the beliefs of the investors, and their preference structures. Section 2.2 also gives algebraic expressions for the equilibrium prices of all assets traded in the economy. Section 2.3 describes how we estimate the process parameters and calibrate the preference parameters of the representative investor. Section 2.4 gives model predicted asset prices and compares them with empirical ones. Section 2.5 concludes.

2.2 The model

Following Longstaff and Piazzesi (2004) we extend the Lucas (1978) and Mehra and Prescott (1985) framework by introducing an explicit model of corporate cash flows into an otherwise standard endowment economy. Corporate cash-flows represent a variable fraction of aggregate consumption. This extension allows us to capture better the high sensitivity of financial income to economic shocks and has crucial asset-pricing implications. We consider an economy populated by an ambiguity averse representative investor.

The agent has two sources of income of a non-storable consumption good. First, the agent receives an exogenous endowment $I_t$ of the consumption good (his nonfinancial income). Second, the agent is also initially endowed with one share of a stock that pays
dividends $D_t$ in the form of the consumption good. The stock is thus a claim to dividends and not aggregate consumption, which in equilibrium is given by the sum of dividends and nonfinancial income, $C_t = D_t + I_t$. The corporate fraction $F_t$ of dividends in the consumption good is specified as

$$F_t = \frac{D_t}{C_t}$$  \hspace{1cm} (2.1)$$

We assume that aggregate log consumption growth, $g_{t+1} = \log C_{t+1} - \log C_t$, is conditionally normally distributed as

$$g_{t+1} = \mu_{t+1} + \sigma_{t+1}^e \epsilon_{t+1},$$  \hspace{1cm} (2.2)$$

where $\epsilon_{t+1}$ is a sequence of standard normals. Both the trend growth rate $\mu_{t+1}$ and the conditional volatility of consumption growth $\sigma_{t+1}^e$ depend on the state of the economy $S_{t+1}$. Time variations in the trend growth rate $\mu_{t+1}$ capture cyclical variations in consumption growth. During booms average consumption growth is higher than during recessions and during depressions aggregate consumption plummets. Time variation in the conditional volatility $\sigma_{t+1}^e$ captures the well known long swings in macroeconomic volatility.\footnote{The importance of the link between macroeconomic volatility and asset prices have been studied by e.g. Lettau et al. (2008). Time variations in consumption volatility is also a key ingredient in Bansal and Yaron (2004).}

Following a large literature in finance and economics, uncertainty in the model is taken to be uncertainty about the underlying state of the economy. Investors do not directly observe the state $S_{t+1}$, nor its manifestations in the trend growth rates or conditional volatilities. They only know that the economy can be in any of $n$ possible states. The state $S_{t+1}$ itself follows a hidden Markov chain characterized by the matrix of state transition probabilities $\Lambda$. The element $i, j$ of the matrix $\Lambda$ gives the conditional probability that the economy will in state $j$ at $t + 1$ if it was in state $i$ at time $t$:

$$\Lambda_{i,j} = \Pr(S_{t+1} = j \mid S_t = i)$$  \hspace{1cm} (2.3)$$

Even though the investors do not observe the state itself, we assume they observe a noisy signal on the state of the economy which is given by:

$$y_{t+1} = \mu_{t+1}^y + \sigma_{t+1}^y \epsilon_{t+1}^y.$$  

The external signal is not a crucial element of the model but it adds some realism and turns out to be helpful when estimating the model parameters. When implementing the model below, we will take the log growth rate of real per capita GDP to be the empirical counterpart of this signal.
Based on the history of consumption growth and external signals, the representative investor arrives at his reference beliefs about the current state of the world. The reference beliefs at time $t$ is a row vector labeled $\pi_t$ where the $j$th element is the conditional probability $\Pr(S_t = j \mid I_t)$ and $I_t$ is the information set of the representative investor at time $t$. Using the Wonham (1964) filter, $\pi_t$ can be found recursively.\(^2\)

Following Longstaff and Piazzesi (2004) we model directly the earnings fraction $F_t$ which we assume follows the generalized AR(1) process:\(^3\)

$$F_{t+1} = \mu^f_{t+1} + \phi F_t + \sigma^f \epsilon^f_{t+1} \tag{2.4}$$

Where the persistence parameter $\phi$ is positive but strictly lower than 1 and $\epsilon_{t+1}$ is a standard normally distributed shock. By allowing the intercept term $\mu^f_{t+1}$ to depend on the state of the economy $S_{t+1}$ the model captures the high cyclicality of corporate earnings relative to aggregate consumption. During booms, the fraction of consumption expenditures that is financed by financial income tends to increase while in recessions and depressions this fraction decreases. In addition to the direct link between the trend consumption growth rate $\mu^c_{t+1}$ and the intercept $\alpha_{t+1}$, the corporate fraction is linked to the consumption process by a positive correlation of the two innovation terms $\epsilon^f_{t+1}$ and $\epsilon^c_{t+1}$ denoted by $\rho$. Consumption growth above the trend tends to be associated with even higher earnings growth, which increases the corporate fraction.\(^4\)

We populate the economy with a representative investor who would rank deterministic consumption plans by

$$(1 - \beta) \sum_{j=0}^{\infty} \beta^j \log C_{t+j}, \tag{2.5}$$

where $\beta$ denotes the investors subjective time-discount factor.

---

\(^2\)Let $\eta_{t+1}$ denote the row vector of joint likelihoods of the realized consumption growth at time $t + 1$ and the observed external signal at time $t + 1$ in each of the $n$ states. By Bayes’ law it follows that

$$\pi_{t+1} = \eta_{t+1} \odot (\pi_t A) \left(\eta_{t+1} \odot (\pi_t A)\right)^{-1},$$

where $\odot$ denotes element by element multiplication. The product $\pi_t A$ gives the vector of state probabilities for $t + 1$ conditional on only the information up to time $t$. The filtering equation gives the updated state probabilities by taking the product of these conditional state probabilities with the joint likelihood of the observations at $t + 1$ and scaling them with the total likelihood of the observations.

\(^3\)The separation of the consumption process into several parts is common with an increasing literature in finance, e.g. Menzly et al. (2004), Santos and Veronesi (2006), Cochrane et al. (2008), Martin (2011), and many others.

\(^4\)[RELATE TO MEHRA/PRESCOTT, DANTHINE/DONALDSON, ETC.]
We make the investor averse to ambiguity about the state of the economy by having him solve a max-min problem where he twists the reference beliefs towards bad states of the world. Following Hansen and Sargent (2001), we formulate ambiguity aversion using multiplier preferences to induce him to select worst case beliefs that are close to his reference beliefs as measured by the divergence of the implied distributions of future endowment growth rates.\footnote{For an axiomatic foundation of these preferences, see Strzalecki (2011).} We denote by $\Pi_t$ the density of future consumption growth rates implied by the reference beliefs and by $\hat{\Pi}_t$ the density of future consumption growth rates implied by some other vector of state probabilities $\hat{\pi}_t$ and the relative entropy of $\hat{\Pi}_t$ with respect to $\Pi_t$ by $\mathcal{R}(\hat{\Pi}_t, \Pi_t)$\footnote{Formally, we think of a probability space $\Omega$ with elements $\omega$. Denote by $g$ the infinite random sequence $g = (g_{t+1}, g_{t+2}, \ldots)$ with a time $t + j$ component $g_{t+j}(\omega)$ that is a measurable function of $\omega$. $\Pi_t(\omega)$ is the distribution that is induced by the reference beliefs $\pi_t$. $\hat{\Pi}_t(\omega)$ is the distribution that is induced by the worst case beliefs $\hat{\pi}_t$. The relative entropy of the density $\hat{\Pi}_t(\omega)$ with respect to the density $\Pi_t(\omega)$ is $\mathcal{R}(\Pi_t, \hat{\Pi}_t) = \int \hat{\Pi}_t(\omega) \log (\hat{\Pi}_t(\omega)/\Pi_t(\omega)) d\omega$}.\footnote{Computational methods for finding the relative entropy will be discussed in the next section.} Computational methods for finding the relative entropy will be discussed in the next section.

\[
\max_{C_t, w_t} \min_{\pi_t} \left\{ E \left[ (1 - \beta) \sum_{j=0}^{\infty} \beta^j \log C_{t+j} | \hat{\pi}_t \right] + \theta \mathcal{R}(\hat{\Pi}_t, \Pi_t) \right\} \tag{2.7a}
\]
subject to the budget constraint

\[
W_{t+1} = (w' R_{t+1}) W_t + I_t - C_t \tag{2.7b}
\]
where $w$ denotes a vector of portfolio weights for the tradable assets in the economy and $\hat{\pi}_t$ denotes a vector of portfolio weights for each tradable asset and $R_{t+1}$ denotes the returns to each of the tradable assets. An equilibrium is a set of processes for consumption and asset prices such that the investors problem is solved and markets clear. I.e. $C_t = D_t + I_t$ at all times and the investor always allocates all his wealth to the share of the dividend process.

### 2.2.1 Worst case beliefs

We solve the representative investor’s problem in (2.7) numerically by imposing the equilibrium conditions that the optimal consumption of the representative investor should always equal the total income of the investor. In the appendix, we show that expected
The present value of future consumption under the worst case beliefs \( \hat{\pi}_t \) is given by:

\[
E \left[ (1 - \beta) \sum_{j=0}^{\infty} \beta^j \log C_{t+j} \mid \hat{\pi}_t \right] = \log C_t + \hat{\pi}_t v
\]

(2.8a)

where \( v \) is the vector

\[
v = (I - \beta \Lambda)^{-1} \beta \Lambda \hat{\mu}^c
\]

(2.8b)

We compute the relative entropy of the distribution of future growth rates under the worst case beliefs relative to that under the reference beliefs by decomposing the total relative entropy into contributions to relative entropy for each period in the future. For given initial worst case beliefs \( \hat{\pi}_t \) and a sequence of growth rates \( \{g_{t+1}, g_{t+2}, \ldots, \} \), we denote the conditional distribution of states at \( t+k \) by \( \hat{\pi}_t^{t+k} \). The corresponding quantity for the reference beliefs \( \pi_t \) and the same sequence of future growth rates is denoted \( \pi_t^{t+k} \)

\[
\mathcal{R}(\hat{\Pi}_t, \Pi_t) = E \left[ \sum_{j=1}^{\infty} \log m_{t+k} \mid \hat{\pi}_t \right],
\]

(2.9)

where \( \log m_{t+k} \) is the log of the ratio of likelihoods of observing the realized growth rate at \( t+k \) conditional on \( \hat{\pi}_t^{t+k-1} \) and \( \pi_t^{t+k-1} \):

\[
\log m_{t+k} = \log f(g_{t+k} \mid \hat{\pi}_t^{t+k-1}) - \log(f(g_{t+k} \mid \pi_t^{t+k-1})
\]

(2.10)

We approximate the expectation in equation (2.9) in the following manner: For each possible state of the economy, we generate 1 Million sequences of future states and growth rates, each of length 400 (100 years). At each point in time \( t \), for the candidate vector of worst case beliefs we compute the filtered state probabilities along each sample path both initialized at the reference beliefs and the worst case beliefs and use these to compute the log likelihood ratios of the observed sequences. The expected relative entropy is found by weighing the expected relative entropy along each sample path with the probability of the initial state of that path according to candidate worst case beliefs. The worst case beliefs are found numerically by searching for the worst case beliefs that solves the investors problem (2.7) using equation (2.8) to compute expected utility and the relative entropy using the equation above.

### 2.2.2 Asset prices

In equilibrium the price of equity satisfies the Euler equation

\[
P_t = E \left[ \sum_{j=1}^{\infty} \beta^j \frac{C_t}{C_{t+j}} D_{t+j} \mid \hat{\pi}_t \right]
\]

(2.11)
Dividing by \( C_t \) and using the definition \( F_{t+j} = D_{t+j}/C_{t+j} \), we can conveniently express the ratio of the price of the aggregate stock market to the consumption level as:

\[
\frac{P_t}{C_t} = E \left[ \sum_{j=1}^{\infty} \beta^j F_{t+j} \mid \hat{\pi}_t \right] \tag{2.12}
\]

In the appendix, we use the linearity of the process for the corporate fraction to show that the price-consumption ratio is given by:

\[
\frac{P_t}{C_t} = \hat{\pi}_t w + \frac{\beta \phi}{1 - \beta \phi} F_t \tag{2.13a}
\]

where the vector \( w \) is given by

\[
w = \frac{1}{1 - \beta \phi} (I - \beta \Lambda)^{-1} \beta \Lambda \hat{\mu} \tag{2.13b}
\]

Dividing through by the corporate fraction \( F_t = D_t/C_t \) we find that the price-dividend ratio is given by

\[
\frac{P_t}{D_t} = \frac{\beta \phi}{1 - \beta \phi} + \hat{\pi}_t w \left( \frac{1}{F_t} \right) \tag{2.14}
\]

The term structure of interest rates is most easily derived from the prices of discount bonds. The equilibrium price of a discount bond maturing in \( n \) periods satisfies the Euler equation:

\[
B^{(n)}_t = E \left[ \beta^n \frac{C_t}{C_{t+n}} \mid \hat{\pi}_t \right] \tag{2.15}
\]

In the appendix, we show that we can express the price of the discount bond which matures at time \( t + n \) as

\[
B^{(n)}_t = \hat{\pi}_t B^{(n)}
\]

with

\[
B^{(n)} = (\Lambda \text{ diag}(\bar{B}))^n \iota
\]

where \( \iota \) is a vector of ones and the vector \( \bar{B} \) is a vector of expectations of the one period discount factor conditional on the state of the economy at \( t + 1 \). The element \( j \) of the vector is given by

\[
\bar{B}_j = E \left[ \beta \frac{C_t}{C_{t+1}} \mid S_{t+1} = j \right] = \beta \exp \left( - \hat{\mu}_j + \frac{1}{2} (\sigma_j^c)^2 \right)
\]
Figure 2.1: State dependent entropy costs
This figure illustrates the higher entropy cost associated with shifting probability to the depression state during low volatility regimes. The solid line in the upper left panel gives the likelihood of observing a particular consumption growth if agents assign almost 100% probability to state 1 under their reference beliefs. The dashed line gives the corresponding likelihood under the worst case beliefs that there is almost 100% probability that the economy is in the depression state. The relative entropy of the worst case beliefs with respect to the reference measure is given by the integral of the curve in the lower left panel. The panels to the right gives the corresponding results for reference beliefs that attach almost 100% probability to the high growth-low volatility regime.
2.2.3 Other shades of robust beliefs

Relation to multiplier preferences

The worst case beliefs that solves the minimization problem in equation (2.7) differ in important ways from the those from the multiplier preferences advocated by Anderson et al. (2003). An investor with multiplier preferences would measure the relative entropy of the worst case beliefs with respect to the reference beliefs by the quantity:

$$\sum_{i=1}^{5} \hat{\pi}_i \log \frac{\hat{\pi}_i}{\pi_i}$$

This is a measure that depends only on how the two probability vectors $\pi$ and $\hat{\pi}$ differ from each other. The predicted relative entropy measure we use depend on how far apart typical future sample paths predicted by $\pi$ and $\hat{\pi}$ are from each other. Arguably a measure of more direct concern to investors.

In terms of pricing, the advantage of using predicted relative entropy is that it provides a link between the perceived volatility state and how much the agent shifts his worst case distribution towards the depression states. The typical sample paths seen during spells of the depression state are much farther away from those observed in normal business cycle states in the low volatility regime than in the high volatility regime.

Figure 2.1 illustrates this for the extreme case where the investor is almost sure that the economy is in a boom and which of the volatility regimes it is in, but contemplates choosing worst case beliefs that attach almost probability 1 to the depression state. To simplify the graphical exposition, we only consider the one period ahead predicted relative entropy. The relative entropy of the worst case beliefs is the divergence of the distribution for consumption growth under the worst case beliefs (dashed green lines) from the distribution of consumption growth under the reference beliefs (solid blue lines). Visually the divergence is much higher for the low volatility state. This is confirmed by the computation of relative entropy. The relative entropies in the two cases are given by the integral under the curves in the lower two panels. In terms of predicted relative entropy, this shift is more than 4 times as big in the low volatility state than in the high volatility state. In terms of the measure used for divergence with multiplier preference the shift would be the same.
2.3 Model estimation

2.3.1 State specification

In the calibration below, we’ll consider a total of 5 possible states of the economy which determines \( \mu_{t+1} \) and the trend and volatility of consumption growth. One of the states is a depression state where consumption growth is on average very low and volatile. The four other states are combinations of mean consumption growth rates from boom and recessions with two volatility regimes. In the normal business cycle states, volatility can be either high or low. The low volatility state captures the decades of low volatility during the great moderation from the mid 1980s to the recent financial crises.

2.3.2 Data

For the model estimation we use the following time-series from the National Income Product Accounts (NIPA), all downloaded from the website of the Bureau of Economic Analysis BEA.\(^7\)

- **Consumption**: Quarterly real per capita consumption of services and non-durables. (Table 7.1, line 13.)
- **GDP**: Quarterly real per capita gross domestic product. (Table 7.1, line 10).
- **Earnings**: Profits after tax with inventory valuation and capital consumption adjustments. (Table 1.10, line 19).

For all series, we use the period 1952:Q1–2010:Q3 for the model estimation. For the calibration of the depression state we use the available annual data from 1929-1932. To translate the nominal earnings series into a real, per-capita series, we deflate it with the ratio of nominal to real consumption expenditures and divide the result with the population series from NIPA table 7.1 (line 18). All series are seasonally adjusted.

Using the quarterly time series from the post world-war period, we estimate the means, transition probabilities and covariances for the 4 regular states of the economy by Gibbs sampling. The underlying assumption is that the economy was never in the depression state during that period. The MCMC algorithm treats both the growth state and the recession state as hidden.

\(^7\)http://www.bea.gov/
Lacking quarterly observations from the depression state, we set the parameter values for this state using annual data from the series above. The NBER dates the beginning of the great depression to August 1929 and its end to March 1933. For consumption and GDP we set the mean quarterly log growth rate in the depression state to 1/4th of its annual average value over the period 1929–1932 and the quarterly standard deviation to 1/2 its annual value over the same interval. Corporate post-tax earnings actually turn negative in 1932, so we only use data up to 1931 to compute the corresponding moments for earnings. Lacking quarterly observations of transitions to the depression state, we set this probability to 0.25 % for all regular business cycle states, so that such transitions take place about every 100 years. All the estimated parameter values are found in table 2.1.

Figure 2.2 shows the filtered beliefs of the business and volatility states implied by our model at the estimated parameter values. The top panel of figure 2.2 shows how investors in our economy slowly attribute more weights to the low volatility state from 1985 onwards. The lower panel of the figure gives corresponding filtered growth state probabilities. In both panels, gray bars mark the periods assigned to recessions by the NBER. The representative investor in our economy uses these filtered beliefs as his reference beliefs. As can be seen from the lower panel, the filtered beliefs pick up well the NBER recessions, even though these dates are picked by the NBER with the benefit of hindsight.

