Essays on pricing kernel estimation, option data filtering and risk-neutral density tail estimation

DISSERTATION
of the University of St. Gallen,
School of Management,
Economics, Law, Social Sciences
and International Affairs
to obtain the title of
Doctor of Philosophy in Economics and Finance

submitted by

Pirmin Meier

from

Willisau (Luzern)

Approved on the application of

Prof. Dr. Francesco Audrino

and

Prof. Dr. Fabio Trojani

Dissertation no. 4390

Difo-Druck GmbH, Bamberg 2015

The University of St. Gallen, School of Management, Economics, Law, Social Sciences and International Affairs hereby consents to the printing of the present dissertation, without hereby expressing any opinion on the views herein expressed.

St. Gallen, December 3, 2014

The President:

Prof. Dr. Thomas Bieger

Acknowledgements

I would like to thank my supervisors Prof. Dr. Francesco Audrino and Prof. Dr. Enrico De Giorgi that they gave me the chance to pursue a PhD under their supervision in a very inspiring research environment. It has been a pleasure working with them and I am particularly thankful that they always took time to answer my questions and to provide helpful inputs for my research.

I would also like to thank Prof. Dr. Matthias Fengler and Prof. Dr. Daniel Buncic for helpful discussions and I am grateful that Prof. Dr. Fabio Trojani accepted to be my external referee.

Moreover, I am thankful for the excellent time I had in St. Gallen with all my fellow PhD students. Many of them became good friends. In particular, I would like to thank Anselm, Kameliya, Simon, Katja, Constantin, Dewi and Jacopo for the interesting and helpful discussions regarding research and beyond, the successful collaboration and all the happy moments that we have spent together at the university and outside it during the last few years.

I do not forget all my other friends (especially Severin and Raphi) and would like to thank them for their support.

Finally, I am enormously grateful to my family and in particular my parents for the constant and unconditional support during my years of study. I owe a lot to you.

December, 2014 Pirmin Meier

Contents

Summary **viii**

Zusammenfassung **ix**

Preface **1**

1 Empirical pricing kernel estimation using a functional gradient descent algorithm based on splines **4**

1.1 Introduction 6

1.2 The role of the pricing kernel in asset pricing theory and its puzzling behavior 8

1.3 Empirical pricing kernel estimation 10

1.4 Empirical analysis 19

1.5 Conclusion 28

2 Modeling the pricing kernel surface dynamics with a boosting approach based on regression trees **30**

2.1 Introduction 32

2.2 The role of the pricing kernel in option pricing 34

2.3 Initial pricing kernel surface estimation 35

2.4 Modeling the time-varying pricing kernel surface 40

2.5 Empirical analysis . . . 45

2.6 Conclusion . . . 61

2.A A short introduction to FGD . . . 62

2.B Additional plots . . . 63

3 Estimating risk-neutral density tails: a comparison 66

3.1 Introduction . . . 68

3.2 Option prices and risk-neutral densities . . . 70

3.3 Methods to estimate the risk-neutral density tails . . . 71

3.4 Experimental Setting . . . 80

3.5 Results . . . 85

3.6 Robustness . . . 95

3.7 Conclusion . . . 100

3.A Univariate extreme value theory . . . 101

3.B Characteristic function of the Bates model . . . 104

3.C Additional plots . . . 107

4 On working with option data: implications of data choice 111

4.1 Introduction . . . 113

4.2 Terminology and data description . . . 114

4.3 Option data filters . . . 116

4.4 Comparison and implication in practice . . . 126

4.5 Conclusion . . . 135

4.A Additional figures . . . 135

4.B Additional tables . . . 140

Bibliography **140**

Curriculum Vitae **148**

Summary

The first chapter introduces a new method to estimate the option-implied empirical pricing kernel. Departing from an adequate and economically motivated initial pricing kernel, we apply a customized functional gradient descent (FGD) algorithm based on B-splines. We empirically illustrate the estimation properties of the method using S&P 500 option data and find that the algorithm yields accurate estimates. In addition, we provide evidence of the superior predictive ability of our method in comparison with other approaches recently introduced in the empirical pricing kernel literature.

The empirical pricing kernel is time-varying, which reflects changes in the relevant pricing kernel state variables. However, little is known about these factors driving the pricing kernel over time. Therefore, I present in the second chapter a time series model for the evolution of the empirical pricing kernel using a boosting approach based on regression trees. Given that trees have the ability to choose among a set of predictors the most relevant ones, I develop a tool to answer the main question regarding the principal pricing kernel driving factors. I show in the empirical part of the chapter that the influence of relevant driving factors such as volatility, financial and macroeconomic variables and sentiment measures substantially varies over time and I provide some insights on how they affect the pricing kernel shape.

In the third chapter, we investigate how to estimate accurately the tails of the option-implied risk-neutral density, which is known to be a challenging problem. We review several methods and additionally introduce a new tail extension approach that combines the idea of price matching with the assumption of tails drawn from a generalized extreme value distribution. Based on a theoretical market model with known implied risk-neutral density, we conduct a performance analysis. We find that the best results are obtained either with implied volatility extrapolation by means of Gatheral's stochastic volatility inspired SVI model, or with our new price matching method.

The last chapter of this thesis is about option data filtering. We discuss why filtering (especially for the tails) might be sensible and present the most common approaches. In addition, and more importantly, we introduce new tail filtering methods that rely on proxies for market activity, which we derive from intraday option data. In the second part of the chapter, we demonstrate that different filters yield to statistically different strike/moneyness ranges and show that these differences can have a non-negligible impact on tail sensitive applications such as moment estimation.

Zusammenfassung

Im ersten Kapitel wird eine neue Methode präsentiert, um den aus Optionspreisen implizierten empirischen Preiskern zu schätzen. Die Methode funktioniert so, dass wir mit einem geeigneten und ökonomisch motivierten Anfangspreiskern starten und dann einen “functional gradient descent “ (FGD) Algorithmus basierend auf B-Splines verwenden, um einen guten Schätzer zu erhalten. Im empirischen Teil des Kapitels benutzen wir S&P 500 Optionsdaten, um die Eigenschaften des Schätzers zu beschreiben. Wir zeigen, dass unsere Methode akkurate Schätzungen liefert und stellen fest, dass die Vorhersageresultate besser sind als mit anderen, vergleichbaren Schätzmethoden.

Der Preiskern ist eine zeitabhängige Grösse, was auf Veränderungen in den zugrunde liegenden Zustandsvariablen zurückzuführen ist. Leider ist wenig bekannt über die Triebfaktoren dieser zeitlichen Dynamik. Daher präsentiere ich im zweiten Kapitel ein Zeitreihenmodell, basierend auf einem Boosting Algorithmus mit Regressionsbäumen, um die zeitliche Entwicklung des Preiskerns zu verstehen. Da Regressionsbäume es erlauben, aus einer Menge von potentiellen Zustandsvariablen die Relevanten auszuwählen, kann unsere Kernfrage bezüglich der Triebkräfte beantwortet werden. Im empirischen Teil zeige ich, dass der Einfluss der Zustandsvariablen stark variiert und ich untersuche, wie die relevanten Variablen die Form des Preiskerns beeinflussen.

Im dritten Kapitel wird analysiert, wie man die Schwänze der aus Optionspreisen implizierten risikoneutralen Dichte möglichst akkurat schätzen kann. Wir begutachten einige kürzlich eingeführte Methoden und entwickeln eine eigene, indem wir die Idee der Preisanpassung mit der Annahme einer generalisierten Extremwertverteilung kombinieren. In unserer Vergleichsstudie, welche auf einem theoretischen Marktmodell mit bekannter risikoneutraler Dichte basiert, finden wir heraus, dass die besten Resultate entweder durch Extrapolation der impliziten Volatilität mit Hilfe von Gatheral’s SVI Modell oder mit unserer neuen Methode erzielt werden.

Das letzte Kapitel ist dem Thema Filtern von Optionsdaten gewidmet. Wir erläutern, wieso das Filtern von Optionsdaten (insbesondere für extreme Strike-Preise) sinnvoll ist und präsentieren die gängigsten Methoden. Zudem entwickeln wir einen neuen Ansatz, gestützt auf Angaben zur Marktaktivität, welche wir aus intraday Optionsdaten erhalten. Wir stellen fest, dass verschiedene Filter zu statistisch signifikant unterschiedlichen Strike/Moneyness Intervallen führen und zeigen, dass diese Unterschiede in der Momentenschätzung einen entscheidenden Einfluss haben.

Preface

This cumulative thesis contains four chapters/papers. The first two are concerned with estimating and modeling the empirical pricing kernel and have been developed within a SNF project entitled “Applying Recent Developments in Computational Statistics to Behavioral Asset Pricing and Portfolio Selection”. Chapters 3 and 4 are about risk-neutral density tail estimation and option data filtering and are both a joint work with Anselm Ivanovas. Hereafter, I provide a short summary of all chapters and show how they are related.

In financial asset pricing, the pricing kernel or stochastic discount factor has a pivotal role, since it summarizes preferences of an investor about risk. Understanding its behavior is therefore particularly important and, consequently, numerous approaches have been developed in the last decade to infer this pricing kernel from observed asset prices, especially options.

In the first chapter, Francesco Audrino and I propose a new method to estimate the option-implied empirical pricing kernel and thus provide a methodological contribution to the existing literature. Departing from an adequate and economically motivated power pricing kernel, we apply a customized functional gradient descent (FGD) algorithm based on B-splines. This approach allows us to locally modify the initial pricing kernel by means of an additive expansion of some relevant B-Spline basis functions and therefore produces an improved final estimate. We empirically illustrate the estimation properties of the method using S&P 500 option data and find that the algorithm yields accurate estimates. In addition, we provide evidence of the superior predictive ability of our method in comparison with other approaches recently introduced in the empirical pricing kernel literature.

In line with other studies, we observe that the empirical pricing kernel is time-varying, which reflects changes in the relevant pricing kernel state variables. Surprisingly, only a few attempts have been made in the recent past to model and understand the observed dynamics.

Therefore, I present in the second chapter a time series model for the evolution of the empirical pricing kernel surface. More specifically, building on some initial estimates of these dynamic surface, I propose a boosting approach based on regression trees. Given that trees have the ability to choose among a set of predictors the most relevant ones, I develop a tool to answer the main question regarding the principal pricing kernel driving factors. In the empirical analysis, I show that the influence of relevant driving factors such as volatility, financial and macroeconomic variables and sentiment measures substantially varies over time. Moreover, I provide some insights on how they affect the pricing kernel shape.

The option-implied risk-neutral density represents a forward-looking view of the underlying asset price distribution. Hence, it cannot only be employed to derive an empirical pricing kernel, but is helpful for various other applications like managing risk or asset allocation. Numerous flexible nonparametric and semiparametric approaches have been developed to recover this density from option prices. However, an important issue often exists when applying such methods. Namely, the risk-neutral density can only be obtained over the strike range of observable prices, which might not extend very far into the tails. Interestingly, there is only a limited number of solutions for that problem even though the tails are particularly important for risk management applications or to estimate higher moments of the risk-neutral distribution.

That's why I investigate together with Anselm Ivanovas in the third chapter how to estimate accurately the risk-neutral density tails. To this end, we review several existing methods to infer the tails from option data and additionally introduce a new tail extension approach which combines the idea of price matching with the assumption of tails drawn from a generalized extreme value distribution. Based on a theoretical market model with known implied risk-neutral density, we conduct a performance analysis. The final outcomes of our study indicate that the best results regarding accuracy and stability are obtained either with implied volatility extrapolation by means of Gatheral's stochastic volatility inspired SVI model, or with our new price matching method. Moreover, we provide evidence that the still very popular constant implied volatility extrapolation approach often yields unsatisfactory results.

As can be seen for example in chapters one to three, option prices are helpful for various applications other than option pricing itself, given that they contain information on the risk factors driving the underlying price process. However, the quantity of option prices available is typically huge, which calls for data selection/filtering. Interestingly, this step of data preparation is often marginalized in the empirical literature.

In the last chapter of this thesis, Anselm Ivanovas and I contribute to this simple but important topic in empirical research. Departing from the full set of intraday S&P 500 option data, we discuss why filtering (especially for the tails) might be sensible and present the most common approaches. In addition, and more importantly, we introduce new tail filtering methods that rely on proxies for market activity, which we derive from intraday option data. In the second part of the chapter, we analyze the effect of different filters. We provide evidence that they yield to statistically different strike/moneyness ranges and show that these differences can have a non-negligible impact on tail sensitive applications such as moment estimation.

Chapter 1

Empirical pricing kernel estimation using a functional gradient descent algorithm based on splines

Francesco Audrino
Pirmin Meier

Abstract

We propose a new methodology to estimate the empirical pricing kernel implied from option data. In contrast to most of the studies in the literature that use an indirect approach, i.e. first estimating the physical and risk-neutral densities and obtaining the pricing kernel in a second step, we follow a direct approach. Departing from an adequate parametric and economically motivated pricing kernel, we apply a functional gradient descent (FGD) algorithm based on B-splines. This approach allows us to locally modify the initial pricing kernel and hence to improve the final estimate. We empirically illustrate the estimation properties of the method and test its predictive power on S&P 500 option data, comparing it as well with other recent approaches introduced in the empirical pricing kernel literature.

Keywords: Empirical pricing kernel, function gradient descent, B-splines, option pricing.

1.1 Introduction

The pricing kernel or stochastic discount factor is a key component of any asset pricing model. It summarizes investor preferences for payoffs over different states of the world and represents an important link between economics and finance. Given its high information content, it is not surprising that several attempts have been undertaken in the past to infer such a kernel from observed (option) market prices.

Seminal papers in the empirical pricing kernel estimation literature include Jackwerth (2000) and Aït-Sahalia and Lo (2000). Both of them adopt a so-called indirect approach, i.e. they first estimate the physical and risk-neutral densities and then obtain the pricing kernel in a second step. Alternatively, it is also possible to estimate the pricing kernel directly using the fundamental asset pricing equation (see Rosenberg and Engle (2002)).

According to classical finance models, one would expect to find a fitted pricing kernel that is a decreasing function of aggregate resources. However, this is contrary to many recent empirical studies (including those mentioned above). The empirical pricing kernel does not seem to be a monotonically declining function, but exhibits instead an upward-sloping region. This phenomenon is known as the pricing kernel puzzle.

In the last ten years, several other estimation methodologies have been proposed. A majority of them rely on the indirect approach: see, among others, Barone-Adesi et al. (2008), Grith et al. (2013a), Detlefsen et al. (2010) and Barone-Adesi and Dallo (2010). In addition, Yang (2009) and Grith et al. (2011) also present some modified versions of the direct method originally introduced in Rosenberg and Engle (2002). The studies have produced mixed results. A large number of them confirm the pricing kernel puzzle. However, Barone-Adesi et al. (2008) and Barone-Adesi and Dallo (2010) find the overall shape of their estimates to be generally decreasing. Furthermore, there are also recent contributions supporting U-shaped pricing kernels (e.g., Bakshi et al. (2010) and Christoffersen et al. (2013)).

Besides the introduction of the above-mentioned estimation techniques, some researchers have performed formal tests to verify the monotonicity of the pricing kernel (see for example Golubev et al. (2011), Härdle et al. (2010) and Beare and Schmidt (2011)). They provide evidence in their empirical applications that the null-hypothesis of non-increasing kernels can typically be rejected. Consequently, several attempts have been made to explain the puzzle. Chabi-Yo et al. (2008) and Grith et al. (2013b) for example consider state-dependent preferences as possible explanations, whereas others such as Shefrin (2008), Ziegler (2007), De Giorgi and Post (2008), Polkovnichenko and Zhao (2013) and Hens and Reichlin (2012) focus on results coming from the behavioral finance literature to solve the puzzle.

In this paper, we propose a new direct estimation methodology. Departing from an adequate and economically motivated power pricing kernel, we apply a customized functional gradient descent (FGD) algorithm based on B-splines. This approach allows us to locally modify the initial pricing kernel by means of an additive expansion of some relevant B-Spline basis functions and therefore produces an improved final estimate. The FGD algorithm (Friedman (2001)) belongs to the class of boosting procedures, which are very popular in the area of machine learning. It can be interpreted as a functional analog of the gradient method applied for parameter optimization. However, our algorithm has some peculiarities not present in the generic FGD procedure. It depends on a set of simulated future returns and must be combined with numerical integration or Monte Carlo methods to compute an expectation, which makes our approach particularly challenging. Recently, FGD or slightly modified versions of it have already been successfully applied to financial function estimation. In particular, Audrino and Bühlmann (2009) have shown that FGD in connection with B-splines yields good results in volatility estimation and forecasting.

Our study contributes to the existing literature along two different lines. First, we bring the idea of boosting into the field of pricing kernel estimation by suggesting a new and rather flexible direct fitting approach that is able to provide accurate estimates. Second, in contrast to almost all studies presented in the literature so far, we also investigate the predictive power of our pricing kernel estimates. Having such accurate forecasts is interesting. They contain helpful information regarding investors' future beliefs and risk behavior and can be used for example to improve option valuation or the performance of option trading/hedging strategies.

In our empirical analysis, we consider S&P 500 option data from 2005 until 2010 to empirically illustrate the estimation properties of our algorithm. Although departing from an initial pricing kernel conforms to classical economic theory, we often observe final estimates showing the puzzling behavior. In agreement with several previous studies, we find that the increasing component is usually located in the area of zero return and resembles a bump. However, the fitted kernel is time-varying and we also get estimates conforming to some other recent contributions supporting the claim that the pricing kernel looks U-shaped. In order to evaluate the accuracy of the fitted kernels, we investigate the in-sample and out-of-sample pricing performance of our methodology. Interestingly, we find that the FGD algorithm based on splines consistently outperforms the parametric specifications introduced in Rosenberg and Engle (2002), which are considered benchmark approaches.

The paper is organized as follows. Section 1.2 contains a review of basic asset pricing theory (mainly to set up notations) and formally introduces the pricing kernel puzzle. Section 1.3 provides a detailed description of our new estimation method using a FGD algorithm based on B-splines. Section 1.4 presents the empirical results and Section 1.5 concludes the paper.

1.2 The role of the pricing kernel in asset pricing theory and its puzzling behavior

1.2.1 Review of some asset pricing theory

Given some general non-arbitrage conditions (see Hansen and Richard (1987)), the time t price $\pi_t(X_{t+1})$ of an asset with payoff $X_{t+1} \in \mathcal{X}_{t+1}$ (the set of payoffs at time $t+1$) can be written as

$$\pi_t(X_{t+1}) = E_t[M_{t,t+1}X_{t+1}], \tag{1.1}$$

where $E_t[\cdot]$ denotes the conditional expectation given investors' information J_t at time t and $M_{t,t+1}$ is the one-period stochastic discount factor (SDF) or pricing kernel. Hence, the price of an asset equals the expected pricing-kernel weighted payoff.

The SDF is a state-dependent function that discounts payoffs using time and risk preferences. Generally, it can depend on many (possibly unknown) state variables. Since there is a considerable debate among researchers about the state variables that enter into the pricing kernel, it is quite common to consider a projected pricing kernel. More precisely, among the admissible SDFs in (1.1), there exists only one that is a function of available payoffs. It is the orthogonal projection of any admissible SDF on the set of payoffs. If we now consider as payoff space $\mathcal{X}_{t+1}$ the set of all squared integrable functions $h(J_t, X_{t+1})$ of some primitive payoff X_{t+1}, we obtain the projected pricing kernel as

$$M^*_{t,t+1} = E_t[M_{t,t+1}|X_{t+1}]. \tag{1.2}$$

Although this projected pricing kernel is not necessarily identical to the original one, it has exactly the same pricing implications for assets with payoffs that depend on X_{t+1} (see for example Cochrane (2005) for a discussion). Moreover, this projected pricing kernel, which is a univariate function of X_{t+1}, can vary over time, reflecting time variation in the pricing kernel state variables.

Hansen and Richard (1987) show that it is sufficient that there exists a particular admissible SDF which is almost surely positive in order to deduce that there are no arbitrage opportunities on $\mathcal{X}_{t+1}$. Inversely, no arbitrage implies that the projected SDF $M^*_{t,t+1}$ is positive with probability one. Therefore, without making any assumptions about market completeness, the absence of arbitrage in the set of contingent claims leads to the existence of a unique and positive projected SDF which is a function of the primitive payoff X_{t+1} (we also refer to Garcia et al. (2010) for a discussion about this point). As a consequence, we can define a risk-neutral probability measure Q such that the price of any contingent claim with payoff $H_{t+1} = h(J_t, X_{t+1})$ is given by

$$\pi_t(H_{t+1}) = E_t[M^*_{t,t+1}H_{t+1}] = E_t[M^*_{t,t+1}]E^Q_t[H_{t+1}]$$

with

$$E^Q_t[H_{t+1}] = E_t\left[\frac{M^*_{t,t+1}}{E_t[M^*_{t,t+1}]}H_{t+1}\right].$$

Thus, pricing is reduced to a riskless discounting using the price at time t of a zero-coupon bond which pays one dollar at time $t+1$ and a distorted (risk-neutral) conditional expectation of the asset payoff H_{t+1}. It is called risk-neutral pricing since it determines prices as if agents were risk neutral. Furthermore, the pricing kernel $M^*_{t,t+1}$ can be seen as the transformation between the risk-neutral and historical measure, that is

$$M^*_{t,t+1} = E_t[M^*_{t,t+1}]\frac{q}{p} = e^{-r_{t,t+1}}\frac{q}{p}, \tag{1.3}$$

where $r_{t,t+1}$ is the (continuously compounded) yield on the zero-coupon bond and q and p designate the risk-neutral or the historical density function. The density q is often also referred to as state price density since it represents the continuous-state counterpart of the so-called Arrow-Debreu securities.

1.2.2 The pricing kernel puzzle

In order to formally introduce the pricing kernel puzzle, let us now consider a representative agent model in which the investor's preferences satisfy the expected utility theory of von Neumann and Morgenstern (1944). A famous example is Lucas' (1978) consumption-based asset pricing model. In this case, the pricing kernel corresponds to the investor's intertemporal marginal rate of substitution, i.e.

$$M_{t,t+1} = \frac{u'(C_{t+1})}{u'(C_t)}, \tag{1.4}$$

where C_t and C_{t+1} are consumption in period t and $t+1$, respectively, and u' is the first derivative of the investor's utility function. Based on assumptions present in many classical finance models that investors satisfy the non-satiation property and are risk averse, we obtain an increasing and concave utility function u. Consequently, the pricing kernel in equation (1.4) should be a decreasing function of aggregate consumption in period $t+1$. Similarly, one would also expect that the projected pricing kernel $M^*_{t,t+1}$ (which coincides with the projection of $\frac{u'(C_{t+1})}{u'(C_t)}$ on the set of payoffs) is a decreasing function. However, this is in contrast with empirical observations. Among others, Jackwerth (2000) and Rosenberg and Engle (2002) find that the pricing kernel is not an overall decreasing function. In other words, they observe a locally increasing pricing kernel, implying a locally increasing marginal utility and convex utility function. We refer to this as the pricing kernel puzzle.

Various attempts have been made to explain the puzzle. Chabi-Yo et al. (2008) and Grith et al. (2013b) for example consider state-dependent preferences as possible explanations whereas others like Shefrin (2008), Ziegler (2007), De Giorgi and Post (2008), Polkovnichenko and Zhao (2013) and Hens and Reichlin (2012) focus on results coming from the behavioral finance literature to try to understand this puzzling behavior of the pricing kernel.

1.3 Empirical pricing kernel estimation

In this section, we will present our FGD method based on splines to estimate an empirical pricing kernel. As opposed to most of the studies in the literature that use an indirect approach, i.e. first estimating the physical and risk-neutral densities and obtaining the pricing kernel in a second step, we follow a direct approach originally introduced in Rosenberg and Engle (2002). Therefore, we start by presenting the general idea of the direct estimation method before continuing with a detailed description of the FGD estimation methodology.

1.3.1 An introduction to the direct estimation approach

The goal is to estimate a projected pricing kernel onto the underlying asset returns using S&P 500 option data and historical returns. Let us now briefly summarize this approach.

Estimation technique

Using the fundamental pricing equation, one can write the price of a derivative with a payoff that depends on the return of the underlying asset r_{t+1} as

$$P_{i,t} = E_t[M^*_{t,t+1}(r_{t+1})g_i(r_{t+1})] = \int M^*_{t,t+1}(r_{t+1})g_i(r_{t+1})f_t(r_{t+1})\,dr_{t+1},$$

where $P_{i,t}$ is the price of the ith asset with payoff function $g_i(r_{t+1})$ and $f_t(r_{t+1})$ designates the probability density of one-period underlying asset returns.

We get the fitted model price $\hat{P}_{i,t}$ using an estimate of the projected kernel $\hat{M}^*_{t,t+1}(r_{t+1})$ together with an accurately estimated return density $\hat{f}_t(r_{t+1})$ such that

$$\hat{P}_{i,t} = E_t[\hat{M}^*_{t,t+1}(r_{t+1})g_i(r_{t+1})] = \int \hat{M}^*_{t,t+1}(r_{t+1})g_i(r_{t+1})\hat{f}_t(r_{t+1})\,dr_{t+1}. \tag{1.5}$$

Whereas the kernel $\hat{M}^*_{t,t+1}(r_{t+1})$ is the object of interest that we have in mind to estimate, we need a model specification of the underlying return process in order to get the density f_t. This question will be addressed in Subsection 1.3.1 below.

The empirical pricing kernel is then defined as the function that makes fitted prices closest to observed prices, using the estimated return density. This is basically a function estimation/optimization problem. But it includes some additional complexity since it must be combined with Monte Carlo methods or numerical integration to calculate the expectation in each step of the optimization procedure. However, the problem can be simplified by assuming a parametric representation $M^*_{t,t+1}(r_{t+1};\theta_t)$ of the projected pricing kernel, where θ_t is an N-dimensional parameter vector.

This leads to a parameter optimization problem and we call empirical pricing kernel the one that solves

$$\min_{\theta_t} \sum_{i=1}^{L} (P_{i,t} - \hat{P}_{i,t}(\theta_t))^2 / Vega^2_{i,t},$$

where L represents the number of asset prices, $\hat{P}_{i,t}(\theta_t)$ is the fitted model price as a function of the pricing kernel parameter vector and $Vega_{i,t}$ is the BS-vega of the option at the market implied level of volatility.

Such vega-weighted pricing errors are an approximation to implied volatility errors, which have desirable statistical properties. In particular, implied volatility errors are proportional to bid-ask spreads and yield a better scaling of the cost functional. Unlike implied volatility errors, they do not require Black-Scholes inversion of model prices at every step in the optimization algorithm, which is therefore favorable from a computational point of view. We refer to Cont and Tankov (2004) and Christoffersen et al. (2013) among others for more details on applications of vega-weighted option valuation errors.

Modeling the underlying return process and numerical approximation of the conditional expectation

As suggested in Rosenberg and Engle (2002), we use an asymmetric GARCH model with empirical innovations to approximate the physical density of S&P 500 returns. More specifically, we model the underlying return process with a GJR-GARCH (Glosten et al. (1993))[1] including a linear autoregressive term for estimating a conditional mean, i.e.,

$$r_t = \ln(S_t/S_{t-1}) = \mu r_{t-1} + \epsilon_t, \quad \epsilon_t \sim h(0, \sigma_t^2) \tag{1.6}$$

and

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha\epsilon_{t-1}^2 + \gamma \max(0, -\epsilon_{t-1})^2. \tag{1.7}$$

This model specification allows it to capture the most important stylized facts of financial return series. The conditional volatility is stochastic and mean-reverting and when $\gamma > 0$ the model also accounts for the leverage effect.[2] The empirical innovation density h is separated into a time-varying component σ_t and the time-invariant distribution of standardized innovations $z_t = \epsilon_t/\sigma_t$. This standardized innovation density incorporates skewness and excess kurtosis that are not captured in a normal density.

We estimate the model using a pseudo maximum likelihood approach with a normal innovation density. Bollerslev and Wooldridge (1992) show that consistent parameter estimates can be obtained under certain conditions even if the true innovation density is not normal.

---

[1] Rosenberg and Engle (2002) fit a number of GARCH models to daily S&P 500 index returns and find that the GJR-GARCH model describes the data best. Since then, this model or similar asymmetric GARCH specifications have been widely used in the empirical pricing kernel literature to model the underlying return process.

[2] The term leverage effect was introduced by Black (1976) and is used to describe the asymmetric reaction of volatility in response to good and bad news (excess returns).

Using the estimated model parameters, the conditional variance and the standardized innovations, we are able to simulate return paths. For a given time point t such a path is obtained by randomly selecting an estimated standardized innovation $z_{[1]}$, updating the conditional variance σ^2_{t+1}, drawing a second innovation $z_{[2]}$, updating the conditional variance σ^2_{t+2}, and continuing up to the desired time horizon. Repeating this procedure many times, we obtain a set of simulated returns and apply a kernel density estimator to get an estimate of the future return density f_t. This method, first introduced in Barone-Adesi et al. (1998) to compute portfolio risk measures, is now widely used and generally called filtered historical simulation (FHS) in the financial econometrics literature (see, for example, Barone-Adesi et al. (2008)).

Finally, we obtain a fitted model price by using the classical midpoint numerical integration rule in order to calculate the integral in equation (1.5). We have

$$\hat{P}_{i,t}(\theta_t) = \frac{\Delta r_{t+1}}{n} \sum_{j=1}^{n} \hat{M}^*_{t,t+1}(\tilde{r}_{j,t+1}; \theta) g_i(\tilde{r}_{j,t+1}) \hat{f}_t(\tilde{r}_{j,t+1}), \tag{1.8}$$

where $\tilde{r}_{j,t+1} = \frac{r_{j-1,t+1}+r_{j,t+1}}{2}$ with $j = 1, \ldots, n$ are n equally spaced mid-points in a partition $\{r_{j,t+1}\}_{j=0}^{n}$ of the integration domain $[r_{0,t+1}, r_{n,t+1}]$ with length $\Delta r_{t+1} = r_{n,t+1} - r_{0,t+1}$. Alternatively, one could also use a Monte Carlo approximation of the integral. We tried both in our empirical analysis. Since the pricing accuracy does not seem to be very sensitive to the choice, we decided to rely on the simple numerical approximation, as it needs less computational time.

Parametric pricing kernel specifications

We will now review the parametric pricing kernel specifications proposed in Rosenberg and Engle (2002). The reason is twofold. First, they can serve as starting models in our FGD algorithm and secondly we can take them as benchmark models when analyzing the predictive power of our FGD approach based on splines in the empirical analysis section.

In the first specification, the kernel is a power function of the underlying asset's gross return, i.e.

$$M^*_{t,t+1}(r_{t+1}; \theta_t) = \theta_{0,t}(r_{t+1})^{-\theta_{1,t}}. \tag{1.9}$$

The first parameter $\theta_{0,t}$ is a scaling factor and the second parameter $\theta_{1,t}$ determines the slope of the kernel at date t. When $\theta_{1,t}$ is positive, the price kernel is negatively sloped, which implies that the value of a unit payoff increases as the underlying asset return decreases.

In the second, more flexible specification, they consider a kernel of the form

$$M^*_{t,t+1}(r_{t+1};\theta_t) = \theta_{0,t} T_0(r_{t+1}) \exp(\theta_{1,t} T_1(r_{t+1}) + \cdots + \theta_{N,t} T_N(r_{t+1})) \tag{1.10}$$

with $N+1$ parameters $(\theta_{0,t}, \dots \theta_{N,t})$ and where $T_0, \dots, T_N$ are Chebyshev polynomials with terms given by $T_n(x) = \cos(n \arccos(x))$ for $x \in [-1, 1]$. To obtain an approximation over a closed interval $[a, b]$, we consider generalized Chebyshev polynomials with $x = (2r_{t+1} - a - b)/(b - a)$.

Whereas the first parametric specification is rather restrictive but popular in financial models,[3] the second is rather flexible. More precisely, if there were an infinite number of polynomial terms in our expansion, we could theoretically approximate any continuous function accurately. However, the number of observed asset prices at time t provides us with an upper bound for the number of polynomial terms. This is the motivation behind using orthogonal polynomials such as Chebyshev polynomials, which provide more precise approximations for lower order expansions. Other reasonable choices could be Hermite or Laguerre polynomials (see for example Yang (2009) and Grith et al. (2011)).

1.3.2 Pricing kernel estimation using a FGD algorithm based on splines

Let us now describe our direct estimation approach that uses a functional gradient descent (FGD) algorithm based on splines. The FGD algorithm (Friedman (2001)) belongs to the class of boosting procedures, which are very popular in the area of machine learning. It can be interpreted as a functional analog of the gradient method used for parameter optimization. The way FGD works is quite intuitive. It takes a simple parametric or non-parametric model as a first approximation and then modifies it in a non-parametric way to improve a pre-specified goodness-of-fit statistic. For conceptual purposes, we start with a short introduction to FGD in a general regression framework and then proceed with a more detailed discussion on how to bring this idea into the field of pricing kernel estimation.

---

[3] Note that we would obtain this parametric form of the pricing kernel assuming that the stochastic process of the underlying stock follows a Geometric Brownian Motion.

A short introduction to FGD

Consider a sample $(X_1, Y_1), \dots, (X_n, Y_n)$, where $Y_i \in \mathbb{R}$ is the response variable and X_i represents a p-dimensional explanatory variable. Based on (X_i, Y_i), we are looking for a function $F \in \mathcal{F} = \{f | f : \mathbb{R}^p \to \mathbb{R}\}$ which minimizes an expected loss of the form $E[\lambda(Y, F(X))]$ with an adequate loss function λ.
The FGD algorithm estimates F by minimizing the empirical risk defined as

$$\Lambda(F) = \frac{1}{n} \sum_{i=1}^{n} \lambda(Y_i, F(X_i)).$$

Starting from an initial function estimate $\hat{F}_0$ (step 1 of the algorithm), the algorithm selects the steepest descent direction in the mth iteration which would be given by the negative functional derivative $-d\Lambda(\hat{F}_{m-1})$. However, due to smoothness and regularization constraints on the minimizer of $\Lambda(\hat{F}_{m-1})$, one must find a function $\hat{f}_m$ which is in the linear span of a class of simple base learners $\mathcal{S}$ and is close to $-d\Lambda(\hat{F}_{m-1})$ in the sense of a functional metric. This is equivalent to fitting the base learner $h(x, \theta) \in \mathcal{S}$ to the negative gradients

$$U_i = - \left. \frac{\partial \lambda(Y_i, Z)}{\partial Z} \right|_{Z = \hat{F}_{m-1}(X_i)}, \quad i = 1, \dots, n.$$

This is often achieved with least squares fitting and we get $\hat{f}_m = h(x, \hat{\theta})$ with

$$\hat{\theta} = \min_{\theta,\, \omega} \sum_{i=1}^{n} (U_i - \omega h(X_i, \theta))^2.$$

This is the second step of the algorithm. In a third step (line search), one finally has to perform a one-dimensional optimization in order to find the best step length $\hat{\omega}_m$ for updating $\hat{F}_{m-1}$ with $\hat{f}_m$. We obtain

$$\hat{\omega}_m = \min_{\omega} \sum_{i=1}^{n} \lambda(Y_i, \hat{F}_{m-1}(X_i) + \omega \hat{f}_m(X_i))$$

and get

$$\hat{F}_m = \hat{F}_{m-1} + \hat{\omega}_m \hat{f}_m.$$

Iterating steps 2 and 3 until $m = M$ produces the FGD estimate

$$\hat{F}_M = \hat{F}_0 + \sum_{m=1}^{M} \hat{\omega}_m \hat{f}_m.$$

Our approach using FGD based on splines

In order to successfully apply the above mentioned boosting technique to pricing kernel estimation, we restrict our kernel to be an additive expansion of the form[4]

$$M^*(r) = M_0^*(r) + \sum_{m=1}^{M} \beta_{d_m} B_{d_m}(r). \tag{1.11}$$

Here, M_0^* denotes an appropriate starting model and each B_{d_m} designs a B-spline basis function with basis index d_m and a corresponding multiplicative coefficient β_{d_m}.[5]

We use B-spline basis functions since B-splines in connection with FGD have already proven to yield good results in volatility forecasting (see Audrino and Bühlmann (2009)). Moreover, B-splines are piecewise polynomial functions and can therefore be used to approximate a general continuous function.[6] Using B-splines will allow for a large flexibility in the shape of the pricing kernel, depending on how we choose the tuning parameters, i.e. the order and the number of breaks (also called knots) of each B-spline basis function. We allow the pricing kernel to be a cubic function of the returns and thus select a spline order of 4. The number of knots is a measure for the approximation accuracy. The higher the number of breaks, the better the approximation we obtain but with a higher variability due to a larger complexity. In our application, we choose as break points empirical α-quantiles of the simulated returns with $\alpha = i/mesh$ (i=1,…,mesh-1) and $mesh \in \mathbb{N}$.[7]

Remind that the projected pricing kernel is a function of future returns only. On the other hand, we observe option prices. That's why we apply the FGD algorithm to the price functional.

---

[4] For simplicity of notation, we remove all time subscripts and just write $M^*(r)$ instead of $M^*_{t,t+1}(r_{t+1})$ for the projected pricing kernel. Where appropriate, we reuse this notational shortcut in subsequent parts of this article as well.

[5] No-arbitrage conditions on the coefficients of the B-spline basis functions have been recently derived by Fengler and Hin (2014).

[6] We refer to de Boor (2001) for an introduction to B-splines.

[7] Generally, one can also use another complexity parameter (the so-called knot's multiplicity) to control the smoothness of the approximation at each knot. So far, we impose our approximation to be continuous and smooth at each break point. This means that we set the knot's multiplicity to be equal to 1 for all knots except the first and last one. For more details refer to de Boor (2001).

Recently, nice asymptotical properties have been shown for L^2-boosting (we refer to Bühlmann and van de Geer (2011) and references therein for a more detailed discussion). Thus, we consider in our FGD procedure a (slightly modified) L^2-loss function λ given by the vega-weighted squared-error loss of observed and fitted option prices, i.e.

$$\lambda(P, \hat{P}) = \frac{1}{2}\left(\frac{P - \hat{P}}{Vega}\right)^2 .$$

We start our algorithm with an initial function estimate

$$\hat{P}_0 = \frac{\Delta r}{n} \sum_{j=1}^{n} \hat{M}_0^*(\tilde{r}_j) g(\tilde{r}_j) \hat{f}(\tilde{r}_j). \tag{1.12}$$

This fitted initial price corresponds to the approximation that has been described in equation (1.8) and obviously depends on the selected starting model for the pricing kernel.

The choice of the starting model used in the FGD algorithm is important since FGD aims at locally improving the empirical loss of an initial model estimate on the basis of non-parametric additive expansions. Hence, one should start from a reasonable initial estimate to obtain a satisfactory performance. In our application, we take the power pricing kernel specification introduced in (1.9) as a starting model. We believe that it represents a good trade-off between adequacy and complexity. Furthermore, this pricing kernel contains a solid economic motivation and conforms to classical finance models (i.e., it does not yield to the empirical pricing kernel puzzle).

After having completed the first step of the algorithm, we start an iterative procedure that always leads to a new additive price component

$$\hat{\beta}_{\hat{d}_m} \frac{\Delta r}{n} \sum_{j=1}^{n} B_{\hat{d}_m}(\tilde{r}_j) g(\tilde{r}_j) \hat{f}(\tilde{r}_j). \tag{1.13}$$

According to the notation used in Subsection 1.3.2, equation (1.13) represents the fitted weak learner multiplied with the estimated optimal step length $\hat{\beta}_{\hat{d}_m}$. In gereral, possible choices of the weak learner are restricted in the following way: it should belong to a given class of statistical procedures that are weak in the sense that they avoid overfitting by limiting the number of parameters involved in the estimation. We are convinced to fulfill this condition using B-spline basis functions. Taking all the above considerations into account, we obtain the following estimation algorithm:

FGD algorithm with B-spline learners

Step 1 (initialization). Estimate the starting function $\hat{M}^*_0$ using the estimated return density $\hat{f}$ and some return values $\{\tilde{r}_j\}_{j=1}^n$. Set $m = 1$.

Step 2 (projection of the negative gradient to the weak learner). Compute the negative gradients

$$U_i = \frac{1}{Vega_i^2}\left(P_i - \frac{\Delta r}{n}\sum_{j=1}^{n}\hat{M}^*_{m-1}(\tilde{r}_j)g_i(\tilde{r}_j)\hat{f}(\tilde{r}_j)\right), \quad i = 1, \dots, L.$$

Then, project the negative gradients onto the weak learner. More precisely, we solve

$$\hat{d}_m = \min_{1\le d\le k}\sum_{i=1}^{L}\left(U_i - \frac{\Delta r}{n}\sum_{j=1}^{n}\hat{\beta}_d B_d(\tilde{r}_j)g_i(\tilde{r}_j)\hat{f}(\tilde{r}_j)\right)^2,$$

where d is a basis index, $\hat{\beta}_d$ denotes the least-squares estimated coefficient when regressing the residuals U_i versus $\frac{\Delta r}{n}\sum_{j=1}^{n} B_d(\tilde{r}_j)g_i(\tilde{r}_j)\hat{f}(\tilde{r}_j)$ (the weak learner) and k is the degree of freedom (number of B-spline basis functions). It is $k = (mesh - 1) + 4$.

Step 3 (line search). Perform a one-dimensional optimization for the step-length $\beta_{\hat{d}_m}$ when updating $\hat{M}^*_{m-1}$

$$\hat{\beta}_{\hat{d}_m} = \min_{\omega}\sum_{i=1}^{L}\lambda(P_i, \hat{P}_i)$$

with

$$\hat{P}_i = \frac{\Delta r}{n}\sum_{j=1}^{n}\left(\hat{M}^*_{m-1}(\tilde{r}_j) + \omega B_{\hat{d}_m}(\tilde{r}_j)\right)g_i(\tilde{r}_j)\hat{f}(\tilde{r}_j).$$

Update the current pricing kernel estimate

$$\hat{M}^*_m(r) = \hat{M}^*_{m-1}(r) + \hat{\beta}_{\hat{d}_m}B_{\hat{d}_m}(r).$$

Step 4 (iteration). Increase m by one and iterate steps 2 and 3, stopping when $m = M$. This produces the estimate

$$\hat{M}^*_M(r) = \hat{M}^*_0(r) + \sum_{m=1}^{M}\hat{\beta}_{\hat{d}_m}B_{\hat{d}_m}(r).$$

The choice of the stopping value M is important and it should be carefully selected to avoid overfitting. Usually, it is estimated by minimizing approximations of the expected prediction error. In our empirical application, we will apply a sample-splitting technique. One part of the data is used to estimate the model (estimation sample) and another part serves to do model evaluation (validation sample). The optimal value M is chosen such that it minimizes the empirical risk in the validation sample.

Furthermore, it is often desirable to make a base learner sufficiently weak (having low complexity). A simple but effective solution to achieve this is via shrinkage toward zero. Hence, the update in Step 3 of the algorithm is replaced by

$$\nu \hat{\beta}_{\hat{d}_m} B_{\hat{d}_m}(\cdot), \quad 0 \leq \nu \leq 1.$$

Obviously, this reduces the variance (a complexity measure) by the factor ν^2. In our empirical analysis, we take a learning rate $\nu = 0.25$ which yields good results.

Finally, we would like to emphasize that our FGD-algorithm, although by nature the same as the classical ones, possesses some additional peculiarities. More precisely, it depends on a set of simulated future returns (and their empirical distribution) and it must be connected to numerical integration or Monte Carlo methods to calculate the expectation in each step of the optimization procedure. Thus an adequate specification of the underlying return process and a correct evaluation of the expectation are important in order to ensure the accuracy of the final pricing kernel estimate. These features are not present in the generic FGD method and their necessity here makes our approach particularly challenging.

1.4 Empirical analysis

1.4.1 Data

In our empirical analysis, we use S&P 500 index option data to derive an empirical pricing kernel. The market for SPX options is one of the most active index options markets in the world. Consequently, it has been the focus of many applications in the empirical pricing kernel literature as for example in Jackwerth (2000), Aït-Sahalia and Lo (2000), Rosenberg and Engle (2002) and many others.

We consider closing prices of SPX European options from January 2005 to October 2010. Option data and all the other necessary data including interest rates and dividend yields are downloaded from OptionMetrics. The averages of bid and ask prices

Table 1.1: Summary of our option data sample used for pricing kernel estimation. The dataset contains 70 cross-sections of SPX European options from 2005 to 2010 with approximately one month until expiration and is obtained after applying the filtering criteria described in the text. We report mean and standard deviation for prices, implied volatilities and number of observations per cross-section according to different moneyness categories. Moneyness is the strike price divided by the spot price minus one, $K/S - 1$.

Moneyness	# Obs. per CS		Price ($)		IV (%)	
($K/S - 1$)	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Less than -0.1	3.2	6.2	5.8	3.6	42.1	12.4
-0.1 to 0	22.4	4.9	10.7	9.9	21.7	9.7
0 to 0.1	20.9	4.1	8.7	10.5	17.0	10.1
More than 0.1	4.9	10.8	3.2	5.0	33.3	13.6
All	51.5	14.5	8.9	9.8	22.2	12.5

are taken as option prices and in order to retain only liquid options in our data sample, we apply the following standard filtering criteria: We merely consider out-of-the money put and call options and discard options with implied volatilities larger than 70%, average price lower than $0.05 or volume equal to 0. Finally, we exclude observations that violate simple non-arbitrage bounds.

Like Rosenberg and Engle (2002), we have in mind to estimate a pricing kernel on a monthly basis. Thus we extract for each month a cross-section of options with approximately one month (20 trading days) until expiration. This procedure yields a sample of 70 cross-sections with a total of 3500 option prices.

Table 1.1 describes some characteristics of the one-month option contracts that we use for pricing kernel estimation. The mean number of options per cross-section is 51.1 with a standard deviation of 14.5. The majority of options have moneyness (defined as $K/S - 1$) from -0.1 to 0.1. This particular moneyness range corresponds to the default setting and coincides with the return domain usually taken into account in our application when estimating an empirical pricing kernel. Outside of this domain, the pricing kernel is equal to its estimated value at -0.1 or 0.1.[8] However, it happens in our sample period (especially during the recent financial crisis) that some considerable mass of the future one-month return distribution lies outside the interval $[-0.1, 0.1]$. In such cases, we consider a larger return domain $[-a, a]$ with $a > 0.1$ such that it contains at least 95% of the future returns. Consequently, we also enlarge the moneyness range in these situations and set it equal to the return domain. Thus we observe cross-sections with options having moneyness outside ±10%, which is reported in the table.

---

[8] This corresponds to the procedure suggested in Rosenberg and Engle (2002) using one month option contracts covering the period 1991 to 1995.

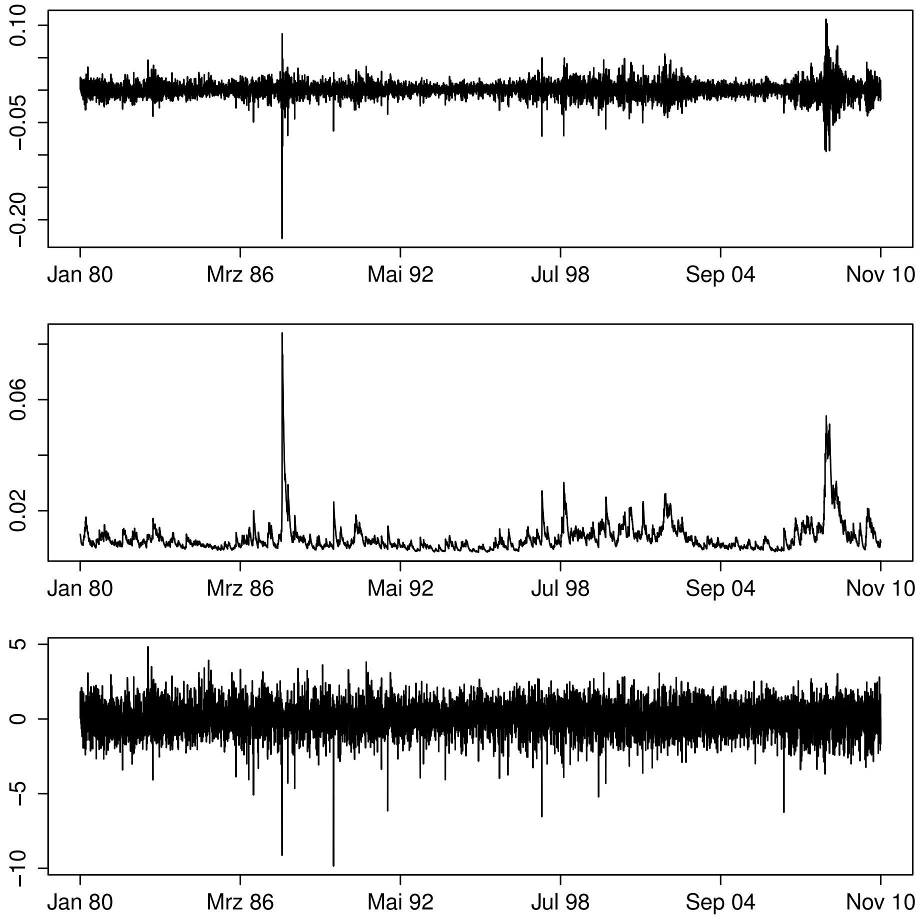


Figure 1.1: The figure shows daily log-returns of the S&P 500 index from January 4, 1980 until November 23, 2010 (top), the estimated conditional volatility σ_t using the AR-GJR-GARCH model (middle) and the corresponding standardized innovations z_t (bottom). We fit the model using pseudo maximum likelihood based on a normal innovation density.

The mean option price is \$8.9 and the price for puts seems to be somewhat higher than the one for calls. Finally, the table also shows average implied volatilities that exhibit the well-known volatility smile pattern.

1.4.2 Empirical results

A first step towards estimating an empirical pricing kernel is to fit the underlying return process. We use S&P 500 daily returns from January 4, 1980 until November 23, 2010 to estimate the AR-GJR-GARCH model introduced above using a pseudo maximum likelihood approach with a normal innovation density. Note that a sufficiently long historical time series is important in order to ensure that the empirical innovation density is accurately estimated. Figure 1.1 shows the S&P 500 daily log-returns, the estimated volatility σ_t and the corresponding standardized innovations z_t. Table 1.2

Table 1.2: Panel A shows the parameter estimates obtained when fitting the AR-GJR-GARCH model using a long historical time series of S&P 500 daily log-returns from January 4, 1980 until November 23, 2010. Panel B reports some characteristics of the standardized innovations z_t. The normality test p-values are obtained using the Anderson-Darling test for normality and the Ljung-Box test with ten lagged values is applied to test for serial correlation.

Panel A: Parameter estimates for AR-GJR-GARCH model					
	$\mu \times 10^2$	$\omega \times 10^6$	β	$\alpha \times 10^2$	γ
Coefficient	2.40	1.74	0.92	1.87	0.11
Std. Error	1.38	0.23	0.01	0.43	0.01

Panel B: Properties of standardized innovations					
Mean	St. dev.	Skewness	Excess Kurtosis	Normality test p-value	Serial correlation test p-value
0.032	1	-0.47	3.50	< 0.001	0.46

summarizes the parameter estimates and provides some characteristics of the standardized empirical innovation density. This density incorporates skewness and excess kurtosis and is highly non-normal.

With the fitted model in hand, we can make use of a filtered historical simulation approach (FHS) to obtain a set of simulated returns and apply a kernel density estimator to get an estimate of the future one-month return density f_t.

After that, we are able to estimate empirical pricing kernels. We use a rolling-window procedure and always consider two consecutive cross-sections of options for our estimation. The first one is used to fit the model whereas the second one serves for doing model evaluation and to choose the optimal complexity parameter M. All other tuning parameters of our algorithm are fixed. We take cubic B-spline basis functions, $mesh = 11$ (10 inner knots), a shrinkage factor $\nu = 0.25$ and use $n = 5000$ midpoints to calculate the integral. Finally, we employ a third cross-section to analyze the predictive power.

The goal is to investigate pricing kernel estimates and to examine the pricing performance of our approach in comparison with the parametric specifications suggested in Rosenberg and Engle (2002). We believe that these are two fair competitors. The power pricing kernel is quite simple but economically motivated. In addition, it is used as a starting model and therefore helps to answer the question whether our FGD algorithm can substantially improve the final estimate. As a second benchmark, we consider the flexible Chebyshev pricing kernel specification. We allow up to 6 parame-

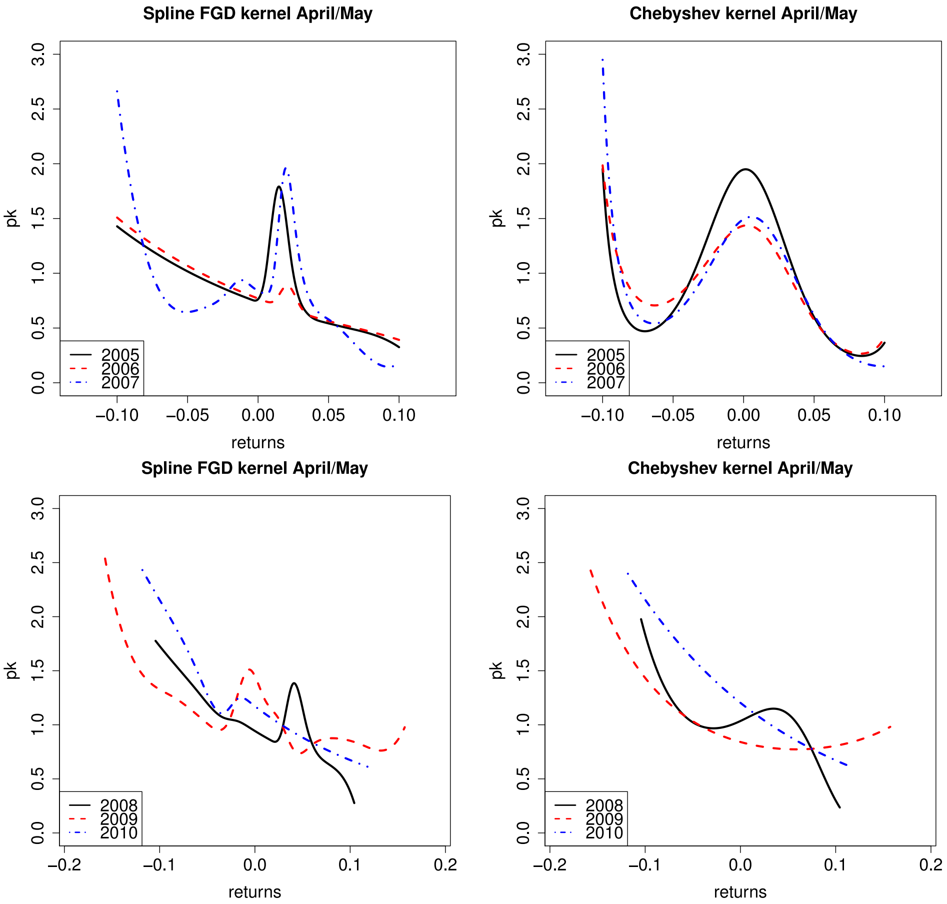


Figure 1.2: The figure shows a selection of estimated pricing kernels using April/May option data each year from 2005 to 2010. We apply the first option cross-section to fit the model whereas the second one is used to do model evaluation and to determine the optimal model complexity. The left part shows results for the B-spline FGD approach and the right part for the Chebyshev model.

ters and determine the optimal complexity using the same sample-splitting technique as in our FGD procedure.

Estimated pricing kernels

Let us start with an inspection of the pricing kernel shape. Figure 1.2 shows a selection of empirical pricing kernels obtained with our FGD methodology as well as with the Chebyshev model. Although we depart from an overall decreasing power pricing kernel in our FGD algorithm, we observe final estimates that contain increasing parts. This feature is present in estimates found with the Chebyshev method too. It is known as the pricing kernel puzzle and has been noticed several times before (see Jackwerth (2000) and Rosenberg and Engle (2002) among others). The increasing component is usually located in the area of zero return and resembles a bump. However, we also

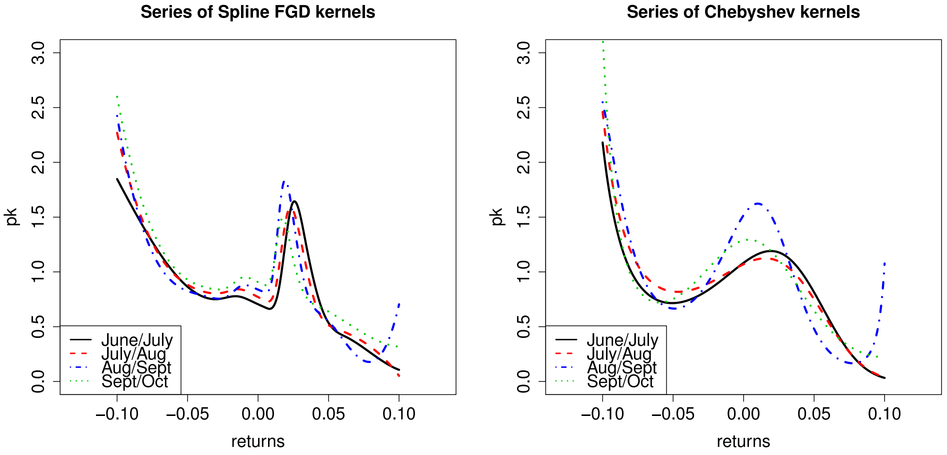


Figure 1.3: The figure shows a series of four consecutively estimated pricing kernels starting with June/July option data from 2006. We apply a rolling-window procedure and always consider two consecutive cross-sections for estimation. The first is used to fit the model whereas the second one serves as a validation sample to determine the optimal model complexity. Left: Spline FGD approach. Right: Chebyshev specification.

get fitted kernels that are rather consistent with another finding recently described in the literature claiming that the pricing kernel looks U-shaped (see for example Bakshi et al. (2010) and Christoffersen et al. (2013)).

An additional point visible in Figure 1.2 is the time-varying pattern of the fitted kernels, reflecting time variation in the pricing kernel state variables. We further investigate this point via Figure 1.3. There we show a series of consecutively estimated pricing kernels, and regardless of the chosen method, we again find time-dependent estimates. Obviously, the observed differences are smaller compared to Figure 1.2 given the higher time frequency.

Furthermore, it might be interesting to see how sensitive these fitted kernels are with respect to the number of inner knots used in our algorithm and their placement. In a small robustness check, we decided to compare three different numbers of break points (i.e. $mesh \in \{7, 11, 15\}$) and two distinct placement strategies to determine their position. The corresponding results are illustrated in Figure 1.4. The left plots show estimates obtained when taking empirical quantiles of the simulated returns as break points whereas we considered equally spaced inner knots to get the figures on the right. Obviously, the pricing kernel shape is more flexible when the number of inner knots is large. Comparing the knot placement strategies, we find somewhat narrower kernels with a higher peak when considering empirical quantiles as break points.

In another robustness check, we look at the sensitivity of the estimated pricing kernels with respect to specification of the underlying return process. Besides the GJR-

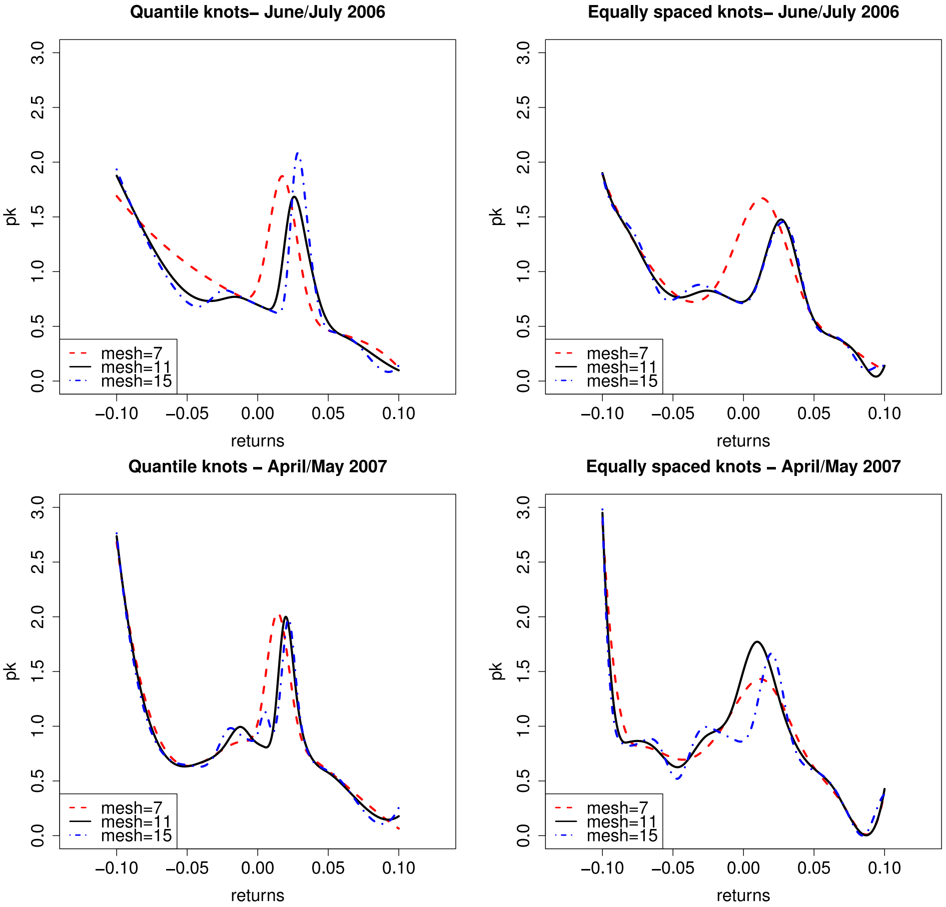


Figure 1.4: Results of a small robustness check concerning the number of inner knots used in our algorithm and their placement. The left plots are obtained using empirical quantiles of the simulated returns as break points whereas in the right figures we considered equally spaced inner knots. The figures show fitted kernels using June/July option data from 2006 (top) and April/May option data from 2007 (bottom).

GARCH model introduced above, we also consider the standard GARCH specification without asymmetry term and the EGARCH model (Nelson (1991)). Results for the EGARCH model are qualitatively similar to those shown for the GJR-GARCH and therefore for the sake of brevity not reported. A comparison of the resulting pricing kernels is illustrated in Figure 1.5. The overall shape is the same in both cases. However, the estimated kernels obtained with the symmetric GARCH model are somewhat bumpier and we observe some differences mainly for large absolute return values.

Pricing performance

Next, we would like to see whether our flexible FGD algorithm based on splines yields better pricing results in comparison with the chosen benchmarks. We take the 70 cross-sections of one-month option contracts from 2005 until 2010 and apply

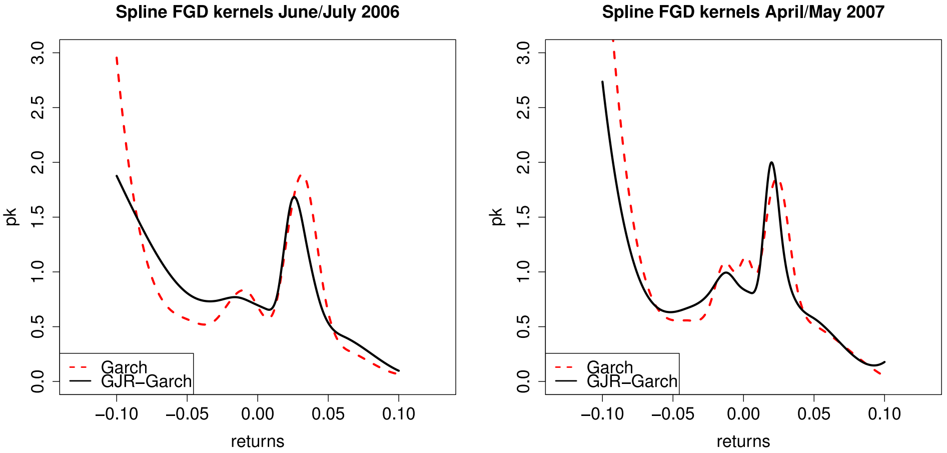


Figure 1.5: Robustness check to explore the sensitivity of the fitted kernels with respect to the specification of the underlying return process. The plots show estimated pricing kernels using June/July option data from 2006 (left) and April/May option data from 2007 (right).

the rolling-window estimation procedure that has been described above. The corresponding results are summarized in Table 1.3.

Table 1.3: Performance results of the different estimation methods using 70 cross-sections of one-month SPX option contracts from 2005 until 2010. We present in-sample (IS) and out-of-sample (OS) errors considering the root mean squared error loss function for implied volatilities (IV RMSE) and prices (Price RMSE). Averages are taken over the 68 observations obtained by applying the rolling-window procedure explained in the text. The table also contains the average optimal stopping value $\hat{M}$ in our FGD algorithm and the average number of parameters for the different estimation methods. Panel A reports the results using the GJR-GARCH model to describe the underlying return process, whereas Panel B indicates the outcomes of a robustness check in which we take a classical GARCH model without asymmetry term.

Model	$\hat{M}_{opt}$	# par	Averaged IS-		Averaged OS-	
			IV RMSE	Price RMSE	IV RMSE	Price RMSE
Panel A: AR-GJR-GARCH model						
Power kernel		2	2.14	1.96	2.75	2.59
Spline FGD	60.82	123.64	1.52	1.24	2.64	2.37
Chebyshev kernel		3.89	1.55	1.28	2.74	2.47
Panel B: AR-GARCH model						
Power kernel		2	3.04	2.85	3.72	3.58
Spline FGD	81.15	164.30	1.96	1.61	3.48	3.23
Chebyshev kernel		4.18	2.00	1.69	3.68	3.38

We present in-sample and out-of-sample pricing errors using the root mean squared error loss function for implied volatilities (IV RMSE) and prices (Price RMSE). Our new estimation methodology consistently outperforms both competitors. In particular, we observe predictive gains over the power pricing kernel specification and the Chebyshev model that range from 4% to 9%, depending on the performance measure. In a robustness test, we consider again the classical GARCH model without asymmetry

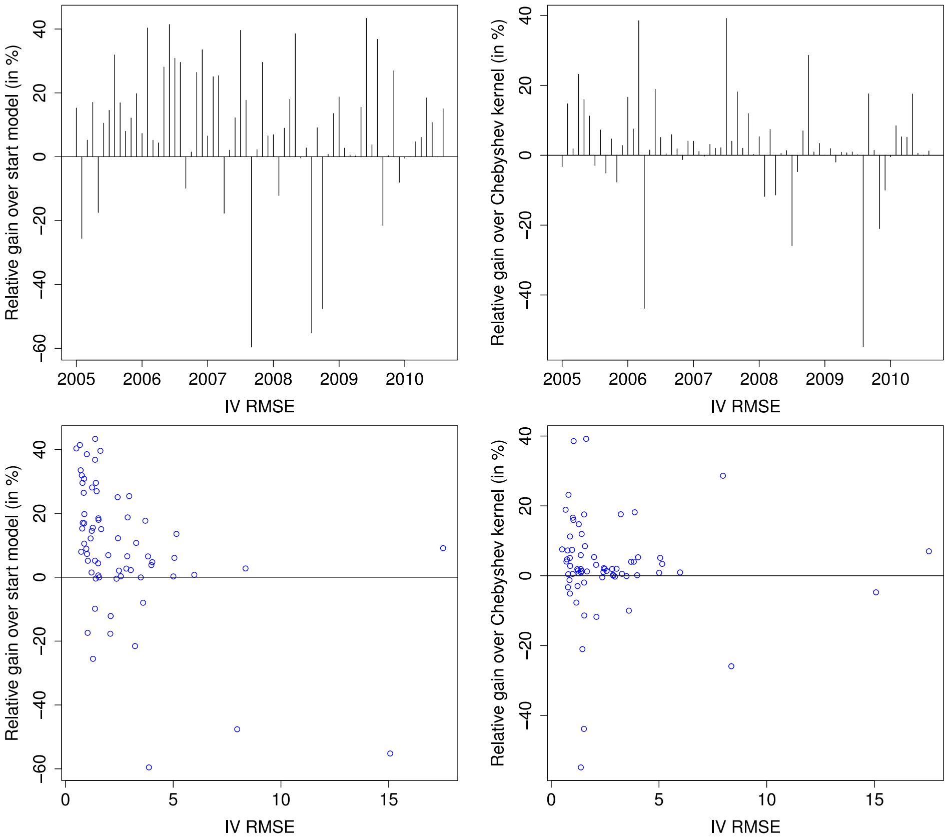


Figure 1.6: The plots show relative forecasting gains of our new method over the power pricing kernel specification (left) and the Chebyshev model (right) using the IV RMSE loss. The upper part of the figure contains the time-series of such gains whereas the lower part plots these gains versus the observed loss.

term. Interestingly, our novel method still performs best and consistently outperforms the chosen benchmarks. However, the observed losses are higher than in the setting with a GJR-GARCH model. Hence, we conclude that the presence of the asymmetry term is necessary in order to improve the pricing accuracy of the method.