2.3.3 Preference parameters

To complete the model we need to assign preference parameters to the representative investor. Economic theory does not currently have much to say about either the time discount factor $\beta$ or how much an ambiguity averse investor should penalize the relative entropy of the probability measure over future endowment growth rates induced by his worst case beliefs with respect to those induced by the filtered beliefs. We tie our hands by picking the combination of $\beta$ and $\theta$ that match estimated average real 3M and 5Y spot rates. (See section 2.4.1 below.) Because changing the time discount factor $\beta$ introduces a level shift of the term structure while changing $\theta$ increases the slope of the average term structure, the parameter combination we choose is unique. The model closely matches the average values for the 3M and 5Y real rates for $\beta = 0.9922$ and $\theta = 0.0598$, so we use

---

8 NBER does not distinguish between recessions and depressions as we do.
Table 2.1: Estimated process parameters

<table>
<thead>
<tr>
<th>Consumption &amp; GDP</th>
<th>Means</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Boom (States 1 &amp; 3)</td>
<td>Recession (States 2 &amp; 4)</td>
<td>Depression (State 5)</td>
<td></td>
</tr>
<tr>
<td>$\bar{\mu}_c$</td>
<td>0.54 %</td>
<td>-0.25 %</td>
<td>-1.72 %</td>
<td></td>
</tr>
<tr>
<td>$\bar{\mu}_y$</td>
<td>0.58 %</td>
<td>-0.54 %</td>
<td>-1.72 %</td>
<td></td>
</tr>
</tbody>
</table>

| Volatilities |  |  |  |
| High (States 1 & 2) | Low (States 3 & 4) | Depression (State 5) |
| $\bar{\sigma}_c$ | 0.49 % | 0.29 % | 1.52 % |
| $\bar{\sigma}_y$ | 1.05 % | 0.46 % | 2.00 % |
| $\rho_{c,y}$ | 0.4415 | 0.4415 | 0.4415 |

| E/C ratio |  |  |  |
| Boom (States 1 & 3) | Recession (States 2 & 4) | Depression (State 5) |
| $\bar{\mu}_f$ | 0.38 % | -0.49 % | -1.30 % |
| $\phi$ | 0.9719 | 0.9719 | 0.9719 |

Transition Probabilities

<table>
<thead>
<tr>
<th>$S_t$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.9483</td>
<td>0.0375</td>
<td>0.0113</td>
<td>0.0004</td>
<td>0.0025</td>
</tr>
<tr>
<td>2</td>
<td>0.2355</td>
<td>0.7503</td>
<td>0.0028</td>
<td>0.0089</td>
<td>0.0025</td>
</tr>
<tr>
<td>3</td>
<td>0.0113</td>
<td>0.0004</td>
<td>0.9483</td>
<td>0.0375</td>
<td>0.0025</td>
</tr>
<tr>
<td>4</td>
<td>0.0028</td>
<td>0.0089</td>
<td>0.2355</td>
<td>0.7503</td>
<td>0.0025</td>
</tr>
<tr>
<td>5</td>
<td>0.0667</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.9333</td>
</tr>
</tbody>
</table>
Figure 2.2: State probabilities
Dashed lines show filtered state probabilities. Solid lines show smoothed state probabilities.
this as our benchmark calibration below.

**Implied detection error probabilities**

To check whether the value for $\theta$ that is needed to explain the term premium gives sensible beliefs we follow Anderson et al. (2003), Hansen and Sargent (2007b), and others and compute the average detection error probabilities implied by the parameter choice.

Assume that the hidden state at time $t$ is drawn from either the reference beliefs $\pi_t$ (model A) or the worst case beliefs $\hat{\pi}_t$ (model B). Let the investor assign equal probability to both models at time $t$. A deception error is said to occur when, after observing the whole sequence of growth rates after time $t$, the investor assigns a higher probability to the state being drawn from the other distribution than the one it was actually drawn from.

Let $p_A$ be the probability that, when the state is drawn from the distribution $\pi_t$, the investor ex-post ends up assigning a higher probability to the state being drawn $\hat{\pi}_t$. By Bayes law,

$$p_A = E\left[ f(\{g_{t+j}\}_{j=1}^{\infty}|\hat{\pi}_t) > f(\{g_{t+j}\}_{j=1}^{\infty}|\pi_t) \mid \pi_t \right]$$

Similarly, we let $p_B$ be the probability that when the state is drawn from $\pi_t$, the investor ends up assigning a higher probability to it being drawn from $\hat{\pi}_t$. Both $p_A$ and $p_B$ depend on how far $\pi_t$ and $\hat{\pi}_t$ are from each other. If $\theta$ is close to 0, $\pi_t$ and $\hat{\pi}_t$ will on average be very different and it will be relatively easy to distinguish from which distribution $S_t$ is drawn. This makes the detection error probability small. As $\theta$ increases $\pi_t$ and $\hat{\pi}_t$ converge and $p_A$ and $p_B$ both tend to 0.5 since it becomes increasingly hard to infer from which distribution $S_t$ is drawn. Because the initial probability attached to models A and B is the same, the total detection error probability is given by the average of $p_A$ and $p_B$

$$p(\theta) = \frac{1}{2}(p_A + p_B)$$

The average detection error probability associated with the choice of $\theta$ that matches the term-structure data in the sample is 0.3232. This is well within the range considered as plausible by in the literature. (Based on introspection a level of 0.2 is typically used to calibrate models, Barillas et al. (2009) argue that even lower values should be considered.)
2.4 Model predictions

2.4.1 Term structure

Table 2.2 compares empirical moments of interest rates with those implied by rational-expectation and robustness, respectively. The reported real rates are the nominal rates from FRED net of expected PCE inflation. Expected PCE inflation rates are those forecasted by an ARMA(1,1) for log changes in the PCE deflator.

A robust model outperforms its rational counterpart in matching the term structure of interest rate. The short term interest rates predicted by log utility and rational expectation are too high. Second, there is no apparent average term premium. The first failure is a reflection of the risk-free rate puzzle (Weil, 1989). Even for the modest time coefficient of relative risk-aversion of 1 implied by the log utility model, the average expected growth rate of consumption in the sample contributes 1.8% to the quarterly risk-free rate. The model could replicate the empirical rate only for $\beta$ larger than 1, but continuation values would be infinite at such parameter values. The second failure of rational expectation, is that the model cannot generate a positive real term premium. In fact, Backus et al. (1989) shows that if investors have power utility functions, a positive average term premium only obtains if consumption growth rates are negatively correlated. In our model economy with growth regimes, they will always be positively correlated. Piazzesi and Schneider (2006) show the same effect for Epstein-Zin preferences.

Robustness breaks the risk-free rate puzzle by endowing the representative investor with worst case beliefs under which expected short term growth rates are typically negative. In models of robustness without learning, this is true because investors price assets with a worst case model with lower average consumption growth than that of their benchmark model. (See e.g. Cagetti et al., 2002). In this model economy, it is true because the investors twist the inferred state probabilities towards states with low consumption growth and thus expect lower short term consumption growth. Also, the model gives average term structures that are upward sloping and thus predicts a term premium, closely comparable with the data. The positive term premium comes about because under the worst case probability distribution expected short term consumption growth is typically lower than expected long run consumption growth. (See Brevik, 2009.) The expected long run consumption growth is asymptotically independent with the worst-case belief, since the hidden Markov chain of the state of the economy eventually converges to the ergodic distribution.
Table 2.2: Model Implied Moments
This table compares model implied moments with their empirical counterparts for the total sample and subsamples before and after the onset of the great moderation.

<table>
<thead>
<tr>
<th>Total sample</th>
<th>3M rate</th>
<th>5Y rate</th>
<th>Equity premium</th>
<th>Sharpe ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empirical</td>
<td>1.61</td>
<td>2.61</td>
<td>6.66</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>(2.10)</td>
<td>(2.20)</td>
<td>(16.97)</td>
<td>—</td>
</tr>
<tr>
<td>Robustness</td>
<td>1.58</td>
<td>2.72</td>
<td>6.12</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>(1.00)</td>
<td>(0.49)</td>
<td>(17.89)</td>
<td>—</td>
</tr>
<tr>
<td>RE ($\theta = \infty$)</td>
<td>4.83</td>
<td>4.65</td>
<td>1.19</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>(0.61)</td>
<td>(0.13)</td>
<td>(6.35)</td>
<td>—</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$E[R_e - R_f]$</td>
<td>$\sigma(R_e)$</td>
<td>SR</td>
<td>$E[R_e - R_f]$</td>
</tr>
<tr>
<td>Empirical</td>
<td>6.84</td>
<td>16.88</td>
<td>0.41</td>
<td>6.27</td>
</tr>
<tr>
<td>Robust</td>
<td>6.44</td>
<td>17.31</td>
<td>0.37</td>
<td>5.41</td>
</tr>
<tr>
<td>RE ($\theta = \infty$)</td>
<td>0.99</td>
<td>5.86</td>
<td>0.17</td>
<td>1.62</td>
</tr>
</tbody>
</table>

2.4.2 Asset price moments
Table 2.2 compares key empirical asset pricing moments with their model counterparts. Our robust model roughly matches the first moment of stock return with empirical estimates, and thus the equity risk premium. While log utility only generates 1.19% excess return, robustness generates much higher excess return by pushing down prices.

We also divide the sample into two pieces 1953-1992 and 1992-2010, where the second period corresponds to a period after the investors learn about the great moderation. The lower panel of table 2.2 describes the results of subsample comparisons. The higher valuation ratios during the great moderation are matched with lower excess returns to
equity, both empirically and in the model economy.

2.4.3 Equity valuation

Figure 2.3 shows the time series of prices predicted by the model when fed with the time-series of earnings, consumption, and GDP from the sample period. In the upper panel, we plot the time series of the model price together with the (inflation adjusted) value of the S&P 500 from the same period. The two price series are not directly comparable, since the model price is the price of the aggregate corporate sector, which also includes privately held companies, while the S&P 500 is constructed to be representative of the traded part of the corporate sector. Still, comparing the two series gives a notion of the fit of the model.

One salient feature that can be taken from the plot is that the two series move closely together most of the time, but that the model implied price series is more volatile. In particular, it drops more during recession periods. Partly this is because the probability that the robust investor assigns to the depression state is very sensitive to negative shocks. But partly it also reflects that any earnings shock is taken to be permanent by the model investors, but real world investors might be able to attribute some of the earnings shocks to transitory events.

Looking at individual periods, the two series diverge in other interesting ways. Model prices rebound sharply after the recession in the mid seventies, while the empirical prices stay subdued. The rebound after the monetary contraction during the early eighties is also much faster in the model implied series than in the empirical one. But the most interesting decoupling of the two series is during the dot-com bubble of the late 90s. Aggregate corporate earnings peak in the 3rd quarter of 1997 and declines to a low in the 4th quarter of 2000. This decline in earnings is reflected in the model-implied price series, while the empirical series keeps soaring. Since the model investors only observes the macroeconomic time series, there is no role for earnings expectations to be driven directly by the technological innovations as they seemed to be during the dot-com bubble. We see something similar in how the model implied price series leads the empirical one in the run-up to the financial crisis. Corporate earnings peak in the 3rd quarter of 2006, the model implied series trend downward with the corporate earnings series, while the real value of the S&P 500 keeps rising for another year. The realized growth rates of the three series during the financial crises are so low that investors assign close to 100 % probability
Figure 2.3: Model implied prices
The upper panel plots the model implied log price of equity (the blue line) together with the empirical log value of the inflation adjusted S&P 500 (the green line) over the sample period. Both series are normalized to 0 at the beginning of the sample. The lower panel plots the model implied price-earning ratio.
to being the depression state even before making a robustness correction.

The lower panel of figure 2.3 shows the price-earnings ratio implied by the model. During NBER recessions there are sharp drops in the ratio which is partly due to the lower expected earnings growth in the recession states, but mostly due to the increased fear of the robust investors that the economy has actually entered a depression state. Apart from the sharp drops in the valuation ratio during periods of crises, the main feature of the series is how it increases during the great moderation. This effect is almost entirely due to a change in how the robust investors shift their beliefs. When volatility is low, a shift of beliefs towards the depression state is more costly in terms of relative entropy. The sample paths predicted by a given shift towards the depression state diverge more from those under the filtered beliefs, because sample paths spread out less when volatility is low.
### 2.5 Conclusion

This paper has documented how robustness to uncertainty about fundamental states of the economy can help account for low frequency fluctuations in valuation ratios and hence realized returns. We find that the model prices and the S&P 500 move closely together most of the time and that the model predicts a run up in valuation ratios during the great moderation similar to those seen in the US stock market. In the model economy this happens both because uncertainty about the state of the economy is lower in the low volatility state and it becomes more expensive in terms of predicted relative entropy to give more weight to the depression state. With investors assigning less weight to the depression state, equity prices increase.

The model predicts reasonable equity premia when calibrated to match sample means of short term interest rates and bond premia, but the predicted Sharpe ratio is only about half that seen in the data, because model equity returns are excessively volatile.
Appendix: Derivations

Equation (2.8). From the process assumptions, it follows that

\[
E[\log C_{t+k} \mid C_t, \hat{\pi}_t] = \log C_t + E[\sum_{j=1}^{k} g_{t+j} \mid \hat{\pi}_t] = \log C_t + \hat{\pi}_t \Lambda \bar{\mu}_c + \hat{\pi}_t \Lambda^2 \bar{\mu}_c + \cdots + \hat{\pi}_t \Lambda^k \bar{\mu}_c
\]

Matching terms and using \((1 - \beta)(\beta^k + \beta^{k+1} + \cdots)\Lambda^k \bar{\mu}_c = \beta^k \Lambda^k \bar{\mu}_c\) we find that

\[
E \left[ (1 - \beta) \sum_{j=0}^{\infty} \beta^j \log C_{t+j} \mid \hat{\pi}_t \right] = \log C_t + \hat{\pi}_t \beta \Lambda \bar{\mu}_c + \hat{\pi}_t \beta^2 \Lambda^2 \bar{\mu}_c + \cdots = \log C_t + \hat{\pi}_t (I - \beta \Lambda)^{-1} \beta \Lambda \bar{\mu}_c
\]

\[\square\]

Equation (2.13). From the specification of the process for the corporate fraction it follows that:

\[
E[F_{t+k} \mid F_t, \hat{\pi}_t] = \phi^k F_t + \phi^{k-1}(\hat{\pi}_t \Lambda \bar{\mu}_f) + \phi^{k-2}(\hat{\pi}_t \Lambda^2 \bar{\mu}_f) + \cdots + \hat{\pi}_t \Lambda^k \bar{\mu}_f
\]

Substituting in equation (2.12) and matching terms, we find that

\[
P_t C_t = E \left[ \sum_{j=1}^{\infty} \beta^j F_{t+j} \mid \hat{\pi}_t \right] = \frac{\beta \phi}{1 - \beta \phi} F_t + \frac{\beta}{1 - \beta \phi} \hat{\pi}_t \Lambda \bar{\mu}_f + \frac{\beta^2}{1 - \beta \phi} \hat{\pi}_t \Lambda^2 \bar{\mu}_f + \cdots = \frac{\beta \phi}{1 - \beta \phi} F_t + \hat{\pi}_t \frac{1}{1 - \beta \phi} (I - \beta \Lambda)^{-1} \beta \Lambda \bar{\mu}_f
\]

\[\square\]

Equation (2.16). The price of the n-period discount bond solves the Euler equation:

\[
B_{t}^{(n)} = E \left[ \beta^n \frac{C_t}{C_{t+n}} \mid \hat{\pi}_t \right] = E \left[ \left( \beta \frac{C_t}{C_{t+1}} \right) \left( \beta \frac{C_{t+1}}{C_{t+2}} \right) \cdots \left( \beta \frac{C_{t+n-1}}{C_{t+n}} \right) \mid \hat{\pi}_t \right]
\]

Where the terms \(\beta C_{t+j}/C_{t+j+1}\) are one period stochastic discount factors. The expectation of the one period discount factor conditional on ending in state \(j\) is given by

\[
\tilde{B}_j = E \left[ \beta \frac{C_t}{C_{t+1}} \mid S_{t+1} = j \right] = \beta \exp \left( -\bar{\mu}_c + \frac{1}{2} \sigma_c^2 \right)
\]
Let $S_{t}^{t+n}$ denote the sequence of states between time $t$ and $t+n$. The probability of the sequence $\{S_{t} = i, S_{t+1} = j, S_{t+2} = k, \ldots S_{t+n-1} = q, S_{t+n} = r\}$ conditional on $\hat{\pi}_{t}$ is $\pi_{t}(i)\Lambda_{i,j}\Lambda_{j,k} \cdots \Lambda_{q,r}$. By the law of iterated expectations:

$$B_{t}^{(n)} = E \left[ E \left[ \left( \beta \frac{C_{t}}{C_{t+1}} \right) \left( \beta \frac{C_{t+1}}{C_{t+2}} \right) \cdots \left( \beta \frac{C_{t+n-1}}{C_{t+n}} \right) \mid S_{t}^{t+n} \right] \mid \hat{\pi}_{t} \right]$$

$$= \sum_{S_{t}^{t+n}} \pi_{t}(i)\Lambda_{i,j}\Lambda_{j,k} \cdots \Lambda_{q,r}(\bar{B}_{j})(\bar{B}_{k}) \cdots (\bar{B}_{r})$$

$$= \sum_{S_{t}^{t+n}} \pi_{t,i}(\Lambda_{i,j}\bar{B}_{j})(\Lambda_{j,k}\bar{B}_{k}) \cdots (\Lambda_{q,r}\bar{B}_{r})$$

$$= \pi_{t}(\Lambda \operatorname{diag}(\bar{B}))^{n}$$

\[\square\]

**Appendix: Relative Entropy**

Regarding the relative entropy of the distribution of future growth rates under the worst case beliefs relative to that under the reference beliefs, we start out with the following decomposition of the likelihood of observing the sequence of growth rates: Let $g_{t}^{k}$ denote the sequence of consumption growth rates between $t+1$ and $t+k$, from the chain rule of probabilities we know that

$$f(g_{t}^{k} \mid \hat{\pi}_{t}) = f(g_{t+k+1} \mid \hat{\pi}_{t+k-1})f(g_{t+k-1} \mid g_{t+k-2}, \hat{\pi}_{t}) \cdots f(g_{t+1} \mid \hat{\pi}_{t}) \quad (2.18)$$

Let $\hat{\pi}_{t}^{k}$ denote the vector of filtered state probabilities given the initial state probabilities $\hat{\pi}_{t}$ and the sequence of growth rates $g_{t}^{k}$. Since the hidden state follows a Markov chain, these filtered state probabilities are sufficient statistics for the distribution of $g_{t+k+1}$ conditional on the initial state probabilities $\hat{\pi}_{t}$ and the observed sequence $g_{t}^{k}$, which allows us to rewrite the equation above as:

$$f(g_{t}^{k} \mid \hat{\pi}_{t}) = f(g_{t+k+k} \mid \hat{\pi}_{t+k-1})f(g_{t+k-1} \mid \hat{\pi}_{t+k-2}) \cdots f(g_{t+1} \mid \hat{\pi}_{t}) \quad (2.19)$$

Let $M_{t+k}$ denote the likelihood ratio:

$$M_{t+k} = \frac{f(g_{t}^{k} \mid \hat{\pi}_{t})}{f(g_{t}^{k} \mid \pi_{t})} \quad (2.20)$$

and letting $m_{t+k} = (f(g_{t+k} \mid \hat{\pi}_{t+k-1}))/f(g_{t+k} \mid \pi_{t+k-1})$ we see that:

$$M_{t+k} = \prod_{j=1}^{k} m_{t+j} \quad (2.21)$$
This allows us to rewrite the representation of the relative entropy of the distribution of future growth rates under the worst case beliefs relative to that under the reference beliefs as

\[
\mathcal{R}(\hat{\Pi}_t, \Pi_t) = E \left[ \log \prod_{j=1}^{\infty} m_{t+k} \mid \hat{\pi}_t \right] = E \left[ \sum_{j=1}^{\infty} \log m_{t+k} \mid \hat{\pi}_t \right] = 0
\]

(2.22)

Since for any given sequence of future growth rates \(\{g_{t+1}, g_{t+2}, \ldots\}\), the influence of the initial state probabilities \(\hat{\pi}_t\) on \(\hat{\pi}_{t+k}\) diminishes with \(k\), we have that \(\pi^k_t\) tends to \(\hat{\pi}_t^k\) and so

\[
\lim_{k \to \infty} \log m_{t+k} = \log \frac{f(g_{t+1} \mid \hat{\pi}_t^k)}{f(g_{t+1} \mid \hat{\pi}_t)} = 0
\]

This means that the sequence on the right hand side of equation (2.22) converges.
Chapter 3

An Anomaly between Two Volatility Markets

3.1 Introduction

Volatility implied from market-traded asset prices is believed to contain important information of future volatilities, since prices are forward looking. From call and put options on the S&P 500 one can measure the expected variance of the S&P500 return between now and the maturity of the options. On the other hand, from the futures contracts on the Chicago Board Options Exchange (CBOE) volatility index (VIX) one can obtain the market expectation of the S&P500 volatility over the 30 day period following the futures maturity. This paper compares variance/volatility implied from S&P500 options and the VIX futures and discovers recurrent inconsistencies between them.