We provide further insights into the pricing performance by means of Figure 1.6. The plots show relative forecasting gains of our new method over the power pricing kernel specification and the Chebyshev model using the IV RMSE loss. The upper part of the figure contains the time-series of such gains whereas the lower part plots these gains versus the observed loss. Again, the better forecasting accuracy of the Spline FGD approach is evident. Note that qualitatively equivalent results could be plotted by taking the other performance measure too.

Finally, there is the issue of whether these gains are statistically relevant. To explore this, we implement some Diebold and Mariano (1995) type tests to measure the superior predictive ability. We consider a t-type test comparing the observed losses

Table 1.4: The table shows results of Diebold and Mariano (1995) type tests to measure the superior predictive ability of our approach over the chosen benchmarks. We consider a t-type test comparing the observed losses (IVRMSE and Price RMSE) of the different models and a sign-type test based on a series of Bernoulli random variables indicating the model with the better forecasting performance. Positive values of the sign-type statistic and negative values of the t-type statistic are in favor of the Spline FGD approach. p-values are reported in parentheses with *, **, *** denoting significance at the 10%, 5% and 1% level, respectively.

Model	IVRMSE		Price RMSE	
	t-type	sign-type	*t*-type	sign-type
Power vs. FGD	-1.280	6.478	-2.217	4.263
	(0.100)	(≈ 0***)	(0.013**)	(≈ 0***)
Chebyshev vs. FGD	-1.622	5.297	-1.283	2.695
	(0.052*)	(≈ 0***)	(0.100*)	(0.004***)
Power vs. Chebyshev	-0.059	3.325	-1.069	2.566
	(0.477)	(≈ 0***)	(0.142)	(0.005***)

(IVRMSE and Price RMSE) of the different models and a sign-type test based on a series of Bernoulli random variables indicating the model with the better forecasting performance. The corresponding outcomes are summarized in Table 1.4. Positive values of the sign-type statistic and negative values of the t-type statistic are in favor of our Spline FGD approach. Table 1.4 confirms the higher predictive ability of our method in comparison with the chosen benchmarks. One might critically argue that the superior performance of the FGD approach in comparison to the method with a Chebyshev kernel is due to the relative performance of the power and Chebyshev kernels and not a consequence of the FGD method. In order to disprove this assumption, we also compare these two benchmarks in Table 1.4. In particular for the sign-type test, we find strong evidence for the superiority of the Chebyshev kernel over the power kernel specification. Hence, it is an interesting result that our Spline FGD approach with a power kernel as starting model performs significantly better than the method with a flexible Chebyshev kernel specification.

1.5 Conclusion

In this study, we proposed the use of a customized functional gradient descent (FGD) algorithm based on B-splines to estimate the empirical pricing kernel. Our model is flexible and computationally feasible, although it involves many parameters. The estimation properties of our methodology are illustrated empirically using S&P 500 index option data. We show that the algorithm yields accurate estimates and we also provide evidence of the superior predictive ability of our method in comparison with the parametric specifications suggested in Rosenberg and Engle (2002).

Having accurate pricing kernel forecasts is interesting from an economic point of view. They contain useful information regarding the investors' future beliefs and risk behavior and one can try to use them to improve option valuation or the performance of option trading/hedging strategies.

Our modeling and computational framework could also be extended in order to fit the complete pricing kernel surface. In this case, a bivariate B-spline basis (i.e. a product of two univariate B-spline basis functions) should be considered. One univariate B-spline basis would be a function of future returns as in our algorithm whereas the other B-spline basis function would depend on the time to maturity, representing the second dimension. Moreover, this generalization may be extended further in a straightforward way to allow the dynamics of the pricing kernel to depend on some additional relevant (endogenous or exogenous) explanatory factors, similarly to what has been done in Audrino and Colangelo (2010).

Surprisingly, the surface modeling perspective has only been considered in the most recent literature. In fact, it is quite common in most approaches to estimate the pricing kernel for each time to maturity separately (so-called slice by slice). Some exceptions to be mentioned are the studies by Giacomini and Härdle (2008) and Fengler and Hin (2014) that consider a two-dimensional modeling approach for the pricing kernel. Given the high estimation and forecasting accuracy shown by the proposed methodology, its extension in a multivariate setting will be the focus of future research.

Chapter 2

Modeling the pricing kernel surface dynamics with a boosting approach based on regression trees

Pirmin Meier

Abstract

The pricing kernel is a key component of any asset pricing model. It represent the link between the physical and risk-neutral density and contains information regarding investor preferences. Important empirical studies show that the option-implied empirical pricing kernel is time-varying, which reflects changes in the relevant pricing kernel state variables. However, little is known about these factors driving the pricing kernel over time. In this paper, we introduce a time series model for the evolution of the empirical pricing kernel surface. More specifically, building on some initial estimates of this dynamic surface, we propose a boosting approach based on regression trees. Given that trees have the ability to choose among a set of predictors the most relevant ones, we develop a tool to answer our main question regarding the principal pricing kernel driving factors. In our empirical analysis, we illustrate the method with European options and return values of the S&P 500 index.

Keywords: Pricing kernel dynamics, driving factors, boosting, regression trees, options.

2.1 Introduction

An equation of fundamental importance in asset pricing states that the price of an asset equals the pricing kernel weighted payoff. Hence, the pricing kernel (also called stochastic discount factor) is a key component of any asset pricing model, and it further contains important information regarding investor risk preferences.

Over the years, starting with Jackwerth (2000), Aït-Sahalia and Lo (2000) and Rosenberg and Engle (2002), many approaches have been developed to infer this kernel from observed asset prices, in particular options. It is a common finding that the resulting function estimate, referred to as empirical pricing kernel, has a non-monotone shape. More specifically, the option-implied pricing kernel features usually either a U-shape or a hump-shape, which is both not conform to classical finance models, where one would expect to find a fitted pricing kernel that is a decreasing function of aggregate resources.

This empirical phenomena is known as “pricing kernel puzzle” (a term ascribed to Brown and Jackwerth (2012) and originally introduced in an earlier version of the paper circulating in 2000). It has evoked a number of research papers trying to solve the puzzle. As mentioned in Barone-Adesi et al. (2014), there are up to date at least half a dozen different explanations. These include: heterogeneous beliefs of agents about the physical return distribution; misspecification of the underlying state space (for example by neglecting to condition on market volatility); preference issues such as ambiguity aversion and rank-dependent utility; and incomplete markets. In addition, the puzzle might be just a statistical artefact, meaning that the monotonicity cannot be rejected when tested statistically. And finally, the pricing kernel puzzle may be simply a consequence of how the conditional physical return density is modeled (see Linn et al. (2014)).

Although there does not seem to exist a consensus regarding the solution of the puzzle, it is commonly accepted that the pricing kernel is time-varying, reflecting changes in the relevant pricing kernel state variables (see for example Rosenberg and Engle (2002)). However, surprisingly few attempts have been made in the recent past to model and understand the observed dynamics. Some exceptions to be mentioned are the studies by Grith et al. (2013a), Polkovnichenko and Zhao (2013), Metaxoglou and Smith (2014) and Barone-Adesi et al. (2014). However, despite these recent advancements, little is still known about the relevant pricing kernel driving factors.

Therefore, we develop in this paper a time series model for the option-implied empirical pricing kernel that allows understanding its variation over time. Contrary to most techniques that model the pricing kernel for each time to maturity separately (so-called slice by slice), we adopt a surface modeling approach. More precisely, we first derive initial estimates of the pricing kernel surface (using separate estimates for both the risk-neutral and historical density surface of the underlying asset return at expiry) and then model the observed time-varying surface with a boosting approach based on regression trees. The aim of boosting procedures, which are very popular in the area of computational statistics and machine learning, is to combine a series of weak learners in order to obtain a powerful model at the end. Recently, boosting approaches have already been successfully applied to financial function estimation. In particular, Audrino and Colangelo (2010) have shown that boosting based on regression trees can be used to model the dynamics of the implied volatility surface. Taking trees as base learner is particularly helpful in our context, given that they automatically select the most relevant predictors among a set of possible candidates. Therefore, we obtain a suitable tool to answer our main question regarding the principal driving factors of the pricing kernel dynamics. In addition, trees model nonlinearity in a parsimonious and interpretable way, which is an advantage over other variable selection techniques.

In our empirical analysis, we illustrate the model with European options and return values of the S&P 500 index and show that our modeling approach can capture the various shapes of the option-implied pricing kernel observed during the sample period, covering January 2002 to December 2010. Further, we provide evidence that the observed dynamics can indeed be explained with variables related to the state of the economy, which is in line with findings presented in Rosenberg and Engle (2002), Grith et al. (2013a) and Metaxoglou and Smith (2014), for example. However, with our modeling approach we are able to quantify the impact of all possible driving factors building on the relative influence measure introduced in Friedman (2001). We find that the relative influence of the single predictors substantially varies over time and we provide some additional insights on how they affect the pricing kernel shape. Finally, we also investigate the predictive performance of our model and provide some evidence that the forecasts are reasonable. However, it is difficult to beat the random walk forecast.

The paper is organized as follows. Section 2.2 reviews the role of the pricing kernel in option pricing. Section 2.3 presents the estimation technique used to infer the option-implied empirical pricing kernel surface. Section 2.4 introduces our time series modeling approach based on a boosting algorithm with regression trees. Section 2.5 presents the empirical results and Section 2.6 concludes the paper.

2.2 The role of the pricing kernel in option pricing

We consider a financial market and assume that a riskless bond with constant interest rate $r > 0$ and a risky stock with price process $\{S_t\}_{t\geq 0}$ (defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$) are traded. By the fundamental theorem of asset pricing, the assumption of non-arbitrage implies the existence of a risk-neutral probability measure $\mathbb{Q}$ equivalent to $\mathbb{P}$ under which the discounted stock price process is a martingale (see Delbean and Schachermeyer (1994)). Therefore, we can calculate the price C_t at time t of an European call option with underlying S, strike K and time-to-maturity τ using the risk-neutral valuation principle, i.e.,

$$C_t = e^{-r\tau} E_{\mathbb{Q}}[\Psi(S_T)|\mathcal{F}_t] = e^{-r\tau} \int_0^\infty \Psi(S_T) q_{t,T}(S_T) dS_T, \tag{2.1}$$

where $\Psi(S_T) = \max(S_T - K, 0)$ is the option payoff at expiry T and $q_{t,T}$ denotes the conditional risk-neutral probability density of the underlying at time T.

This risk-neutral pricing density is often also referred to as state price density since it represents the continuous-state counterpart of the so-called Arrow-Debreu securities. It inherently differs from the historical density of the underlying price process as it does not only contain information about the market participants' expectations regarding the evolution of the underlying price process, but also depends on their perceptions of risk and ambiguity.

Alternatively, the option price can be equivalently expressed using the probability density $p_{t,T}$ of S_T under the historical measure $\mathbb{P}$. We have

$$C_t = E_{\mathbb{P}}[\Psi(S_T)|\mathcal{F}_t] = \int_0^\infty \Psi(S_T) M_{t,T}(S_T) p_{t,T}(S_T) dS_T \tag{2.2}$$

with $M_{t,T}$ denoting the pricing kernel or stochastic discount factor (SDF). Strictly speaking, $M_{t,T}$ represents a projected version of a general SDF, which may depend on multiple state variables. Since there is considerable debate among researchers about the relevant pricing kernel state variables, it is quite common to consider the projection $M_{t,T}(S_T)$ or equivalently the projection onto returns $M_{t,T}(S_T/S_t)$. Although this projected pricing kernel is not necessarily identical to the original one, it has exactly the same pricing implications for assets with payoffs depending on S_T such as European put and call options considered in this paper (see Cochrane (2005) for a discussion).

Combining equations (2.1) and (2.2), we can obtain the following relationship:

$$M_{t,T}(S_T) = e^{-r\tau} \frac{q_{t,T}(S_T)}{p_{t,T}(S_T)}. \tag{2.3}$$

Hence, the pricing kernel is equal to the discounted ratio of the risk-neutral and the physical probability density and therefore characterizes the change of measure between $\mathbb{P}$ and $\mathbb{Q}$.

Building on this important result, several approaches have been developed to derive empirical pricing kernels using separate estimates of $p_{t,T}$ and $q_{t,T}$. It is a common finding that the resulting kernel is time-varying, which reflects variation in the relevant state variables. The principal goal of this paper is to understand the economic content and to build a model for this time-variation. In addition, we consider a surface modeling perspective and describe the complete pricing kernel surface as a function of both future return and time-to-maturity.

2.3 Initial pricing kernel surface estimation

In the next section, we will show how to model the time-varying empirical pricing kernel surface with a flexible boosting approach. For this purpose, we need initial estimates of this dynamic surface. As is common practice in the literature, we recover empirical kernels building on a so-called indirect estimation approach, i.e. using separate estimates of the risk-neutral and historical density of the future underlying asset return (see equation (2.3)). The following two subsections contain a description of how we estimate both of these density surfaces.

2.3.1 Risk-neutral density surface estimation

To estimate the option-implied risk-neutral density, one can build on the famous result originally presented in Breeden and Litzenberger (1978). It states that the discounted risk-neutral probability density of the underlying asset at expiry is equal to the second derivative of the European call price function with respect to the strike.

However, fitting the price curve or surface directly is numerically challenging as explained for instance in Brunner and Hafner (2003). Consequently, it is quite common to transform option prices using the famous Black-Scholes formula in order to get the corresponding implied volatilities. Interpolation and smoothing is then performed in this implied volatility space, which has desirable statistical properties. After all, the resulting continuous implied volatility function is mapped back to the original price space where one can rely on numerical differentiation techniques to derive the risk-neutral density.[1]

In our application, we employ the stochastic volatility inspired (SVI) model introduced in Gatheral (2004) to describe the implied volatility smile. Although being a parametric model, it is rather flexible and particularly popular among practitioners. It depends on five parameters $\theta = (a, b, \rho, m, \sigma)$ and describes the total implied variance $\omega(k; \theta)$ for a fixed time to maturity τ as a function of log-moneyness k. We have

$$\omega(k; \theta) = a + b \left[\rho(k - m) + \sqrt{(k - m)^2 + \sigma^2} \right],$$

with $a \in \mathbb{R}$, $b \geq 0$, $|\rho| < 1$, $m \in \mathbb{R}$, $\sigma > 0$ and the obvious condition $a + b\sigma\sqrt{1 - \rho^2} \geq 0$ to ensure that $\omega(k; \theta) \geq 0$ for all $k \in \mathbb{R}$.

The model parameters have the following impact on the total variance function: a determines the general level of total variance, b gives the angle between the left and right asymptotes, increasing ρ induces a counter-clockwise rotation of the smile, σ describes the at-the-money curvature and changing m translates the smile horizontally. For an illustration we refer to Figure 2.1.

An interesting property of the SVI parametrization is its consistency with arbitrage bounds for extreme strikes. More precisely, for a fixed time to expiry τ, the implied Black-Scholes variance $\sigma^2_{BS}(k) = \omega(k)/\tau$ is linear in the log-moneyness k as $|k| \to \infty$ which is consistent with Roger Lee's moment formula (see Lee (2004)).

Our application requires not only single density slices but the whole surface. Thus we rely on the results recently presented in Gatheral and Jacquier (2014). In this paper, they show how to extend the SVI model to get an implied volatility surface free of static arbitrage. Following the ideas mentioned therein, we fit the SVI model for each maturity slice using observed market option prices and thereby ensure that no

---

[1] Note that it is not necessary to assume that the Black and Scholes (1973) model correctly specifies the underlying asset dynamics. It simply serves as computational device to transform data into a measurement space more conductive to apply curve-fitting techniques.

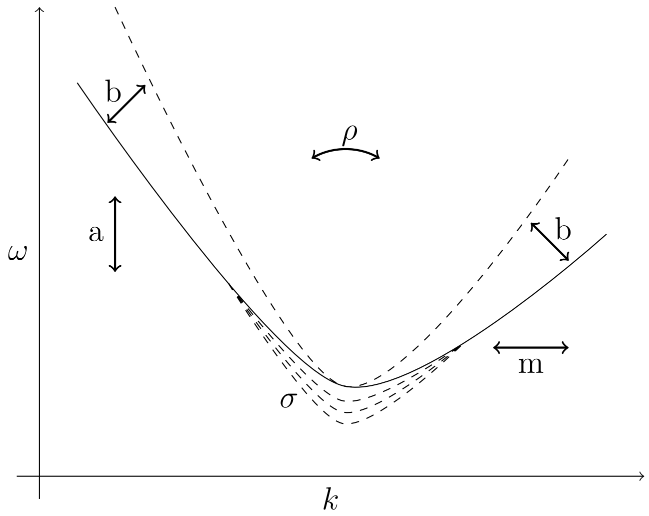


Figure 2.1: Illustration of how the SVI model parameters affect the total variance function.

butterfly arbitrage exists (i.e. the resulting density is positive). In addition, one can avoid calendar spread arbitrage by introducing a big penalty in the objective function if there are crossed lines in the total variance plot.[2]

After having obtained the SVI parametrization for a discrete set of expiries, we apply the interpolation algorithm suggested in Gatheral and Jacquier (2014). More precisely, given two total variance functions $\omega(k, \tau_1)$ and $\omega(k, \tau_2)$ with $\tau_1 < \tau_2$, we use the Black-Scholes formula to calculate undiscounted call prices $C_i = C(F_i, K_i, \tau_i)$ with strike price $K_i = F_i e^k$ and forward F_i $(i = 1, 2)$. The price $C_\tau = C(F_\tau, K_\tau, \tau)$ of a call with maturity $\tau_1 < \tau < \tau_2$ and strike $K_\tau = F_\tau e^k$ is then defined through the equation

$$\frac{C_\tau}{K_\tau} = \alpha_\tau \frac{C_1}{K_1} + (1 - \alpha_\tau) \frac{C_2}{K_2}$$

with

$$\alpha_\tau = \frac{\sqrt{\theta_{\tau_2}} - \sqrt{\theta_\tau}}{\sqrt{\theta_{\tau_2}} - \sqrt{\theta_{\tau_1}}} \in [0, 1]$$

and where θ_τ represents any monotonic interpolation of the at-the-money implied total variance $\omega(0, \tau)$. Gatheral and Jacquier (2014) show that the resulting call respectively implied volatility surface is arbitrage-free if both smiles $\omega(k, \tau_1)$ and $\omega(k, \tau_2)$ are free of butterfly arbitrage and $\omega(k, \tau_2) \geq \omega(k, \tau_1)$ for all $k \in \mathbb{R}$.

The question remains how to extrapolate the surface beyond the range of observed maturities $0 < \tau_1 < \cdots < \tau_n$ with $n \geq 1$. For expiries $\tau < \tau_1$, one may interpolate with the procedure described above and using the fact that for $\tau = 0$ the value of

---

[2] Note that a volatility surface is free of calendar spread arbitrage if and only if $\partial_\tau \omega(k, \tau) \geq 0$ for all $k \in \mathbb{R}$ and $\tau > 0$.

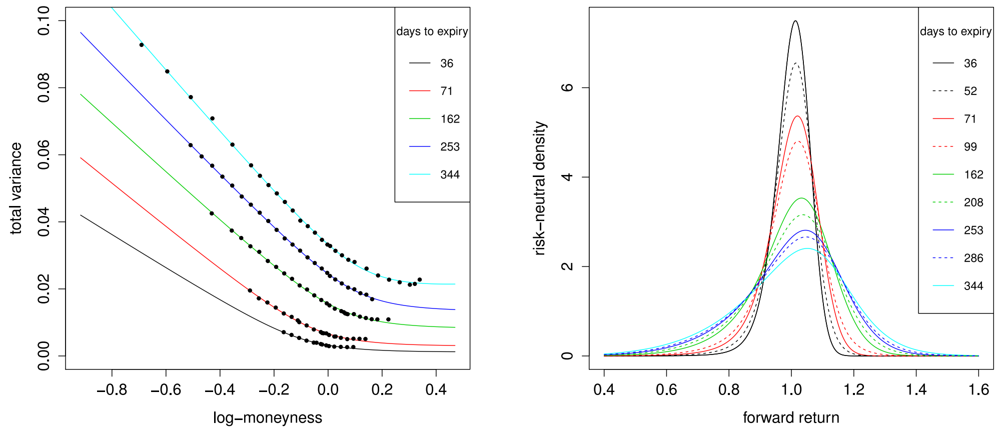


Figure 2.2: The figure illustrates the SVI model fit using option data on July 9, 2003. The left panel shows the total variance functions for all available maturities and the right panel presents the corresponding risk-neutral densities. Densities with dashed lines are obtained using the interpolation algorithm mentioned in the text.

a call is just its intrinsic value. For extrapolation beyond τ_n, Gatheral and Jacquier (2014) suggest to use a monotonic increasing extrapolation of θ_τ (asymptotically linear in time seems to be reasonable) and then to define for $\tau > \tau_n$

$$\omega(k,\tau) = \omega(k,\tau_n) + \theta_\tau - \theta_{\tau_n}.$$

We provide an illustration of this estimation procedure in Figure 2.2. The left panel shows the fit of the total variance function for all option maturities available on the arbitrarily selected day July 9, 2003. The right panel presents the risk-neutral densities obtained using the estimated SVI parameters and the interpolation algorithm mentioned above. According to our experience does this approach yield very reasonable risk-neutral density surface estimates. Only when performing the extrapolation for expiries below τ_1, we had some problems to obtain accurate and smooth densities. For this reason, we will not consider this extrapolation in our empirical application.

For robustness purposes and to further check the accuracy, we compared the estimated density surfaces with those obtained using the semi-nonparametric estimator of the call-option price surface presented in Fengler and Hin (2014), which takes into account strike and time-to-expiry no-arbitrage constraints. We usually get very similar results. However, we sometimes find bimodal shapes (in particular for larger maturities) when using the call price surface estimator of Fengler and Hin (2014). As stated in Fengler and Hin (2014), bimodality is possibly an artefact from the sparse data distribution. Therefore, we prefer the SVI approach and will use it in the empirical application later on.

2.3.2 Historical density surface estimation

The historical density p is typically estimated using past return observations of the underlying index. For example, Grith et al. (2013a) employ a nonparametric kernel density estimator similar to Aït-Sahalia and Lo (2000) based on historical return realizations. Alternatively, it is also popular to make additional modeling assumptions regarding the evolution of the underlying stock price process and then to use a simulation-based approach to recover an estimate of p.

Similar to Polkovnichenko and Zhao (2013), Barone-Adesi et al. (2014) and others, we favor the latter approach and thus model the daily log-returns of the underlying S&P 500 index with an asymmetric GJR-GARCH (Glosten et al. (1993)).[3] It is

$$r_t = \ln(S_t/S_{t-1}) = \mu + \epsilon_t, \quad \epsilon_t \sim h(0, \sigma_t^2) \tag{2.4}$$

and

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha\epsilon_{t-1}^2 + \gamma \max(0, -\epsilon_{t-1})^2. \tag{2.5}$$

This model specification captures the most important stylized facts of financial return series. That is, the conditional volatility is stochastic and mean-reverting and when $\gamma > 0$ the model also accounts for the leverage effect.[4] The empirical innovation density h is separated into a time-varying component σ_t and the time-invariant distribution of standardized innovations $z_t = \epsilon_t/\sigma_t$. This standardized innovation density incorporates skewness and excess kurtosis and other extreme return behaviors which are not captured in a normal density.

To estimate the model parameters, we apply a pseudo maximum likelihood approach under the assumption of normal innovations. Bollerslev and Wooldridge (1992) show that this technique yields consistent parameter estimates even if the true innovation distribution is not normal.

Using the estimated model parameters, the conditional variance and the standardized innovations, we can simulate return paths. For a given time point t such a path is obtained by randomly selecting an estimated standardized innovation $z_{[1]}$, updating the conditional variance σ^2_{t+1}, drawing a second innovation $z_{[2]}$, updating the conditional variance σ^2_{t+2}, and continuing up to the desired time horizon. Repeating this procedure a sufficient number of times, we obtain a set of simulated returns and apply a kernel density estimator to get an estimate of the future return density p. This method, first introduced in Barone-Adesi et al. (1998) to compute portfolio risk measures, is now widely used and generally called filtered historical simulation (FHS) in the financial econometrics literature (see, for example, Barone-Adesi et al. (2008)).

---

[3] Rosenberg and Engle (2002) fit a number of GARCH models to daily S&P 500 index returns and find that the GJR-GARCH model describes the data best. Since then, this model or similar asymmetric GARCH specifications have been widely used in the empirical pricing kernel literature to model the underlying return process.

[4] The term leverage effect was introduced by Black (1976) and is used to describe the asymmetric reaction of volatility in response to good and bad news (excess returns).

2.4 Modeling the time-varying pricing kernel surface

In this section, we introduce our approach to model the time-varying pricing kernel surface. We start with the model formulation and then show how to estimate it in a second part. Finally, we present the exogenous factors used to explain the dynamics of the modeled surface.

2.4.1 Model formulation

Our goal is to describe the time-series of the projected pricing kernel surface $M_t(r, \tau)$ which is a function of both future gross return r and time-to-maturity τ. For this purpose, we consider the following model:

$$M_t(r, \tau) = K(r, \tau, \Phi_{t-1}) + \epsilon(r, \tau), \tag{2.6}$$

where $K(\cdot)$ represents a flexible functional specification able to capture the shapes of the observed surfaces and ϵ_t is an error term. We make the assumption that the surface dynamics can be explained using a set of time-dependent exogenous factors Φ_{t-1}. Time-lagged factors are used since we also have in mind to investigate the predictive performance of the fitted kernels in our empirical application.

Note that this model formulation should be considered as an approximation because, strictly speaking, there is no realization of the pricing kernel surface available. However, we can use the estimates of this surface introduced in the previous section similar to what has been done in Grith et al. (2013a).[5]

Our model (2.6) with possibly many explanatory variables will be estimated using a classical boosting algorithm. Thus, we restrict the function $K(\cdot)$ to be an additive expansion of the form

$$K(r,\tau,\Phi_{t-1}) = K_0(r,\tau) + \sum_{j=1}^{J} B_j(r,\tau,\Phi_{t-1}),$$

where K_0 denotes an appropriate starting model (assumed to be time-invariant and further explained in the next section) and each B_j represents a pre-specified statistical procedure also known as base learner in the machine-learning literature. Common examples of such base learners are regression trees or splines. However, one has to ensure that the chosen base learner is “weak” (i.e. it does not involve a too large number of parameters to be estimated) in order to avoid an immediate overfit within the first iterations of the algorithm.

In this study, we take simple regression trees as base learners. A regression tree creates a binary partition of the predictor space using a sequence of logical if/then conditions and then models the dependent variable in each partition cell with a constant. Hence we get

$$B_j(r,\tau,\Phi_{t-1}) = \sum_{l=1}^{L} a_j^{(l)}\, \mathbb{I}_{\left[\Psi_t \in \mathcal{R}_j^{(l)}\right]}, \quad \Psi_t = \{r,\tau,\Phi_{t-1}\}$$

with L denoting the number of end nodes and $\mathcal{R}_j^{(l)}$ the partition cells of the regression trees and where $a_j^{(l)}$ design some constant location parameters.

Since the influential work of Breiman et al. (1984), trees became very popular in computational statistics and in the machine-learning community mainly because of their simplicity and interpretability. A feature particularly appealing for our application is the trees ability to choose the most relevant predictors among a set of candidates. Therefore, we will be able to answer our main question regarding the driving factors of the pricing kernel surface. Additionally, trees model nonlinearity in a parsimonious and interpretable way, which is an advantage over other variable selection techniques.

---

[5] Or expressed differently, we model the empirical pricing kernel which serves as an approximation for the true stochastic discount factor.

2.4.2 Model estimation

We estimate the model with a classical boosting algorithm, which aims at combining weak predictors in order to obtain a powerful model at the end. Thereby we exploit the functional gradient descent (FGD) representation of boosting as described in Friedman (2001).[6] More specifically, we apply the FGD technique with a squared error loss function which yields a simple boosting approach typically referred to as L_2-boosting. The idea behind L_2-boosting is rather intuitive. Having estimated an appropriate starting model, the algorithm performs a stage-wise residual fitting using the base learners (simple regression trees in our case). Recently, Audrino and Colangelo (2010) have successfully applied boosting based on regression trees to model the dynamics of the implied volatility surface. In addition, nice asymptotical properties have been shown for L^2-boosting and we refer to Bühlmann and van de Geer (2011) and references therein for a more detailed discussion.

For a formal description of the algorithm, let us consider time-varying estimates of the pricing kernel surface obtained using the methods given in Section 2.3. We write $(r_{ti}, \tau_{ti}, m_{ti})$ to denote the return, time-to-maturity and pricing kernel observations at day t with $t \in \{1, \dots, T\}$ and $i \in \{1, \dots, N\}$. Hence we assume a constant number of daily observations N. In addition, we suppose to have a time-series of exogenous factors Φ_t.

In a first step, we fit the starting model K_0. Choosing K_0 appropriately is important and it must obey two restrictions. On the one hand, it should be not too complex in order to avoid an immediate overfit. On the other hand, it must be reasonable since the boosting algorithm is going to improve it locally using regression trees as base learners. As in Chapter 1 we use a power pricing kernel parametrization for K_0 conform with a neoclassical framework featuring constant relative risk aversion (CRRA), i.e.

$$K_0(r, \tau) = \theta_0 r^{-\theta_1}, \tag{2.7}$$

where θ_0 is a discount factor measuring the degree of impatience and θ_1 designs the coefficient of relative risk aversion.

Our starting kernel is time-invariant and thus may be best interpreted as an “average” pricing kernel over time. The time-variability will be taken into account in the second step using the regression trees. Moreover, we assume K_0 to be constant for different maturities. Obviously, this is a simplification. But again, such differences (if existent) will be modeled in the second step of the algorithm.

---

[6] For a short introduction to FGD we refer the reader to the appendix.

To obtain estimates of θ_0 and θ_1, we consider the logarithmic version of equation (2.7)

$$\log(K_0(r,\tau)) = \log(\theta_0) - \theta_1 \log(r)$$

and thus regress the logarithm of the empirical pricing kernel $\log(m_{ti})$ on a constant and the log-return $\log(r_{ti})$. The fitted intercept and slope parameters can then be employed to get estimates $\hat{\theta}_0$ and $\hat{\theta}_1$.

In a second step of the algorithm, we perform a stage-wise residual fitting using simple regression trees. The set of predictors thereby includes r, τ and the exogenous factors Φ. Hence we are able to capture arbitrary shapes and also time-variability by means of the time-dependent exogenous factors.

The estimation procedure as described previously can be summarized in the following algorithm:

L_2-boosting with tree learners

Step 1 (initialization). Estimate the initial pricing kernel $\hat{K}_0$ using the realizations $(r_{ti}, \tau_{ti}, m_{ti})$ for $t \in \{1, \ldots, T\}$ and $i \in \{1, \ldots, N\}$. Set $j = 1$.

Step 2 (residual fitting). Compute the residuals

$$u_{ti} = m_{ti} - \hat{K}_{j-1}(r_{ti}, \tau_{ti}, \Phi_{t-1}), \quad t \in \{1, \ldots, T\}, i \in \{1, \ldots, N\}.$$

Then, fit the regression tree B_j to the current residuals by least squares, i.e. determine the best partition $\hat{\mathcal{R}}_j^{(l)}$ and the best coefficients $\hat{a}_j^{(l)}$ (for $1 \leq l \leq L$) in order to get $\hat{B}_j$. Update the current pricing kernel estimate

$$\hat{K}_j(r, \tau, \Phi_{t-1}) = \hat{K}_{j-1}(r, \tau, \Phi_{t-1}) + \hat{B}_j(r, \tau, \Phi_{t-1}).$$

Step 3 (iteration). Increase j by one and iterate step 2 until $j = J$. This produces the estimate

$$\hat{M}_t(r, \tau) = \hat{K}(r, \tau, \Phi_{t-1}) = \hat{K}_0(r, \tau) + \sum_{j=1}^{J} \hat{B}_j(r, \tau, \Phi_{t-1}).$$

The number of iterations J is a complexity parameter of the estimator. It must be carefully selected to avoid overfitting. Typically, it is estimated by minimizing approximations of the expected prediction error. In our empirical application, we will use standard cross-validation to determine an optimal value for J.

In addition, it is often desirable to make a base learner sufficiently weak (having low complexity). A simple but effective solution to achieve this is via shrinkage toward zero. Hence, the update in Step 2 of the algorithm is replaced by

$$\nu \hat{B}_j(r, \tau, \Phi_{t-1}), \quad 0 \leq \nu \leq 1,$$

which reduces the variance (a complexity measure) of the base learner by the factor ν^2.

2.4.3 Pricing kernel driving factor candidates

Principally, one could include an arbitrary and even big number of possible pricing kernel driving factors, since our boosting algorithm based on trees has the ability to automatically select the most relevant ones. Given previous empirical evidence, see for example Rosenberg and Engle (2002), Grith et al. (2013a) and Metaxoglou and Smith (2014), we will try to relate the shape of the pricing kernel to measures of volatility, financial and macroeconomic variables, sentiment indicators and past returns of the underlying S&P 500 index. Specifically, we include in the pool of possible explanatory variables Φ the following measures:

Volatility: Following the arguments in Chabi-Yo (2012) and Christoffersen et al. (2013), volatility is a particularly relevant pricing kernel state variable. Thus, we take the CBOE Volatility Index (VIX) as a proxy and use it as a potential pricing kernel driving factor in our study.

Financial and macroeconomic variables: Similar to Rosenberg and Engle (2002) and Grith et al. (2013a), we include several variables known to be related to the state of the economy: the level of the short-term interest rate (ShortRate) using Federal Reserve's tree-month constant maturity Treasury yield; the credit spread (CredSpread) defined as the difference between Moody's long-term Baa corporate bond yield and the Federal Reserve's 30-year constant maturity Treasury yield; the yield curve slope (YCSlope) referring to the 30-year constant maturity Treasury yield minus the short-term interest rate; and the logarithm of the S&P 500 index as a proxy for consumption. Furthermore, we use the seasonally adjusted Consumer Price Index (CPI) for all urban consumers and the unemployment rate (UNEMPL) and then compute year-on-year log differences to construct a measure for inflation and the unemployment growth rate, respectively.

Sentiment factors: According to Shefrin (2008), the logarithm of the pricing kernel can be decomposed into a sum of a fundamental component and a behavioral component, which reflects investor sentiment. Consequently, we also employ sentiment factors to explain the pricing kernel dynamics. We consider three different measures: the Baker-Wurgler sentiment index (BWSent), the Yale/Shiller crash confidence index (CrashConfInd) and the Michigan Consumer Confidence Index (CSent).

Past returns: Finally, we add S&P 500 log-returns over different time horizons (one month (M1Ret), half a year (M6Ret), one year (Y1Ret) and two years (Y2Ret)) as possible explanatory variables in our algorithm.

Figure 2.3 shows the weekly time series of our explanatory variables Φ, covering the period January 2002 to December 2010. We take observations each Wednesday and consider an adjacent day when no data is available. Some variables like the sentiment factors, CPI and UNEMPL are only monthly updated. In this case, we assume the series to be constant within one month, i.e. no interpolation is performed. We find a significant variation over time for all variables. Based on this time-variation, our boosting approach with trees will determine different regimes, which allows modeling the pricing kernel dynamics.

2.5 Empirical analysis

In this section, we present the outcomes of our empirical analysis. We start by introducing the S&P 500 option data used to infer the conditional risk-neutral density surfaces. In a second step, we show estimates of the empirical pricing kernel surface obtained with the estimation procedure presented in Section 2.3. The principal results of this study are then contained in Section 2.5.3. There, we illustrate our dynamic model for the pricing kernel, determine the main driving factors and show their impact on the pricing kernel shape. Finally, we close this section with a short investigation regarding the predictive performance of our model.

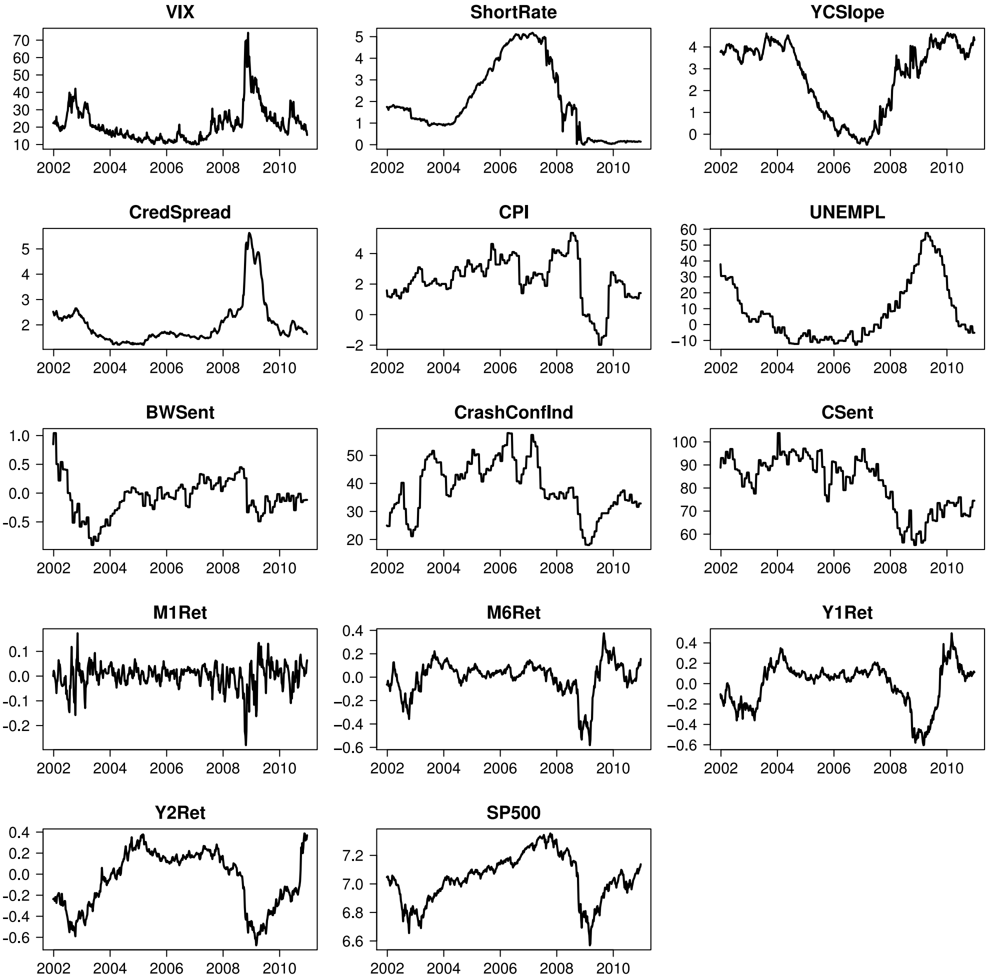


Figure 2.3: The figure shows the weekly time series of our explanatory variables employed to describe the dynamics of the empirical pricing kernel surface. The considered time period is January 2002 to December 2010 and the variables are defined as mentioned in the text.

2.5.1 Option Data

We use European options written on the S&P 500 index (SPX) to calibrate the SVI implied volatility surfaces introduced in Section 2.3.1. The market for SPX options is one of the most active index options markets in the world. At any time, expiry dates of traded options are the three near-term months and three additional months from the March, June, September, December, quarterly cycle. Strike price intervals are 5 points and 25 points for far months.

In our analysis, we use option cross-sections for each Wednesday from January 2002 to December 2010 downloaded from OptionMetrics. In case a particular Wednesday was a holiday, we take data from an adjacent trading day. The closing mid quote, defined as the average of best bid and ask quote at market closing, is taken as option price. For liquidity reasons, we discard in-the-money (ITM) options, options with expiries less than 10 or more than 365 calendar days and options with bids below 0.5 dollar.

We use interest rates as obtained from OptionMetrics and then apply the procedure suggested in Aït-Sahalia and Lo (1998) to address the problems of non-synchronous prices between the option and underlying asset and the unobserved dividend process. The idea is to use the put-call parity

$$C + Ke^{-r\tau} = P + Fe^{-r\tau},$$

which holds independently of any option pricing model. In this equation, C and P denote prices of a call and put with strike K and time to maturity τ, r is the risk-free interest rate and F the forward. We derive implied forwards for all dates and expiries separately using a put-call pair closest to at-the-money. Finally, one can back out the unobserved dividend yield δ via the spot-forward parity

$$F = Se^{(r-\delta)\tau}.$$

Table 2.1 summarizes the characteristics of our option data set for different forward-moneyness (K/F) and time to maturity categories. In total, we observe 102'114 OTM option prices thereof 62% puts and 38% calls. We further report the mean and standard deviation for prices and implied volatilities. In particular for higher maturities, puts are somewhat more expensive than calls. In addition, we observe the well-known volatility smile pattern and its corresponding term structure (i.e. the smiles tend to become flatter for larger expiries).

2.5.2 Empirical pricing kernel estimation

As mentioned previously, we derive estimates of the empirical pricing kernel surface $\hat{M}_t(r,\tau)$ using separate estimates of the risk-neutral and historical density surfaces of the future underlying asset return denoted $\hat{q}_t(r,\tau)$ and $\hat{p}_t(r,\tau)$, respectively. Whereas τ designs the time to maturity, r represents the forward gross return. We take forward gross returns instead of spot returns because Polkovnichenko and Zhao (2013) argue that these are better proxies for the returns on the total wealth process in case of stochastic dividend processes.

Using the option cross-sections described above, we calibrate each week SVI volatility smiles for all available maturities and then derive an estimate of the complete risk-

Table 2.1: Summary of our option data sample used to infer the risk-neutral density surfaces. The data set contains 470 weekly cross-sections of SPX European options from January 2002 to December 2010 with expiries between 10 and 365 days and is obtained after applying the filtering criteria described in the text. We report mean and standard deviation for prices and implied volatilities and the number of observations according to different categories regarding forward moneyness and option maturity.

Moneyness		Maturity (days)					
		Less than 60		60 to 160		More than 160	
K/F		Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
<0.85	Put Price ($)	3.04	3.55	5.29	5.56	11.52	11.40
	IV (%)	45.03	15.68	37.18	11.46	31.27	10.21
	Observations	5552		9321		11019	
0.85-1.0	Put Price ($)	10.83	10.51	24.49	15.90	46.36	22.80
	IV (%)	24.62	10.36	23.68	8.22	22.08	6.80
	Observations	17481		11433		8400	
1.0-1.15	Call Price ($)	10.89	10.80	20.38	16.72	40.42	25.08
	IV (%)	19.27	9.53	19.15	7.59	18.27	6.45
	Observations	11847		10430		7555	
>1.15	Call Price ($)	3.87	3.95	4.93	5.78	9.37	10.34
	IV (%)	36.68	9.33	25.16	8.13	20.28	6.25
	Observations	867		2440		5769	

neutral density surface based on the algorithm presented in Section 2.3.1. The accuracy of these densities is obviously strictly related to the fitting quality of the SVI model. To assess its quality, we show in Table 2.2 root mean squared errors for implied volatilities and prices (averaged over the whole sample period) for the same moneyness/maturity categories as in Table 2.1. We find that the fit is very good independently of moneyness and time to maturity. Hence, we conclude that our estimates $\hat{q}_t(r, \tau)$ are sufficiently accurate. Additionally, it is worth to be mentioned that we only consider time to maturities in the interval $[30, 240]$ (in calendar days) for our application. The lower bound is chosen since we had some difficulties to get accurate densities when applying the extrapolation for maturities below the lowest observed expiry, whereas the upper bound is somewhat arbitrary (i.e. other upper bounds could be used too).

The second input, needed to derive the empirical pricing surface, is an estimate of the conditional historical return density p_t. For this purpose, we calibrate the GJR-GARCH model presented in (2.4)-(2.5) every week using observed daily returns within a rolling window of length 3500 trading days, similarly to what has been done in Barone-Adesi et al. (2008). The mean and standard deviation of the estimated model parameters over the period January 2002 and December 2010 are shown in Table 2.3. They are in line with values obtained in Barone-Adesi et al. (2014), which consider

Table 2.2: This table illustrates the performance of the SVI model fit. We report root mean squared errors for implied volatilities (IV RMSE, in percent) and prices (Price RMSE, in dollars) using SPX option cross-sections each Wednesday from January 2002 to December 2010. Errors are shown for the same moneyness/maturity categories as in Table 2.1.

Moneyness	Maturity (days)					
	Less 60		60 to 160		More than 160	
K/F	Price RMSE	IV RMSE	Price RMSE	IV RMSE	Price RMSE	IV RMSE
<0.85	0.01	0.18	0.03	0.09	0.12	0.07
0.85-1.0	0.03	0.04	0.08	0.02	0.23	0.02
1.0-1.15	0.08	0.09	0.37	0.06	0.34	0.02
>1.15	0.12	0.73	0.04	0.09	0.12	0.04

Table 2.3: This table shows parameter estimates of the GJR-GARCH model estimated on each Wednesday from January 2002 to December 2010 using a pseudo maximum likelihood approach with normal innovations and 3500 historical return observations.

GJR-GARCH model parameter estimates					
	$\mu \times 10^4$	$\omega \times 10^6$	β	$\alpha \times 10^3$	γ
Mean	2.70	1.26	0.93	2.71	0.12
Std. dev.	0.80	0.27	0.01	3.23	0.02

a time period close to ours. Note that we do not report here means and standard deviations of the standardized empirical innovations, but they are close to 0 and 1, respectively. This means that the GJR-GARCH models are correctly identified. Additionally, we observe that the empirical innovation densities incorporate skewness and excess kurtosis and are highly non-normal. We omit further details for brevity but descriptive statistics are available from the author upon request.

Based on the obtained parameter values and the empirical innovations, we then employ the filtered historical simulation approach explained in Section 2.3.2 to generate returns for any desired time-horizon τ. The corresponding density is finally obtained via an application of a kernel density estimator.

Given estimates $\hat{q}_t$ and $\hat{p}_t$ for both density surfaces and every week t, estimates of the pricing kernel surface can be determined by calculating their discounted ratio. Consequently, we show in Figure 2.4 the time series of the empirical pricing kernel obtained over the period January 2002 to December 2010 for three different maturities (50 days, 100 days and 200 days). Note that for clarity of presentation, we do not plot the kernel for every week, but only once per month.

Interestingly, the shape of the pricing kernel is rather consistent between different maturities, meaning that investors' preferences do not change a lot with respect to their

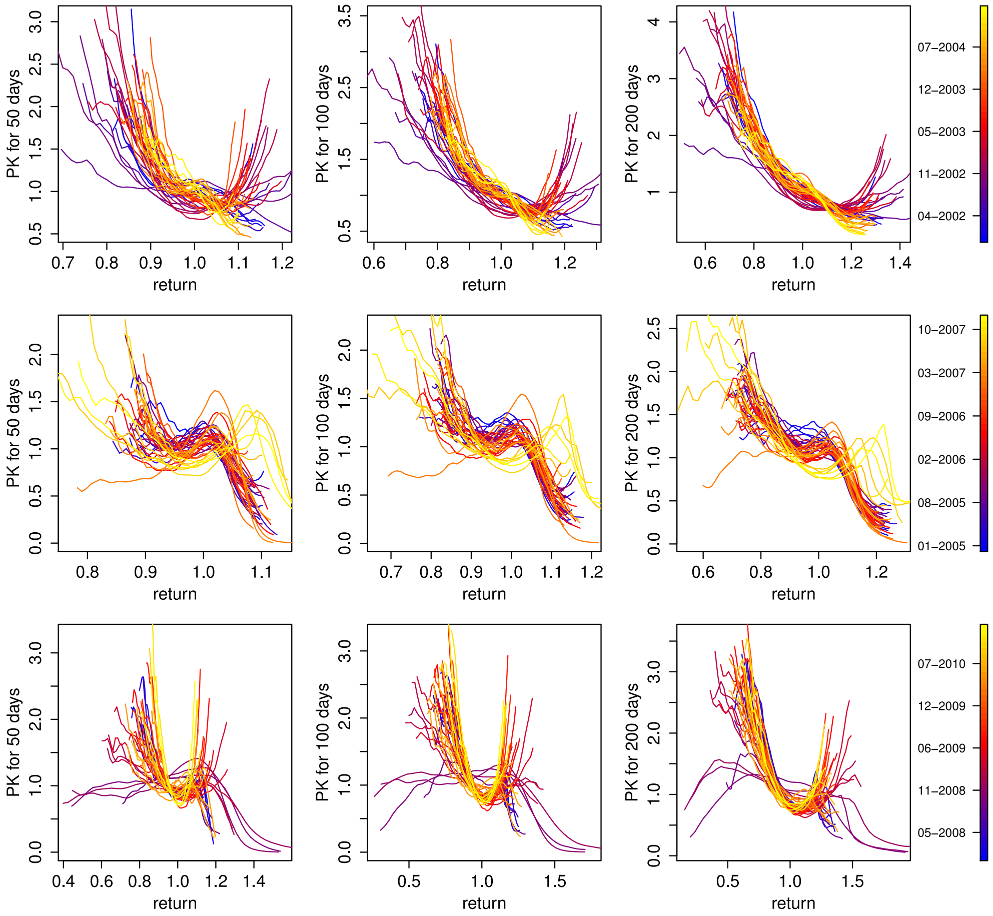


Figure 2.4: This figure shows empirical pricing kernel estimates for three different maturities: 30 days (left), 100 days (middle) and 200 days (right). They were obtained with the techniques mentioned in the text and cover the whole sample period ranging from January 2002 to December 2010. For clarity of presentation, we do not plot the kernel for every week, but only once per month.

investment horizon. In contrast, there is a substantial time variation, i.e. investors change their risk attitude over time. In line with findings presented in Polkovnichenko and Zhao (2013), we observe declining functions at the beginning of 2002 and then U-shaped kernels until early 2004. During 2005 and 2007 we find another typical shape. Namely, pricing kernel curves having a bump at around one on the gross return scale in nearly all cases. Finally, towards the end of the sample period, the kernel features again an U-shape. Furthermore, one can also see the effect of the recent financial crisis, which is reflected by very particular pricing kernel shapes (see for example the inverted U-shaped curves with very large support at the end of 2008 or beginning 2009).

2.5.3 Fitting the time-series model

Given the estimates of the empirical pricing kernel surface presented above, we can now model the observed dynamics with our boosting approach based on trees. For that purpose, we construct a set of observations $(r_{ti}, \tau_{ti}, m_{ti})$ with $i \in \{1, \dots, N\}$ and $m_{ti} = \hat{M}_t(r_{ti}, \tau_{ti})$ for every week t in our sample.[7] Specifically, we extract each week pricing kernel slices for 8 different maturities to represent the whole surface, which is preferred over sampling uniformly the two-dimensional domain of definition simply for computational reasons. To get the specific maturities, we divide the considered maturity interval $[30, 240]$ (in calendar days) into 8 subintervals of equal length and then randomly select an expiry in each of them. In comparison with arbitrarily taking expiries in $[30, 240]$, this procedure has the advantage that there are never "clustered" maturities. Consequently, we obtain a picture over the whole maturity range every week, while still having some randomness in the procedure ensures that the whole time interval will be covered when aggregating observations over time. For each of the resulting 8 expiries, the corresponding empirical pricing kernel is evaluated in 30 equally spaced return values with bounds determined by quantiles of the return distribution for this particular time horizon. We take the 1%- and 99%-quantile under the historical distribution to avoid unwanted numerical effects in the tails. Using this procedure, we finally get $N = 240$ pricing kernel evaluations per week and repeating these calculations for all 470 weeks in our sample period yields a total of 112'800 observations.

With these pricing kernel realizations at hand, it is possible to fit our dynamic model. Since we have in mind to investigate whether and how the impact of the relevant driving factors changes over time, we employ a rolling window estimation approach. More precisely, we take a window length of one year. For robustness reasons, we repeat the estimation with other widow sizes too. Some additional results for two years are thereby contained in the appendix.

In a first step, the algorithm always fits the initial kernel K_0 as specified in equation (2.7). This kernel does not depend on time t and expiry τ (inside the rolling window) and thus may be best interpreted as an average power pricing kernel over all weeks and time to maturities. Afterward, the algorithm continues with a stage-wise residual fitting using simple regression trees. In our application, we take trees with $L = 6$ end nodes and use a shrinkage value $\nu = 0.1$ to further reduce the complexity of the base learner.[8] Given that the regression tree employ beside r and τ also the time-varying exogenous factors Φ as predictors, we can capture the dynamics of the observed kernels.

---

[7] Here, $\hat{M}_t$ refers to the empirically derived option-implied pricing kernel obtained in Section 2.5.2. However, from now on, we consider m_{ti} to be observations of the true pricing kernel M_t and use $\hat{M}_t$ to denote its approximation obtained with our boosting algorithm, which is consistent with the notation in Section 2.4.

As mentioned previously, determining an optimal value for the complexity parameter J (the number of boosting iterations) is important in order to avoid overfitting. We use 4-fold cross-validation to obtain an optimal value $\hat{J}$. To be precise, it is a sort of blocked cross-validation given that we arbitrarily allocate the block of weekly observations to one of the 4 complementary subsamples used for cross-validation.

In order to see how the algorithm works and to describe the performance, we have estimated our model separately for each year from 2002 to 2010. The corresponding outcomes are presented in Figure 2.5 by means of an arbitrarily selected day per year. More precisely, the observed pricing kernel and its fitted counterpart are shown for three different maturities. These are the expiries randomly selected in the first, forth and seventh subinterval (see explanation above) and thus differ from week to week. We find that although departing from a rather simple starting kernel, our model is able to capture the various shapes observed over the period January 2002 to December 2010. Additionally, we check for a fixed time to maturity $\tau = 100$ days whether the algorithm can account for the dynamics observed within a year. The results presented in Figure 2.6 provide evidence that this is indeed possible. Certainly, these figures only allow getting a qualitative picture of the fit. However, quantifying the error by means of pricing kernel differences is difficult to interpret. Therefore, we will analyze implied option price differences in our forecasting application later on.

However, for the moment, let us put the focus on our main question regarding the relevant driving factors of the pricing kernel surface. For that purpose, we estimate the model week by week from January 2003 to December 2010 using a rolling window of length one year (i.e. 52 weeks). The relative influence of each predictor variable is then determined by using the measure introduced in Friedman (2001). That is, the importance $\mathcal{I}_i$ of predictor i for J boosted trees B_j $(1 \leq j \leq J)$ is defined as

$$\mathcal{I}_i = \left(\frac{1}{J} \sum_{j=1}^{J} \mathcal{I}_i^2(B_j) \right)^{0.5} ,$$

---

[8] These values are conform to the requirements mentioned earlier and close to those employed in similar studies (see for instance Audrino and Colangelo (2010)). It would also be possible to determine optimal values using for example cross-validation. Because of computational reasons and since we do not expect large improvements, we do not do that.

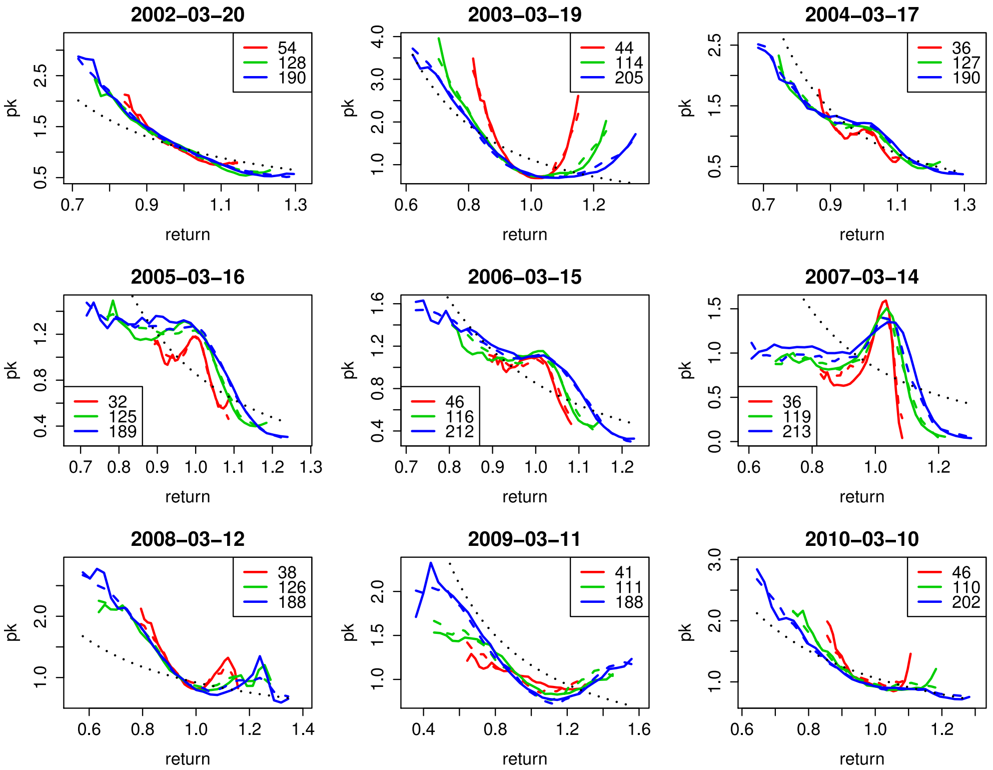


Figure 2.5: The figure presents the fitting quality obtained with our tree-based boosting algorithm. Estimation has been done for each year separately. For some randomly selected dates, we plot the estimated initial pricing kernel (solid) and its fitted counterpart (dashed) for three different maturities whose values are resumed in the legend. The black dotted line indicates the starting power pricing kernel.

where $\mathcal{I}_i(B)$ designs the measure of relevance proposed by Breiman et al. (1984). It is

$$\mathcal{I}_i^2(B) = \sum_{k=1}^{L-1} \hat{\iota}_k \mathbb{I}_{[v_k=i]},$$

where the summation goes over the nonterminal nodes k of the tree B with L leaves, v_k is the splitting variable associated with node k and $\hat{\iota}_k$ denotes the corresponding empirical improvement in squared error as a result of the split.

Figure 2.7 shows the obtained variable importance measures over time. Note that the values are scaled such that they sum up to 100 over all considered predictors. As expected, we observe that the relative influence of the single predictors substantially varies over time. Of course, the return is always the most important variable, which is a consequence of the fact that we start from a rather simple starting model. Its influence ranges from about 40% at the beginning of our considered time period (with decreasing kernels relatively close to the starting model) and goes up to 85% during the crisis 2008. Interestingly, apart from 2004, the time to maturity has a relatively

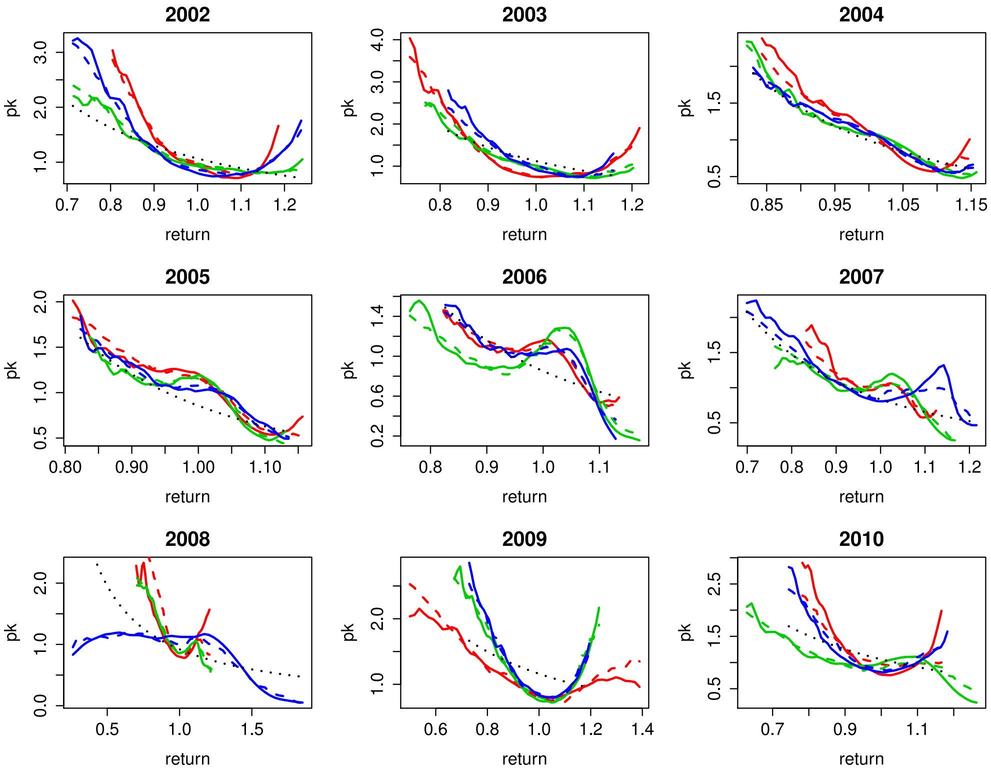


Figure 2.6: The figure shows how the tree-based boosting algorithm can capture the pricing kernel dynamics. For each year and a fixed time to maturity 100 days, we plot the estimated initial pricing kernel (solid) and its fitted counterpart (dashed) for three different weeks of a year: first week (red), 26th week (green) and 50th week (blue). The black dotted line indicates the starting power pricing kernel.

low influence, which conforms with the findings presented in Figure 2.4.

If we now turn towards the time-dependent predictor variables, we can identify the relevant driving factors. Not surprisingly, past returns seem to explain a part of the pricing kernel variation. The figure shows that the return over the last six months is particularly important at the beginning and towards the end of the sample period, whereas the one month return has an influence peak during the crisis. As stated in Shefrin (2008), the pricing kernel can be decomposed into a fundamental part and sentiment part. Thus we may expect that sentiment factors are also relevant drivers. The figure indicates that they have indeed an influence, however, not all the time. The Yale/Shiller crash confidence index is particularly relevant before the recent crisis starts, which is understandable given that it measures the probability of a market crash.[9] In contrast, the Baker-Wurgler sentiment series seems to explain some time-variability on the empirical pricing kernel at the beginning of the considered time period. Referring to Figure 2.3, we observe a decreasing series at that time, meaning

---

[9] A detailed description of the index can be found on the Yale School of Management web page under the rubric data.

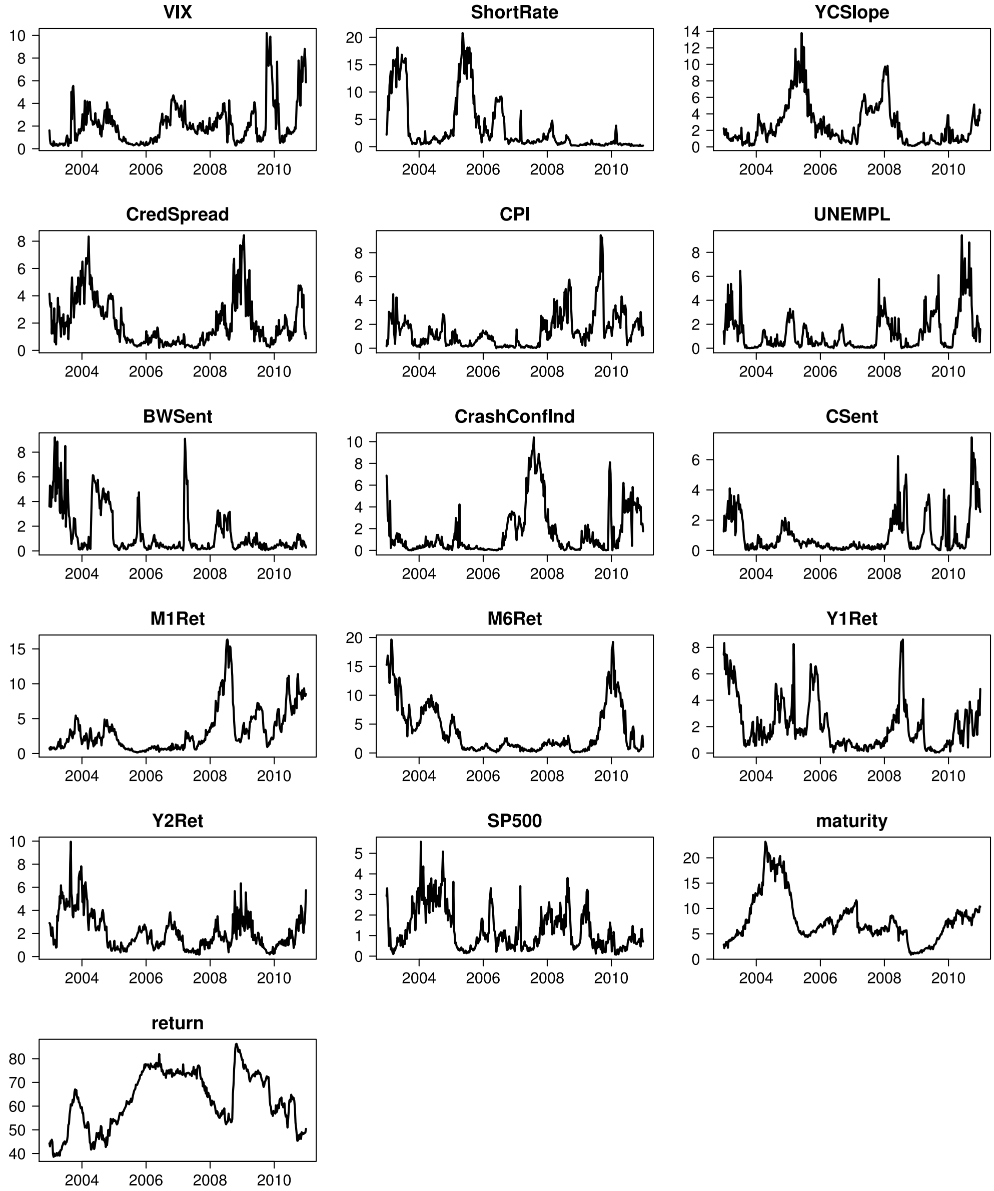


Figure 2.7: This figure shows the relative influence (measured as explained in the text) of the different explanatory variables obtained from our boosting algorithm based on trees. To estimate the model, we use a rolling window of length one year and the time axis indicates the endpoint of this period.

that excessive optimism for stocks diminishes (see Barone-Adesi et al. (2014) for an empirical justification of this result). Moreover, we find that financial and macroeconomic variables also play a pivotal role in shaping/driving the pricing kernel's dynamics (i.e. forming investor's preferences) over time. At the beginning of the sample period (2003-2006), for example, it seems that investors used the negative correlation between bonds (short rate + yields) and stocks (S&P 500) to constantly rebalance their portfolios toward less risky and more liquid securities. The recent financial crisis and the (almost) zero interest rates during the last years forced investors to seek alternative sources of assessing uncertainty. Indeed, Figure 2.7 clearly indicates that since late 2008, investors started to pay more attention to macroeconomic fundamentals such as inflation (CPI) and unemployment (UNEMPL) (see, for example, David and Veronesi (2013) for a discussion on that topic). Interestingly, we observe that the relative importance of the VIX has grown significantly after the financial crisis, revealing that investors became more protective/risk averse in the last years.