The methodology that we use for measuring the expected variances is the same as in the construction of the VIX. The VIX was first designed in 1993 by Whaley (1993) with two goals, one to estimate the expected short-term market volatility and the other to provide a volatility benchmark upon which derivatives such as volatility futures and options can be written. The original VIX was based on the implied volatilities of at-the-money S&P100 Index (OEX) option prices. The definition of VIX was revised in 2003 to become based on out-of-money S&P500 Index options to reflect the market’s expectation of the S&P500 volatility over the next 30 day period. The theoretical foundation for the new VIX is that the risk neutral expectation of forward realized variance can be represented as a weighted average of prices of the out-of-money European call and put options on the
same underlying and same maturity date. Hence it is possible to measure the expected variance of the S&P500 up to a certain option maturity date. The expected variances up to two option maturities nearest to 30 days are further weighted to obtain the VIX index. By the market convention, VIX is published in volatility percentages.

In fact, these risk neutral expected variances should be equal to the variance swap rates of variance swap contracts that run from now to the option maturities. A variance swap is an OTC derivatives instrument that allows one to pay a fixed amount, which is called the variance strike or variance swap rate, in exchange to receiving the realized variance from the inception to the maturity of the swap contract. Aït-Sahalia et al. (2012) and Filipovic et al. (2013) model the term structure of the variance swap rates on the S&P500. However, in this study hypothetical variance swaps written on the S&P500 will be synthesized using S&P500 options by the theory that underlies the VIX. The difference is that the actual variance swaps in their studies have fixed terms, while the option-synthesized variance swaps have fixed maturities. The feature of the latter allows the option-synthesized variance swap rates to be compared with the prices of the VIX futures.

An alternative way for an investor to have exposure to market volatility is through VIX futures. On March 26, 2004, the trading of futures on the VIX began on the CBOE Futures Exchange. Clearly the volatility implied from the VIX futures should be closely linked to the variance swap rate implied from the S&P500 options. Still, we will see that the futures prices augment the information set offered by the option-implied variance swap rates in an important way. Intuitively, this is due to the concavity of the square root function, since the value of the VIX futures is the expectation of the square root of the swap rate of a forward variance swap that starts at the maturity of the futures contract.

Volatility trading with variance swap and VIX futures has become popular. In addition to allowing directional bets on future volatilities, and their obvious use as a volatility hedge, these contracts can be attractive for asset allocation because innovations to the level of the VIX are strongly negatively correlated with returns to the S&P 500 index (see, e.g., Briere et al. (2010) and Szado (2009)). Despite their growing popularity, little is known about how prices of VIX futures relate to the option-implied variance swap rates.

In the papers by Neuberger (1994), Neuberger (1990) and Dupire (1993), they discover
the insight that the P&L of holding a contract paying the log of the price of the underlying asset, when delta-hedged, is the difference between the realized variance and the fixed variance in delta hedging. Generalizing their results, Carr and Madan (1998a) provide a review of methods for trading realized volatility which is based on the insight of Breeden and Litzenberger (1978) that the log contract can be replicated with a static position in options. These results are obtained under the assumption that the path of the price of the underlying asset is continuous. Many follow-up researches are dedicated to the pricing and hedging issues of volatility derivatives, either keeping to or relaxing this assumption. Examples are the Heston stochastic volatility model (see, e.g., Brockhaus and Long (2000) and Zhang and Zhu (2006)), affine jump diffusion models (see, e.g., Lin (2007) and Zhu and Lian (2012)), pure jump models (Carr et al. (2005)), and exponential Levy models (see, e.g., Carr et al. (2012) and Itkin and Carr (2010)). On the other hand, there is a new area of research exploring the pricing implication of variance risk premium which is defined as the difference between expected variances under the risk neutral measure and the physical measure. Bollerslev et al. (2009) find that the variance risk premium is time-varying and able to predict aggregate stock market returns at a quarterly horizon, in contrast to long-horizon predictors such as the price-dividend ratio. Drechsler and Yaron (2011) relate the variance risk premium to the equity premium in a long run risk model.

This paper centers on a different concern whether the variance swap rates implied from S&P options are consistent with the volatility implied from VIX futures. As already pointed out in Carr and Wu (2006), the concavity of the square root function means that the price of VIX futures is bounded from above by the forward-starting variance swap rate. In fact, the difference between the forward variance swap rate and the square of the price of VIX futures is the risk-neutral variance of the VIX. Using this idea, I construct time series of the implied variance of the VIX. While measurements of the variance of the VIX should be interesting per se, for example for risk-management and asset pricing, the use I will make of them in this paper is to document inconsistencies between the pricing in the index-options and VIX-futures markets. The time series reveals recurring observations of negative variances, a mathematical impossibility. This does not seem to be due purely to measurement errors, because simple trading strategies designed to exploit these inconsistencies produce very high Sharpe ratios. The analysis suggests an anomaly between the variances implied from options and volatilities implied from VIX futures.

The structure of the rest of the paper is the following: Section 3.2 outlines the the-
oretical foundation for the option-implied variance swap rate. Section 3.3 describes how I measure the variance of the VIX. Section 3.4 shows the design and the results of the trading strategies. Section 3.5 concludes.

3.2 Theoretical Underpinnings

3.2.1 Option-synthesized Variance Swap Rate

By no-arbitrage, the variance swap rate reflected by S&P500 options should be equal to the risk-neutral expectation of the realized variance from now up to a certain option maturity. It is shown in Carr and Madan (1998a) that the variance swap contract can be replicated, up to a higher-order term, using a static position in a continuum of out-of-money European options and a dynamic position in futures trading. Since the risk-neutral expected value of the P&L of the dynamic futures trading is zero, the variance swap rate is equal to the initial cost of the options in the replication. They show that, if the log S&P follows a diffusion process:

$$\log S_t = \mu_t dt + \sigma_t dt$$

the following relation holds

$$ E_t^Q \left[ \frac{1}{T} \int_t^{t+T} \sigma_s^2 ds \right] = \frac{2e^{rT}}{T} \left\{ \int_0^{F_t} \frac{1}{K^2} \text{put}_T(K)dK + \int_{F_t}^{\infty} \frac{1}{K^2} \text{call}_T(K)dK \right\} \quad (3.1) $$

where the symbol $E_t^Q$ indicates that the conditional expectation is to be computed under the risk neutral probability measure given information available at time $t$, $F_t$ is the forward index level, $r$ is the risk free rate to expiration, $K$ is the strike price, put$_T(K)$ and call$_T(K)$ are prices of European options with strike $K$ and remaining time-to-maturity $T$.

If the underlying process contains a discontinuous jump component,

$$\log S_t = \mu_t dt + \sigma_t dt + J_t dN_t$$

where $N_t$ is a Poisson jump process and $J_t$ is the random jump size, relation (3.1) still holds, up to an approximation error of cubic order of jumps in returns.
The option-synthesized variance swap rate $\text{Var}(t, t + T)$ represents as a discrete approximation of relation (3.1):

$$E_t^Q \left[ \frac{1}{T} \int_t^{t+T} \sigma_s^2 ds \right] \approx \frac{2}{T} \sum_i \frac{\Delta K_i}{K_i^2} e^{rT} Q(K_i) - \frac{1}{T} \left[ \frac{F}{K_0} - 1 \right]^2 \approx \text{Var}(t, t + T) \quad (3.2)$$

$Q(K_i)$ is the midpoint of the bid-ask for an option with strike $K_i$. $\Delta K_i$ is defined as half of the difference between the strikes on either side of $K_i$. At the upper and lower edges of the strip of options, $\Delta K_i$ is simply the difference between the end strike price and the strike next to it. $F$ is the forward index level derived from index option prices via put-call parity, and $K_0$ is first strike below $F$. The term $\frac{1}{T} \left[ \frac{F}{K_0} - 1 \right]^2$ represents the adjustment term to convert the in-the-money call with strike $K_0$ into an out-of-money put via the put-call parity.\(^1\)

### 3.2.2 The VIX and VIX Futures

In practice, it is volatility rather than variance that is more often quoted. The CBOE VIX index is designed to reflect the market’s expectation of S&P500 volatility over the next 30 day period. On most trading days there are no options with exactly 30 day time-to-maturity. The CBOE then linearly interpolates the option-synthesized variance swap rates over the nearest two maturities to obtain a 30-day estimate. Because options with time to maturity less than one week are not used, sometimes it is an extrapolation with two option maturities over 30 days, instead of an interpolation.

The VIX is defined as:

$$\text{VIX} = 100 \sqrt{\frac{365}{30} \left[ \frac{T_1 \text{Var}(t, t + T_1) \frac{T_2 - 30/365}{T_2 - T_1} + T_2 \text{Var}(t, t + T_2) \frac{30/365 - T_1}{T_2 - T_1}} \right]} \quad (3.3)$$

where $t + T_1$ and $t + T_2$ are the nearest two option maturity dates, and $\text{Var}(t, t + T)$ is as defined in equation (3.2).

VIX futures have been offered on the CBOE Futures Exchange since 2004, as a tool to trade the market volatility. The underlying value for the futures is based off the VIX

\(^1\)For more details please refer to VIX white paper http://www.cboe.com/micro/vix/vixwhite.pdf
The futures contract is settled against the value of VIX at the maturity. By the risk neutral pricing approach the fair price of VIX futures with time to maturity \( T \) is:

\[
F^T_t = E^Q_t VIX_{t+T}
\]

The expectation above is taken under the risk neutral measure \( Q \).

In light of the definition of VIX, the price of the VIX futures with time to maturity \( T \) reflects the annualized volatility between \([t + T, t + T + \frac{30}{365}]\), which is:

\[
\sqrt{\frac{365}{30}} E^Q_t \left[ \sqrt{E_{t+T} \left( \int_{t+T}^{t+T + \frac{30}{365}} \sigma_s^2 ds \right)} \right] \tag{3.4}
\]

### 3.3 Empirical Measurements

#### 3.3.1 Data and Empirical Features

Based on equation (3.2) I construct the option-implied variance swap rates using closing quotes for the European options on the S&P 500 and risk-free rates via Optionmetrics. The data ranges from March 26th, 2004 through September 30th, 2011. Observations of options are eliminated if the time to maturity is less than one week, to minimize pricing anomalies that might occur close to expiration. Following the description in the CBOE VIX white paper, the selected options are out-of-the-money calls and out-of-the-money puts with non-zero bid prices. Furthermore, the strike price range of options with nonzero bids is determined in such a way that once two options with consecutive strike prices are found to have zero bid prices, no options with further out-of-money strikes are considered for inclusion. The strike range can expand or contract day to day, due to the rise and fall of implied volatilities. To keep a minimal precision, I do not calculate the variance swap rate if the number of options in the valid strike range is less than ten.

A similar exercise has been done in Travis (2012). However, there the option-synthesized variance swap rates are interpolated to obtain a fixed term. To validate the exercise, I also build the VIX using synthesized option-implied variance swap rates and compare it with the published VIX. Figure 3.1 plots their difference. As it shows, the discrepancy is small except for the two outliers which will be deleted in later analysis.
Data of closing prices of VIX futures for the same sample period is downloaded from the CBOE website. I filter out the observations with zero volume. On March 26, 2007, CBOE revised the specification of the VIX futures by dividing the contract by 10 and increasing the multiplier from 100 to 1000. As a result, the traded futures price after March 26, 2007 is reduced by a factor of ten. To keep the consistency of prices in my dataset, I reduce the futures prices before March 26, 2007 by ten. The detailed information of futures prices and volumes is provided in table 3.6.

According to the CBOE, the Exchange may list for trading up to nine near-term serial months and five months on the February quarterly cycle for the VIX futures contract. In fact, the number of futures listed for trading has grown from 7 in 2004 to 12 in 2011. This means that we can construct the VIX futures term structure. On the other hand, the term structure of variance swap rate can be obtained as well, by varying the time to maturity of the options. Time series illustrations of the two term structures are shown in figure 3.2 and 3.3. I report variance swap rates up to six front maturities, and assign zero values if the number of options available for use is less than ten. Values of the variance swap rate are in annualized variance percentage, and values of the VIX futures price are in annualized volatility percentage. The variance swap rates are comparable to the volatility implied from the VIX futures. The average of the variance swap rate implied from the front options is 5.8%, and the average of the annualized volatility implied from the front futures is 21.3%. In addition, the variance swap rate and the VIX futures volatility tend to cycle together, both spiking in the fall of 2008.

Notice the illustrated term structure of variance swap rate is a spot curve, as the variances are for the time periods between the time of observation and the option maturities, while the term structure of the VIX futures is a forward curve as the volatilities are for the 30 day period after the futures maturities. Because of the linear additivity of expected variance over time if the variance is a Markov process, I can also measure the expected variances between two option maturities, a procedure transforming the spot curve of variance swap rate into a forward curve. Though the forward curve of variance swap rates is not directly comparable to the forward curve of volatility, by looking at the directions of their slopes we can have a general sense of whether their predictions of the change of long run variance/volatility are consistent with each other. As a motivation to suggest
the possibility of an existing anomaly between them, I take a first step by visualizing the slope directions of the two forward curves. For the variable $S$, I let

$$S = \begin{cases} 
1 & \text{if both of the two forward curves are upward sloping} \\
0 & \text{if the two forward curves slope in opposite directions} \\
-1 & \text{if both of the two forward curves are downward sloping}
\end{cases}$$

As it is difficult to picture a situation where the market expects volatility to decrease but variance to increase, recurrent occurrence of $S = 0$ as shown in figure 3.4 suggests a potential anomaly between the option-implied variance and futures-implied volatility. For a more rigorous investigation, I will not use the slopes and instead search for a statistic that is information sufficient to tell whether the option-implied variance is consistent with the futures-implied volatility, which turns out to be the risk neutral variance of the VIX at the futures maturity.

### 3.3.2 Matching Maturities of Options and Futures

For knowing the risk neutral variance of the VIX at the futures maturity, we need to know the first and second moments of the VIX under the risk neutral probability. The first moment is directly revealed from the futures price. Neglecting the interpolation error in the construction of the VIX, the second moment of the VIX at the futures maturity date $t + T$ is

$$E_t^Q VIX^2_{t+T} = \frac{1}{T_1 - T} E_t^Q E_{t+T} \int_{t+T}^{t+T_1} \sigma_s^2 ds = \frac{1}{T_1 - T} E_t^Q \int_{t+T}^{t+T_1} \sigma_s^2 ds$$

(3.5)

where $T_1$ is 30 days after $T$. The first equality is by the valuation of VIX and the second equality is by the law of iterated expectations.

Therefore the second moment of the VIX is the expected variance over the 30-day period following the futures maturity date. By the linear additivity of variance, if there exist options maturing at $t + T$ and $t + T_1$, the second moment can be approximated by option-synthesized variance swap rates as:
$$E_t^Q \text{VIX}^2_{t+T} = \frac{1}{T_T - T_t} E_t^Q \int_t^{t+T_T} \sigma_s^2 ds - \frac{1}{T_T - T_t} E_t^Q \int_t^{t+T_T} \sigma_s^2 ds$$  \tag{3.6}$$

$$\frac{T_T \text{Var}(t, T_T) - T \text{Var}(t, T)}{T_T - T}$$ \tag{3.7}$$

can be read as the swap rate for the forward variance swap that starts at $t + T$ and matures at $t + T_T$.

However, this approximation is complicated by the fact that there are no option maturity dates satisfying the above relation. This is due to the two-day gap between the VIX futures settlement date and the S&P500 options settlement date. The S&P500 options are often settled on the third Friday of the expiration month, while the final settlement date for the VIX futures is defined as the Wednesday that is thirty days prior to the third Friday of the calendar month immediately following the month in which the contract expires.

On the other hand, the specification of the settlement dates for the options and futures suggests that 30 days after the VIX futures settlement date is the S&P500 option settlement date. Also considering that there is often a four-week distance between two option settlement dates, I propose to approximate the expected variance over the 30-day period by the expected variance over the four-week period between two option maturities. The assumption is that the variation of the variance over the two-day gap is small. To provide more clarity, figure 3.5 illustrates a situation where the thirty-day horizon $[T, T_T]$, over which the VIX futures that settles at $T$ predicts the volatility, consists of the Wednesday-to-Friday gap between the futures settlement date $T$ and the option settlement date $T_T$, and the twenty-eight days between option maturities $T_T$ and $T_T$. I shall approximate the expected variance over $[T, T_T]$ by the expected variance over $[T_T, T_T]$ after properly scaling the length of time, and hence approximate the second moment of the VIX$_{t+T}$.

So each day I check the possibility of matching the maturities of options and futures under the above guideline. Sometimes there is a five-week distance between two option maturities, in which case I do not consider the matching. A successful match means that one VIX futures maturity and two S&P options maturities can be grouped in the way
For considerations of liquidity, I only consider the contracts with time to maturity more than 14 days. In case there are multiple successful matches on a single day, the one with nearest maturities is selected for the simplicity of my analysis.

3.3.3 Variance of the VIX

For ease of exposition, in the rest of the paper I will continue to consider the situation in figure 3.5 and follow the notations there. For each matching, we have knowledge of \( E_t^Q VIX_{t+T}^2 \) derived from options prices and \( E_t^Q VIX_{t+T} \) from the futures price. Hence the risk neutral variance of \( VIX_{t+T} \) becomes a natural statistics to compare the two moments. As for any random variable, the variance is equal to the difference between the second moment and the square of the first moment, or:

\[
\text{Var}_t VIX_{t+T} = E_t^Q VIX_{t+T}^2 - (E_t^Q VIX_{t+T})^2 \approx \frac{T_1 \text{Var}(t, T_1) - T \text{Var}(t, T)}{T_1 - T} - (F_t^T)^2 \quad (3.8)
\]

\( F_t^T \) is the price of the VIX futures with time to maturity \( T \), \( \frac{T_1 \text{Var}(t, T_1) - T \text{Var}(t, T)}{T_1 - T} \), which approximates \( E_t^Q VIX_{t+T}^2 \) by equation (3.7), is the forward variance swap rate.

As \( E x^2 \geq (E x)^2 \), the price of the VIX futures is bounded from above by the forward variance swap rate, and the risk neutral variance of \( VIX_{t+T} \) should not be negative. Because the forward variance swap rate and the futures price are observable, I am able to construct the time series of the risk neutral variance of the VIX using the option and futures prices. A surprising finding demonstrated in figure 3.6, is that many observations of the variance of the VIX are negative, and these negative observations show up throughout the whole sample. Table 3.1 reports the summary statistics of the measured variance of the VIX. 23.33% observations are negative numbers, suggesting that the futures prices frequently break the upper bound implied from option prices. The sample is also divided into two parts, before and after January 1st, 2008. The number of observations in the two samples are close, while the first subsample contains a higher proportion of negative observations (27.60%) than the second subsample (17.41%). In addition, the average level of the variance in the first subsample is only 0.0012, substantially lower than the average of the second subsample (0.0058). The differences are consistent with the observation that VIX behaves more volatile after the credit crunch.
<table>
<thead>
<tr>
<th>Number of Obs</th>
<th>Proportion of Negative Obs</th>
<th>Mean</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>320 /1433</td>
<td>22.33%</td>
<td>0.0036</td>
<td>0.0086</td>
</tr>
<tr>
<td></td>
<td><strong>Before 01/01/2008</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>191 /692</td>
<td>27.60%</td>
<td>0.0012</td>
<td>0.0029</td>
</tr>
<tr>
<td></td>
<td><strong>After 01/01/2008</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>129 /741</td>
<td>17.41%</td>
<td>0.0058</td>
<td>0.0111</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CBOE VIX</th>
<th>Number of Obs</th>
<th>Sample Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entire Sample</td>
<td>1881</td>
<td>0.0052</td>
</tr>
<tr>
<td>Before 01/01/2008</td>
<td>941</td>
<td>0.0002</td>
</tr>
<tr>
<td>After 01/01/2008</td>
<td>940</td>
<td>0.0083</td>
</tr>
</tbody>
</table>

Table 3.1: Variance of the VIX
Combining the S&P options and the VIX futures, whenever their maturities can be matched, results in the measurement of the risk neutral variance of the VIX at the futures maturity date. For example, for the entire sample, there are 1433 days on which I can measure the variance of the VIX, and 320 of them (22.33%) are negative measurements. The table also reports the mean and standard deviation of the measured variance of the VIX. Especially the mean should be comparable to the realized sample variance of the CBOE VIX.

To gauge the effectiveness of those measurements, I compare the mean of the measured risk neutral variance of the VIX with the sample variance of the CBOE VIX index. The mean of my measurements reflects the true mean of the VIX under the risk neutral measure by the law of large number, if the volatility process is ergodic. On the other hand, the sample variance of the VIX is the realized variance of the VIX for the sample path that we observe. Without any prior, the two should be comparable. Table 3.1 also shows that for the entire sample the sample variance of the VIX is 0.0052, and that the mean of the measured variance is 0.0032. For the first subsample, the sample variance of the VIX is at a low level of 0.0002, consistent with the low mean of the measured variance. In the
second subsample, the sample variance rises to 0.0083. In general, the sample variance of the VIX shows the same pattern as the measured variance, and the two are of comparable magnitude.

### 3.4 Trading On Negative Observations of the Variance of the VIX

#### 3.4.1 Long-Short Strategy

To explore the economic significance of the finding of the negative observations of the variance of the VIX, in this section I design a trading strategy and show that the inconsistency documented is economically significant.

As already noted, negative observations of variance of the volatility indicate a pricing anomaly between the price of VIX futures and the forward variance swap rate.

If the observation of $\text{Var}_t \text{VIX}_{t+T}$ is negative, then

$$\text{Var}(t, T, T_1) = \frac{T_1 \text{Var}(t, T_1) - T \text{Var}(t, T)}{T_1 - T} < (F_t^T)^2$$  \hspace{1cm} (3.9)

A natural way to exploit this anomaly is to long the forward variance swap and short the VIX futures. The weights allocated to the swap and futures is determined in a way so that the portfolio is vega neutral to a certain type of volatility shock. Instead of setting up a parametric volatility model which is possibly misspecified, I consider a simple type of volatility shock that raises all the future volatility from $\sigma_t$ to $(1 + \alpha\%)\sigma_t$. Correspondingly, the change of the swap rate would be $(\alpha\%)^2 \text{Var}(t, T, T_1)$ and the change of the VIX futures price would be $\alpha\% F_t^T$. As a result, a portfolio consisting of a long position in $F_t^T / \text{Var}(t, T, T_1)$ units of variance swap and a short position in 2 units of VIX futures is vega neutral in the sense that the rate of change of the portfolio value with respect to $\alpha$ is zero, as:

$$\frac{\partial}{\partial \alpha} \left[ F_t^T / \text{Var}(t, T, T_1)(\alpha\%)^2 \text{Var}(t, T, T_1) - 2\alpha\% F_t^T \right]_{\alpha=0} = 0$$

This allocation rule resembles the theoretical arbitrage strategy for this anomaly which sells 2 units of the VIX futures, buys $1 / F_t^T$ units of the variance swap, and holds the positions till the futures maturity date. At the VIX futures maturity $t + T$, the futures
price converges to $VIX_{t+T}$, and the forward variance swap rate converges to $VIX^2_{t+T}$. This implies the P&L of the theoretical arbitrage strategy at $t + T$ becomes

$$P&L = \frac{1}{F^T_t}(VIX^2_{t+T} - \text{Var}(t, T, T_1)) - 2(VIX_{t+T} - F^T_t)$$

$$= \frac{1}{F^T_t}[VIX^2_{t+T} - 2VIX_{t+T}F^T_t + (F^T_t)^2 + (F^T_t)^2 - \text{Var}(t, T, T_1)]$$

$$= \frac{1}{F^T_t}[(VIX_{t+T} - F^T_t)^2 + (F^T_t)^2 - \text{Var}(t, T, T_1)] > 0$$

So for shorting each 2 units of VIX futures, the arbitrage strategy takes a long position in $1/F^T_t$ units of the variance swap, and my long-short strategy takes a long position in $F^T_t/\text{Var}(t, T, T_1)$ units of variance swap. Both are inversely proportional to the level of the volatility. I do not consider the theoretical arbitrage strategy because in the analysis above the variance swap contract is only hypothetical. Instead of tracing the hypothetical variance swap rate, my trading strategy will use option prices to approximate the variance swap rate. This approximation has the merit that it allows to study an actual trading strategy, since S&P options are tradable. Also it is because that the option maturity date is two days after the futures maturity date, and option prices may show irregular behavior as options approach to their maturity dates.

Instead I consider a self-financing trading strategy which is rebalanced on a daily basis. Each day I measure the risk neutral variance of the VIX if matching of S&P options and VIX futures is possible. Upon a negative measurement of the variance, the portfolio wealth is invested in the option portfolio which mimics the forward variance swap. Correspondingly, the strategy also enters into a short position in VIX futures, with the weight as described earlier to make the portfolio vega neutral. The assumption here is that it is costless to enter into the futures position. Denote the wealth available at time $t$ as $W_t$. The long position in the option-synthesized variance swap is $W_t/\text{Var}(t, T, T_1)$, and the short position in the VIX futures is $2W_t/F^T_t$. As the strategy is self-financing, the positions are closed and the portfolio value is marked to the market at the end of the second day. If again a negative measurement of the variance of the VIX is found, new positions in the variance swap and the VIX futures will be taken.

I run this strategy from March 26th, 2004 to September 30th, 2011. Plotted in figure 3.7, the final return of the portfolio accumulates to a surprising high of 88.12. To provide a
measurement of the return-risk tradeoff, I calculate the Sharpe ratio for the realized daily returns by dividing the difference between the averages of the returns and the Fama-French one month bill rates by the standard deviation of the returns. The relevant statistics can be found in the first row of table 3.3. The standard deviation of the daily returns is 0.062, and the annualized Sharpe ratio is 1.50.\footnote{The annualized Sharpe ratio is obtained by multiplying the daily Sharpe ratio by $\sqrt{260}$.} 64.38% of the returns on trading negative observations of variance of volatility are positive, and there are a few large positive realizations that can be clearly spotted in figure 3.7. Together they make a gigantic accumulated return.

It is important to mention that the transaction cost is not included in the calculation of the returns of this long-short strategy. In fact, most of the time the transaction cost of the options portfolio that mimics the variance swap dominates the overnight change of its value. Consequently, the positive profits would disappear if the bid-ask spreads of options were subtracted from the returns. Despite this, arguably the positive returns of long-short strategy still suggest the inconsistency between option-implied variance swap rates and VIX futures prices, considering that the transaction cost may not be a friction with first order effects if there were actual variance swap replacing the options portfolio. More importantly, in the next section I show that a strategy that only involves trading VIX futures can make a profit as well.

### 3.4.2 Short Strategy

In this section I show returns of a strategy that only shorts VIX futures when the observed variance of the VIX is negative. This helps address the concern of the transaction cost of the options portfolio that is overlooked in the long-short strategy. Compared to the long-short strategy, the results of the short strategy are demonstrated using the close prices of VIX futures, and therefore should be cleaner as the futures prices are more easily accessible at the market.

The strategy is implemented for the same period as the long-short strategy. Upon a negative measurement of the variance of the VIX, it enters into a short position of $W_t/F_t^T$. 

\[ \text{Annualized Sharpe ratio} = \frac{\text{Daily Sharpe ratio} \times \sqrt{260}}{\sqrt{260}} \]
in the VIX futures. The position is unwound at the end of the second day. In this way I can decompose the returns of the long-short strategy into a futures part and an option part. Shorting VIX futures, conditional on the observation of negative variance, also makes a positive profit. Figure 3.8 plots the accumulated return. Though less striking as the achievement of the long-short strategy, the accumulated return of only shorting futures is still as high as 484%. The standard deviation of daily returns is 0.067, and the annualized Sharpe ratio is 0.45. 57.79% of the returns on trading negative observations are positive. These statistics give further evidence in support of that the volatility anomaly is economically significant.

3.5 Robustness

In this section I provide three additional analyses to support the robustness of the trading strategies which claim the volatility anomaly is economic significant. First I control the number of options for approximating the variance swap rate. In the second subsection I offer a benchmark of evaluating the magnitude of returns. In the end I use an alternative choice of matching when there are multiple successful matchings on a day.

3.5.1 Number of Options for Approximating the Variance Swap Rate

As mentioned earlier, the number of options included for calculating the variance swap rate tend to vary. I first analyze the sensitivity of the variation of variance swap rates if the two options at the lowest and highest strikes are removed. In general, the changes are negative. 98.99% of the changes are between 0 and -1% (in terms of annualized variance). The greatest falling off leads to a negative measurement of variance swap rate (-1.97%), which occurs on Oct 8th, 2008 for the options with time to maturity of 87 days. The reason is due to that there are only 11 options in the valid strike range, and that an insufficient number of options leads to a severe downward bias in the estimation of the variance swap rate. At the shorter maturities, the changes of variance swap rate after removing the two end options are smaller. Confining the time of maturities within 60
days, 91.04% of the changes are declination that is less than 0.1%. On the other hand, removing the two options at the end strikes can also occasionally increase the swap rate, if there are very few options and the changes of $\Delta K$ at the two ends take a dominant effect.

Secondly, I also control the minimal number of options that allows the calculation of the variance swap rate and hence also the variance of the VIX. Denote the number as $N$. In all previous analysis, $N$ is set to be 10. Here $N$ is also configured to be 20, 30, and 40. If there are fewer options than $N$ in the valid strike range, I do not consider approximating the variance swap rate by option prices.

As $N$ increases there are fewer days for which I calculate the variance of the VIX. For the entire sample, the proportion of negative observations seems to decrease as $N$ increases, as seen in table 3.2. Still, more than 15% of the observations of the variance are negative numbers. For the first subsample, the number of total observations are cut down quickly as $N$ increases, while the proportion of negative observations remains stable. For the second subsample, the number of total observations decreases less than the first subsample, while the proportion of negative observations declines more. On the other hand, for all the samples the mean and the standard deviation of the variance of the VIX stay relatively stable against the variation of the $N$.

I implement the long-short strategy and the short strategy for various values of $N$. Increasing $N$ reduces the accumulated returns as the trading occurs less frequent. As $N$ increases to 40, the accumulated return of the long-short strategy is significantly cut down, while the accumulated return of the short strategy is less affected. Nevertheless, the annualized Sharpe ratios remain high, as reported in table 3.3.

### 3.5.2 A benchmark for evaluating returns

As the returns reaped by the trading strategies are strikingly high, I provide a benchmark here that allows to evaluate the magnitude of the returns. This benchmark strategy shorts the VIX futures whenever the risk neutral variance of the VIX can be measured, regardless of whether it is negative or not. To make it comparable with the long-short strategy and the short strategy, in the benchmark strategy the number of futures shorted
<table>
<thead>
<tr>
<th>N</th>
<th>Number of Obs</th>
<th>Proportion of Negative Obs</th>
<th>Mean</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Entire Sample</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>320 /1433</td>
<td>22.33%</td>
<td>0.0036</td>
<td>0.0086</td>
</tr>
<tr>
<td>20</td>
<td>284 /1321</td>
<td>21.50%</td>
<td>0.0038</td>
<td>0.0089</td>
</tr>
<tr>
<td>30</td>
<td>182 /1025</td>
<td>17.76%</td>
<td>0.0043</td>
<td>0.0088</td>
</tr>
<tr>
<td>40</td>
<td>114 /734</td>
<td>15.53%</td>
<td>0.0047</td>
<td>0.0077</td>
</tr>
<tr>
<td></td>
<td>Before 01/01/2008</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>191 /692</td>
<td>27.60%</td>
<td>0.0012</td>
<td>0.0029</td>
</tr>
<tr>
<td>20</td>
<td>164 /598</td>
<td>27.42%</td>
<td>0.0011</td>
<td>0.0030</td>
</tr>
<tr>
<td>30</td>
<td>88 /359</td>
<td>24.51%</td>
<td>0.0011</td>
<td>0.0033</td>
</tr>
<tr>
<td>40</td>
<td>49 /182</td>
<td>26.92%</td>
<td>0.0009</td>
<td>0.0022</td>
</tr>
<tr>
<td></td>
<td>After 01/01/2008</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>129 /741</td>
<td>17.41%</td>
<td>0.0058</td>
<td>0.0111</td>
</tr>
<tr>
<td>20</td>
<td>120 /723</td>
<td>16.60%</td>
<td>0.0060</td>
<td>0.0112</td>
</tr>
<tr>
<td>30</td>
<td>94 /666</td>
<td>14.11%</td>
<td>0.0060</td>
<td>0.0102</td>
</tr>
<tr>
<td>40</td>
<td>65 /552</td>
<td>11.78%</td>
<td>0.0059</td>
<td>0.0084</td>
</tr>
</tbody>
</table>

Table 3.2: Variance of the VIX
As a sensitivity analysis I control the number of options used for calculating the variance swap rate. The variance of the VIX is not reported if there are fewer options than N. As N increases the number of observation decreases. For example, for the entire sample, the number of days on which I can measure the risk neutral variance of the VIX falls from 1433 to 734, and the number of negative observation also falls from 320 to 114.
Table 3.3: Trading on Negative Observations of Variance of Volatility

<table>
<thead>
<tr>
<th>N</th>
<th>Accumulated Return</th>
<th>Sharpe Ratio</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>88.12</td>
<td>1.50</td>
<td>0.062</td>
</tr>
<tr>
<td>20</td>
<td>51.25</td>
<td>1.36</td>
<td>0.064</td>
</tr>
<tr>
<td>30</td>
<td>36.02</td>
<td>1.49</td>
<td>0.067</td>
</tr>
<tr>
<td>40</td>
<td>6.90</td>
<td>1.04</td>
<td>0.065</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>Accumulated Return</th>
<th>Sharpe Ratio</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>4.84</td>
<td>0.45</td>
<td>0.067</td>
</tr>
<tr>
<td>20</td>
<td>4.11</td>
<td>0.42</td>
<td>0.066</td>
</tr>
<tr>
<td>30</td>
<td>2.86</td>
<td>0.35</td>
<td>0.072</td>
</tr>
<tr>
<td>40</td>
<td>2.73</td>
<td>0.39</td>
<td>0.076</td>
</tr>
</tbody>
</table>

The variable N controls the minimal number of options in the strike range for calculating variance swap rate. With insufficient number of options, the variance of VIX is not calculated. Therefore increasing N reduces the number of days on which negative observation of the variance of the VIX is found. The long-short strategy takes a long position in the option-synthesized option swap and a short position in the VIX futures. The short strategy only takes a short position in the VIX futures.

is inversely proportional to the portfolio wealth.

As shown in figure 3.9, shorting VIX futures unconditionally does not make a profit over the sample period. Starting from 1 dollar, the portfolio value gradually diminishes, though occasionally rising, and finally ends up at zero.

### 3.5.3 An alternative choice of matching

In the previous analysis, I use the one with the shortest maturities whenever there are multiple successful matchings on a day. Here I instead use the matching with the second shortest maturities, if there are more than one successful matchings. Results are reported in table 3.4 and table 3.5. Overall the proportions of the negative measurements
become smaller, and associated returns are also reduced. But the changes are small. Still a significant number of negative observations of the variance of the VIX are discovered, and the returns of the trading strategies are much higher than typical equity returns.

<table>
<thead>
<tr>
<th>N</th>
<th>Number of Obs</th>
<th>Proportion of Negative Obs</th>
<th>Mean</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Entire Sample</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>291 /1433</td>
<td>20.31%</td>
<td>0.0036</td>
<td>0.0090</td>
</tr>
<tr>
<td>20</td>
<td>273 /1321</td>
<td>20.67%</td>
<td>0.0037</td>
<td>0.0094</td>
</tr>
<tr>
<td>30</td>
<td>169 /1025</td>
<td>16.49%</td>
<td>0.0042</td>
<td>0.0093</td>
</tr>
<tr>
<td>40</td>
<td>110 /734</td>
<td>14.99%</td>
<td>0.0049</td>
<td>0.0077</td>
</tr>
<tr>
<td></td>
<td>Before 01/01/2008</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>156 /692</td>
<td>22.54%</td>
<td>0.0016</td>
<td>0.0030</td>
</tr>
<tr>
<td>20</td>
<td>147 /598</td>
<td>24.58%</td>
<td>0.0014</td>
<td>0.0031</td>
</tr>
<tr>
<td>30</td>
<td>70 /359</td>
<td>19.50%</td>
<td>0.0013</td>
<td>0.0034</td>
</tr>
<tr>
<td>40</td>
<td>47 /182</td>
<td>25.82%</td>
<td>0.0011</td>
<td>0.0023</td>
</tr>
<tr>
<td></td>
<td>After 01/01/2008</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>135 /741</td>
<td>18.82%</td>
<td>0.0055</td>
<td>0.0119</td>
</tr>
<tr>
<td>20</td>
<td>126 /723</td>
<td>17.43%</td>
<td>0.0056</td>
<td>0.0120</td>
</tr>
<tr>
<td>30</td>
<td>99 /666</td>
<td>14.86%</td>
<td>0.0057</td>
<td>0.0110</td>
</tr>
<tr>
<td>40</td>
<td>63 /552</td>
<td>11.41%</td>
<td>0.0061</td>
<td>0.0084</td>
</tr>
</tbody>
</table>

Table 3.4: Variance of the VIX

The matching scheme sometimes leads to multiple matchings of options and futures. In this table and the next table 3.5, the exact same exercises as in the previous analysis are taken, but using the combination of options and futures that has the second shortest maturity. This table reports the variance of the VIX measured using the alternative choice of matching.
Table 3.5: Trading on Negative Observations of Variance of Volatility

Using the alternative choice of matching produces different measurements of the risk neutral variance of the VIX, whenever there are multiple matchings. The long-short strategy and the short strategy are implemented again to reflect the changes. The table reports the statistics of the returns.