As a robustness check, we repeat the same analysis with a rolling window size of two years. The respective results are shown in Figure 2.12, contained in the appendix. The outcomes are rather similar, but obviously, the resulting influence series are more persistent.

In order to analyze the effect of the different predictors on the pricing kernel shape, we present so-called partial dependency plots (see Friedman (2001)). These plots show the effect of a variable on the response (pricing kernel in our case) after accounting for the average effects of all other variables in the model. We select some of the most influential variables and calculate their marginal effects for that time window where they had the highest relative influence (according to Figure 2.7). The resulting partial dependencies are summarized in Figure 2.8. Again, we do the same analysis for a two year estimation window too. These results are presented in Figure 2.13, which can be found in Appendix 2.B.

We show the effects for the left ($r < 0.9$), central ($0.9 \leq r \leq 1.05$) and right part ($r > 1.05$) of the pricing kernel, separately. Not surprisingly, the effects of all predictors are particularly pronounced at the left tail. Further, we observe that the impact of the predictors is typically quite similar in the tails and differs for the central part. Moreover, we find that the response of the pricing kernel to the various explanatory variables is often non-linear and can therefore not be captured by a simple linear model.

2.5.4 Out-of-sample performance

Contrary to other approaches recently introduced in the literature to explain the pricing kernel dynamics (see Grith et al. (2013a) and Polkovnichenko and Zhao (2013)), we are able to forecast the kernel. Hence, we continue with a short analysis how accurate these predictions are. As already mentioned previously, differences between the observed and modeled pricing kernel are difficult to interpret. Therefore, we use

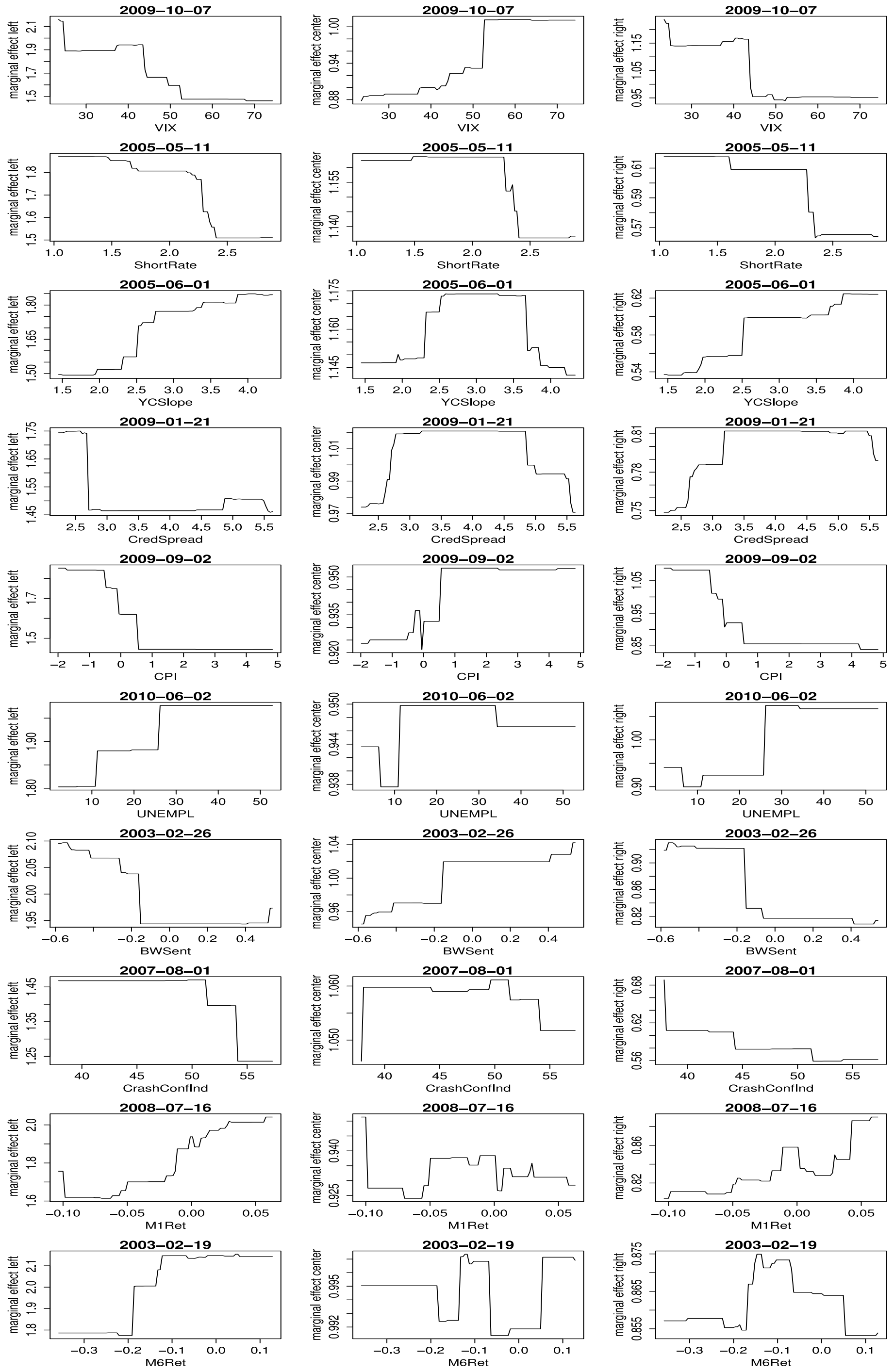


Figure 2.8: The figure shows partial dependency plots for some selected explanatory variables. The rolling window size for estimation is one year and the corresponding endpoint of the time period is indicated in the figure title. It is the period where the predictor had its highest influence. Marginal effects on the pricing kernel are shown for the left ($r < 0.9$), central ($0.9 \leq r \leq 1.05$) and right part ($r > 1.05$) of the pricing kernel, separately.

implied price differences to quantify the accuracy of pricing kernel forecasts.

To be precise, we calculate at time t the "true price" of a call option with strike K and time to maturity $\tau = T - t$ (similarly for puts) as

$$C_t(K,\tau) = e^{-r_{t,T}\tau} \int_{\frac{K}{F_{t,T}}}^{\infty} M_t(r,\tau)(rF_{t,T} - K)\hat{p}_t(r,\tau)dr,$$

where $F_{t,T}$ is the forward price, $r = S_T/F_{t,T}$ the forward gross return, $\hat{p}_t$ the estimated conditional historical density and M_t the initially derived empirical pricing kernel (see Section 2.5.2). Strictly speaking, M_t is also an estimate but we do not use the hat notation in order to avoid confusion with the fitted pricing kernel obtained with our boosting approach. Namely, the corresponding price implied by the predicted kernel $\hat{M}_{t|t-1}$ obtained from our estimated time series model is

$$\hat{C}_t(K,\tau) = e^{-r_{t,T}\tau} \int_{\frac{K}{F_{t,T}}}^{\infty} \hat{M}_{t|t-1}(r,\tau)(rF_{t,T} - K)\hat{p}_t(r,\tau)dr,$$

with $\hat{M}_{t|t-1}(r,\tau) = \hat{K}(r,\tau,\Phi_{t-1})$ and where $\hat{K}$ is calculated using an estimation window of length one year ending at time $t-1$.

The forecast error at time t is finally defined as the price root mean squared error

$$RMSE_t = \sqrt{\frac{1}{nm}\left[\sum_{i=1}^{n}(C_t(K_i,\tau_i) - \hat{C}_t(K_i,\tau_i))^2 + \sum_{j=1}^{m}(P_t(K_j,\tau_j) - \hat{P}_t(K_j,\tau_j))^2\right]},$$

for n calls and m puts with expiries $\tau_i, \tau_j \in [30, 240]$ (in calendar days) available at time t.

Given that we consider an estimation window of size one year, we repeat this forecast exercise every Wednesday from January 2003 to December 2010. The corresponding outcomes (pricing errors) are presented in Figure 2.9. Note that we do neither price options with relatively low time to maturities (less than 30 calendar days) nor options with observed bid prices below 0.5 dollar. Thus we conclude that the obtained pricing performance from 2003 to the beginning of 2007 is rather good (mean error of 0.91 dollar), meaning that also the pricing kernel forecasts are accurate. However, during the crisis we typically get quite high errors.

Additionally, we present in Figure 2.10 the option pricing performance for November 30, 2005. On that day, we get a price RMSE of 0.78 dollar, which is close to the median error observed during the period January 2003 to January 2007. In general, the performance is rather satisfactory. Only for options with higher time to maturities, and in particular for OTM call options, we observe some substantial price differences.

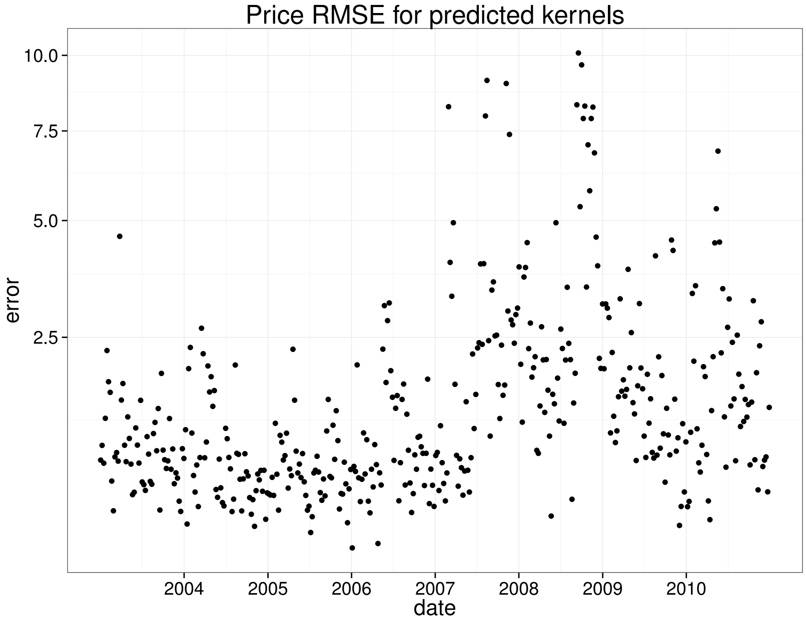


Figure 2.9: This figure shows the performance of our predicted kernel for every week (Wednesday) between January 2003 and December 2010. The pricing errors are calculated as explained in the text.

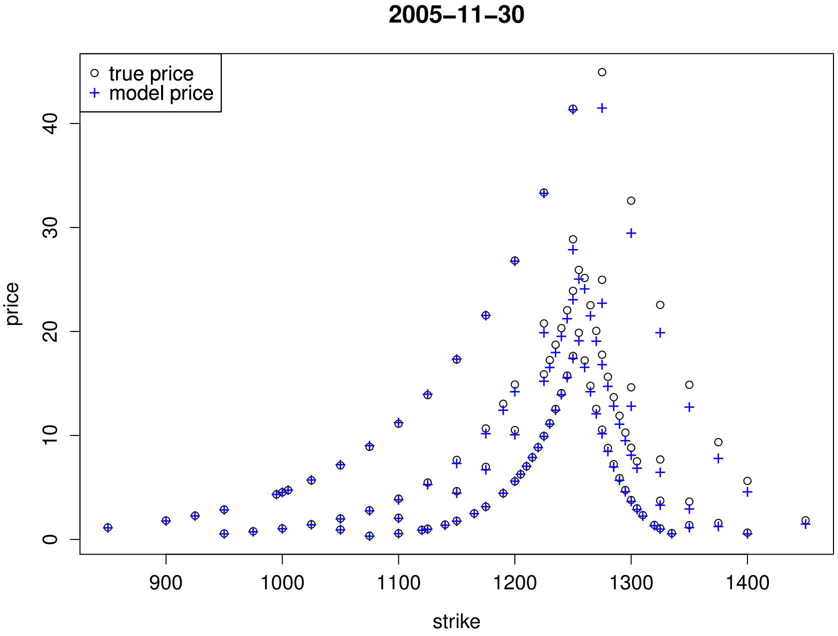


Figure 2.10: This figure shows the pricing performance for November 30, 2005. The "true prices" are obtained with the observed empirical pricing kernel on that date and the model prices are calculated using the kernel forecast of our model. The price RMSE for that day is 0.78 dollar.

Since, at least to our best knowledge, there is no comparable model available yet, we decide to benchmark our forecasting performance against a simple random walk prediction of the pricing kernel, which we expect to be a tough benchmark. That is,

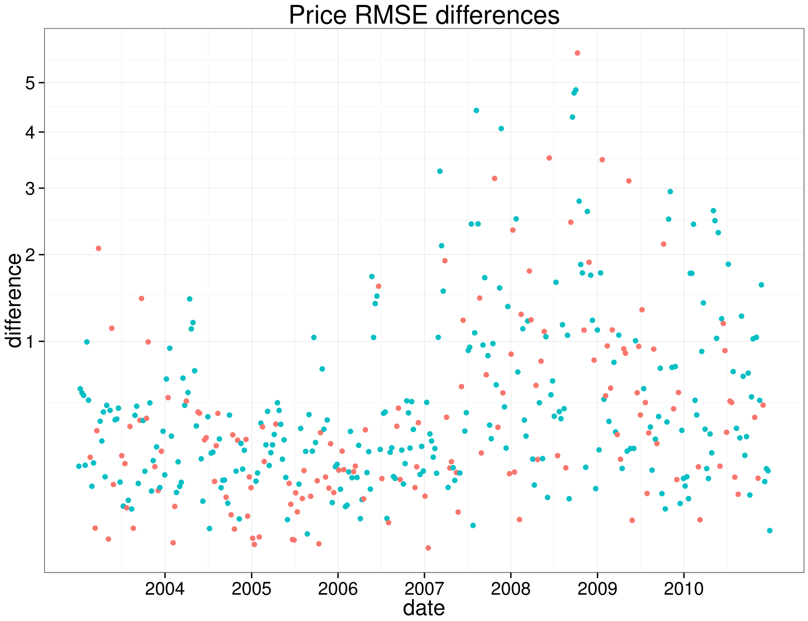


Figure 2.11: This figure presents absolute RMSE differences between the two forecast methods. Red points indicate cases where the tree-boosting algorithm performs better and the blue-green points those where the random walk forecast performs better.

we do the same calculations but take $\hat{M}_{t|t-1} = M_{t-1}$. Here again, M_{t-1} refers to the initially estimated empirical pricing kernel, which has been presented in Section 2.5.2.

Figure 2.11 presents the differences between the two forecasting methods and indicates that the random walk performs better compared to the predictions obtained with our model. This is confirmed statistically in Table 2.4, where we show the outcomes of Diebold and Mariano (1995) type tests to measure the equal predictive ability. We identify two reasons for this result. First, the recent crisis. It seems that one can hardly do better than a simple random walk strategy during that period. Second, we find that the inferior performance is mainly due to higher pricing errors for options with longer time to maturity. In contrast, if we only consider the pre-crisis period and options with time to maturities no longer than 100 days, we find that predictions from our model are even slightly better, but not statistically different. We conclude that our predicted kernels are reasonable, however, it is difficult to beat the random walk forecast.

Table 2.4: The table shows the mean difference in RMSE between both kernel forecasting methods (negative values are in favor of our model) together with p-values (reported in parentheses) of Diebold and Mariano (1995) type tests to measure the equal predictive ability, i.e. the null hypothesis is that the models have equal predictive ability.

Time period	Maturity range (in days)	
	30 to 100	30 to 240
Jan 2003 to Dec 2006	-0.004	0.095
	(0.45)	($\approx$ 0)
Jan 2003 to Dec 2010	0.069	0.186
	(0.07)	($\approx$ 0)

2.6 Conclusion

It is commonly accepted that the asset pricing kernel is time-varying, which reflects variation in the relevant pricing kernel state variables. However, despite recent advancements in the literature, little is still known about these driving factors. As a consequence, we develop in this paper a time series model for the option-implied empirical pricing kernel that allows to understand its variation over time. Contrary to most techniques that model the pricing kernel for each time to maturity separately (so-called slice by slice), we adopt a surface modeling approach. More precisely, we derive first initial estimates of the empirical pricing kernel surface (using separate estimates for both the risk-neutral and historical density surface of the underlying asset return at expiry) and then model the observed time-varying surface with a boosting approach based on regression trees. Taking trees as base learner is particularly helpful in our context, given that they automatically select the most relevant predictors among a set of possible candidates. Therefore, we obtain a suitable tool to answer our main question regarding the principal driving factors of the pricing kernel dynamics.

In our empirical analysis, we illustrate the model with European options and return values of the S&P 500 index and show that our modeling approach can capture the various shapes of the option-implied pricing kernel observed during the sample period, covering January 2002 to December 2010. Further, we provide evidence that the observed dynamics can indeed be explained with variables related to the state of the economy, which is in line with findings presented for example in Rosenberg and Engle (2002), Grith et al. (2013a) and Metaxoglou and Smith (2014). However, with our modeling approach we are able to quantify the impact of all possible driving factors building on the relative influence measure introduced in Friedman (2001). We find that the relative influence of the single predictors substantially varies over time and we provide some additional insights on how they affect the pricing kernel shape. Finally, we also investigate the predictive performance of our model and provide some evidence that the forecasts are reasonable. However, it is difficult to beat the random walk forecast.

2.A A short introduction to FGD

We present here the main idea of functional gradient descent (FGD) in the framework of general regression. Hence we consider a sample $(X_1, Y_1), \dots, (X_n, Y_n)$, where $Y_i \in \mathbb{R}$ is the response variable and X_i represents a p-dimensional explanatory variable. Based on (X_i, Y_i), we are looking for a function $F \in \mathcal{F} = \{f | f : \mathbb{R}^p \to \mathbb{R}\}$ which minimizes an expected loss of the form $E[\lambda(Y, F(X))]$ with an adequate loss function λ.
The FGD algorithm estimates F by minimizing the empirical risk defined as

$$\Lambda(F) = \frac{1}{n} \sum_{i=1}^{n} \lambda(Y_i, F(X_i)).$$

Starting from an initial function estimate $\hat{F}_0$ (step 1 of the algorithm), the algorithm selects the steepest descent direction in the mth iteration which would be given by the negative functional derivative $-d\Lambda(\hat{F}_{m-1})$. However, due to smoothness and regularization constraints on the minimizer of $\Lambda(\hat{F}_{m-1})$, one must find a function $\hat{f}_m$ which is in the linear span of a class of simple base learners $\mathcal{S}$ and is close to $-d\Lambda(\hat{F}_{m-1})$ in the sense of a functional metric. This is equivalent to fitting the base learner $h(x, \theta) \in \mathcal{S}$ to the negative gradients

$$U_i = - \left. \frac{\partial \lambda(Y_i, Z)}{\partial Z} \right|_{Z = \hat{F}_{m-1}(X_i)}, \quad i = 1, \dots, n.$$

This is often achieved with least squares fitting and we get $\hat{f}_m = h(x, \hat{\theta})$ with

$$\hat{\theta} = \min_{\theta, \omega} \sum_{i=1}^{n} (U_i - \omega h(X_i, \theta))^2.$$

This is the second step of the algorithm. In a third step (line search), one finally has to perform a one-dimensional optimization in order to find the best step length $\hat{\omega}_m$ for updating $\hat{F}_{m-1}$ with $\hat{f}_m$. We obtain

$$\hat{\omega}_m = \min_{\omega} \sum_{i=1}^{n} \lambda(Y_i, \hat{F}_{m-1}(X_i) + \omega \hat{f}_m(X_i))$$

and get

$$\hat{F}_m = \hat{F}_{m-1} + \hat{\omega}_m \hat{f}_m.$$

Iterating steps 2 and 3 until $m = M$ produces the FGD estimate

$$\hat{F}_M = \hat{F}_0 + \sum_{m=1}^{M} \hat{\omega}_m \hat{f}_m.$$

2.B Additional plots

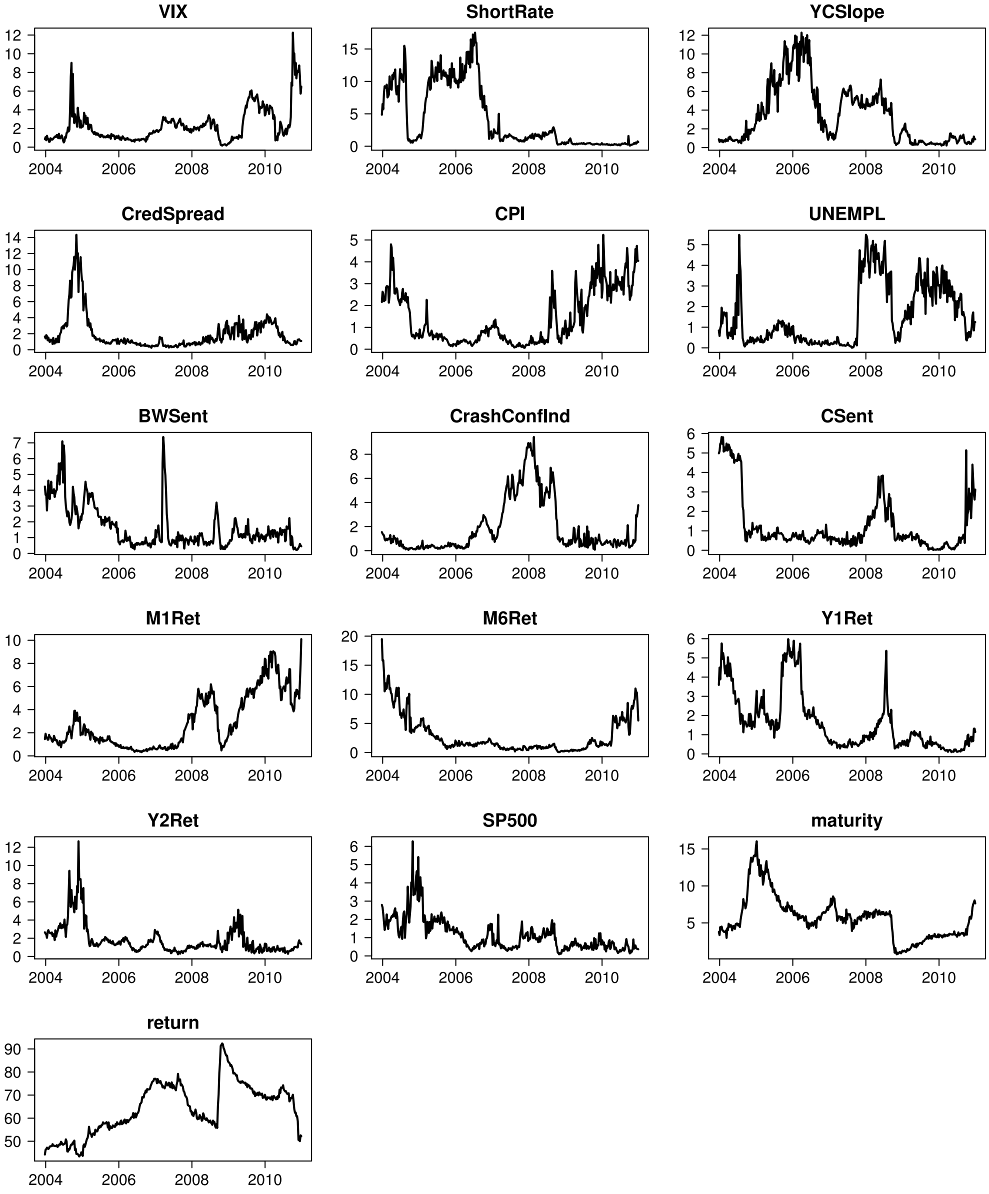


Figure 2.12: This figure shows the relative influence (measured as explained in the text) of the different explanatory variables obtained from our boosting algorithm based on trees. To estimate the model, we use a rolling window of length two years and the time axis indicates the endpoint of this period.

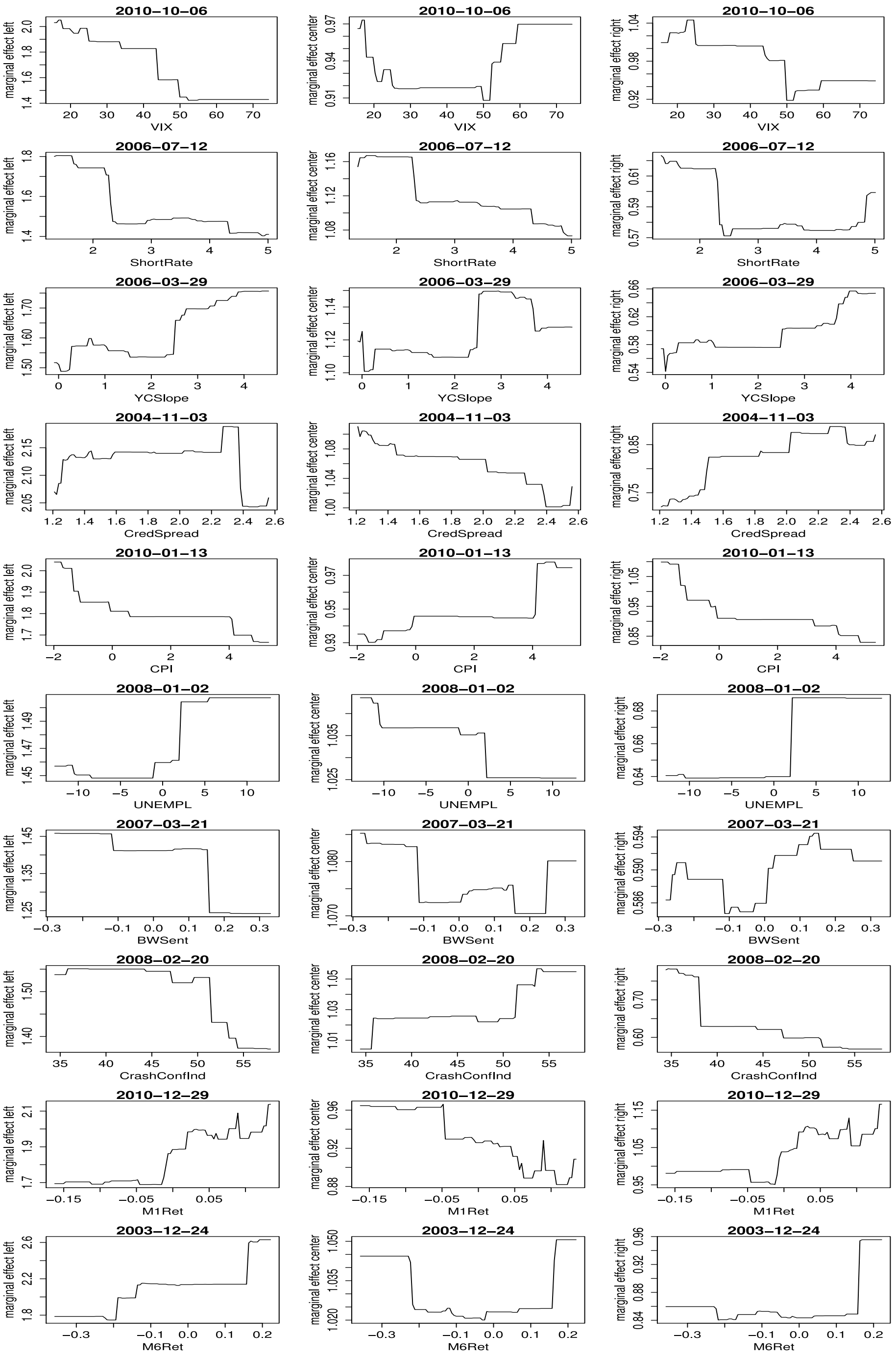


Figure 2.13: The figure shows partial dependency plots for some selected explanatory variables. The rolling window size for estimation is two years and the corresponding endpoint of the time period is indicated in the figure title. It is the period where the predictor had its highest influence. Marginal effects on the pricing kernel are shown for the left ($r < 0.9$), central ($0.9 \leq r \leq 1.05$) and right part ($r > 1.05$) of the pricing kernel, separately.

Chapter 3

Estimating risk-neutral density tails: a comparison

Anselm Ivanovas
Pirmin Meier

Abstract

In this paper, we review several methods to estimate risk-neutral density tails from option data. In addition, we introduce a new tail extension approach that combines the idea of price matching with the assumption of tails drawn from a generalized extreme value distribution. Based on a theoretical market model with known implied risk-neutral density, we conduct a performance analysis. The final outcomes of our study indicate that the best results regarding accuracy and stability are obtained either with implied volatility extrapolation by means of Gatheral's stochastic volatility inspired SVI model, or with our new price matching method. Moreover, we provide evidence that the still very popular constant implied volatility extrapolation approach often yields unsatisfactory results.

Keywords: Risk-neutral density, tail estimation, IV extrapolation, density matching, price fitting.

3.1 Introduction

The risk-neutral density (RND) of asset prices is an essential tool for the valuation of financial derivatives. It represents a forward-looking view of the asset price distribution and thus is particularly helpful for various other applications too. Examples include estimating investors risk preferences (Jackwerth, 2000; Rosenberg and Engle, 2002), managing risk (Aït-Sahalia and Lo, 2000) and asset allocation (Kostakis et al., 2011). Furthermore, it enables the study of how market participants react to new information and how risk perceptions evolve over time (Birru and Figlewski, 2012).

Numerous flexible nonparametric and semiparametric approaches have been developed to recover the RND from option prices.[1] However, an important issue often exists when applying such methods. Namely, the risk-neutral density can only be obtained over the strike range of observable prices, which might not extend very far into the tails.

In particular the left tail of the RND is important for any risk management application. An example is the calculation of an economic value at risk (Aït-Sahalia and Lo, 2000; Markose and Alentorn, 2008). Further, both RND tails are important to estimate higher RND moments. Such moments (in particular skewness and kurtosis) have gained popularity among researchers in the recent past. For instance, Han (2008) and Conrad et al. (2013) show that they play an important role in portfolio choice and the pricing of assets. Other applications include density estimation by moment matching (Rompolis and Tzavalis, 2008; Eriksson et al., 2009).

In this paper we review several methods to estimate such risk-neutral density tails from option data. They can be categorized into three groups: the first two use tails drawn from a suitable parametric distribution and apply density-matching or price-fitting respectively, whereas approaches in the third category extrapolate the implied volatility smile.

From the first two categories we specifically review the density-matching method of Figlewski (2010) and Birru and Figlewski (2012) (using generalized extreme value, generalized Pareto and normal distribution tails) and the price-fitting method of Hamidieh (2013), which assumes generalized Pareto tails.

From the third category we consider the popular method of constant implied volatility extrapolation (Bliss and Panigirtzoglou, 2004; Jiang and Tian, 2005; Polkovnichenko and Zhao, 2013; Shimko, 1993), linear implied volatility extrapolation as shown in (Jiang and Tian, 2007), and extrapolation by means of the parametric "stochastic volatility inspired" model introduced in (Gatheral, 2004).

---

[1] See Aït-Sahalia and Lo (1998), Bliss and Panigirtzoglou (2002), Fan and Mancini (2009), Fengler and Hin (2014) and Ludwig (2014) among many others.

We further present a new tail extension approach that combines the distributional assumption of Figlewski (2010) and the idea of price fitting used in Hamidieh (2013). For this purpose, we employ the closed-form option pricing model developed in Markose and Alentorn (2011) under the assumption of a generalized extreme value RND. In addition, we use a modified version of the estimation approach of Figlewski (2010), which puts more emphasis on the shape of the RND on the observed part of the tail.

The principal goal of this paper is to analyze the numerical performance of the different methods. Based on a known data-generating process for the underlying asset, we compute option prices and the corresponding risk-neutral density. Subsequently, we analyze how accurately the different methods can recover the tail of this density when option prices are only used up to a certain maximum or minimum strike price.

There are some studies similar to ours. For instance, Bondarenko (2003), Bu and Hadri (2007) and Lai (2011) all rely on a comparable experimental setting to assess the performance of different RND estimation methods. However, to the best of our knowledge, this is the first study investigating the performance of tail estimation methods.[2]

The final outcomes of our survey show that the best performance regarding accuracy and stability is achieved either with price matching and tails drawn from a generalized extreme value distribution, or implied volatility extrapolation by means of the stochastic volatility inspired model of Gatheral (2004). Moreover, we provide evidence that parametric tails fitted based on price-matching approaches generally outerperform those using density-matching, independently of the chosen parametric specification. In addition, we observe that the still very popular constant implied volatility extrapolation approach often leads to unsatisfactory results.

We believe that the results of this study and the included robustness discussion provide helpful insights for researchers who depend on accurate tail estimates for their application.

The remainder of this paper is organized as follows. Section 3.2 shows how to derive the implied RND from observed option prices and discusses data limitation problems that appear in practice. Section 3.3 contains a summary of the different RND tail estimation techniques. Section 3.4 explains the experimental setting to compare the performance of these methods. Section 3.5 presents the outcomes of this comparison study. Section 3.6 discusses robustness issues and Section 3.7 concludes the paper.

---

[2] Birru and Figlewski (2012) provide in their appendix a short comparison of Figlewski (2010)'s tail extension method under three different distributional assumptions. But they use market data for which the true RND is unknown.

3.2 Option prices and risk-neutral densities

Ross (1976) and Cox and Ross (1976) show that in a dynamically complete, arbitrage-free market and when using the risk-neutral valuation principle, the price of an option is given by the expected present value of the payoff computed under the risk-neutral density (RND). Formally, we obtain for the price at time t of an European call option $C_t(K)$

$$C_t(K) = e^{-r_t\tau} \int_0^\infty \max(S_T - K, 0) q_t(S_T) dS_T, \tag{3.1}$$

where q_t denotes the conditional risk-neutral probability density of the underlying price S at time T, K the strike price, T the expiry date, $\tau = T - t$ the time to maturity and r_t the instantaneous risk-free interest rate for that maturity.
Differentiating equation (3.1) with respect to the strike price K yields

$$\frac{\partial C_t}{\partial K} = -e^{-r_t\tau}(1 - Q_t(K)).$$

Here, Q_t denotes the conditional risk-neutral probability distribution function of S_T. Calculating the second derivative with respect to K, we finally obtain

$$\left.\frac{\partial^2 C_t}{\partial K^2}\right|_{K=S_T} = e^{-r_t\tau} q_t(S_T). \tag{3.2}$$

This is the famous result originally presented in (Breeden and Litzenberger, 1978) and (Banz and Miller, 1978). It states that the second derivative of the European call price function with respect to the strike is equal to the discounted risk-neutral probability density of S_T.[3]

This option implied density (which is often also referred to as state price density, since it represents the continuous-state counterpart of the so-called Arrow-Debreu securities) inherently differs from the historical density of the underlying price. It not only contains information about the market participants' expectations regarding the evolution of the underlying price, but also depends on their perceptions of risk.

---

[3] A similar expression can also be derived for European put options.

Departing from the aforementioned theoretical concepts, numerous methods have been developed to extract the implied RND from option prices observed on the market. Bliss and Panigirtzoglou (2002), Jackwerth (2004) and Figlewski (2010) for instance provide comprehensive reviews of this literature. As mentioned in Ludwig (2014), it is possible to make a distinction between approaches that model the implied density directly (for example by means of mixture methods or assuming generalized distributions) and those which first estimate the implied volatility or price function and then rely on equation (3.2) to derive the state price density.

In particular for the methods in the second group, a successful estimation of the RND has to address the following data limitation problems which normally appear in practice (see (Bondarenko, 2003)):

1. Using equation (3.2) to calculate the risk-neutral density requires options with a continuum of exercise prices. However, in reality, options are only available for a discrete set of strikes.
2. Observed option prices contain microstructure noise from various sources such as nonsynchronous trading, price discreteness and bid-ask bounce.
3. Strikes of available options do not cover the entire support of S_T. Due to the rules that an exchange uses to introduce new option contracts, very low and high strikes are often unavailable.

It is common in the literature to address the first two issues by using a combination of interpolation and smoothing techniques, while there is only a limited number of solutions for the third point. We present several such approaches in the next section.

3.3 Methods to estimate the risk-neutral density tails

In this section we provide a detailed description of the tail estimation methods that we subsequently benchmark. Some of them rely on results in extreme value theory, which we briefly summarize in the appendix.

3.3.1 Figlewski's tail extension approach and some variants thereof

Figlewski (2010) introduces a new approach to extract a well-behaved risk-neutral density from observed option prices and thereby addresses all three data limitation problems mentioned in Section 3.2. As it is common practice, he suggest applying a curve-fitting approach for implied volatilities (IVs). More precisely, the usage of a 4th degree spline with a single knot placed at-the-money to model IV as a function of strike. He argues that at least a polynomial of degree 4 is needed to ensure that the resulting RND has no kink. Furthermore, a single inner knot at-the-money seems to provide enough flexibility to capture the shapes of IV curves typically observed in practice. By transforming implied volatilities back to prices and applying formula (3.2), one can finally infer the middle portion of the RND.

Addressing the third data limitation issue, Figlewski (2010) suggests a way to extend and complete the empirical RND beyond the range of available strikes, using a suitable parametric distribution. This can be considered the main contribution of the paper. The idea is to determine the parameters of the selected probability distribution such as to match the shape of the estimated RND over the observed portion of the tail region. We now summarize the approach for the right end of the density with tails drawn from a generalized extreme value (GEV) distribution and thus following Figlewski's recommendation.

After having obtained the middle portion of the RND by means of observed out-of-the-money (OTM) put and call options (see Figure 3.1), two points $K_{c,i}$ and $K_{c,c}$ must be selected. The first designs the connection point to the GEV distribution, whereas the second point serves to match the GEV tail shape to that of the observed RND. In order to identify the three GEV parameters by optimization, Figlewski (2010) imposes the following conditions:

$$\begin{aligned} Q_{GEV}(K_{c,i};\theta) &= \hat{Q}(K_{c,i}) \\ q_{GEV}(K_{c,i};\theta) &= \hat{q}(K_{c,i}) \\ q_{GEV}(K_{c,c};\theta) &= \hat{q}(K_{c,c}) \end{aligned}$$

The first condition requires that the right tail probability at the connection point is equal for both the empirical distribution $\hat{Q}$ and the approximating GEV distribution Q_{GEV}. The other two equations serve to fit the shape of the observed RND $\hat{q}$ in the

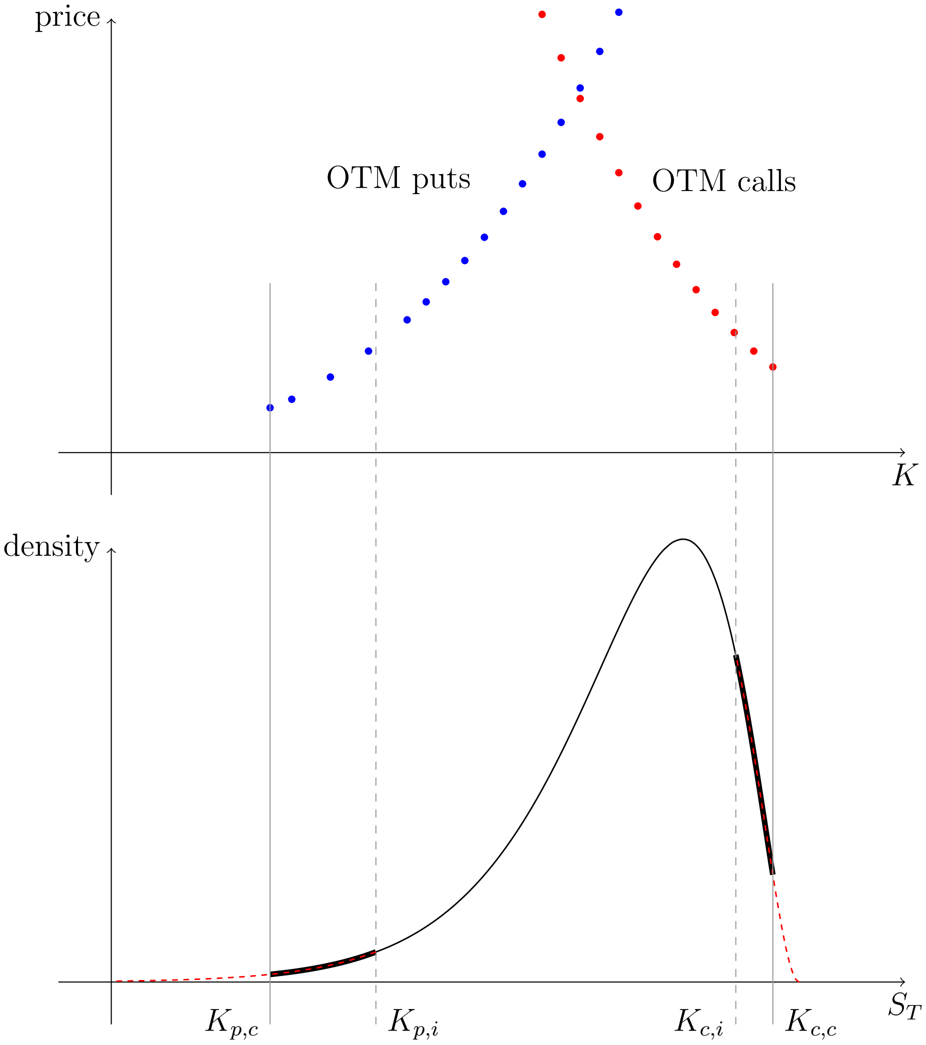


Figure 3.1: Schematic illustration of the typical data situation when fitting parametric tails.

observed tail area.

We implement several modifications with respect to the initial suggestion by Figlewski (2010). On the one hand, we connect the parametric density at the point $K_{c,c}$ instead of $K_{c,i}$ since we believe that the maximum possible part of the initially estimated RND should be used. Thus, any potential parametric misspecification of the density is restricted to the strike range where no data is available. On the other hand, we suggest the following "overidentified" optimization problem

$$\arg\min_{\theta} \sum_{j=1}^{n} [w_1(x_j)(q_{GEV}(x_j;\theta) - \hat{q}(x_j))]^2 + [w_2(x_j)(Q_{GEV}(x_j;\theta) - \hat{Q}(x_j))]^2$$

with $K_{c,i} = x_1 < x_2 < \cdots < x_{n-1} < x_n = K_{c,c}$ and two functions w_1 and w_2 specified such that the impact on the optimization objective function is of the same magnitude for every point x_j as well as both losses (PDF and CDF).

With this approach we match the curvature of the fitted RND on the whole interval $[K_{c,i}, K_{c,c}]$ and reduce the dependence of the method on the (relatively arbitrary) choice of the inner point $K_{c,i}$.[4] This increases the robustness of the method since the original version can be sensitive to the data quality around the single inner point.

An important issue is the selection of an appropriate parametric distribution for the tail extension. The GEV distribution suggested in (Figlewski, 2010) is rather flexible and thus can handle various tail behaviors. However, from a theoretical point of view the generalized Pareto distribution would be a correct specification under the assumption that the points $K_{c,c}$ and $K_{c,i}$ are far enough out in the tails (see Theorem 2 in the appendix). Hence, we also consider tails drawn from this distribution later on in our performance comparison, and determine the parameters similarly as for the GEV distribution. Additionally, distributions with two parameters can be used as well. In this case, no interior point $K_{c,i}$ is required. The two parameters can be determined merely relying on the PDF and CDF conditions at the connection point $K_{c,c}$. We employ both the normal and lognormal distribution in our subsequent comparison study.

3.3.2 Hamidieh's GPD approach

Another method to infer the RND tail has recently been introduced in Hamidieh (2013). He makes a parametric generalized Pareto distribution (GPD) assumption about the tail. In contrast to Figlewski's idea, the method directly depends on observed (deep) out-of-the-money call and put option prices and does not need any prior estimate of the RND. We present the approach for the right tail using OTM calls, and refer to Appendix 3.A for a short summary of relevant extreme value theory results.

Departing from the fundamental pricing equation (3.1), we obtain

$$C = e^{-r\tau} E[\max(S_T - K, 0)] = e^{-r\tau} E[S_T - K | S_T > K] P(S_T > K),$$

where the expectation is taken under the risk-neutral measure Q.[5]
If the distribution of S_T is in the maximum domain of attraction (MDA) of a distribution G, we can consider a high strike price K_0 such that the Pickands-Balkema-de

---

[4] See Section 3.6 for a discussion on the choice of the point $K_{c,i}$.

[5] We omit the time dependence t for simplicity of notation.

Haan theorem holds (see Theorem 2 in Appendix 3.A). It follows that the excess $S_T - K_0 | S_T > K_0$ has an approximate generalized Pareto distribution with parameters $\sigma(K_0)$ and ξ. Hence, we obtain

$$C_0 = e^{-r\tau} E[S_T - K_0 | S_T > K_0] P(S_T > K_0) \approx e^{-r\tau} \frac{\sigma(K_0)}{1-\xi} P(S_T > K_0) = C_0^*.$$

The above approximation (emphasized through the notation *) uses the expected value of a generalized Pareto distribution $H_{\sigma,\xi}$. Therefore, we have to assume that $\xi < 1$ in order to avoid an infinite option value.[6]

Making use of the closure property of GPD (see Lemma 1 in Appendix 3.A), we can get a similar approximation for all calls with strike prices $K_i > K_0$. It is

$$C_i^* = e^{-r\tau} \frac{\sigma(K_i)}{1-\xi} P(S_T > K_i),$$

where the new scale parameter is given by

$$\sigma(K_i) = \sigma(K_0) + \xi(K_i - K_0). \tag{3.3}$$

Using equation (3.3) and

$$P(S_T > K_i) \approx P(S_T > K_0) \left[\frac{\xi}{\sigma(K_0)} (K_i - K_0) + 1 \right]^{-1/\xi}$$

it is possible to express C_i^* in terms of C_0^*. Indeed

$$C_i^* = C_0^* \left[\frac{\xi}{\sigma(K_0)} (K_i - K_0) + 1 \right]^{1-1/\xi}.$$

Obviously, one can obtain a similar expression for put options under the assumption that the distribution of $-S_T$ is in the MDA of G. For brevity, we omit the derivation here and just state the final formula. For puts P_i and P_0 with strike prices $K_0' > K_i'$, we obtain

$$P_i^* = P_0^* \left[\frac{\xi'}{\sigma(K_0')} (K_0' - K_i') + 1 \right]^{1-1/\xi'}.$$

Hamidieh (2013) demonstrates that the absolute relative error of the GPD approximation tends to zero for high strikes (calls) and low strikes (puts). Therefore, one can get the approximated GPD tails through the solution of the following optimization problems

---

[6] A random variable X with distribution $H_{\sigma,\xi}$ has an expected value $E(X) = \frac{\sigma}{1-\xi}$ for $\xi < 1$. For $\xi \geq 1$ the first moment is infinite.

$$\min_{\sigma(K_0),\xi} \sum_{i=0}^{n} \left| \frac{C_i - C_i^*}{C_i} \right| \quad \text{and} \quad \min_{\sigma(K_0'),\xi'} \sum_{j=0}^{m} \left| \frac{P_j - P_j^*}{P_j} \right|,$$

where C_i $(i = 0, \ldots, n)$ and P_j $(j = 0, \ldots, m)$ denote OTM call and put options with strikes $K_i < K_{i+1}$ and $K_j' > K_{j+1}'$.
Note that Hamidieh (2013) assumes $C_0 = C_0^*$ and $P_0 = P_0^*$ such that he does not need estimates of the risk-free rate r and the probabilities $P(S_T > K_0)$ and $P(S_T < K_0')$.

3.3.3 Price fitting with closed-form GEV prices

In addition to the two previously mentioned approaches, we propose a further method that combines the distributional assumption of Figlewski (2010) and the idea of price fitting used in Hamidieh (2013).

We build on the closed-form option pricing model introduced by Markose and Alentorn (2011), which is based on the GEV distribution. They model simple losses, i.e. $-\frac{S_T - S_t}{S_t} = 1 - \frac{S_T}{S_t}$ to be GEV distributed. Therefore, the left tail of the RND corresponds to the right GEV tail, with the economic argument that the probability of extreme losses is higher than the one of extreme gains (under the Fréchet type distribution i.e. $\xi > 0$).

Note that only $\xi = 0$ implies infinite support of the density on both sides. Bounded support (for $\xi \neq 0$) is theoretically not admissible for a RND. The authors argue that the GEV is still a sufficient approximation, given the empirical distribution limits of past years.

They derive the following option pricing formulas for the case $\xi \neq 0$:[7]

$$C(K) = e^{-r\tau} \left\{ S \left[(1 - \mu + \sigma/\xi) e^{-H^{-1/\xi}} - \frac{\sigma}{\xi} \Gamma\left(1 - \xi, H^{-1/\xi}\right) \right] - K e^{-H^{-1/\xi}} \right\} \tag{3.4}$$

---

[7] They also derive the exact case for $\xi = 0$, but with numerical optimization in mind, we only use the results for $\xi \neq 0$ and consider $\xi \to 0$ as a limiting case. We numerically confirmed that we can come sufficiently close to the exact case of $\xi = 0$.

and

$$P(K) = e^{-r\tau}\Big\{K\left(e^{-h^{-1/\xi}} - e^{-H^{-1/\xi}}\right) - S\Big[(1-\mu+\sigma/\xi)\left(e^{-H^{-1/\xi}} - e^{-h^{-1/\xi}}\right) \\ - \frac{\sigma}{\xi}\Gamma\left(1-\xi, h^{-1/\xi}, H^{-1/\xi}\right)\Big]\Big\} \tag{3.5}$$

with $H = 1 + \frac{\xi}{\sigma}\left(1 - \frac{K}{S} - \mu\right)$, $h = 1 + \frac{\xi(1-\mu)}{\sigma}$, $\Gamma(a,b) = \int_b^\infty y^{a-1}e^{-y}dy$ the incomplete Gamma function and $\Gamma(a,b,c) = \Gamma(a,b) - \Gamma(a,c) = \int_b^c y^{a-1}e^{-y}dy$ the generalized Gamma function.

To fit a GEV tail to observed option data, we consider the following optimization problem for calls (the put case is analogous)

$$\hat{\theta} = \arg\min_{\theta} \sum_{K} f(C(K), \hat{C}(K,\theta)), \tag{3.6}$$

with $f(\cdot,\cdot)$ some loss function[8] and $\hat{C}(K,\theta)$ the model prices. Note that compared to other methods described in this section, the fitted tail types for calls and puts are actually different (since they are priced using the left and right tail of a GEV distribution respectively).

One could now directly optimize over all three GEV parameters, i.e. $\theta = (\mu, \sigma, \xi)$, but there are two issues with this optimization problem and the resulting parameters:

1. The probability mass under the tail might not be equal to the CDF at the cut-off point. Thus, the combined object might not integrate to unity.
2. The three parameter optimization is numerically rather unstable and does often result in unrealistic fits.

By imposing a further condition, we can overcome both issues. Specifically, we impose that the empirical and model CDF coincide in one point, i.e. $Q(K^*) = \hat{Q}(K^*, \theta)$ for some strike K^*. While K^* could in principle be chosen arbitrarily from the range of available data, we always choose the outermost available strike for this condition

---

[8] In our implementation of (3.6) we use the loss function $f(C(K), \hat{C}(K,\theta)) = \left|1 - \frac{\hat{C}(K,\theta)}{C(K)}\right|$, as in Hamidieh (2013).

(i.e. $K_{p,c}$ or $K_{c,c}$, as shown in Figure 3.1). This implies that by construction the probability mass under the estimated tail and the empirical CDF match at the cut-off point. Further, by rearranging the condition

$$Q(K^*) = \hat{Q}(K^*, \theta) = \exp\left\{ - \left[1 + \xi \left(\frac{K^* - \mu}{\sigma} \right) \right]^{-1/\xi} \right\},$$

we obtain

$$\mu = K^* - \frac{\sigma \left((-\ln(Q(K^*)))^{-\xi} - 1 \right)}{\xi}.$$

The value of the empirical CDF $Q(K^*)$ is known from the data (computed as the first numerical derivative of the observed prices with respect to their strike). This way, we have internalized one GEV parameter and we obtain $\mu\big(Q(K^*), K^*, \tilde{\theta}\big)$ with a new parameter vector $\tilde{\theta} = (\sigma, \xi)$. This reduces (3.6) to a two-parameter problem.

For simplicity, the GEV above is stated in terms of strikes directly, while we perform the optimization in simple returns. The density transformation is trivial and simply requires S as additional parameter.

3.3.4 Implied volatility fitting and extrapolation

Fitting the price curve directly and building on equation (3.2) to derive the corresponding RND is difficult. This is due to the non-linearity of this function as explained for instance in (Brunner and Hafner, 2003). Moreover, the second derivative of the price function with respect to strikes reacts sensitively to small data perturbations present in observed market prices and coming from various sources such as for example nonsynchronous trading and price discreteness. This fact is known as "curse of differentiation".

To overcome this issue, alternative approaches have been developed. One possibility is to transform option prices using the famous Black-Scholes formula in order to get the corresponding implied volatilities. Interpolation and smoothing is then performed in this implied volatility space, which has desirable statistical properties. Finally, the resulting continuous IV function is mapped back to the original price space where one can rely on numerical differentiation techniques to derive the RND.[9]

---

[9] Note that it is not necessary to assume that the Black and Scholes (1973) model correctly specifies the underlying asset dynamics. It simply serves as computational device to transform data into a

Since the implied volatility function is also unknown beyond the range of available strike prices, there remains the question of how to extend and complete the estimated RND. While Figlewski (2010) advises to consider tails drawn from a parametric distribution, it is also very common to extrapolate the implied volatility smile.

By far the most popular way to address this issue is constant implied volatility extrapolation. The procedure was first presented in (Shimko, 1993) and is still applied frequently in the literature (see (Bliss and Panigirtzoglou, 2004), (Jiang and Tian, 2005) and (Polkovnichenko and Zhao, 2013) among others). It is equivalent to the assumption of lognormally distributed tails (for prices) and does therefore not conform with features of typically observed asset distributions.

Another possibility is to linearly extrapolate the implied volatility function as shown in (Jiang and Tian, 2007). However, in this case it matters whether the IV function is expressed in terms of moneyness or log-moneyness. Thus we will consider both linear extrapolation schemes in our numerical analysis.

Finally, implied volatility extrapolation can be achieved with a parametric specification too. A popular and rather flexible parametric approach is the stochastic volatility inspired (SVI) model introduced in (Gatheral, 2004). It depends on five parameters $\theta = (a, b, \rho, m, \sigma)$ and describes the total implied variance (which is implied variance times time to maturity) $\omega(k;\theta)$ as a function of log-moneyness k. We have

$$\omega(k;\theta) = a + b\left[\rho(k-m) + \sqrt{(k-m)^2 + \sigma^2}\right],$$

with $a \in \mathbb{R}$, $b \geq 0$, $|\rho| < 1$, $m \in \mathbb{R}$, $\sigma > 0$ and the obvious condition $a + b\sigma\sqrt{1-\rho^2} \geq 0$ to ensure that $\omega(k;\theta) \geq 0$ for all $k \in \mathbb{R}$.

The model parameters have the following impact on the total variance function: a determines the general level of total variance, b gives the angle between the left and right asymptotes, increasing ρ induces a counter-clockwise rotation of the smile, σ describes the at-the-money curvature and changing m translates the smile horizontally. For an illustration we refer to Figure 3.2.

An interesting property of the SVI parameterization is its consistency with arbitrage bounds for extreme strikes. More precisely, for a fixed time to expiry τ, the implied Black-Scholes variance $\sigma_{BS}^2(k) = \omega(k)/\tau$ is linear in the log-moneyness k as $|k| \to \infty$ which is consistent with Roger Lee's moment formula (see (Lee, 2004)). We therefore

---

measurement space more conductive to apply curve-fitting techniques.

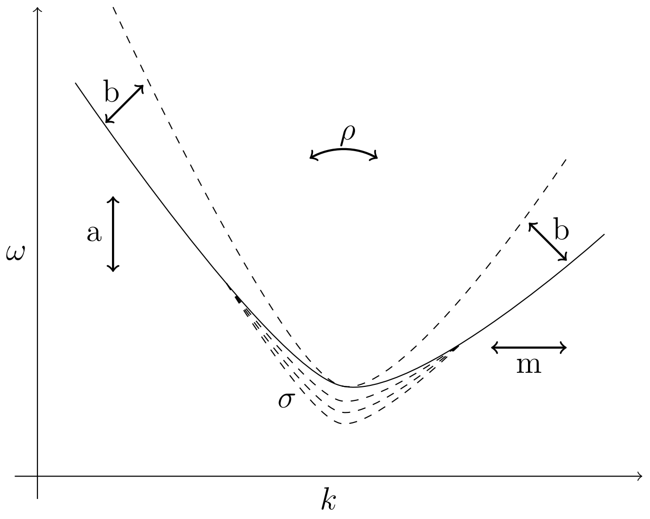


Figure 3.2: Illustration of how the SVI model parameters affect the total variance function.

consider it as an additional extrapolation method.

3.4 Experimental Setting

The main contribution of this paper is a comparison of the above-mentioned estimation techniques. More specifically, we analyze and contrast the performance (accuracy and stability) of the different methods in recovering the missing RND tail. As schematically illustrated in Figure 3.3, we depart from a theoretical market model with a known data-generating process. With this model specification at hand, we calculate option prices and derive the corresponding true risk-neutral density. The computed prices are then used as necessary input for the different methods to obtain an estimate of the unavailable tail. We finally evaluate the performance by comparing the fitted tails with the theoretical one. Note that this would not be possible relying on traded options where the true RND is unknown.

We continue now with a more detailed description of the selected market model and the loss functions employed to measure the performance of the competing tail estimation approaches.

3.4.1 Market model and implied densities

To mimic different market scenarios, we use several densities for which we examine the performance of the different methods. Instead of assuming some arbitrary densities,

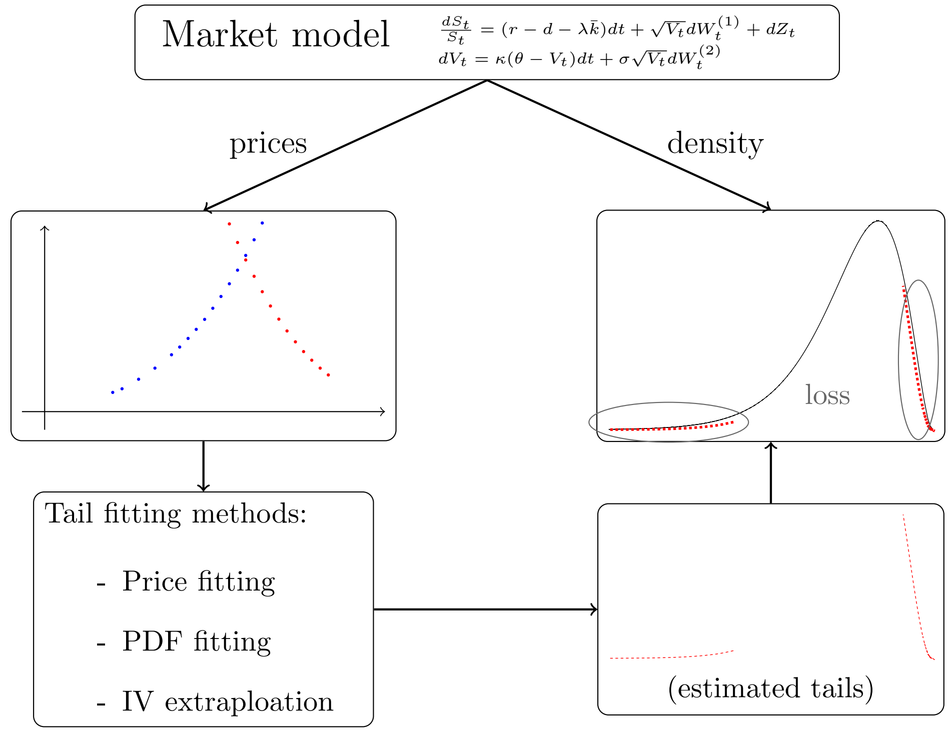

Market model
$\frac{dS_t}{S_t} = (r - d - \lambda\bar{k})dt + \sqrt{V_t}dW_t^{(1)} + dZ_t$
$dV_t = \kappa(\theta - V_t)dt + \sigma\sqrt{V_t}dW_t^{(2)}$
prices
density
loss
Tail fitting methods:
- Price fitting
- PDF fitting
- IV extraploation
(estimated tails)

Figure 3.3: Schematic illustration of the experimental setting.

we consider cases that were generated by a flexible model for the underlying asset dynamics. Specifically, we employ the stochastic volatility with jumps (SVJ) model suggested by Bates (1996).

This data-generating process is frequently used in the literature in order to obtain an empirically relevant price process of the underlying asset (see among others (Jiang and Tian, 2005), (Rompolis, 2010), (Lai, 2011) and (Fengler and Hin, 2014)). The characteristic function of this model is known in closed form, which is convenient in our experimental setting.

We employ it as a tool to generate risk-neutral densities with features typically observed in practice, such as negative skewness and excess kurtosis. In this spirit, we believe that our model choice represents a good trade-off between complexity and tractability.[10]

The SVJ model assumes that the risk-neutralized stochastic process of the underlying stock price S_t and its variance V_t are defined by

---

[10] We also refer to (Bu and Hadri, 2007). They provide a similar line of argument in order to motivate the use of a Heston (1993) model specification (which is actually nested by the SVJ) for the underlying asset, when comparing the performance of two common methods to estimate the RND.

$$dS_t/S_t = (r - d - \lambda\bar{k})dt + \sqrt{V_t}dW_t^{(1)} + dZ_t \tag{3.7}$$
$$dV_t = \kappa(\theta - V_t)dt + \sigma\sqrt{V_t}dW_t^{(2)}, \tag{3.8}$$

where $W_t^{(1)}$ and $W_t^{(2)}$ design two Brownian motions with correlation coefficient ρ, Z_t is a compound Poisson process with jump intensity λ and percentage jump size J with

$$\ln(1 + J) \sim N\left(\ln(1 + \bar{k}) - \frac{1}{2}\delta^2, \delta^2\right)$$

and κ, θ and σ denote, respectively, the rate of reversion, the long-run mean and the coefficient of variation (also known as “volatility of volatility”) of V_t.

Model-implied densities The characteristic function of this model can be used for option pricing as well as for the computation of the corresponding PDF and CDF by Fourier inversion. For our specific purpose it is important to know the densities sufficiently far in the tails. Since it is a known issue of (numerical) Fourier inversion techniques that the results can become unstable for events far out in the tails, we specifically check the resulting densities and option prices for such effects.

We get the numerically most stable results with the following setup (technical details can be found in the appendix):

Option prices	Pricing by fast Fourier transform (FFT), following Carr and Madan (1999). Interpolation to a finer strike grid in IV-strike space
CDF	Numerical first derivative of option prices
PDF	Direct FFT of the characteristic function

Calculating the CDF using direct FFT needs an integral of the form $\int_0^\infty \Re\left[\frac{e^{-izx}\Phi(z)}{z}\right] dz$ to be (numerically) computed. The z in the denominator increases the oscillating effect in the resulting CDF for $z \to 0$. We find that the CDF obtained as the numerical first derivative of option prices behaves better in the left tail.

We consider parameter values for which the model generates empirically plausible risk-neutral densities (see Table 3.1).

Table 3.1: Different SVJ parameter sets used for the numerical analysis.

Set	κ	θ	ρ	r	σ	$\bar{k}$	δ	λ
1	1	0.034	0.000	0.020	0.250	-0.075	0.000	0.500
2	4	0.170	-0.550	0.020	1.390	-0.030	0.000	0.130
3	4	0.040	-0.600	0.020	1.500	-0.075	0.000	0.500

Scenario 1 is borrowed from (Bakshi et al., 1997) and represents a situation with almost no skewness and excess kurtosis. The second scenario with parameter values taken from (Borak et al., 2005) is in line with features usually observed in reality. The implied density is negatively skewed and shows moderate excess kurtosis. Finally, the third parameter set is selected such that it yields skewed densities with rather extreme excess kurtosis. Nevertheless, this specification is not unrealistic at all. Similar or even more extreme cases can be obtained when fitting the model of (Bates, 1996) to observed market data during the recent crises or shortly after (i.e. between 2007 and 2010).

Figure 3.4 shows the implied risk-neutral densities of log-returns $\log(S_T/S_t)$ with three different maturities (30 days, 180 days and 360 days) whereas Table 3.2 contains a summary of the corresponding moments.

Table 3.2: The table shows the volatility σ_{IV}, skewness κ_3 and excess kurtosis κ_4 of log-returns for different maturities and genrated by the SVJ model with parameter sets as in Table 3.1.

σ_{IV}	κ_3	κ_4	τ	σ_{IV}	κ_3	κ_4	τ
Set 1				Set 3			
0.06	-0.12	0.28	30	0.06	-1.69	5.76	30
0.14	-0.08	0.56	180	0.16	-2.59	14.10	180
0.19	-0.11	0.79	360	0.22	-2.44	12.72	360
Set 2							
0.12	-0.74	1.24	30				
0.31	-1.22	3.18	180				
0.44	-1.17	2.93	360				

All in all we conclude that with these scenarios and all three maturities, we obtain a wide range of different shapes for the RND reflecting various empirical features of asset distributions. Consequently, a fair comparison of the different tail estimation methods should be possible.