<table>
<thead>
<tr>
<th>N</th>
<th>Accumulated Return</th>
<th>Sharpe Ratio</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>41.57</td>
<td>1.27</td>
<td>0.064</td>
</tr>
<tr>
<td>20</td>
<td>42.08</td>
<td>1.29</td>
<td>0.065</td>
</tr>
<tr>
<td>30</td>
<td>20.63</td>
<td>1.21</td>
<td>0.072</td>
</tr>
<tr>
<td>40</td>
<td>6.30</td>
<td>1.00</td>
<td>0.065</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>Accumulated Return</th>
<th>Sharpe Ratio</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>3.24</td>
<td>0.34</td>
<td>0.067</td>
</tr>
<tr>
<td>20</td>
<td>2.47</td>
<td>0.26</td>
<td>0.065</td>
</tr>
<tr>
<td>30</td>
<td>3.37</td>
<td>0.42</td>
<td>0.073</td>
</tr>
<tr>
<td>40</td>
<td>2.64</td>
<td>0.39</td>
<td>0.076</td>
</tr>
</tbody>
</table>

3.6 Conclusion

I compare the variance swap rates implied from S&P500 options with the prices of the VIX futures. The two are connected with each other through the volatility of S&P500 Index, and combining them leads to the measurement of the risk neutral variance of the VIX. However, I find recurrent negative observations of the variances. To demonstrate the economic significance of this anomaly, trading strategies are designed and found to reap high returns. These results suggest inconsistency between the variance implied from S&P500 options and the volatility implied from VIX futures.
Table 3.6: VIX Futures Close Prices and Daily Volume

<table>
<thead>
<tr>
<th>Code</th>
<th>No Obs</th>
<th>Start Date</th>
<th>End Date</th>
<th>VIX Futures Prices Mean</th>
<th>Volume Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>K4</td>
<td>37</td>
<td>26-3-2004</td>
<td>18-5-2004</td>
<td>18.55</td>
<td>152</td>
</tr>
<tr>
<td>M4</td>
<td>54</td>
<td>26-3-2004</td>
<td>15-6-2004</td>
<td>18.84</td>
<td>177</td>
</tr>
<tr>
<td>N4</td>
<td>32</td>
<td>24-5-2004</td>
<td>13-7-2004</td>
<td>16.88</td>
<td>210</td>
</tr>
<tr>
<td>Q4</td>
<td>91</td>
<td>26-3-2004</td>
<td>17-8-2004</td>
<td>19.29</td>
<td>148</td>
</tr>
<tr>
<td>U4</td>
<td>41</td>
<td>19-7-2004</td>
<td>14-9-2004</td>
<td>17.97</td>
<td>162</td>
</tr>
<tr>
<td>V4</td>
<td>36</td>
<td>23-8-2004</td>
<td>12-10-2004</td>
<td>15.76</td>
<td>131</td>
</tr>
<tr>
<td>X4</td>
<td>153</td>
<td>26-3-2004</td>
<td>16-11-2004</td>
<td>18.91</td>
<td>155</td>
</tr>
<tr>
<td>F5</td>
<td>60</td>
<td>21-10-2004</td>
<td>18-1-2005</td>
<td>14.39</td>
<td>78</td>
</tr>
<tr>
<td>G5</td>
<td>162</td>
<td>21-6-2004</td>
<td>15-2-2005</td>
<td>17.08</td>
<td>129</td>
</tr>
<tr>
<td>H5</td>
<td>36</td>
<td>24-1-2005</td>
<td>15-3-2005</td>
<td>12.70</td>
<td>159</td>
</tr>
<tr>
<td>K5</td>
<td>159</td>
<td>20-9-2004</td>
<td>17-5-2005</td>
<td>15.46</td>
<td>166</td>
</tr>
<tr>
<td>M5</td>
<td>58</td>
<td>21-3-2005</td>
<td>14-6-2005</td>
<td>14.34</td>
<td>145</td>
</tr>
<tr>
<td>Q5</td>
<td>170</td>
<td>23-11-2004</td>
<td>16-8-2005</td>
<td>14.68</td>
<td>189</td>
</tr>
<tr>
<td>V5</td>
<td>80</td>
<td>21-6-2005</td>
<td>18-10-2005</td>
<td>14.24</td>
<td>71</td>
</tr>
<tr>
<td>X5</td>
<td>181</td>
<td>22-2-2005</td>
<td>15-11-2005</td>
<td>15.01</td>
<td>134</td>
</tr>
<tr>
<td>Z5</td>
<td>40</td>
<td>24-10-2005</td>
<td>20-12-2005</td>
<td>12.35</td>
<td>91</td>
</tr>
<tr>
<td>F6</td>
<td>34</td>
<td>22-11-2005</td>
<td>17-1-2006</td>
<td>12.45</td>
<td>61</td>
</tr>
<tr>
<td>G6</td>
<td>171</td>
<td>23-5-2005</td>
<td>14-2-2006</td>
<td>14.95</td>
<td>159</td>
</tr>
<tr>
<td>H6</td>
<td>40</td>
<td>24-1-2006</td>
<td>21-3-2006</td>
<td>12.58</td>
<td>69</td>
</tr>
<tr>
<td>J6</td>
<td>37</td>
<td>22-2-2006</td>
<td>18-4-2006</td>
<td>12.47</td>
<td>69</td>
</tr>
<tr>
<td>K6</td>
<td>171</td>
<td>22-8-2005</td>
<td>16-5-2006</td>
<td>14.62</td>
<td>307</td>
</tr>
<tr>
<td>M6</td>
<td>41</td>
<td>24-4-2006</td>
<td>20-6-2006</td>
<td>14.92</td>
<td>292</td>
</tr>
<tr>
<td>N6</td>
<td>40</td>
<td>22-5-2006</td>
<td>18-7-2006</td>
<td>15.81</td>
<td>182</td>
</tr>
<tr>
<td>Q6</td>
<td>159</td>
<td>22-12-2005</td>
<td>15-8-2006</td>
<td>15.14</td>
<td>481</td>
</tr>
<tr>
<td>U6</td>
<td>41</td>
<td>24-7-2006</td>
<td>19-9-2006</td>
<td>14.14</td>
<td>422</td>
</tr>
<tr>
<td>V6</td>
<td>41</td>
<td>21-8-2006</td>
<td>17-10-2006</td>
<td>13.56</td>
<td>718</td>
</tr>
<tr>
<td>X6</td>
<td>171</td>
<td>9-3-2006</td>
<td>14-11-2006</td>
<td>14.80</td>
<td>627</td>
</tr>
<tr>
<td>Z6</td>
<td>63</td>
<td>21-9-2006</td>
<td>19-12-2006</td>
<td>12.62</td>
<td>550</td>
</tr>
<tr>
<td>F7</td>
<td>55</td>
<td>24-10-2006</td>
<td>16-1-2007</td>
<td>12.36</td>
<td>200</td>
</tr>
<tr>
<td>G7</td>
<td>218</td>
<td>9-3-2006</td>
<td>13-2-2007</td>
<td>14.90</td>
<td>394</td>
</tr>
</tbody>
</table>
VIX Futures Close Prices and Daily Volume (Continued)

<table>
<thead>
<tr>
<th>Code</th>
<th>No Obs</th>
<th>Start Date</th>
<th>End Date</th>
<th>VIX Futures Price Mean</th>
<th>Volume Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>H7</td>
<td>87</td>
<td>24-10-2006</td>
<td>20-3-2007</td>
<td>13.47</td>
<td>468</td>
</tr>
<tr>
<td>J7</td>
<td>95</td>
<td>24-10-2006</td>
<td>17-4-2007</td>
<td>13.84</td>
<td>363</td>
</tr>
<tr>
<td>K7</td>
<td>261</td>
<td>23-3-2006</td>
<td>15-5-2007</td>
<td>15.20</td>
<td>290</td>
</tr>
<tr>
<td>M7</td>
<td>96</td>
<td>11-1-2007</td>
<td>19-6-2007</td>
<td>14.10</td>
<td>426</td>
</tr>
<tr>
<td>N7</td>
<td>102</td>
<td>22-1-2007</td>
<td>17-7-2007</td>
<td>14.59</td>
<td>466</td>
</tr>
<tr>
<td>Q7</td>
<td>228</td>
<td>26-6-2006</td>
<td>21-8-2007</td>
<td>16.07</td>
<td>727</td>
</tr>
<tr>
<td>U7</td>
<td>107</td>
<td>30-3-2007</td>
<td>18-9-2007</td>
<td>18.19</td>
<td>586</td>
</tr>
<tr>
<td>V7</td>
<td>111</td>
<td>3-5-2007</td>
<td>16-10-2007</td>
<td>18.58</td>
<td>685</td>
</tr>
<tr>
<td>X7</td>
<td>238</td>
<td>22-11-2006</td>
<td>20-11-2007</td>
<td>17.54</td>
<td>895</td>
</tr>
<tr>
<td>Z7</td>
<td>146</td>
<td>21-5-2007</td>
<td>18-12-2007</td>
<td>20.20</td>
<td>815</td>
</tr>
<tr>
<td>F8</td>
<td>116</td>
<td>23-7-2007</td>
<td>15-1-2008</td>
<td>22.30</td>
<td>559</td>
</tr>
<tr>
<td>H8</td>
<td>120</td>
<td>24-9-2007</td>
<td>18-3-2008</td>
<td>24.15</td>
<td>744</td>
</tr>
<tr>
<td>J8</td>
<td>116</td>
<td>24-10-2007</td>
<td>15-4-2008</td>
<td>24.96</td>
<td>776</td>
</tr>
<tr>
<td>K8</td>
<td>343</td>
<td>25-4-2006</td>
<td>20-5-2008</td>
<td>20.30</td>
<td>352</td>
</tr>
<tr>
<td>M8</td>
<td>222</td>
<td>21-6-2007</td>
<td>17-6-2008</td>
<td>22.52</td>
<td>457</td>
</tr>
<tr>
<td>N8</td>
<td>133</td>
<td>31-12-2007</td>
<td>15-7-2008</td>
<td>24.24</td>
<td>589</td>
</tr>
<tr>
<td>Q8</td>
<td>226</td>
<td>23-8-2007</td>
<td>19-8-2008</td>
<td>23.39</td>
<td>468</td>
</tr>
<tr>
<td>U8</td>
<td>118</td>
<td>25-3-2008</td>
<td>16-9-2008</td>
<td>23.45</td>
<td>739</td>
</tr>
<tr>
<td>V8</td>
<td>127</td>
<td>22-4-2008</td>
<td>21-10-2008</td>
<td>26.75</td>
<td>831</td>
</tr>
<tr>
<td>X8</td>
<td>232</td>
<td>26-11-2007</td>
<td>18-11-2008</td>
<td>27.59</td>
<td>556</td>
</tr>
<tr>
<td>Z8</td>
<td>195</td>
<td>30-1-2008</td>
<td>16-12-2008</td>
<td>29.87</td>
<td>589</td>
</tr>
<tr>
<td>F9</td>
<td>163</td>
<td>22-4-2008</td>
<td>20-1-2009</td>
<td>33.56</td>
<td>262</td>
</tr>
<tr>
<td>G9</td>
<td>188</td>
<td>25-2-2008</td>
<td>17-2-2009</td>
<td>33.85</td>
<td>225</td>
</tr>
<tr>
<td>H9</td>
<td>153</td>
<td>11-7-2008</td>
<td>17-3-2009</td>
<td>37.14</td>
<td>256</td>
</tr>
<tr>
<td>J9</td>
<td>146</td>
<td>16-7-2008</td>
<td>14-4-2009</td>
<td>38.94</td>
<td>248</td>
</tr>
<tr>
<td>K9</td>
<td>167</td>
<td>1-8-2008</td>
<td>19-5-2009</td>
<td>36.81</td>
<td>267</td>
</tr>
<tr>
<td>M9</td>
<td>151</td>
<td>12-9-2008</td>
<td>16-6-2009</td>
<td>37.04</td>
<td>309</td>
</tr>
<tr>
<td>N9</td>
<td>165</td>
<td>30-9-2008</td>
<td>21-7-2009</td>
<td>36.01</td>
<td>408</td>
</tr>
<tr>
<td>Q9</td>
<td>150</td>
<td>18-12-2008</td>
<td>18-8-2009</td>
<td>34.03</td>
<td>581</td>
</tr>
</tbody>
</table>
### VIX Futures Close Prices and Daily Volume (Continued)

<table>
<thead>
<tr>
<th>Code</th>
<th>No Obs</th>
<th>Start Date</th>
<th>End Date</th>
<th>VIX Futures Price Mean</th>
<th>Volume Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>U9</td>
<td>173</td>
<td>3-12-2008</td>
<td>15-9-2009</td>
<td>33.79</td>
<td>669</td>
</tr>
<tr>
<td>V9</td>
<td>188</td>
<td>12-1-2009</td>
<td>20-10-2009</td>
<td>32.43</td>
<td>845</td>
</tr>
<tr>
<td>X9</td>
<td>200</td>
<td>27-1-2009</td>
<td>17-11-2009</td>
<td>31.31</td>
<td>871</td>
</tr>
<tr>
<td>Z9</td>
<td>145</td>
<td>21-5-2009</td>
<td>15-12-2009</td>
<td>27.73</td>
<td>1219</td>
</tr>
<tr>
<td>F10</td>
<td>148</td>
<td>18-6-2009</td>
<td>19-1-2010</td>
<td>27.23</td>
<td>1186</td>
</tr>
<tr>
<td>G10</td>
<td>143</td>
<td>22-7-2009</td>
<td>16-2-2010</td>
<td>27.13</td>
<td>1386</td>
</tr>
<tr>
<td>H10</td>
<td>142</td>
<td>20-8-2009</td>
<td>16-3-2010</td>
<td>26.03</td>
<td>1373</td>
</tr>
<tr>
<td>J10</td>
<td>149</td>
<td>16-9-2009</td>
<td>20-4-2010</td>
<td>25.08</td>
<td>1548</td>
</tr>
<tr>
<td>K10</td>
<td>155</td>
<td>25-9-2009</td>
<td>18-5-2010</td>
<td>25.04</td>
<td>2070</td>
</tr>
<tr>
<td>M10</td>
<td>161</td>
<td>23-10-2009</td>
<td>15-6-2010</td>
<td>25.93</td>
<td>1957</td>
</tr>
<tr>
<td>N10</td>
<td>160</td>
<td>24-11-2009</td>
<td>20-7-2010</td>
<td>26.42</td>
<td>1697</td>
</tr>
<tr>
<td>Q10</td>
<td>162</td>
<td>22-12-2009</td>
<td>17-8-2010</td>
<td>26.38</td>
<td>1556</td>
</tr>
<tr>
<td>U10</td>
<td>160</td>
<td>26-1-2010</td>
<td>14-9-2010</td>
<td>27.49</td>
<td>1831</td>
</tr>
<tr>
<td>V10</td>
<td>166</td>
<td>23-2-2010</td>
<td>19-10-2010</td>
<td>27.79</td>
<td>2270</td>
</tr>
<tr>
<td>X10</td>
<td>167</td>
<td>23-3-2010</td>
<td>16-11-2010</td>
<td>27.87</td>
<td>3072</td>
</tr>
<tr>
<td>Z10</td>
<td>167</td>
<td>27-4-2010</td>
<td>21-12-2010</td>
<td>27.53</td>
<td>3742</td>
</tr>
<tr>
<td>F11</td>
<td>165</td>
<td>25-5-2010</td>
<td>18-1-2011</td>
<td>27.94</td>
<td>3499</td>
</tr>
<tr>
<td>G11</td>
<td>166</td>
<td>22-6-2010</td>
<td>15-2-2011</td>
<td>26.97</td>
<td>3861</td>
</tr>
<tr>
<td>H11</td>
<td>162</td>
<td>26-7-2010</td>
<td>15-3-2011</td>
<td>26.38</td>
<td>5043</td>
</tr>
<tr>
<td>J11</td>
<td>167</td>
<td>23-8-2010</td>
<td>19-4-2011</td>
<td>25.16</td>
<td>5234</td>
</tr>
<tr>
<td>K11</td>
<td>167</td>
<td>20-9-2010</td>
<td>17-5-2011</td>
<td>24.05</td>
<td>5001</td>
</tr>
<tr>
<td>M11</td>
<td>161</td>
<td>25-10-2010</td>
<td>14-6-2011</td>
<td>22.84</td>
<td>5368</td>
</tr>
<tr>
<td>N11</td>
<td>165</td>
<td>22-11-2010</td>
<td>19-7-2011</td>
<td>22.48</td>
<td>6533</td>
</tr>
<tr>
<td>Q11</td>
<td>161</td>
<td>28-12-2010</td>
<td>16-8-2011</td>
<td>22.75</td>
<td>8562</td>
</tr>
<tr>
<td>U11</td>
<td>167</td>
<td>24-1-2011</td>
<td>20-9-2011</td>
<td>24.53</td>
<td>8159</td>
</tr>
<tr>
<td>X11</td>
<td>130</td>
<td>21-3-2011</td>
<td>22-9-2011</td>
<td>24.96</td>
<td>3264</td>
</tr>
<tr>
<td>Z11</td>
<td>105</td>
<td>26-4-2011</td>
<td>22-9-2011</td>
<td>24.73</td>
<td>3559</td>
</tr>
<tr>
<td>F12</td>
<td>86</td>
<td>23-5-2011</td>
<td>22-9-2011</td>
<td>26.23</td>
<td>1790</td>
</tr>
<tr>
<td>G12</td>
<td>67</td>
<td>20-6-2011</td>
<td>22-9-2011</td>
<td>27.04</td>
<td>890</td>
</tr>
</tbody>
</table>
VIX Futures Close Prices and Daily Volume (Continued)

<table>
<thead>
<tr>
<th>Code</th>
<th>No Obs</th>
<th>Start Date</th>
<th>End Date</th>
<th>VIX Futures Price</th>
<th>Volume Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>H12</td>
<td>43</td>
<td>25-7-2011</td>
<td>22-9-2011</td>
<td>28.65</td>
<td>547</td>
</tr>
<tr>
<td>J12</td>
<td>23</td>
<td>22-8-2011</td>
<td>22-9-2011</td>
<td>31.07</td>
<td>156</td>
</tr>
<tr>
<td>K12</td>
<td>21</td>
<td>24-8-2011</td>
<td>22-9-2011</td>
<td>30.77</td>
<td>82</td>
</tr>
</tbody>
</table>

Note: The sample spans from Mar 26th, 2004 to Sep 30th, 2011. Start and end dates are the first and last trading day documented in my dataset. The futures contract code is the expiration month code followed by a digit representing the expiration year. The expiration month codes follow the convention for all commodities futures, which is defined as follows: January - F, February - G, March - H, April - J, May - K, June - M, July - N, August - Q, September - U, October - V, November - X, and December - Z.
Figure 3.1: The difference between two VIX
This figure plots the difference between the VIX constructed with the synthesized variance swap rates and the CBOE VIX. Values are in volatility percentage units. The two outliers are excluded in the later analysis.