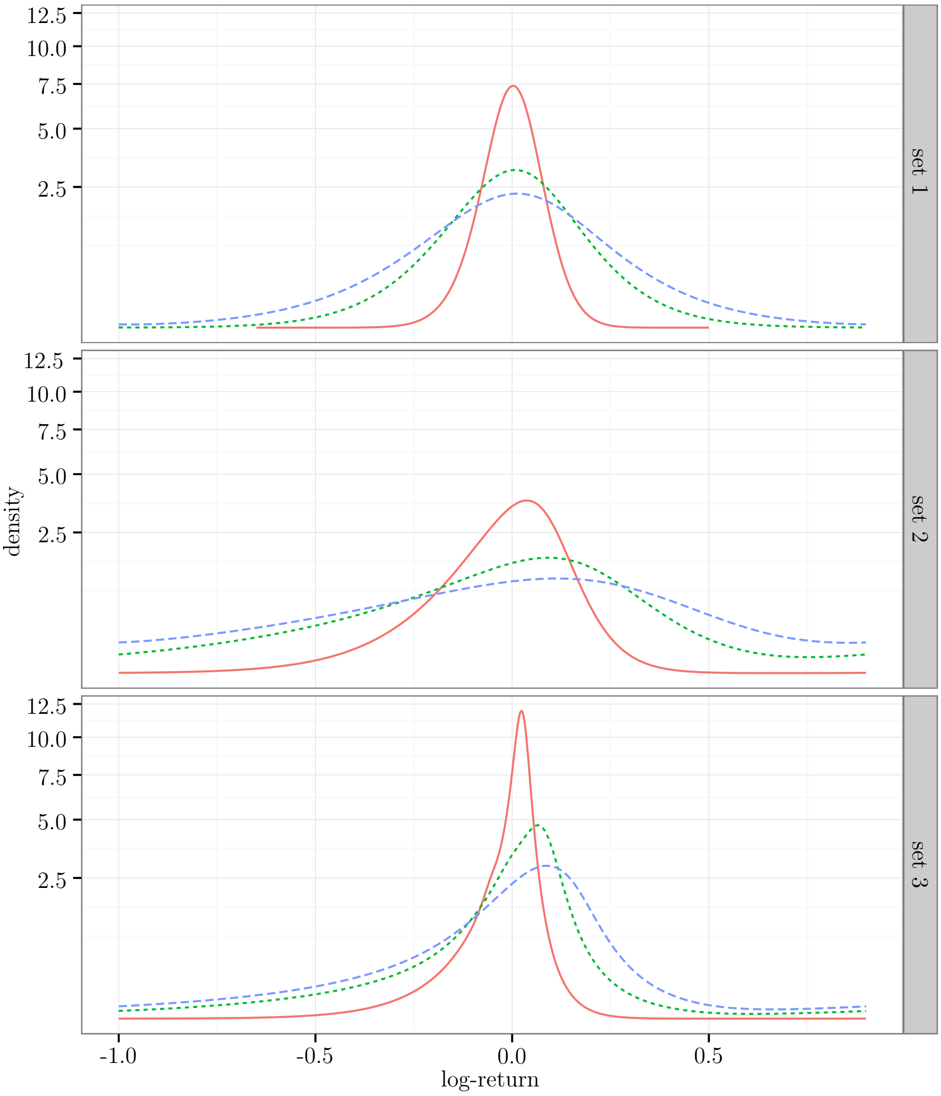


Figure 3.4: The figure shows the implied densities of log-returns for three different parameter specifications of the SVJ model. For each set we consider three maturities: 1 month (solid, orange), 6 months (dotted, green) and 12 months (dashed, blue).

3.4.2 Loss functions

In order to measure the performance of the considered estimation techniques, we use the following three loss functions:

1) **PDF loss:** Mean absolute error of the fitted and model RND in the corresponding tail area. That is in the region above (below) the cut-off point for the right (left) tail respectively.

2) **Higher moment loss:** The ratio of the estimated non-central 4th moment under consideration of the fitted tail, over the true 4th moment of the density

(computed analytically, see Appendix 3.B). Values larger than one therefore indicate an estimated tail heavier than the true one. For each call and put loss we assume the respective other tail to be completely observable.

3) **Price loss:** Mean absolute price error between prices fitted and model prices. Similar as for the PDF loss only options in the tails (i.e. with strikes above $K_{c,c}$ (right tail) or below $K_{p,c}$ (left tail)) are taken into account.

The first loss specification represents a standard choice since we have in mind to fit the missing RND tail. In addition, the two other loss functions help to quantify the impact of a possibly inaccurate tail fit regarding higher moment and price estimates. Higher moments are known to be particularly tail sensitive. Thus it is interesting to see whether the true moments can be recovered with the inferred tail. Moreover, pricing options by means of a given or estimated RND is an obvious application. The accuracy of such RND induced prices can be analyzed through the third loss specification.

We are convinced that all three loss functions together provide a detailed view about the quality of the estimated RND tail. Furthermore, a good fitting technique should be able to deliver satisfactory results for all considered losses.

3.5 Results

3.5.1 Numerical performance under perfect conditions

In this section we present the results of the performance analysis under "perfect conditions". Specifically, we use the exact option prices as generated by the SVJ model, with strike increments of one, i.e. $\Delta K = 1$.[11] This sufficiently reduces any error due to discretely observable option prices and completely eliminates the issue of microstructure noise. The goal is to compare the performance of the different methods, assuming that the only source of error is the limited available tail data.

For a comparison of the tail estimation methods, we truncate available call and put option prices at varying $K_{p,c}$ and $K_{c,c}$, repeating the fitting for all methods in each point. We choose $K_{p,i}$ and $K_{c,i}$ such that $\hat{Q}(K_{c,c})-\hat{Q}(K_{c,i}) = \hat{Q}(K_{p,i})-\hat{Q}(K_{p,c}) = 7.5\%$ with $\hat{Q}$ being the empirical CDF function. In other words our in-sample data for the fittings always covers 7.5% probability mass of the available tail on each side (see also Section 3.6).[12]

---

[11] As in all other settings in this paper we use $S = 1300$, so $\Delta K = 1$ can be compared to the magnitude of one S&P 500 index point.

We show the results in terms of all three introduced loss functions in Figures 3.5 and 3.6 (for puts and calls respectively), with the cut-off points reported as quantiles of the respective densities. The results shown are for options with $\tau = 180$ days to maturity and all three SVJ parameter sets. Note that the maturity has only influence on the shape of RND generated by the model and no further impact on the actual methods.

In both figures we separately plot the results for IV extrapolation methods (solid lines) and methods that perform price- or density matching (solid lines and dashed lines respectively, but same color for the tail type).

Figures 3.5 and 3.6 report a rich set of results. We summarize the most important stylized facts and some specifics below. The full details can best be viewed in the figures. We specifically want to point out that some of the tail fitting methods do not appear in each of the panels in Figure 3.5. This is because of extremely high values of the loss function, with the implication that these methods should not be considered in such cases.

Generally, we find that the results vary along three main dimensions:

1. Assumed parametric tail
2. Fitting approach
3. Kurtosis of the density at hand (increasing from set 1 to set 3, see Table 3.2)

The effects are generally slightly different, depending on the considered loss function. Further, due to negative skewness the effects on the left tail are much larger than on the right (with the exception of set 1, which features an almost symmetrical density).

Results for the left tail Not surprisingly, the assumption of normal (or constant implied volatility extrapolation respectively) or log-normal tails works best for little excess kurtosis, but has the worst relative performance with heavier tails. We also want to point out that density matching of a normal tail performs better or at least equally good compared to constant IV extrapolation.

---

[12] No inner points are required for implied volatility extrapolation methods.

On the contrary, we find that the assumption of GEV or GPD tails works well when departing from the "close to normal" case. Only in the case of little kurtosis (set 1), the assumption of a GEV with price matching actually overestimates the tail-heaviness (this is only true for the put side, as on the call side the less heavier tail of the GEV is used). With the same exception, the results show that price matching is always superior to density matching, both in terms of performance (level of losses) and stability (smoothness of the curves). In most cases this is even true for the PDF loss, which is the loss function actually minimized by density-matching methods.

A further important finding is that the GPD assumption cannot be applied for densities with high kurtosis. In such cases, the fitted tails are extremely heavy and the resulting $E[X^4]$ can be several hundred times higher than the true moment (note the missing results for e.g. set 3).[13] For density matching, this effect does even occur with GEV tails, if the kurtosis is too high.

Among the methods that extrapolate IV, we find that the SVI model performs best in all situations. It slightly outperforms also the best candidate of the price matching methods (GEV). We see the largest difference for set 3 and the fourth moment loss if the tail is truncated at larger quantiles. Also, we want to point out that linear extrapolation in log-moneyness implies exponential extrapolation in regular moneyness. This effect can be observed in the $E[X^4]$-loss. For set 2, linear extrapolation in log-moneyness already overestimates the fourth moment (the effect is even stronger for set 3, results not reported due to very high losses). We thus recommend linear IV extrapolation in simple moneyness or strikes on the put side. Finally, the results indicate that the popular choice of constant IV extrapolation is inferior to almost all other considered methods.

---

[13] In some cases the fourth moment will actually not be defined. We only obtain finite results since we perform numerical integration on a finite interval.

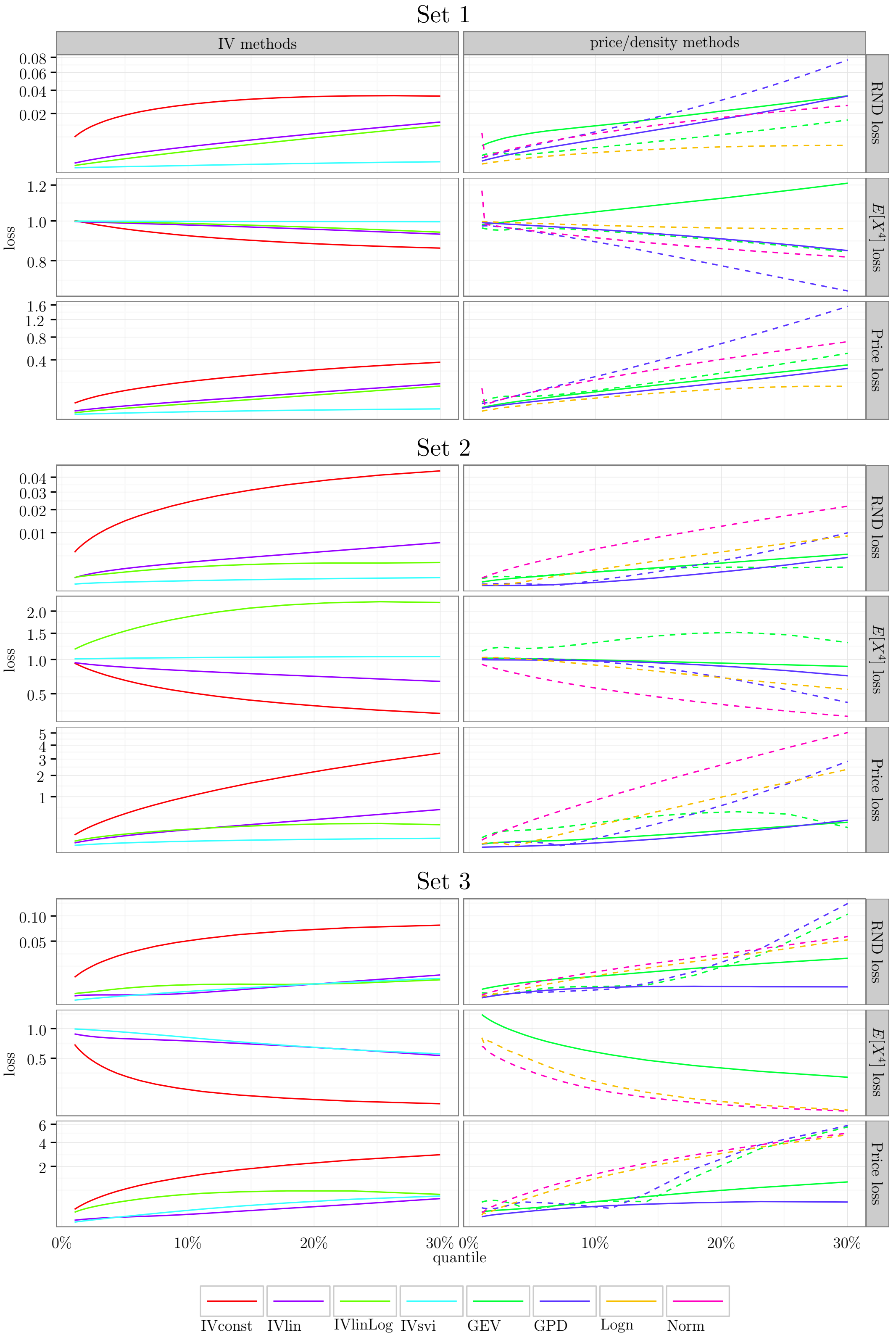


Figure 3.5: Loss comparison for the *left* tail with 180 days time to maturity including all sets and methods. In the right panels, solid lines represent fits in price dimension, dashed lines in PDF dimension. Details about the loss functions can be found in the text.

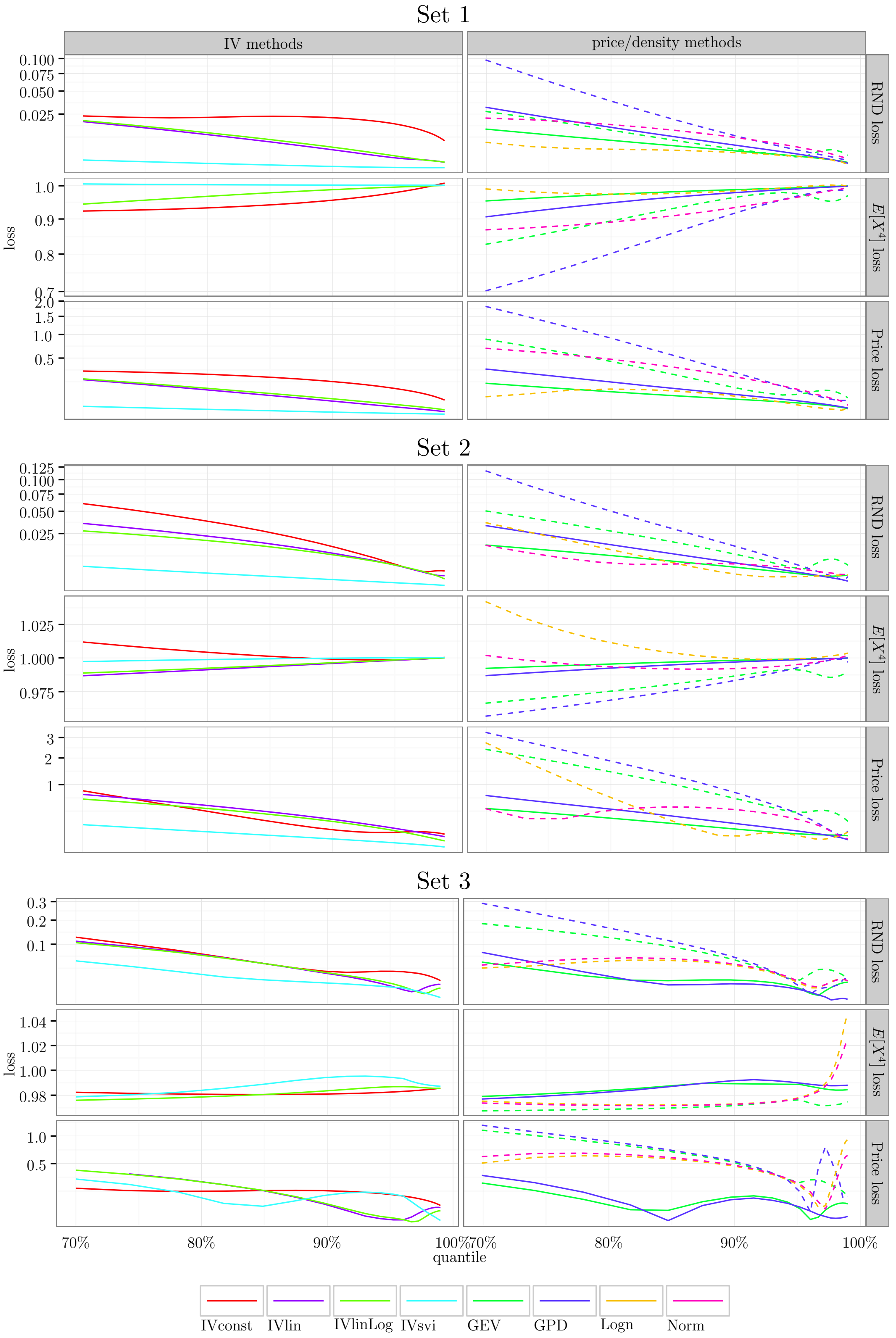


Figure 3.6: Loss comparison for the *right* tail with 180 days time to maturity including all sets and methods. In the right panels, solid lines represent fits in price dimension, dashed lines in PDF dimension. Details about the loss functions can be found in the text.

Results for the right tail Similar results can be obtained for tail-fits using call data (Figure 3.6). In general, the ranking of methods remains the same. At the same time, the magnitude of losses is significantly smaller on the right side of the density compared to the left side. This means that the correct choice of a tail fitting method for calls is not as crucial, especially if the total error from puts *and* calls is important.

Due to negative skewness, methods that do not imply heavy tails perform noticeably better (this includes GEV with price matching on the call side). Although constant IV extrapolation also falls in this category, it still performs worst in most cases.

Finally, on the call side we can observe a further issue of density matching methods. With very thin tails (high kurtosis and negative skewness), obtaining a good fit becomes more difficult. As a result, the fits are numerically very unstable and sometimes deliver spurious results. This leads to noisy loss functions for high quantiles (best observed for set 3 in Figure 3.6). We comment further on this effect in Section 3.6.

Deviating effects for shorter time to maturity As already mentioned above, the time to maturity per-se has no impact on the numerical experiments in this paper. Still, the SVJ model produces different densities for different values of τ. For a comparison, the corresponding loss plots with $\tau = 30$ days can be found in the appendix. We omit results for $\tau = 360$ days, since they are almost identical to the ones presented in this section.

An interesting observation between the results in this section and the appendix is that GPD price matching still works well for an excess kurtosis of 3.18 (set 2 and $\tau = 180$ days), but already results in too heavy tails for an excess kurtosis of 5.76 (set 3 and $\tau = 30$ days). We can therefore provide a rough limit (in terms of expected kurtosis) up to which this method can still be used.

We conclude that it is possible to compensate missing data up to 15% probability mass in the tails, by fitting a parametric tail or by IV extrapolation. However, the tail assumption in combination with an appropriate estimation method is important.

1. For densities with low kurtosis, price matching and a GPD assumption, or linear and SVI IV extrapolation work well.

2. For high kurtosis, price matching and a GEV assumption, or IV extrapolation with the SVI work well.

In all considered cases, the SVI method is the best choice (slightly outperforming the GEV with price matching). On the other hand, it has the drawback that in theory the resulting combination of tails and any other density estimation in the center, might not deliver a valid density object (that integrates to unity). In any case, the choice of a suitable method is much more important for the put tail, since the fitting errors are much larger than for the call tail. The common practice of constant IV extrapolation is clearly inferior to all other methods.

The previous discussion is based on the performance under all three loss functions. Depending on the requirements of the application for which the fitted tails are subsequently used, a different ranking of the methods might be possible, especially if other factors such as computational simplicity play a role.

3.5.2 Numerical performance under realistic market conditions

In the previous section, we investigate the performance of the different methods under the assumption that option prices are observed without noise and with fine strike increments of $\Delta K = 1$. However, such a scenario is rather unlikely in reality.

Consequently, we provide a supplementary analysis. The aim is to check the performance of the tail estimation methods using "realistic" put and call option slices. For this purpose, we consider maturity-dependent strike price intervals and cut-off points $K_{c,c}$ and $K_{c,p}$. Further, we randomly remove some options and disturb a part of the remaining true option prices in order to mimic microstructure noise.

Table 3.3 shows the corresponding values used in our experiment. They reflect characteristics of observed market data. That is to say, more data is available for the left tail (OTM puts) compared to the right tail (OTM calls) and the quantile range of observed strikes decreases with increasing maturity. The strike price intervals vary between 5 and 25 (roughly corresponding to S&P 500 index points) and are larger for longer time to maturity. Further we add noise to a fraction of option prices, where the magnitude of noise is distributed as

$$\epsilon \sim \mathcal{N}\left(0, \frac{s(P)P}{7}\right).$$

The term $\frac{s(P)P}{7}$ denotes the standard deviation of the price error and depends on the option price itself. The function $s(P)$ denotes the relative bid-ask spread of an option (option spread divided by its price). We estimate $s(P)$ from empirical option data. More precisely, we pool options with different expiries during a period of ten years (2000 until 2010) and then fit a cubic smoothing spline through the observed $(P_i, (A_i - B_i)/P_i)$ pairs with A_i and B_i denoting the bid and ask price of single options respectively. We use 1/7 as scaling factor for the standard deviation, such that less than 0.1% of the noisy prices lie outside the empirical bid-ask spread.

Table 3.3: Description of the realistic experiment: τ gives the expiry in days, $\hat{Q}(K_{c,p})$ and $\hat{Q}(K_{c,c})$ are the quantiles of the cut-off strikes, and ΔK denotes the strike price interval. The term *remove* indicates the fraction of options that are removed and *noise* the percentage of remaining options with an added noise component.

τ	$\hat{Q}(K_{c,p})$	$\hat{Q}(K_{c,c})$	ΔK	noise	remove
30	0.02	0.95	5	35%	15%
180	0.05	0.90	20	35%	15%
360	0.10	0.85	25	35%	15%

We use this procedure to construct a set of put and call option prices, and then apply the different methods to get estimates of the risk-neutral density tails. In addition, we repeat the whole exercise (i.e. data generation and tail estimation) 500 times to obtain an indication about the distribution of the different losses.

The corresponding outcomes for expiries 30 days and 180 days are presented in Figures 3.7 and 3.8 respectively.[14] They summarize the accuracy and stability for all methods in both the right and left tail, and all parameter sets respectively.

The accuracy is measured through the median loss and shown with points in the plot. As expected, the results are conform with the outcomes presented in the previous analysis for the corresponding cut-off points. That is, the most accurate results overall are obtained using either price fitting methods with tails drawn from a GPD or GEV

---

[14] Results for expiry 360 days are again qualitatively similar to those shown for time to maturity 180 days and therefore not reported for the sake of brevity.

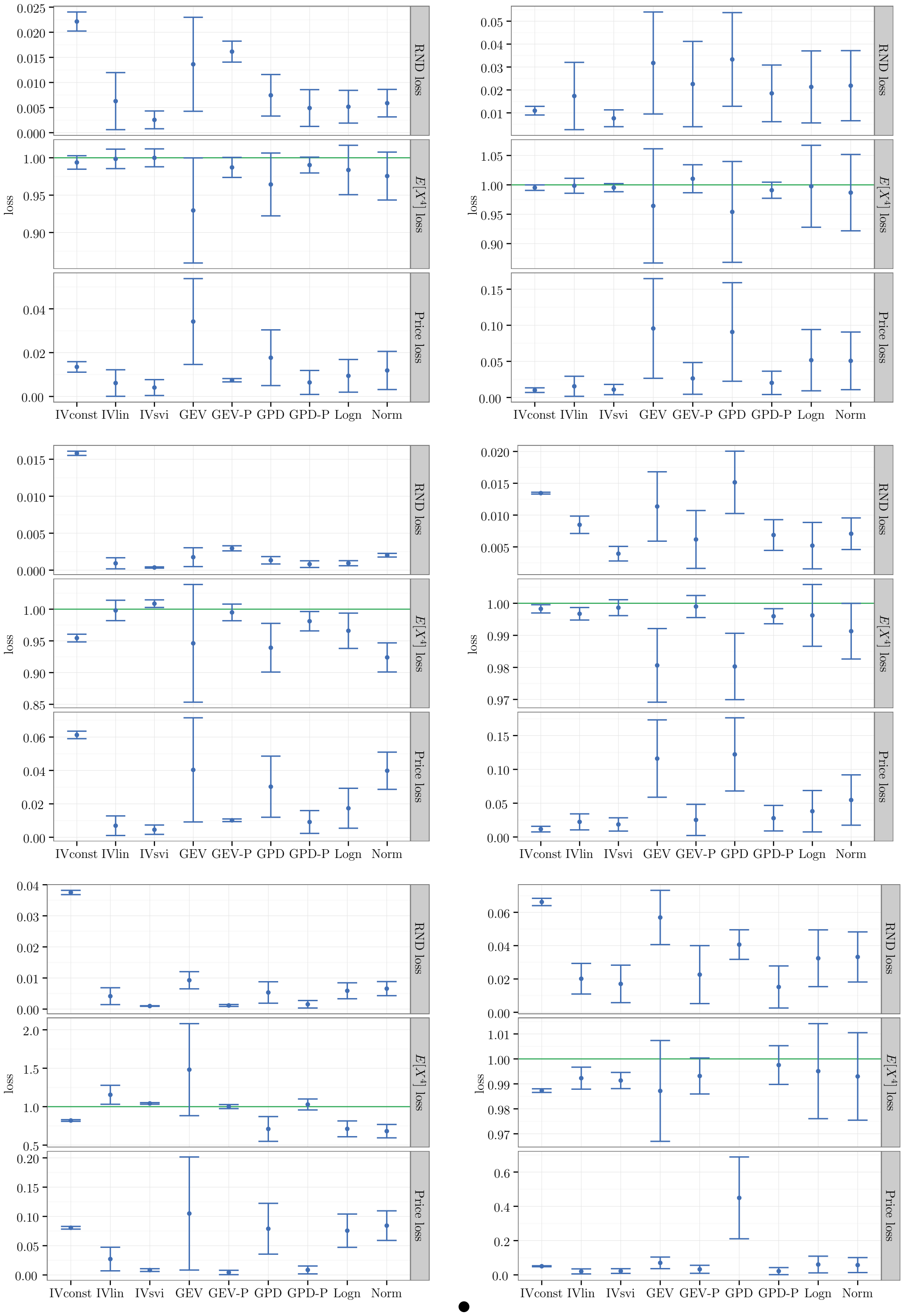


Figure 3.7: Accuracy and stability results for all methods and time to maturity 30 days. Parameter sets $1-3$ arranged vertically, each with put and call tail results on the left and right side respectively (note the different y-scales!). Points are median losses whereas bars indicate the median absolute deviation (over 500 repetitions). The suffix "-P" indicates price-fitting.

distribution or implied volatility extrapolation by means of the SVI model of Gatheral (2004).

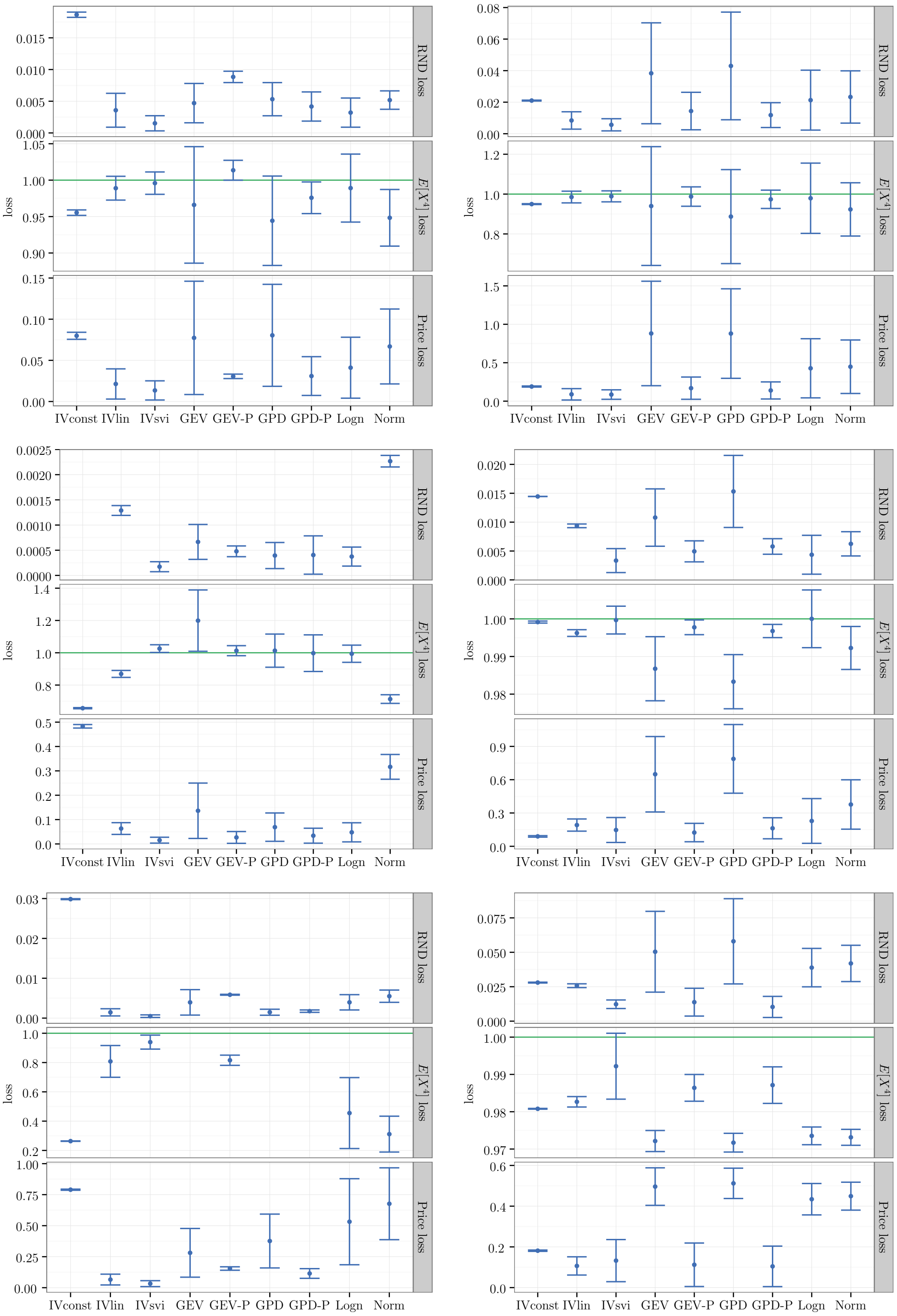


Figure 3.8: Accuracy and stability results for all methods and time to maturity 180 days. Parameter sets $1 - 3$ arranged vertically, each with put and call tail results on the left and right side respectively (note the different y-scales!). Points are median losses whereas bars indicate the median absolute deviation (over 500 repetitions). The suffix "-P" indicates price-fitting.

An advantage of this second and more realistic experiment is its capacity to express the robustness/stability of the tail estimation methods. It is measured using the median absolute deviation of the losses, and indicated with bars in the plot. We find that the density-matching approaches usually show the highest variance. This is not surprising since these methods rely on a prior estimate of the RND in order to fit the corresponding tails. In contrast, the other methods merely depend on observed put and call prices or their Black-Scholes implied volatilities. The stability of the best approaches with respect to accuracy is rather similar. However, fitting prices with a GEV assumption seems to be slightly more robust than modelling prices with a GPD. Finally, we notice that constant implied volatility extrapolation is very stable. Yet, the median loss is typically so high that it still performs weakly overall.

3.6 Robustness

In this section we present our findings regarding the actual implementation of the methods presented in Section 3.3, in particular the ones that match parametric tails (i.e. IV extrapolation is not considered here). This includes issues that we encountered with the data and optimization routines, but do not appear in the final results (since properly taken care of) in Section 3.5.

The contents of this section should be considered a practical guideline for researchers who are interested in using the methods examined in this paper.

3.6.1 How much data to use

For all methods, the cut-off point ($K_{p,c}$ or $K_{c,c}$) is typically either dictated by the data or is fixed by construction. On the other hand, the “inner” point ($K_{p,i}$ or $K_{c,i}$), which limits the in-sample data, can in principle be chosen arbitrarily. Hamidieh (2013) uses 10% of the available data in the tails, while Figlewski (2010) chooses $K_{p,i}$ such that $\hat{Q}(K_{p,i}) - \hat{Q}(K_{p,c}) = c$ (for calls analogously), with c some probability mass (in the case of the specific paper $c = 3\%$). Another alternative would be a specific difference in strike, return, etc.

In this paper we also use the method suggested by Figlewski (2010), because it is applicable rather independently of the market conditions (while e.g. strike or moneyness ranges can change). The actual amount of used data will of course still depend on the density and position of $K_{p,c}$ and $K_{p,i}$. See Figure 3.1 for an illustration of this effect (difference between call and put side).

Similar to a bandwidth parameter in non-parametric regressions, the choice of the parameter c allows for a bias-variance tradeoff for the tail fittings. With an increasing c there is more in-sample data available for the estimation. On the other hand, this additional data comes (by construction) from the more central part of the distribution and is therefore less relevant for the tail.

In our analysis $c \in [5\%, 10\%]$ delivered good results, specifically in terms of reduced variance, with moderately increased bias. The figures reported in Section 3.5 are all obtained with $c = 7.5\%$.

As with the losses reported in Section 3.5, we observe that effects vary depending on the method, and more importantly for put and call tails. In general, due to the steep right tail (i.e. for calls) in the case of negatively skewed densities, smaller values for c are more appropriate on the right side, while for the longer left tail (i.e. for puts) a higher c should be used.

We illustrate this bias-variance tradeoff in Figure 3.9, for SVJ set 2 and $\tau = 180$ days. This figure shows the same experiment as in Section 3.5.2, but repeated for different values of c (we only report the methods that actually require the choice of a c). The main observations are:

- Bias *and* variance increase with c if we use density matching under GEV or GPD assumption.
- GEV and GPD with price matching exhibit the classical bias/variance tradeoff. For the latter, bias first decreases and increases slowly for $c > 15\%$.
- The effects slightly differ, depending on the loss function.

We report the corresponding plot for the right tail in the appendix for completeness. In this case, all methods exhibit the expected behavior with increasing c (decreasing variance, increasing bias). Again, the effect on the call side of the data is significantly smaller.