Figure 3.2: Term structure of option-synthesized variance swap rate
Values are in annualized variance percentage units. The illustrated term structure of variance swap rate is a spot curve, and the variances are for the time periods between the time of observation and the option maturities.
Figure 3.3: Term structure of the VIX futures price
Values are in annualized volatility percentage units. The illustrated term structure of the VIX futures is a forward curve, and the volatilities are for the 30 day period after the futures maturities.

Figure 3.4: The Sum of the Slope Signs of the Two Term Structures
By measuring the expected variances between two option maturities I transform the spot curve of variance swap rate into a forward curve. I then compare the slopes of the forward curve of variance swap rate and the forward curve of the VIX futures. 1 indicates that both of the term structures are upward sloping, -1 indicates that both of the term structures are downward sloping, and 0 indicates that the two term structures slope in opposite directions.
$t + T$ : VIX futures Maturity Date  
$t + T_1$ : Option Maturity Date  
$t + T_2$ : Option Maturity Date  
$T_2 - T = 2$ Days  
$T_1 - T_2 = 28$ Days  

As futures are often settled on the third Wednesday and options are often settled on the third Friday, there is a two-day gap between the maturities of the VIX futures and the S&P options. The next option maturity is often four weeks from the coming Friday.

Figure 3.5: Matching the Maturities of Options and Futures
The figure plots the measured risk neutral variances of the VIX at the VIX futures maturity dates, as implied from the S&P500 options and the VIX futures.

Figure 3.6: The Variance of the VIX
Trade occurs whenever the measured risk neutral variance of the VIX is negative. The strategy takes a short position in the VIX futures and mimics a long position in the variance swap rate using S&P500 options. The returns are shown in log scale. The absolute accumulated return reaches to 88.12 at the end of the sample.
Trade occurs whenever the measured risk neutral variance of the VIX is negative. The strategy takes a short position in the VIX futures.
Figure 3.9: Returns of the Benchmark Strategy

The benchmark strategy shorts the VIX futures whenever the risk neutral variance of the VIX can be measured. As the figure shows, shorting VIX futures unconditionally does not make a profit over the given sample period. This provides evidence in support of that the positive returns earned by the long-short strategy and the short strategy conditional on the volatility anomaly are economic significant.
Chapter 4

Testing Affine Stochastic Volatility Models with the VIX

4.1 Introduction

A variance swap is a derivatives instrument that allows investors to trade the volatility of an asset directly. At expiry, the long party pays a predetermined variance swap rate. In return, the long side receives the realized variance of the underlying over the life of the swap contract. Before the variance swap was actually traded in market, the theory of how to price a variance swap had already been established in a working paper by Neuberger (1990) and later an independent article by Dupire (1993). These papers show that, for a continuously-sampled variance swap on an underlying asset with continuous path, the variance swap rate is simply twice the forward price of a contract that pays the log of the asset price at the same maturity date.

Moreover, for a continuous stochastic volatility model, the forward price of the log contract can be replicated from prices of European options on the same underlying with the same maturity (see Dupire (1993) and Carr and Madan (1998b)). This theoretical result is brought further to a successful application, which is the Chicago Board of Options Exchange (CBOE) volatility index (VIX). The VIX is defined as a weighted average of out-of-money option prices, to reflect the market’s expectation of S&P500 volatility over the next 30 day period. The square of the VIX represents the variance swap rate with the fixed time to maturity of 30 days. Nowadays the VIX is widely tracked and often referred to as the fear gauge. Arguably the success of the VIX is a recognition of the significance
of the theory for pricing variance swaps.

The insight that the price of the log contract can be represented as a function of European option prices provides a somewhat model-free approach to pricing variance swap. However, it rests on the assumption that the underlying price process is continuous, which can be questioned especially in light of the recent crisis, and may also have failed to hold prior to the crisis. Concerned with the possibility that the path of an asset price can contain discontinuous jumps, researchers have studied volatility models which permit jumps in the price. Meanwhile, demand for these models also comes from the innovation of more sophisticated volatility derivatives, such as VIX futures, VIX options, and exotic variance swaps. To price these volatility derivatives, two lines of research emerge. One prices variance swaps on a general exponential Levy process which is further stochastically time-changed by a continuous integrable clock (see Carr et al. (2012)). In this approach, the variance swap rate is still a multiple of the price of the log contract. However the multiplier is not 2 as in the theory of Neuberger (1990) and Dupire (1993), but a constant dependent on the characteristics of the Levy process. As one of the implications, the design of the VIX is flawed if the multiplier is not 2.

The other approach goes along the line of parametric models that augment the Black-Scholes model with stochastic volatility and jumps. Typical examples are pure jump models and affine jump diffusion models. The pure jump models treat the volatility process as a pure jump process that has only discrete movements and no small continuous movements (see Carr et al. (2005)). Using pure jump models for the volatility is supported by the findings of Todorov and Tauchen (2011), though the high-frequency VIX data in their study is noisy and empirical conclusions are not unambiguously clear cut. In fact, if the volatility process were a pure jump process, the VIX is also subject to a systemic bias in the approximation of the variance swap rate of which the consequences, until now, are empirically unknown.

Our analysis in this study will be focused on the affine jump diffusion models (AJD) as formulated in Duffie et al. (2000). The reason is multifold. First of all, the affine models have a clear implication ready to be examined. They imply that the option-synthesized variance swap rate for S&P500 is simply an affine function of the instantaneous variance. Secondly, they are natural extensions of the continuous stochastic volatility models that underly the theory for the VIX. Hence, whether the affine volatility model is valid or not
matters for the justification of the VIX index. Last but not least, it is because of their special significance as tools in pricing options.

Popular AJDs are Stochastic Volatility(SV), Stochastic Volatility with Jumps in returns(SVJ), Stochastic Volatility with Independent Jumps in returns and volatility(SVIJ), and Stochastic Volatility with Common Jumps in returns and volatility(SVCJ). They are widely adopted in the option pricing literature, for their computational tractability and clarity of the endowed economic interpretation. There have been a series of applications of AJDs in describing market dynamics of returns and identifying risk premiums for stochastic volatility and for jumps in the underlying process (see Bakshi et al. (1997), Bates (2000), Chernov and Ghysels (2000), Pan (2002), Eraker et al. (2003), Jones (2003), Eraker (2004), Jones (2006), and Broadie et al. (2007)). The motivation of these studies is to understand how risk factors, namely the diffusive Brownian risk representing small continuous changes and the jump risk representing rapid and substantially large market changes, are priced in the market. In this approach option pricing is done in the risk-neutral description of the model, with risk premiums under the Q measure added to the real world P measure. Almost exclusively, it relies on returns data to identify the P measure and option prices data of a narrow strike range to infer the Q measure. One exception is Andersen et al. (2011) that exploits the information of option panels which include a large number of observations in the moneyness-maturity(cross-sectional) dimension.

To provide an evaluation of using affine models to model variance swap rates, we use option-synthesized variance swap rates and the VIX index. As mentioned earlier, with prices of European options on the S&P500 it is possible to infer the variance swap rate up to option maturities. The variance swap rate over the remaining life of the options at a certain maturity is also a key quantity for option pricing, as it is the integrated variance that matters for calculating the option values. Perhaps this relevance can be best appreciated in the context of the Hull and White (1987) model, where the option price is the conditional expectation of the Black-Scholes price with the integrated variance replacing the Black-Scholes variance.

In this paper, the evaluation of the AJD models is done by relating the instantaneous variance to the VIX and variance swap rates. According to the affine volatility model, the VIX square, which is the variance swap rate with a constant 30 day term to maturity, depends linearly on the instantaneous variance. In the data we find a quasi linear
relationship between the VIX square and the instantaneous variance. Regressing the VIX square on the estimated instantaneous variance produces a R-square of 61.5%. On the other hand, adding volatility lags is found to significantly improve the R-square. This shows that volatility persistence matters for pricing variance swaps rate.

We also form a vega neutral portfolio with two option-synthesized variance swaps. Following the affine volatility models, the portfolio value should be constant and represent the long run mean of the variance in the risk neutral measure. However, we find that the portfolio is still subject to volatility risk and the portfolio value tends to correlate with the instantaneous volatility. These findings together indicate the inefficiency of using affine volatility models to price and hedge variance swaps.

4.2 AJD Models and their Implication for Expected Forward Integrated Variance

The particular type of affine jump diffusion models we consider follows the standard in the literature. We fix a probability space \((\Omega, \mathcal{F}, P)\) and an information filtration \((\mathcal{F}_t)\) satisfying the usual conditions (see, for example, Protter (1990)), and assume the joint dynamics of the S&P500 log price and the associated variance under the equivalent martingale measure are of the form:

\[
\begin{align*}
\left( \frac{dS_t}{dV_t} \right) &= \left( \frac{\alpha_t}{\theta(k - V^-)} \right) dt + \sqrt{V^-} \begin{pmatrix} 1 & 0 \\ \rho \sigma_v & \sqrt{1 - \rho^2} \sigma_v \end{pmatrix} \begin{pmatrix} dW^1_t \\ dW^2_t \end{pmatrix} + \left( \xi^s_t dN^s_t + \xi^v_t dN^v_t \right) 
\end{align*}
\] (4.1)

where \(S_t\) is the log price, \(V_t\) is the instantaneous variance, \(V^- = \lim_{s \uparrow t} V_s\), \(W^1_t\) and \(W^2_t\) are two independent standard Brownian motions, \(\rho\) is between 0 and 1, and models the correlation between \(S_t\) and \(V_t\) through the Brownian motions, \(\alpha_t\) is the risk neutral drift of \(S_t\), \(k\) is the long run average variance, \(\theta\) is the rate at which \(V\) converts to \(k\), \(N^s_t\) and \(N^v_t\) are Poisson processes with intensities \(\lambda_s\) and \(\lambda_v\), and \(\xi^s_t\) and \(\xi^v_t\) are the random jump sizes in returns and volatility, respectively.

Unlike many studies, we omit the specification of the data generating processes under the physical measure \(P\). There are two reasons. First, the analysis will only involve the VIX and option-synthesized variance swap rates, both of which are determined by the Q
measure. As the analysis does not touch on the underlying index, there is no necessity to make the specification under the P measure. Second, the instantaneous variance \( V \) has to be the same for both P and Q to ensure the two measures are equivalent, otherwise P and Q would be distinguishable by measuring the quadratic variation of \( S_t \). And later we will take a non-parametric approach to estimate the instantaneous variance, which is actually regardless of the specifications of P and Q.

The specification in (1) nests many popular models. Without jumps, \( \lambda_s = 0, \lambda_v = 0 \), it reduces to the Heston stochastic volatility model. Based on the Heston model Grubichler and Longstaff (1996) discuss valuation for volatility futures and volatility options. Zhang and Zhu (2006) calibrate the Heston model to the VIX futures data. With jumps in returns but not in volatility, the specification becomes the stochastic volatility model with jumps (SVJ). For example, in Bates (1996) the SVJ model has normally distributed jumps in returns \( \xi_s \sim N(\mu_s, \sigma_s^2) \), but no jumps in volatility, \( \lambda_v = 0 \). With jumps both in the underlying and the volatility, this specification can lead to the stochastic volatility model with independent jumps in returns and volatility (SVIJ), if the jumps in returns and volatility arrive independently, or the stochastic volatility model with common jumps (SVCJ), if it has contemporaneous arrivals of jumps, \( N_t^s = N_t^v \).

The quadratic variation (QV) process associated with the log return \( S_t \) is defined as

\[
[S_t] = \lim_{n \to \infty} \{ S_{\Delta^n} \}
\]

where \( \{ S_{\Delta^n} \} = \sum_{j=0}^{n-1} (S_{t_{j+1}} - S_{t_j})^2 \), for any sequence of partitions \( t_0 = 0 < t_1 < ... < t_n = t \) with \( \sup_j \{ t_{j+1} - t_j \} \to 0 \) for \( n \to \infty \). It can be split into continuous and discontinuous components:

\[
[S]_t = [S^c]_t + [S^d]_t
\]

In the case of affine jump diffusion models, the continuous component \( [S^c]_t = \int_0^t V_s \, ds \) and the discontinuous component \( [S^d]_t = \sum_{j=1}^{N_t} (\xi_j^s)^2 \).

We are interested in the conditional expectation of the quadratic variation from the time of observation \( t \) up to a certain time \( T \) in the future \( (t < T) \), \( E_t[S^T]_t \). In the following we consider several cases:

**A. The Heston stochastic volatility model.**
In the Heston model, \([S]_T^T = [S^c]_T^T = \int_t^T V_s ds\), and \(E_t[S]_t^T = E_t \int_t^T V_s ds\). Let \(W^u_t = \rho W^1_t + \sqrt{1 - \rho^2} W^2_t\). Notice that

\[
dV_t = \theta(k - V_t)dt + \sigma_v \sqrt{V_t} dW^v_t
\]

\[
\Rightarrow dV_t + \theta V_t dt = \theta k dt + \sigma_v \sqrt{V_t} dW^v_t
\]

\[
\Rightarrow e^{-\theta t} d(e^{\theta t} V) = \theta k dt + e^{\theta t} \sigma_v \sqrt{V_t} dW^v_t
\]

It further gives

\[
e^{\theta(t+u)} V_{t+u} = e^{\theta t} V_t + \int_t^{t+u} e^{\theta s} \theta k ds + \int_t^{t+u} e^{\theta s} \sigma_v \sqrt{V_s} dW^v_s
\]

\[
V_{t+u} = V_t e^{-\theta u} + e^{-\theta(t+u)} \int_t^{t+u} e^{\theta s} \theta k ds + e^{-\theta(t+u)} \int_t^{t+u} e^{\theta s} \sigma_v \sqrt{V_s} dW^v_s
\]

Taking expectation on both sides leads to

\[
E_t(V_{t+u}) = V_t e^{-\theta u} + \kappa(1 - e^{-\theta u})
\]

This allows us to compute \(E_t[S]_t^T\)

\[
E_t[S]_t^T = E_t \int_t^T V_s ds = \int_t^T E_t V_s ds
\]

After calculating the integral we have

\[
E_t[S]_t^T = \frac{1 - e^{-\theta(T-t)}}{\theta} (V_t - k) + k(T-t)
\]

(4.2)

**B. Heston stochastic volatility model with random jumps.**

Now we consider the case that in the return and volatility processes there are jumps with constant intensity. Examples include SVJ in which \(\lambda_v = 0\) and \(\lambda_s\) is constant, SVCJ, and SVIJJ, in which both of \(\lambda_v\) and \(\lambda_s\) are constant. In the literature the distributions of the jump size \(\xi_v\) and \(\xi_s\) are set to be exogenously fixed. Denote \(E \xi_v\) by \(\mu_v\), and \(E(\xi_s)^2\) by \(\mu_s^2\).

An alternative way of writing

\[
dV_t = \theta(k - V_t)dt + \sqrt{V_t - \sigma_v} dW^\nu_t + \xi^\nu dN^\nu_t
\]

(4.3)
\[ dV_t = \theta \left( \frac{\theta k + \lambda_v \mu_v}{\theta} - V_{t^-} \right) dt + \sqrt{V_t} \sigma_v dW_t^v + [\xi^v dN_t^v - \lambda_v \mu_v dt] \]

Applying Ito’s lemma to \( e^{\theta t} V_t \), we have

\[ de^{\theta t} V_t = (\theta k + \lambda_v \mu_v) e^{\theta t} dt + \sigma_v e^{\theta t} dW_t^v + e^{\theta t} [\xi^v dN_t^v - \lambda_v \mu_v dt] \]

Hence

\[ V_{t+u} = V_t e^{-\theta u} + e^{-\theta(t+u)} \int_t^{t+u} (\theta k + \lambda_v \mu_v) e^{\theta s} ds + e^{-\theta(t+u)} \int_t^{t+u} \sigma_v e^{\theta s} \sqrt{V_s} dW_s^v \\
+ e^{-\theta(t+u)} \int_t^{t+u} e^{\theta s} \xi^v dN_s^v - \mu_v \lambda_v ds \]

Because \( dW_s^v \) and \( \xi^v dN_s^v - \mu_v \lambda_v ds \) are increments of Martingales,

\[ E_t V_{t+u} = V_t e^{-\theta u} + e^{-\theta(t+u)} \int_t^{t+u} (\theta k + \lambda_v \mu_v) e^{\theta s} ds \\
= V_t e^{-\theta u} + \frac{\theta k + \lambda_v \mu_v}{\theta} (1 - e^{-\theta u}) \]

It further implies

\[ E_t [S_{T-t}^c] = \frac{1 - e^{-\theta(T-t)}}{\theta} (V_t - \frac{\theta k + \lambda_v \mu_v}{\theta}) + \frac{\theta k + \lambda_v \mu_v}{\theta} (T-t) \]

On the other hand, \( E_t [S_{T-t}^d] = E_t \sum_{j=N_t}^{N_{T-t}} (\xi_j^s)^2 \).

\[ E_t [S_{T-t}^T] = E_t [S_{T-t}^c] + E_t [S_{T-t}^d] \]

\[ = \frac{1 - e^{-\theta(T-t)}}{\theta} (V_t - \frac{\theta k + \lambda_v \mu_v}{\theta}) + \frac{\theta k + \lambda_v \mu_v}{\theta} (T-t) + E_t \sum_{j=N_t}^{N_{T-t}} (\xi_j^s)^2 \]

For SVIJ and SVCJ in which the jump intensities are constant, we have:

\[ E_t [S_{T-t}^T] = \frac{1 - e^{-\theta(T-t)}}{\theta} (V_t - \frac{\theta k + \lambda_v \mu_v}{\theta}) + \frac{\theta k + \lambda_v \mu_v}{\theta} (T-t) + \lambda_s (T-t) \mu_s^2 \] (4.4)

C. Heston stochastic volatility model with state dependent jumps.
In the literature often the jump intensity is a linear function of the \( V_t \). If \( \lambda_v = 0 \) and \( \lambda_s(t) \) is a linear function of \( V_t \): \( \lambda_s = \lambda_s^0 + \lambda_s^1 V_t \),

\[
E_t[S]_t^T = (1 + \lambda_s^1 \mu_s^2) \frac{1 - e^{-\theta(T-t)}}{\theta} (V_t - k) + k(T - t) + \lambda_s^0 (T - t) \mu_s^2
\]

If there are jumps in both returns and volatility, in this case simultaneous jumps (SVCJ) are assumed as the standard in the literature, \( N_t^v = N_t^s \), and \( \lambda_s = \lambda_s^0 + \lambda_s^1 V_t \).

We can rewrite equation 4.3 as

\[
dV_t = (\theta - \lambda_1 \mu_v) \left( \frac{\theta k + \lambda_0 \mu_v}{\theta - \lambda_1 \mu_v} - V_{t-} \right) dt + \sqrt{V_t - \sigma_v dW_t^v} + [\xi_v dN_t^v - \mu_v (\lambda_0 + \lambda_1 V_{t-})] dt
\]

By applying Ito’s lemma to \( e^{(\theta - \lambda_1 \mu_v)t} V_t \), we have

\[
de^{(\theta - \lambda_1 \mu_v)t} V_t = (\theta k + \lambda_0 \mu_v) e^{(\theta - \lambda_1 \mu_v)t} V_{t-} dt + \sigma_v e^{(\theta - \lambda_1 \mu_v)t} \sqrt{V_t} dW_t^v + e^{(\theta - \lambda_1 \mu_v)t} \xi_v N_t^v - \mu_v (\lambda_0 + \lambda_1 V_{t-}) dt
\]

Hence

\[
V_{t+u} = V_{t-} e^{-(\theta - \lambda_1 \mu_v)u} + e^{-(\theta - \lambda_1 \mu_v)(t+u)} \int_u^{t+u} (\theta k + \lambda_0 \mu_v) e^{(\theta - \lambda_1 \mu_v)s} ds
\]

\[
+ e^{-(\theta - \lambda_1 \mu_v)(t+u)} \int_u^{t+u} \sigma_v e^{(\theta - \lambda_1 \mu_v)s} \sqrt{V_s} dW_s^v
\]

\[
+ e^{-(\theta - \lambda_1 \mu_v)(t+u)} \int_u^{t+u} e^{(\theta - \lambda_1 \mu_v)s} [\xi_v dN_s^v - \mu_v (\lambda_0 + \lambda_1 V_s^-) ds]
\]

Because \( dW_s^v \) and \( \xi_v dN_s^v - \mu_v (\lambda_0 + \lambda_1 V_s^-) ds \) are increments of Martingales,

\[
E_t V_{t+u} = V_{t-} e^{-(\theta - \lambda_1 \mu_v)u} + e^{-(\theta - \lambda_1 \mu_v)(t+u)} \int_u^{t+u} (\theta k + \lambda_0 \mu_v) e^{(\theta - \lambda_1 \mu_v)s} ds
\]

\[
= V_{t-} e^{-(\theta - \lambda_1 \mu_v)u} \frac{\theta k + \lambda_0 \mu_v}{\theta - \lambda_1 \mu_v} (1 - e^{-(\theta - \lambda_1 \mu_v)u})
\]

It further implies that

\[
E_t[S]_t^T = 1 - e^{-(\theta - \lambda_1 \mu_v)(T-t)} \frac{(V_t - \theta k + \lambda_0 \mu_v)}{\theta - \lambda_1 \mu_v} + \frac{\theta k + \lambda_0 \mu_v}{\theta - \lambda_1 \mu_v} (T-t)
\]

As jumps in returns and volatility are simultaneously,

\[
E_t[S]_t^T = (\lambda_1 E_t[S]_t^T + \lambda_0 (T-t)) \mu_s^2
\]

Therefore

\[
E_t[S]_t^T = (1 + \lambda_1 \mu_s^2) \frac{1 - e^{-(\theta - \lambda_1 \mu_v)(T-t)}}{\theta - \lambda_1 \mu_v} (V_t - \theta k + \lambda_0 \mu_v) + (1 + \lambda_1 \mu_s^2) \frac{\theta k + \lambda_0 \mu_v}{\theta - \lambda_1 \mu_v} (T-t) + \lambda_0 (T-t) \mu_s^2
\]

(4.5)

As the analysis shows, the affine volatility models imply that the conditional expectation of forward quadratic variation is a linear function of the instantaneous variance. By setting the difference between \( t \) and \( T \) to be constant, the parameters in the linear function are fixed and therefore allows for regression analysis.
4.3 VIX

A natural proxy for $E_t [S]_{t+30 \text{ days}}$ is the square of the CBOE VIX index. Often referred to as the fear gauge, the VIX index measures the market’s expectation of S&P500 volatility over the next 30 day period under the Q measure where option pricing is done. The square of VIX hence has the economic interpretation as the fair strike of a variance swap contract that changes a predetermined amount of money for realized variance of S&P500 over the next 30 day period. This fair strike, or the square of VIX, can be tractably determined in an affine jump diffusion model (AJD) that has the structure as in equation (1). Carr and Madan (1998b) show that $E_t [S]_{t+T}$ is equal to the value of a continuum basket of European out-of-the-money options across all the strikes $K$ and maturing at time $t + T$.

\[
E_t [S]_{t+T} = E_t \int_t^{t+T} V_u du + E_t \sum_{N_i}^{N_T} \xi_{t+i}^2
\]

where $r$ denotes the risk free rate, $\epsilon$ denotes the approximation error, and $P_t$ denotes the time-t value of an out-of-the-money option with strike $K$ and time-to-maturity $T$ (a call option when $K$ is greater than put-call-parity implied forward value, and a put option otherwise). $\epsilon$ is zero for the SV model. With the presence of jumps, there arises an approximation error, which is of the cubic order of the jumps in the realized path of asset price. However, as argued in Jiang and Tian (2005) and Carr and Wu (2009b), it is only of a small magnitude.

\[
\epsilon = -2E_t \int_t^{t+T} (e^{\xi_u^s} - 1 - \xi_u^s - \frac{1}{2} (\xi_u^s)^2) dN_u^s
\]

The CBOE VIX results from discretization of $\int_0^{\infty} \frac{2P_t(K,T)}{K^2} dK$ using S&P500 option prices.

\[
\text{VIX}_t(T) = 100 \sqrt{\frac{2}{T} \sum_i \frac{\Delta K_i}{K_i^2} e^{rT} \left( Q(K_i) - \frac{1}{T} \left[ \frac{F}{K_0} - 1 \right]^2 \right)}
\]

$Q(K_i)$ is the midpoint of the bid-ask for an option with strike $K_i$. $\Delta K_i$ is defined as half of the difference between the strikes on either side of $K_i$. At the upper and lower edges of the strip of options, $\Delta K_i$ is simply the difference between the end strike and the
strike next to it. F is the forward index level derived from index option prices via put-call parity, and \( K_0 \) is first strike below F. The term \( \frac{1}{T} \left[ \frac{F}{K_0} - 1 \right]^2 \) represents the adjustment term to convert the in-the-money call with strike \( K_0 \) into an out-of-money put via the put-call parity. \(^1\)

### 4.4 Instantaneous Variance

In a sequence of papers (see, e.g., Barndorff-Nielsen and Shephard (2006)), Barndorff-Nielsen and Shephard define the bipower variation (BPV) process. The realized 1,1-order BPV process is based on sampling the log returns over intervals of time of length \( \Delta^n \):

\[
\{S_{\Delta^n}\}_t^{[1,1]} = \sum_{j=2}^{[t/\Delta^n]} |\Delta S_{j-1}||\Delta S_j|
\]

where \( \Delta S_j = S_{j\Delta^n} - S_{(j-1)\Delta^n} \), and \( [t/\Delta^n] \) is the integer part of \( t/\Delta^n \).

The 1,1-order BPV process, if it exists, is defined as the limit of \( \{S_{\Delta^n}\}_t^{[1,1]} \)

\[
[S]_t^{[1,1]} = \lim_{\Delta_n \to 0} \sum_{j=2}^{[t/\Delta_n]} |\Delta S_{j-1}||\Delta S_j|
\]

Applying the obtained statistical results for BPV to the AJD models, we have \( S_t^{[1,1]} = \mu_1^2 \int_0^t \sigma_s^2 ds = \mu_1^2 [S^c]_t \), where \( \mu_1 = \sqrt{2}/\sqrt{\pi} \). Clearly, \( [S]_t^{[1,1]} \) can be consistently estimated by the realized BPV process.

As the instantaneous variance \( V \) is the instantaneous increment to the \( [S^c]_t \),

\[
V = \lim_{h \to 0} \frac{[S^c]_t - [S^c]_{t-h}}{h}
\]

Naturally an estimator for \( V \) is

\[
\hat{V} = \frac{\{S_{\Delta^n}\}_t^{[1,1]} - \{S_{\Delta^n} \}_t^{[1,1]}}{h^2 \mu_1^2 h_n \Delta_n}
\]

with \( h = h_n \Delta^n \to 0 \) as \( n \to \infty \).

On the other hand, the BPV allows to estimate the jump component in the realized quadratic variation, which is \( \mu_1^{-2} \{S_{\Delta^n}\}_t^{[1,1]} - \{S_{\Delta^n}\}_t \). For testing Poisson-type jumps,

\(^1\)For more details please refer to VIX white paper http://www.cboe.com/micro/vix/vixwhite.pdf
Barndorff-Nielsen and Shephard (2006) show that under the null hypothesis that the process has only a continuous Brownian component,

$$G = \frac{\sqrt{\Delta n}(\mu_1^{-2}\{S_{\Delta n}\}_t^{[1,1]} - \{S_{\Delta n}\}_t)}{\sqrt{\mu_1^{-4}\{S_{\Delta n}\}_t^{[1,1,1,1]}}} \to N(0, v)$$

where $$\{S_{\Delta n}\}_t^{[1,1,1,1]} = \frac{1}{\Delta n} \sum_{j=4}^{[t/\Delta n]} |\Delta S_{j-3}||\Delta S_{j-2}||\Delta S_{j-1}||\Delta S_j| \to \mu_1^4 \int_0^t \sigma_s^4 ds$$ and $$v = \frac{\pi^2}{4} + \pi - 5 \approx 0.6090.$$

For a certain confidence level $$\alpha$$, we have an estimate of the jump component of the realized quadratic variation:

$$J2(t) = I(G_t > \phi_\alpha)(\mu_1^{-2}\{S_{\Delta n}\}_t^{[1,1]} - \{S_{\Delta n}\}_t)$$

where $$I(.)$$ denotes the indicator function and $$\phi_\alpha$$ refers to the appropriate critical value from the standard normal distribution.

### 4.5 Data

The sample period of our empirical exercises ranges from January 2nd, 2004 to September 31st, 2011. To measure the instantaneous variance, we use the tick-by-tick transaction prices for the S&P500 Index futures contracts traded on the Chicago Mercantile Exchange. The earlier data is not taken for (i) an over lengthy sample raises issues of potential non-stationarity and/or regime changes; (ii) trading back in the nineties and early 2000s was not as liquid as now, limiting the ability to estimate the instantaneous variance using high frequency data. We also need the data of the VIX and option-synthesized variance swap rates. The data of the end-of-the-day volatility index VIX are obtained from Chicago Board of Options and Exchange, and the data of SPX options are obtained from Option-Metrics.

The theory of high-frequency econometrics for estimating instantaneous variance is based on the notion of “fill-in” asymptotics which requires increasingly finer sampled high-frequency returns. However, the actual trading frequency is finite and the tick-by-tick prices are contaminated by market microstructure frictions rather than perfect continuous time semi-martingale as claimed by no arbitrage theory. We follow the dominant approach in the literature and convert the tick-by-tick prices to equally spaced five-minute observations. Sampling over five-minute returns is believed to strike a reasonable balance
between the sufficiency of number of observations and the alleviation of microstructure frictions. To have an estimation of the end-of-the-day instantaneous variance, we choose a window that is the day trading time from 8:35am to 15:15pm (Chicago Time). We first calculate the \( \{ S_{\Delta n} \}^{[1,1]} \) over the day trading time and then estimate the instantaneous variance and the part of the daily quadratic variation due to jumps as described in the previous section. As the time unit is one year, the VIX square and the instantaneous variance are expressed in terms of annualized variance percentage. Correspondingly, we also multiply the estimated jump component of the realized daily quadratic variation by 365.

The summary statistics are reported in table 4.1. The average of the VIX square is greater than that of the instantaneous variance, suggesting a positive variance risk premium that has already been widely documented in the literature. The empirical distribution of VIX square and instantaneous variance are highly skewed and leptocurtic, but of different degrees. Their daily series are displayed in figure 4.1. The dynamics of the two series display a close connection. They tend to move in the same direction most of the time, and both spike in the recent financial crisis.

With 1% critical value, within our sample we find that on 20.27% of the days there are jumps. The correlation between realized daily \([S^c]\) and \([S^g]\) is 41.84%. As also seen in figure 4.1, jumps arrive more frequently when volatility is high.
Table 4.1: Summary Statistics. Variables are annualized.

The VIX Square is the risk neutral expectation of the variance over the next 30 day period. The Instantaneous Variance (IV) is measured with high frequency S&P500 futures price and expressed in terms of annualized variance percentage. J2 is the part of the daily realized quadratic variation due to jumps, and is multiplied by 365 for being rendered comparable with VIX and IV.

\[
\begin{array}{cccccc}
\text{Variable} & \text{Mean} & \text{Std} & \text{Median} & \text{Skewness} & \text{Kurtosis} \\
\hline
\text{VIX square} & 0.0552 & 0.0723 & 0.0315 & 3.8260 & 21.3809 \\
\text{IV} & 0.0415 & 0.1056 & 0.0150 & 8.2483 & 94.4601 \\
\text{J2} & 0.0083 & 0.0264 & 0.0033 & 12.9388 & 213.4847 \\
\end{array}
\]

4.6 Test AJDs

This section tests AJD models with the VIX and the measured instantaneous variance. The AJD models imply a one to one linear mapping between the instantaneous variance and the VIX square. This forms the basis of our tests.

4.6.1 Simple OLS

We run a simple regression as the following:

\[
\text{VIX}^2 = \text{constant} + \beta_v \times \text{IV} + \beta_J \times \text{J2} + \epsilon
\]

The results of the OLS regressions are reported in table 4.2, and the scatter plot in figure 4.2 provides a visual illustration. Simply regressing the VIX square on the instantaneous variance results in a R square of 61.5%. A high degree of fit means that the instantaneous variance is capable to explain a large part of the variation of the VIX square. According to the estimated \( \beta_v \) of the instantaneous variance, one percentage increase in real-time instantaneous variance leads up to about 54% increase in the square of the VIX. The fact that \( \beta_v \) is less than 1 suggests that part of the volatility shock is
Figure 4.1: Time series of logarithmic VIX square, logarithmic IV and logarithmic J2

All series keep at modest levels from 2004 to the summer of 2008, then spike up in the fall of 2008, and continue to move with greater volatility after falling back from their peaks.
Adding the estimated jump component in the realized variance does not improve the R square. In fact, the jump effect is not significant. This supports the prediction of the AJD models that the instantaneous variance is the only state variable for the VIX.

<table>
<thead>
<tr>
<th></th>
<th>const</th>
<th>IV</th>
<th>J2</th>
<th>Adj $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.033</td>
<td>0.537</td>
<td></td>
<td>0.615</td>
</tr>
<tr>
<td></td>
<td>0.002790***</td>
<td>0.066641***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.033</td>
<td>0.551</td>
<td>-0.241</td>
<td>0.617</td>
<td></td>
</tr>
<tr>
<td>0.002683***</td>
<td>0.049336**</td>
<td>0.483024</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.2: OLS Regression of the VIX Square on Instantaneous Variance (IV)

IV is the instantaneous variance, J2 is the jump component of the realized variance.*** indicates significance at 5% level, and ** indicates significance at 10% level.
4.6.2 Volatility Persistence

A well known fact of volatility is that volatility exhibits persistence. The first order serial autocorrelation for the variance in the Heston model is $e^{-\theta \Delta t}$. Correspondingly, the square of VIX should exhibit the same autocorrelation as the variance, if their relationship is a one-to-one linear mapping. However, figure 4.3 indicates that the VIX square is more persistent than the instantaneous variance. In table 4.3 we report the statistics of the sample autocorrelation and partial autocorrelation. It is immediately clear that there is substantially less serial correlation in the instantaneous variance. Over a week horizon, the first-order autocorrelation of the VIX square is still high as 89.04%, while that of the instantaneous variance is only 54.30%. On the other hand, the partial autocorrelation of the VIX square seems to decay faster than that of the instantaneous variance.

It is necessary to note that these sample serial correlations for the VIX square and instantaneous variance are for the two different measures $P$ and $Q$, respectively. The differences in their serial correlation mean that the variance risk premium associated with the change of measure from $P$ to $Q$ induces a change in the serial persistence. Certainly this empirical feature is against the AJD models that are based on the Heston type volatil-
Adding lags of the instantaneous variance into the regressors results in better fitting of the linear model. Table 4.4 reports the regression statistics. As the seven-day lag, fourteen-day lag, and twenty-first-day lags are sequentially added into the regressors, the adjusted R-square gradually increases up to more than 82.1%.

### Table 4.3: Autocorrelation in the VIX Square and Instantaneous Variance

<table>
<thead>
<tr>
<th>Lag Length (Days)</th>
<th>VIX Square Partial Autocorrelation</th>
<th>Instantaneous Variance Partial Autocorrelation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.9682</td>
<td>0.7058</td>
</tr>
<tr>
<td>2</td>
<td>0.9480</td>
<td>0.7138</td>
</tr>
<tr>
<td>3</td>
<td>0.9343</td>
<td>0.5902</td>
</tr>
<tr>
<td>4</td>
<td>0.9198</td>
<td>0.6664</td>
</tr>
<tr>
<td>5</td>
<td>0.9158</td>
<td>0.6368</td>
</tr>
<tr>
<td>6</td>
<td>0.9032</td>
<td>0.5601</td>
</tr>
<tr>
<td>7</td>
<td>0.8904</td>
<td>0.5430</td>
</tr>
</tbody>
</table>
Table 4.4: Regressing the VIX square on the instantaneous variance (IV) and the lags

<table>
<thead>
<tr>
<th></th>
<th>IV</th>
<th>IV(-7)</th>
<th>IV(-14)</th>
<th>IV(-21)</th>
<th>Adj $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>const</td>
<td>0.033</td>
<td>0.537</td>
<td></td>
<td></td>
<td>0.615</td>
</tr>
<tr>
<td></td>
<td>0.002790***</td>
<td>0.066641***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.028</td>
<td>0.389</td>
<td>0.271</td>
<td></td>
<td>0.726</td>
</tr>
<tr>
<td></td>
<td>0.002246***</td>
<td>0.053044***</td>
<td>0.045504****</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.025</td>
<td>0.335</td>
<td>0.187</td>
<td>0.209</td>
<td>0.787</td>
</tr>
<tr>
<td></td>
<td>0.001842***</td>
<td>0.042616***</td>
<td>0.033094***</td>
<td>0.039553***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.023</td>
<td>0.315</td>
<td>0.154</td>
<td>0.151</td>
<td>0.157</td>
</tr>
<tr>
<td></td>
<td>0.001543***</td>
<td>0.033506***</td>
<td>0.027763***</td>
<td>0.026932***</td>
<td>0.029900***</td>
</tr>
</tbody>
</table>
4.7 Hedging Volatility

The exactly same methodology of constructing the VIX index can also be applied to obtain the variance swap rate of the S&P500 up to the option maturities. In addition to the regression analysis in the previous section, here we provide an alternative way of evaluating the affine volatility models with the option-synthesized variance swap rates.

Notice that the instantaneous variance is the only state variable for the variance swap rate in the affine volatility models, and that variance swaps of different maturities possess different vega. Hence combining variance swap of different maturities, if properly weighted, can hedge out the volatility risk. The vega of the variance swap rate with time to maturity $\tau$ is \( \frac{1-e^{-\theta\tau}}{\theta\tau} \). The fact that the vega is only dependent on the time to maturity means that a vega neutral portfolio can be made with two variance swaps. A necessary step for knowing the vega is to estimate the mean reversion rate, which can be derived from the OLS coefficient.

However, the values of $\theta$ that are reported in the literature on VIX are far from close. Using prices of the VIX futures, Zhang and Zhu (2006) find the estimated value for the parameter $\theta$ is not stable, depending on the choice of the sample period. In the estimation of the Heston stochastic volatility model, they report different values for $\theta$: 5.8313 for the sample period of January 2nd, 1990 - March 1st, 2005, and 11.7360 for the sample period of January 2nd - December 31st, 2003. In Zhu and Lian (2011), the estimated $\theta$ have values ranging from 1.7640 to 2.2680, depending on whether jumps are included in returns and volatility. With closing quotes on variance swap rates, the estimated $\theta$ in Egloff et al. (2010) is only 0.1547. In a second step, they add a stochastic long-run mean into the volatility process and as a result, the new estimate of mean reversion rate $\theta$ becomes 4.3730. Using the same model, Aït-Sahalia et al. (2012) find $\theta = 3.8628$.

In our benchmark OLS regression, the loading of the VIX square on the instantaneous variance $\beta_v$ is 0.537. Solving the equation

\[
\frac{1 - e^{-\theta T}}{\theta T} = \beta_v \text{ where } T = 30/365
\]

gives $\theta \approx 17.1004$.

Let $W_1$ and $W_2$ denote the portfolio weights for the two variance swaps. $W_1$ and $W_2$
are determined in the way that the weighted vega for the portfolio is zero:

\[ W_1 + W_2 = 1 \]
\[ W_1 \beta(\tau_1) + W_2 \beta(\tau_2) = 0 \]

We have

\[ W_1 = -\frac{1 - e^{-\theta \tau_2}}{\theta \tau_2} / \left( \frac{1 - e^{-\theta \tau_2}}{\theta \tau_1} - \frac{1 - e^{-\theta \tau_2}}{\theta \tau_2} \right) \] (4.9)
\[ W_2 = -\frac{1 - e^{-\theta \tau_1}}{\theta \tau_1} / \left( \frac{1 - e^{-\theta \tau_2}}{\theta \tau_2} - \frac{1 - e^{-\theta \tau_1}}{\theta \tau_1} \right) \] (4.10)

In light of equations 4.2, 4.4, and 4.5, the weighted average of swap rates should be constant. For example, in the case of the most complicated affine volatility models with state dependent jumps,

\[ W_1 E_t[S]^{\tau_1} + W_2 E_t[S]^{\tau_2} = (1 + \lambda_1 \mu_s^2) \frac{\theta k + \lambda_0 \mu_v}{\theta - \lambda_1 \mu_v} + \lambda_0 \mu_v^2 \]

In fact, this is the long run average of the variance under the risk neutral measure. The same conclusion also applies for the Heston model and the Heston model with random jumps.

We form vega neutral portfolios for the set of various values of \( \theta \) that have been reported in the literature. The portfolio invests on the blend of the option-synthesized variance swaps with the two nearest maturities beyond two weeks. If the weighted average of the swap rates really represents the long run variance mean, the portfolio should possess a constant value. It starts with 1 dollar from March 26th, 2004, rebalanced on a daily basis, and is closed when the portfolio value falls below 0.001 dollar.

The P&L of the strategy for different values of \( \theta \) are shown in figure 4.4 and figure 4.5. For \( \theta \) greater than 3, the portfolio is almost been extinguished by January 2006. For smaller theta, the portfolio is able to continue longer at a low level. The failure of the hedged portfolio to survive indicates the failure of the affine volatility model.

The failure of the affine models is perhaps reflected in an even deeper sense in the fact that the weighted average of variance swap rates is time varying. We expect the weighted average to manifest the mean of the variance under the risk neutral measure. In contrast,
For a given value of $\theta$, we roll a combination of two option-synthesized variance swaps with time-to-maturity over 14 days in a way that the portfolio vega is zero as implied by the affine models. The blue line (P&L) represents the Profit&Loss of the trading strategy, and the green line (IV) represents the estimated instantaneous variance.

the figure 4.6 shows that it is time varying and seems to correlate with the instantaneous variance. Particularly the portfolio value is very sensitive to the sharp volatility rise. A possible reason is that the jump risk in volatility is not completely hedged, which would imply the jump risk in volatility is not correctly modeled by the affine class.
Figure 4.5: Profit&Loss of the Hedging Strategy (Continued)

For a given value of $\theta$, we roll a combination of two option-synthesized variance swaps with time-to-maturity over 14 days in a way that the portfolio vega is zero as implied by the affine models. The blue line (P&L) represents the Profit&Loss of the trading strategy, and the green line (IV) represents the estimated instantaneous variance.
Figure 4.6: The Weighted Average of Variance Swap Rates for $\theta = 17.1004$

The weighted average of swap rates, according to the affine volatility models, should have a constant value that represents the long run mean of the variance in the risk neutral measure. However, as demonstrated by the blue line, it is time varying and appears to correlate with the instantaneous variance.
4.8 Conclusion

Affine volatility models are popular candidates in the literature for pricing variance swaps. We evaluate them with the volatility index VIX and variance swap rates synthesized by S&P500 options. Two pieces of empirical evidence are found against the affine models. First, these models imply a linear one-to-one mapping between the VIX square and instantaneous variance, while we find that the perfect linear relationship is broken off by volatility persistence. Volatility lags help explain the variation of the VIX square. Second, we form a vega neutral portfolio whose value is supposed to be constant and represent the long run average of the variance under the risk neutral measure. However, the hedging of volatility risk according to the affine models is not successful. Instead of showing the constant value of long run variance mean, the portfolio’s P&L reflects a time varying mean that tends to vary with the instantaneous variance. These findings indicate the inefficiency of using affine volatility models to price and hedge variance swaps.
Chapter 5

Conclusion

At the closing part of the dissertation are concluding summaries. Chapter 2 studies an asset pricing model that can simultaneously explain the three important features in the asset price data, namely, the high equity premium, the low risk free rate, and the positive slope of the term structure of the real interest rates at average. Many important asset pricing models can explain one or two of the features, but not all three. The model considers a hidden Markov economy in which the dynamics of the fundamental economic structure is subject to the regime of the economy. The regime that is invisible to investors switches among a finite number of states, and one of the states represents the great depression. By considering robust investors who are wary of the rational expectation estimation of the underlying state, and alternatively assign a greater probability weight to the depression state, the model generates a higher risk premium than in the standard rational expectations models. As robust investors are more pessimistic of the outlook of the economic growth in the next period, the short term real interest rate is low. However, in the long run the state of the economy is still decided by the stationary distribution of the fundamental state, rather than the investors’ belief of the current state. As a result, the long term real rate is much less affected by the robust belief than the short term rate. Therefore the model also produces a positive real term structure.

Chapter 3 discovers a pricing anomaly between S&P500 options and the VIX futures. On the one hand, by combining S&P500 option prices at different maturities and strike prices, one can infer the expected variance in the risk neutral measure up to the option maturities. On the other hand, prices of futures contracts written on the volatility index VIX measures the expected volatility in the risk neutral measure over the thirty day horizon after the futures maturities. If options and futures markets are integrated, the same
pricing kernel should apply to both, and the option-implied variance and futures-implied volatility can be combined to give the risk-neutral variance of the VIX. But taking them together often implies a negative measurement of the variance of the VIX, a mathematical impossibility. The recurrent inconsistencies are not simply due to possible measurement errors, because simple trading strategies designed to exploit these inconsistencies produce very high Sharpe ratios. The analysis suggests an anomaly between the variances implied from options and volatilities implied from VIX futures.

Chapter 4 examines the affine stochastic volatility models with the volatility index VIX and variance swap rates synthesized by S&P500 options. In affine volatility models, the expected variance implied by the VIX square is linear with the instantaneous variance. Based on this implication we carry out two exercises. First, we regress the VIX square on the instantaneous variance and find that there is a quasi linear relationship between the VIX square and the instantaneous variance, and that the perfect linear relationship is ruined by volatility persistence. Second, we form a vega neutral portfolio that should have a constant value representing the long run average of the variance under the risk neutral measure. However, the volatility hedging by the affine models is not successful and the portfolio is still subject to volatility risk. These findings indicate the inefficiency of using affine volatility models to price and hedge variance swaps.

This dissertation opens up a few potential topics for future research. Compared to previous studies on aggregate equity returns whose focus is on matching empirical moments, the model in chapter 2 is able to produce a more plausible description of the historical time series of equity prices. However, the model time series generated by robust preference is more volatile than the historical counterpart. This feature suggests to search for a smoothing modification of robustness preferences. One possibility is to have a dynamic structure of forming robustness that smooths the divergence from rational expectations in respect to large shocks. It can be some kind of stickiness in the adjustment of the agent’s formulation of robustness. Another observation as a byproduct in chapter 2 is that corporate earnings do not always move in the same direction as market returns. This suggests that either we should look for alternative proxies of aggregate cash flows, or that we should add more state variables that matter for pricing and can change differently than earnings which, however, introduces more complications.

The empirical study of volatility derivatives can be extended with the results that
have already been obtained. For example, it would be interesting to compare the implied volatility of the VIX with the Black-Scholes volatility of VIX options, as both of them represent the volatility of the VIX under the risk neutral measure. Possibly we can search for a certain pattern in their relationship, and try to build volatility forecast models. Another follow-up research that can be attempted is to link the implied volatility of the VIX to the modeling of realized variance, as the implied volatilities are forward looking and therefore informative.
Samenvatting (Summary in Dutch)

Als afsluiting volgt een bekopte samenvatting van het proefschrift.

Hoofdstuk 2 onderzoekt een asset pricing model dat summiertaan de drie belangrijkste eigenschappen van “asset pricing data” kan verklaren, namelijk, de hoge risicopremie, de lage risicovrije rente en de positieve helling van de termijnstructuur van de reële rente, gemiddeld gezien. Belangrijke asset pricing modellen kunnen meestal n of twee van deze belangrijke eigenschappen verklaren, maar niet alle drie de eigenschappen. Het model beschouwt een verborgen Markov economie waarin de dienamiek van de fundamentele economische structuur onderhevig is aan het regiem van de economie. Het regiem is niet waarnembaar voor de investeerders en kan zich in een eindig aantal verschillende toestanden bevinden, waarbij n van de toestanden de Grote Depressie is. Door robuuste investeerders in beschouwing te nemen, die kritisch zijn tegen rationele verwachtingen schatten voor de onderliggende toestanden, en hierdoor een grotere kansmaat toekennen aan de Grote Depressie, genereert het model een hoger risico premie dan in de standaard rationele verwachtingen modellen. Omdat robuuste investeerders meer pessimistisch zijn over de vooruitzichten van de economische groei in de volgende periode, is de korte termijn reële interestvoet laag. Hoewel, op lange termijn wordt de toestand van de economie nog seeds bepaald door de stationaire verdeling van de fundamentele economische toestand, en niet door de huidige toestands overtuiging van de investeerder. Het resultaat hiervan is dat de lange termijn reële rente in vele mindere mate wordt beïnvloed door de overtuiging van de robuuste investeerders dan de korte termijn rente. Daardoor produceert het model ook een positieve reële termijn structuur.

In hoofdstuk 3 wordt een ontdekte anomalie in het prijzen van S&P500 opties and VIX futures beschreven. Enerzijds, door het combineren van S&P500 optieprijzen voor tijden tot expiraties en uitoefenprijzen, kan men de verwachte variantie afleiden tot de expiratie van de optie, berekend in de risico neutrale waardering. Anderzijds, prijzen
van futures contracten, geschreven op de volatiliteit index VIX, zijn een maat voor de verwachte volatiliteit, berekend in de risico neutrale waardering met een tijdshorizon van 30 dagen nadat de futures zijn gexpireerd. Als de optie- en de futures- markt volledig geintegreerd zijn, zou dezelfde stochastische discount factor bij beide markten moeten horen. De impliciete variantie van de optie en de impliciete volatiliteit van de futures kan dan gecombineerd worden om de risico neutrale variantie van de VIX af te leiden. Dit leidt echter vaak tot een negatieve variantie van de VIX, wat matematisch gezien onmogelijk is. Deze terugkerende inconsistenties ontstaan niet simpelweg door mogelijke benaderingsfouten, omdat simpele handelsstrategieën, ontwikkeld om te profiteren inconsistenties, zeer hoge Sharpe ratios produceren. De analyse suggereert dat er een anomalie tussen de impliciete varianties van de opties en volatiliteiten gempliceerd door de VIX futures is.

Hoofdstuk 4 onderzoekt affiene stochastische volatiliteit modellen op de volatiliteit index VIX en op variantie swap rentes verkregen door S&P 500 opties. In affiene volatiliteit modellen, is de verwachte variantie gempliceerd door de VIX lineair met de geschatte variantie van de S&P 500. Gebaseerd op deze implicatie, verrichten wij twee onderzoeken. Als eerste regresseren we de VIX square op de variantie van de S&P 500 en vinden we een quasi lineaire relatie tussen deVIX square en de variantie S&P 500. Ook zien we dat de perfecte lineaire relatie wordt verstoord door volatiliteit persistentie. Als tweede, vormen we een vega-neutrale portfolio, die dan over tijd genomen een constante waarde zou moeten hebben, die voor de gemiddelde variantie onder de risico neutrale maatstaf staat. Echter, het hedgen van volatiliteit door affiene modellen is niet succesvol en de portfolio is nog steeds blootgesteld aan volatiliteit risico. Deze bevindingen wijzen op de inefficintie van het gebruik van affiene volatiliteit modellen voor het prijzen en hedgen van variantie swaps.

Dit proefschrift opent een paar potentiele onderwerpen voor toekomstig onderzoek. In vergelijking tot eerdere studies naar geaggregeerde rendementen van aandelen, waarbij de focus ligt op het matchen van de empirische bewegingen van aandeelprijzen, is het model in hoofdstuk 2 in staat om een meer plausibele beschrijving te geven van de historische tijdreeks van aandeelprijzen. Echter, de model tijdreeksen die gegenereerd zijn door robuuste preferenties zijn meer volatiel dan de tegenhangende historische tijdreek- sen. Dit kenmerk suggereert dat men op zoek moet naar een gesmoothe modificatie van de robuuste preferenties. Een mogelijkheid is om een dynamische structuur voor het vormen van robuuste preferenties te hebben, die de divergentie van rationale verwachtingen
in verhouding tot grote shocken smooth. Dit zou een een of andere standvastigheid in de aanpassing van de agent zijn formulatie van robustheid kunnen zijn. Een andere observatie als bijproduct in hoofdstuk 2 is dat de bedrijfsinkomsten niet altijd in dezelfde richting lopen als de marktrendementen. Dit suggereert dat we ofwel naar alternatieve benaderingen voor het aggregeren van geldstromen zouden moeten kijken of dat we meer toestand variabelen zouden moeten toevoegen die van belang zijn voor het prijzen, wat echter meerder compliacties met zich meebrengt.

De empirische studie van volatiliteit derivaten kan worden uitgebreid met de resultaten die al zijn verkregen in dit proefschrift. Het zou bijvoorbeeld interessant zijn om de gempliceerde volatiliteit van de VIX te vergelijken met de gempliceerde Black-Scholes volatiliteit van VIX opties, omdat beiden de volatiliteit van de VIX onder de risico neutrale maatstaf aangeven. We zouden mogelijk op zoek kunnen gaan naar een bepaald patroon in hun relatie, en hiermee een model opbouwen om de volatiliteit te voorspellen. Een ander vervolgonderzoek dat kan worden uitgevoerd is om de gempliceerde volatiliteit van de VIX te linken aan de gerealiseerde variantie, omdat de gempliceerde varianties vooruit kijkend zijn en daarom informatief zijn.
Bibliography


The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and VU University Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

532. T. WILLEMS, Essays on Optimal Experimentation
533. Z. GAO, Essays on Empirical Likelihood in Economics
534. J. SWART, Natural Resources and the Environment: Implications for Economic Development and International Relations
535. A. KOTHIYAL, Subjective Probability and Ambiguity
536. B. VOOGT, Essays on Consumer Search and Dynamic Committees
537. T. DE HAAN, Strategic Communication: Theory and Experiment
538. T. BUSER, Essays in Behavioural Economics
539. J.A. ROSERO MONCAYO, On the importance of families and public policies for child development outcomes
540. E. ERDOGAN CIFTCI, Health Perceptions and Labor Force Participation of Older Workers
541. T.WANG, Essays on Empirical Market Microstructure
542. T. BAO, Experiments on Heterogeneous Expectations and Switching Behavior
543. S.D. LANSDORP, On Risks and Opportunities in Financial Markets
544. N. MOES, Cooperative decision making in river water allocation problems
545. P. STAKENAS, Fractional integration and cointegration in financial time series
546. M. SCHARTH, Essays on Monte Carlo Methods for State Space Models
547. J. ZENHORST, Macroeconomic Perspectives on the Equity Premium Puzzle
548. B. PELLOUX, the Role of Emotions and Social Ties in Public On Good Games: Behavioral and Neuroeconomic Studies
549. N. YANG, Markov-Perfect Industry Dynamics: Theory, Computation, and Applications
550. R.R. VAN VELDHUIZEN, Essays in Experimental Economics
551. X. ZHANG, Modeling Time Variation in Systemic Risk
552. H.R.A. KOSTER, The internal structure of cities: the economics of agglomeration, amenities and accessibility
553. S.P.T. GROOT, Agglomeration, globalization and regional labor markets: micro
evidence for the Netherlands

554. J.L. MHLMANN, Globalization and Productivity Micro-Evidence on Heterogeneous Firms, Workers and Products

555. S.M. HOOGENDOORN, Diversity and Team Performance: A Series of Field Experiments

556. C.L. BEHRENS, Product differentiation in aviation passenger markets: The impact of demand heterogeneity on competition

557. G. SMRKOLJ, Dynamic Models of Research and Development

558. S. PEER, The economics of trip scheduling, travel time variability and traffic information

559. V. SPINU, Nonadditive Beliefs: From Measurement to Extensions

560. S.P. KASTORYANO, Essays in Applied Dynamic Microeconometrics

561. M. VAN DUIJN, Location, choice, cultural heritage and house prices

562. T. SALIMANS, Essays in Likelihood-Based Computational Econometrics

563. P. SUN, Tail Risk of Equidity Returns

564. C.G.J. KARSTEN, The Law and Finance of M&A Contracts

565. C. OZGEN, Impacts of Immigration and Cultural Diversity on Innovation and Economic Growth

566. R.S. SCHOLTE, The interplay between early-life conditions, major events and health later in life

567. B.N. KRAMER, Why don’t they take a card? Essays on the demand for micro health insurance

568. M. KILI, Fundamental Insights in Power Futures Prices

569. A.G.B. DE VRIES, Venture Capital: Relations with the Economy and Intellectual Property

570. E.M.F. VAN DEN BROEK, Keeping up Appearances


572. F.T. ZOUTMAN, A Symphony of Redistributive Instruments

573. M.J. GERRITSE, Policy Competition and the Spatial Economy

574. A. OPSCHOOR, Understanding Financial Market Volatility

575. R.R. VAN LOON, Tourism and the Economic Valuation of Cultural Heritage

576. I.L. LYUBIMOV, Essays on Political Economy and Economic Development

577. A.A.F. GERRITSEN, Essays in Optimal Government Policy


579. E. RAVIV, Forecasting Financial and Macroeconomic Variables: Shrinkage, Dimen-
580. J. TICHEM, Altruism, Conformism, and Incentives in the Workplace

581. E.S. HENDRIKS, Essays in Law and Economics

582. X. SHEN, Essays on Empirical Asset Pricing