3.6.2 Fat tails

When working with GPD or GEV tails, it is by construction possible to obtain densities with “fat tails”. For $\xi \geq \frac{1}{n}$, $n \in \mathbb{N}$, the n^{th} order moment of both distributions is not

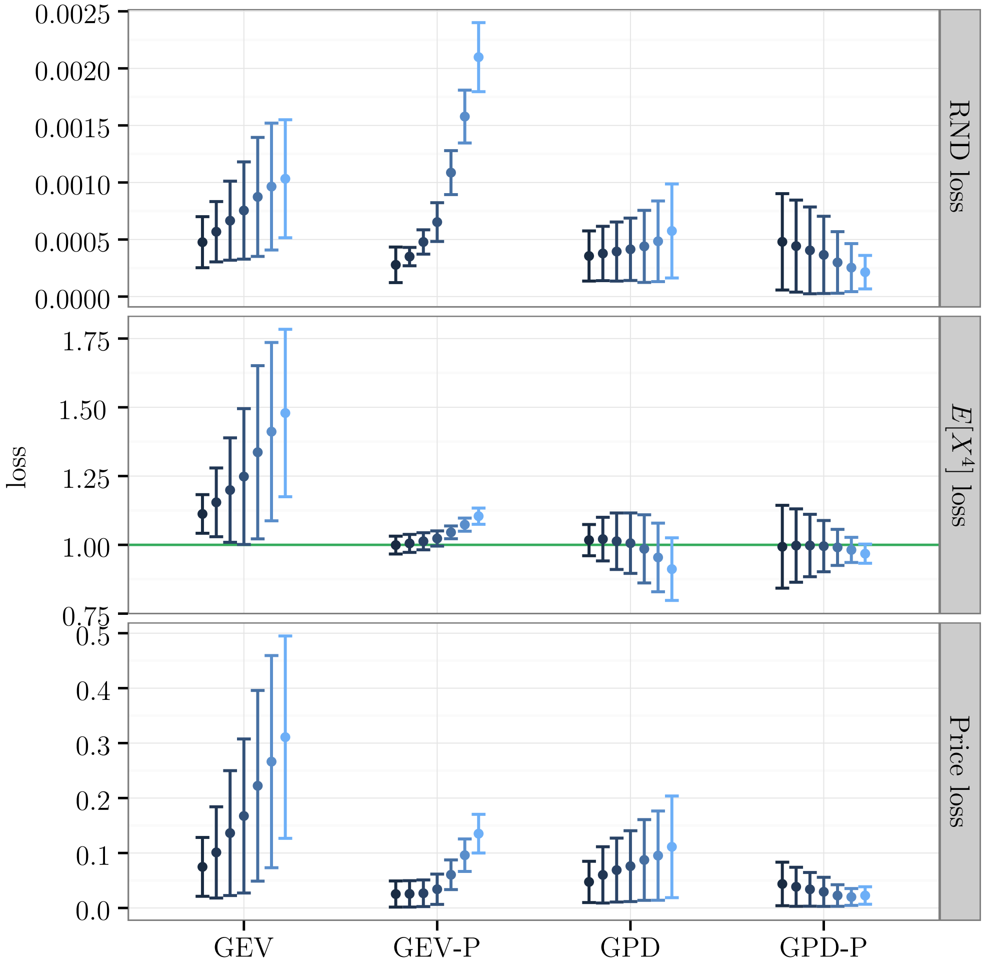


Figure 3.9: Mean/Variance comparison of all tail fitting methods that require a choice of $c = \hat{Q}(K_{p,i}) - \hat{Q}(K_{p,c})$, for different values of c. The results are shown for *puts*, set 2 and 180 days to maturity. The error bars represent the median and median absolute deviation of losses from 500 Monte Carlo fits including noise. Results are shown for $c \in \{0.025, 0.05, 0.075, 0.1, 0.15, 0.2, 0.25\}$, from left to right, colored according to their size. The suffix "-P" indicates price-fitting.

finite. This property is transferable to constructed densities which only use these parametric specifications in the tails.[15]

In practice, this leads to cases in which we significantly overestimate the fourth moment when using fitted GEV or GPD tails.[16]

We want to point out that for applications that do not require the (higher) moments to be finite (or that numerically only evaluate the density on a bounded interval), the fit in such cases can still be reasonably good. We can observe this by comparing the different loss types in Section 3.5.

---

[15] For the GEV, this is only true if one uses the *right* GEV tail, since the left GEV tail has bounded support for $\xi > 0$.

[16] As mentioned in Section 3.3, we often obtain moments that are several hundred times higher than the true one. In fact, they might even not defined at all.

More importantly, we want to stress that in *all* cases in which we observed such fits, this was due to unsuitable parametric assumptions or issues with the data. More specifically

1. When fitting GPD tails to densities with high kurtosis, the necessary in-sample data covers far more than what could be considered the "tail" of the distribution (see also (Ivanovas, 2014) for further results with actual market data).[17]
2. When option prices are observed with noise (in practice or as generated in Section 3.5.2) and the tail is fitted far OTM, spurious results are possible due to local optima in the optimization (see also the following subsections).

Therefore, results of GEV or GPD fits with $\xi > 0$ and especially with $\xi > 0.1$ should always be checked carefully.

3.6.3 Less can be more

Depending on the density at hand, the cut-off points $K_{p,c}$, $K_{c,c}$ and the chosen amount of data (determined by c as introduced above), it is possible to run into another numerical difficulty, namely "weak identification".

Too far out in the tails, either the density or OTM option prices are almost zero. The resulting effect can be loosely compared to a "flat likelihood surface", i.e. there is (too) little information to distinguish an optimum in the loss function. This can lead to spurious results in the fitted parameters and therefore densities. Besides the discussed point of fat tails (i.e. the tail decay can not be recovered correctly from the data) we have also observed cases with even increasing tails, leading to multi-modal densities.

Further, with lower prices in the tails there is a worse "signal-to-noise ratio", i.e. the relative size of noise compared to the prices becomes large. With far OTM options, it can e.g. happen that the spread is as large as the price itself. This effect is even amplified for the second derivative of the prices (i.e. the density). This means that fitting tails becomes increasingly sensitive to (microstructure) noise in the data (see Sections

---

[17] As a validation mechanism of a GPD choice, one can compare the fitted tail shape parameter ξ to the one of a GEV fit to the same data. These should be similar (say ±0.05) in cases where the GPD assumption is still valid.

3.2 and 3.5.2), if the cut-off point is chosen too far out in the tail. Or in other words, if the rest of the in-sample data is relatively flat, optimization might deliver results that simply represent the "best fit to the noise". This result can also be observed in Figures 3.5 and 3.6, in the ends of the tails.

Thus, it can be strongly beneficial to discard some portion of the data in the tails, up to a point where microstructure noise can be ignored (at least for most of the in-sample data).

It is already common practice in empirical option pricing literature to discard some data in the tails by some rule (e.g. minimum price, or a pre-defined interval in moneyness). This is known as "data filtering" (see Chapter 4 for a review of typical data filters and further arguments in favor of this practice).

Following our results from Section 3.5, we can recommend cut-off points such that the fitted tails cover probability mass in the range of 1% to 5% (sometimes even up 10% or 15% is possible). The bias increase due to less data in the tails is small, compared to the variance improvement.

Finally we would like to add that IV extrapolation methods are relatively robust with respect to these data issues, since the fitting is done in a transformed space with better properties in the tails.

3.6.4 Optimization issues

When optimizing the loss function of methods that perform price or density matching, the global minimum can be difficult to achieve. Often, several parameter combinations are locally optimal and provide a challenge to the optimization algorithm.

In our applications we tried two optimization routines, namely Nelder-Mead Simplex and Differential evolution (a stochastic, global optimizer).

The second can still be fooled into local minima. To achieve robust results, one needs to increase search parameters (e.g. number of population members) to a point where it becomes relatively slow. The Simplex algorithm on the other hand is fast, but often converges into non-optimal solutions unless good starting parameters are provided. This is especially true for methods using the GEV density, followed by GPD. We achieve the best results (in terms of speed and reliability) by using the Simplex with preceding grid search for starting parameters.

Grid values of the tail shape parameter ξ are the most critical. Parameter σ depends on the particular space in which the optimization is performed (log-returns, simple-returns, strikes). For ξ we find values between -0.6 and -0.05 for puts and between -1.6 and -0.2 for calls to work best. We always include some points with $\xi > 0$, although they rarely turn out to be optimal (see Section 3.6.2). The grid values of σ should be chosen in the range of typical market volatility. Additionally, a very small value for σ (say 0.03 for log-returns) should also be included. The typical grid size in our optimizations was in the range of 20 - 30 points. We finally use the 3-4 best parameter combinations of the grid search as starting parameters for the Simplex algorithm.

3.7 Conclusion

In this paper, we summarize and review several methods to estimate risk-neutral density tails from option data. They can be categorized into three groups: methods that apply density-matching, price-fitting approaches and methods that extrapolate the implied volatility smile. Moreover, we introduce a new tail extension approach that combines the idea of price matching used in (Hamidieh, 2013) with the distributional assumption of Figlewski (2010) (i.e. generalized extreme value tails).

To compare the different estimation techniques, we conduct a performance analysis. By imposing a data-generating process for the underlying asset, we compute option prices and the corresponding risk-neutral density. We then analyze how accurately the different methods can recover the tail of this density, when the computed option prices are only used up to a certain maximum or minimum strike price.

Overall, we find that the best performance regarding accuracy and stability is achieved either with price matching and tails drawn from a generalized extreme value (GEV) distribution, or implied volatility extrapolation by means of the stochastic volatility implied (SVI) model of Gatheral (2004). In addition, we provide evidence that the employed loss function plays an important role. More specifically, parametric tails fitted based on price-matching approaches generally outerperform those using density-matching, independently of the chosen parametric specification.

The assumption of a GEV tail is only inappropriate when the true density is close to a normal with small standard deviation. In such cases, we observe that the fitted

GEV distribution implies a too heavy tail, and the assumption of another parametric distribution such as the normal or log-normal provide good performance instead.

The generalized pareto distribution (GPD) is best justified from a theoretical standpoint and typically yields results very similar to the GEV. But for densities with high kurtosis, we often observe that GPD tails substantially overestimate the tail-heaviness of the true distribution, which particularly affects the corresponding higher moment estimates.

Implied volatility extrapolation using the SVI model (slightly) outperforms even the best choice of parametric tails. However, this method does not impose any explicit condition on the cumulative distribution function, and therefore the resulting density might not integrate to unity when the SVI is only used to obtain the tails of the distribution.

Finally, the commonly applied method of constant implied volatility extrapolation performs rather badly in almost all considered cases. While its results have very small variance (i.e. it is robust to noise in the data), it strongly underestimates the tails of the distribution and consequently the (higher) moments.

3.A Univariate extreme value theory

The Fisher-Tippett-Gnedenko theorem is a fundamental result in extreme value theory (EVT), as it provides a description of the limiting behavior of appropriately normalized sample maxima.

Let us assume to have a sequence $\{X_i\}_{i\geq 1}$ of independent and identically distributed random variables with distribution function F and define $M_n = \max(X_1, \ldots, X_n)$. If there exist sequences of normalizing constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that

$$P((M_n - b_n)/a_n \leq x) \rightarrow G(x), \quad \text{for } n \rightarrow \infty, \tag{3.9}$$

where $G(x)$ is a non-degenerate distribution function, we say that the distribution F is in the maximum domain of attraction of G and we write $F \in MDA\ (G)$.

Fischer and Tippett (1928) and Gnedenko (1943) have shown the following theorem:

Theorem 1 (Fisher-Tippett-Gnedenko)**.** *If $F \in MDA\ (G)$, then G must be of the same type as one of the following distributions:*[18]

---

[18] Two distributions G_1 and G_2 are said to be of the same type if there exist constants $a > 0$ and

Type I (Gumbel):

$$\Lambda(x) = \exp\left[-\exp(-x)\right], \quad x \in \mathbb{R},$$

Type II (Fréchet):

$$\Phi_\alpha(x) = \begin{cases} 0, & x \leq 0, \\ \exp(-x^{-\alpha}), & x > 0, \end{cases}$$

Type III (Reversed Weibull):

$$\Psi_\alpha(x) = \begin{cases} \exp\left[-(-x)^\alpha\right], & x < 0, \\ 1, & x \geq 0, \end{cases}$$

for some parameter $\alpha > 0$*.*

It is possible to summarize the three types mentioned in Theorem 1 above in a single parametric family known as the generalized extreme value distribution (GEV). The corresponding distribution function is

$$G_{\mu,\sigma,\xi}(x) = \begin{cases} \exp\left\{-[1+\xi(x-\mu)/\sigma]^{-1/\xi}\right\}, & \text{if } \xi \neq 0, \\ \exp\left\{-\exp[-(x-\mu)/\sigma]\right\}, & \text{if } \xi = 0, \end{cases}$$

for all x such that $1+\xi(x-\mu)/\sigma > 0$. There are three parameters: a location parameter $\mu \in \mathbb{R}$, a scale parameter $\sigma > 0$ and a shape parameter $\xi \in \mathbb{R}$. The parameter ξ is important since it determines the distribution type. If $\xi < 0$, the support of G is bounded from above at $x = \mu - \sigma/\xi$ and we recover a reversed Weibull distribution with $\alpha = -\xi^{-1}$. For $\xi = 0$, G corresponds to a Gumbel distribution with unbounded support and whose upper tail shows an exponential decay. Finally, for $\xi > 0$, the support of the distribution has a lower bound at $x = \mu - \sigma/\xi$ and we get a Fréchet distribution with $\alpha = \xi^{-1}$. In this case, the shape parameter also controls the decay of the upper tail and thus the potential severity of extreme events. In particular, the rth moment of the GEV distribution is only finite if $r\xi < 1$.

Remember that Theorem 1 only makes a statement about the limiting behavior of appropriately normalized maxima but does not provide conditions when such a limit law exists. However, it can be shown that just about any continuous distribution used in statistics and finance is in the MDA of G.[19]

A further important family of probability laws in EVT is the so-called generalized

---

$b \in \mathbb{R}$ such that $G_1(ax+b) = G_2(x)$ for all $x \in \mathbb{R}$.

[19] For a more detailed theoretical discussion about that point and some proofs, we refer to Resnick (1987).

Pareto distribution (GPD). The GPD is usually expressed as a two-parameter distribution of the form

$$H_{\sigma,\xi}(x) = \begin{cases} 1 - (1 + \xi x/\sigma)^{-1/\xi}, & \text{if } \xi \neq 0, \\ 1 - \exp(-x/\sigma), & \text{if } \xi = 0, \end{cases}$$

with $\sigma > 0$. The support is $x \geq 0$ when $\xi \geq 0$ and $0 \leq x \leq -\sigma/\xi$ when $\xi < 0$. The GPD subsumes some other distributions under its parameterization. For $\xi > 0$ we obtain a reparameterized version of the usual Pareto distribution, if $\xi < 0$, a Pareto type II distribution is recovered and the case $\xi = 0$ corresponds to the exponential distribution.

We can extend the two parameter family by adding a location parameter μ. We get $H_{\mu,\sigma,\xi}(x)$ which is defined as $H_{\sigma,\xi}(x - \mu)$.

The next theorem (see Balkema and De Haan (1974) and Pickands (1975)) is often referred to as the second theorem in EVT. It highlights the importance of GPD when modeling extreme events as values above a high threshold.

Theorem 2 (Pickands-Balkema-De Haan)**.** *Assume that X is a random variable with distribution function $F \in MDA\ (G)$ and let $x^* = \sup\{x \in \mathbb{R} : F(x) < 1\}$ be the finite or infinite upper endpoint of the distribution F. Then, we have*

$$\lim_{x \to x^*} \sup_{0 \leq x \leq x^* - u} \left| P(X - u \leq x | X > u) - H_{\sigma(u),\xi}(x) \right| = 0,$$

for a positive measurable function $\sigma(u)$ and a fixed shape parameter ξ.

We conclude that for a large class of distributions, i.e. $F \in MDA\ (G)$, GDP can accurately approximate the probability of an excess above a sufficiently high threshold.

If GPD is an appropriate choice to model the conditional distribution of the excesses above some high threshold u, we may use the same distributional assumption for the tail above this high threshold too. Let us demonstrate that this is indeed true. We have for $x \geq u$

$$F(x) = P(X \leq x) = (1 - P(X \leq u))P(X - u \leq x | X > u) + P(X \leq u)$$

and we know that $P(X - u \leq x | X > u)$ can be approximated with $H_{\mu,\sigma,\xi}(x)$ for u large. In addition, an estimate of $P(X \leq u)$ can be obtained using the empirical distribution function F_n. Hence, we get the following approximation of $F(x)$:

$$\hat{F}(x) = (1 - F_n(u))H_{\mu,\sigma,\xi}(x) + F_n(u).$$

It is relatively easy to show that $\hat{F}(x)$ is also generalized Pareto distributed with identical shape parameter ξ but with a different scale parameter $\tilde{\sigma} = \sigma(1 - F_n(u))^{\xi}$ and new location parameter $\tilde{\mu} = \mu - \tilde{\sigma}((1 - F_n(u))^{-\xi} - 1)/\xi$.

We close this short introduction of EVT with a result known as closure property of GPD. The following lemma formalizes this:

Lemma 1. *Let X be a random variable with distribution function $F \in MDA\ (G)$ and further define $x^* = \sup\{x \in \mathbb{R} : F(x) < 1\}$. If it is true that for some high threshold u_0 and $0 \leq x \leq x^* - u_0$, we have*

$$P(X - u_0 \leq x | X > u_0) = H_{\sigma(u_0),\xi}(x)$$

with $\xi \in \mathbb{R}$ and $\sigma(u_0) > 0$. Then we obtain for any higher threshold $u_1 > u_0$

$$P(X - u_1 \leq x | X > u_1) = H_{\sigma(u_1),\xi}(x)$$

with

$$\sigma(u_1) = \sigma(u_0) + \xi(u_1 - u_0).$$

3.B Characteristic function of the Bates model

The characteristic function of the SVJ model of Bates (1996) can be found in the original paper itself, or with slightly different notations in e.g. Bakshi et al. (1997) or Borak et al. (2005). The latter provides the most concise notation as follows (for the log-price x_t under the risk-neutral measure):

$$\Phi_{x_t}(z) = \Phi^D_{x_t}(z)\Phi^J_{x_t}(z) \tag{3.10}$$

where

$$\Phi^D_{x_t}(z) = \frac{\exp\left\{\frac{\kappa\theta t(\kappa - i\rho\sigma z)}{\sigma^2} + izt(r - \lambda\bar{k}) + izx_0\right\}}{\left(\cosh\frac{\gamma t}{2} + \frac{\kappa - i\rho\sigma z}{\gamma}\sinh\frac{\gamma t}{2}\right)^{\frac{2\kappa\theta}{\sigma^2}}} \cdot \tag{3.11}$$

$$\exp\left\{-\frac{(z^2 + iz)v_0}{\gamma\coth\frac{\gamma t}{2} + \kappa - i\rho\sigma z}\right\}, \tag{3.12}$$

and

$$\Phi^J_{x_t}(z) = \exp\left\{t\lambda\left[\exp\left(\frac{-\delta^2 z^2}{2} + i(\ln(1+\bar{k}) - \frac{1}{2}\delta^2)z\right) - 1\right]\right\} \tag{3.13}$$

give the diffusion and jump part respectively. The parameters are as described in Section 3.4.1. x_0 denotes the initial log-price, v_0 the initial variance and i the imaginary number.

For an extensive review of characteristic functions and their applications in option pricing we refer to Schmelzle (2010). We briefly list some properties that are needed in our application.

The characteristic function of a real random variable X is defined as

$$\Phi_X(z) = E\left[e^{izX}\right] = \int_{-\infty}^{\infty} e^{izx} g_X(x)dx, \tag{3.14}$$

where $g_X(x)$ is the density function of the variable and the integral above is the Fourier transform $\mathcal{F}[g_X(x)]$ of the density function. We therefore obtain the density function by Fourier inversion of the characteristic function

$$g_X(x) = \mathcal{F}^{-1}[\Phi_X(z)] = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-izx}\Phi_X(z)dz. \tag{3.15}$$

For the CDF we have (following Schmelzle (2010) and Gil-Pelaez (1951) respectively)

$$G_X(x) = \frac{1}{2} - \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{-izx}\Phi_X(z)}{iz}dz. \tag{3.16}$$

Since PDF and CDF are real-valued and due to symmetry in the real part of $\Phi_X(z)$, we can simplify to

$$g_X(x) = \frac{1}{\pi} \int_0^\infty \Re[e^{-izx}\Phi_X(z)]dz. \tag{3.17}$$

and

$$G_X(x) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \Re\left[\frac{e^{-izx}\Phi_X(z)}{iz}\right] dz. \tag{3.18}$$

To actually compute the PDF, we use the fast fourier transform (FFT) algorithm, which is an efficient means to compute sums of the type (discrete fourier transform)

$$w_u = \sum_{j=0}^{N-1} e^{-i\frac{2\pi}{N}ju} x_j \tag{3.19}$$

for $u = 0, \dots, N-1$. This can be used to compute a discretized version of (3.17).

The approach for the CDF is similar. But as pointed out in Section 3.4.1, the results of the FFT applied to (3.18) start oscillating (albeit on a very small scale) for negative log-returns far in the tail of the density. This is most likely due to z in the denominator in (3.18). We obtain a numerically better behaved CDF function in this range, by taking the first numerical derivative from computed option prices (see below) with respect to their strikes. We therefore omit the expression for the CDF for brevity.

Moments For our higher moment loss function we need the fourth, non-centralized moment of the respective densities. As discussed in Schmelzle (2010), we can obtain the n^{th} non-central moment as

$$E[X^n] = \frac{1}{i^n} \frac{d^n \Phi_X(z)}{dz^n}\bigg|_{z=0},$$

which we compute analytically in Mathematica.

Option prices As shown by Carr and Madan (1999), the FFT algorithm for the discrete fourier transform in equation (3.19) can also be used to obtain option prices.

Additionally to the fourier inversion, the pricing method requires the use of a “dampening factor” α, since the call price function in terms of log-returns is not square integrable (for details we refer to the original paper or the review in (Borak et al., 2005)). For our application we choose $\alpha = 1.2$, $N = 2^{11}$ discretization points, and discretization intervals of $\eta = 0.2$ in the frequency domain. We compute call prices by application of FFT and put prices by using the put-call parity.

3.C Additional plots

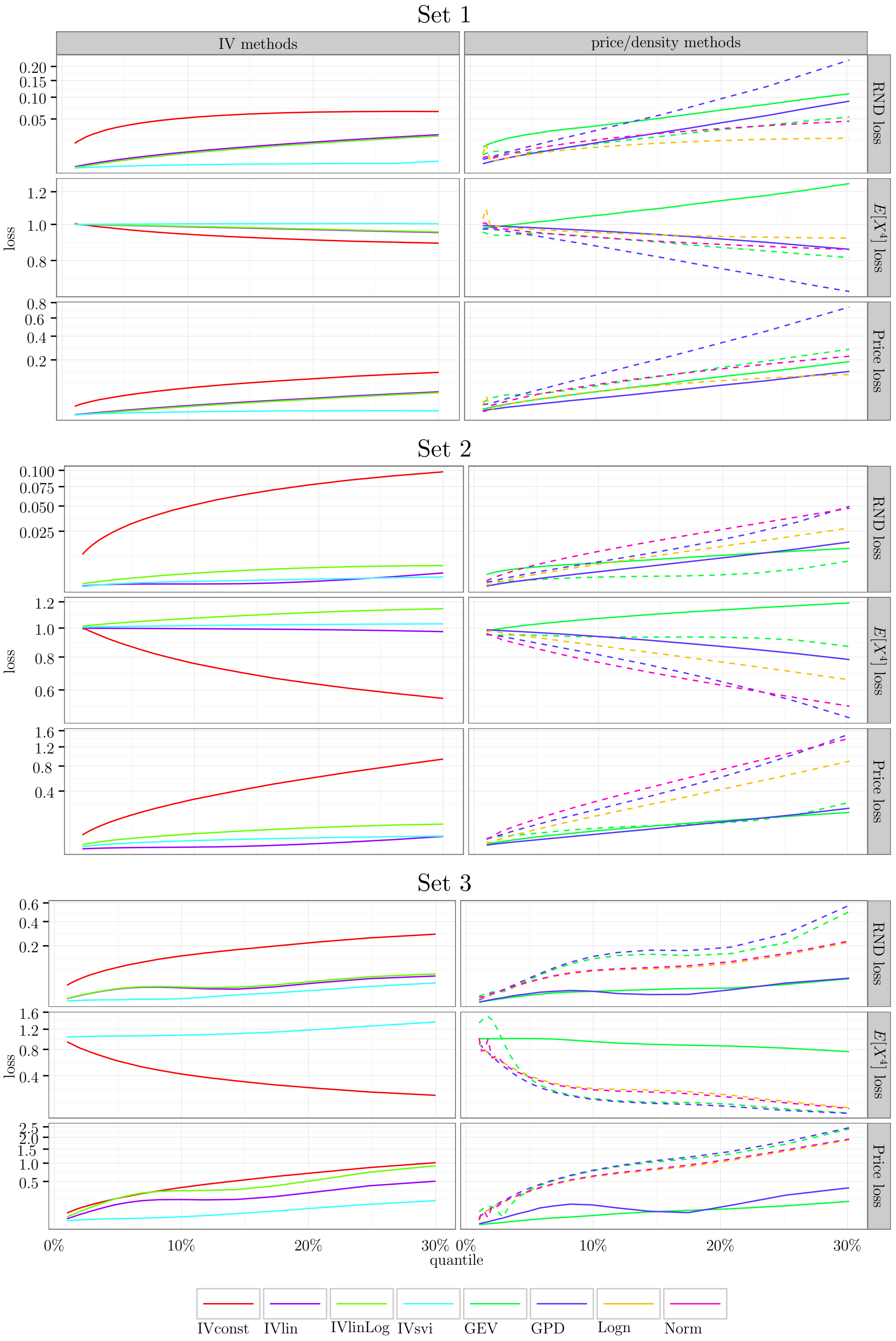


Figure 3.10: Loss comparison for the left tail with 30 days time to maturity including all sets and methods. In the right panels, solid lines represent fits in price dimension, dashed lines in PDF dimension.

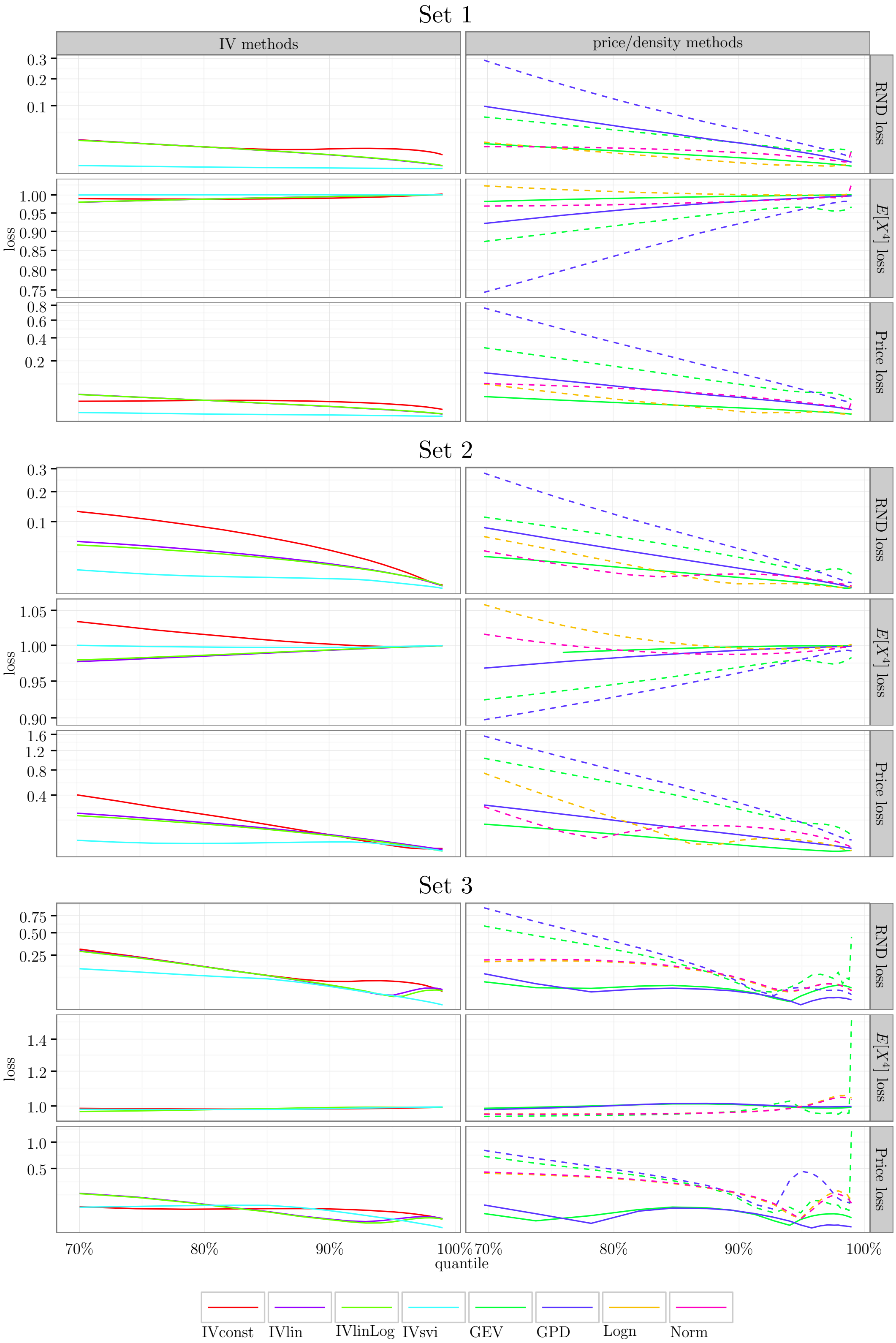


Figure 3.11: Loss comparison for the right tail with 30 days time to maturity including all sets and methods. In the right panels, solid lines represent fits in price dimension, dashed lines in PDF dimension.

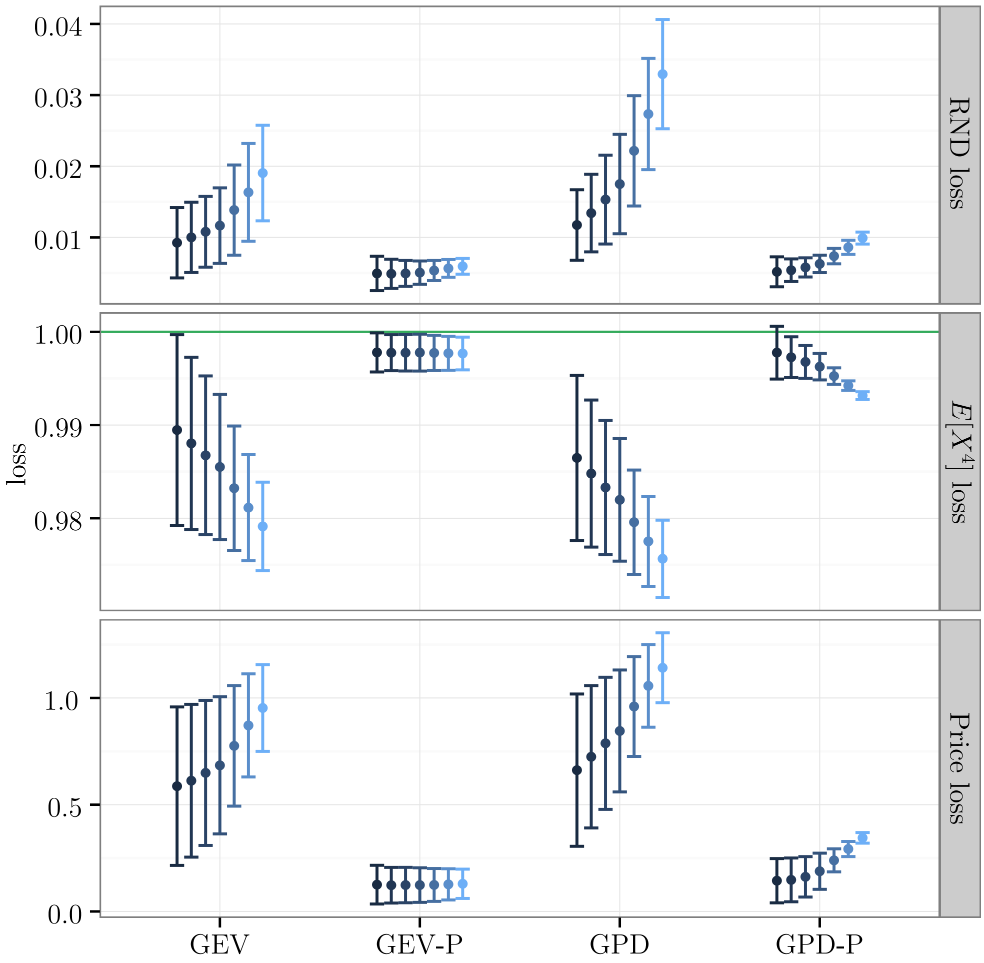


Figure 3.12: Mean/Variance comparison of all tail fitting methods that require a choice of $c = \hat{Q}(K_{p,i}) - \hat{Q}(K_{p,c})$, for different values of c. The results are shown for *calls*, set 2 and 180 days to maturity. The error bars represent the median and median absolute deviation of losses from 500 Monte Carlo fits including noise. Results are shown for $c \in \{0.025, 0.05, 0.075, 0.1, 0.15, 0.2, 0.25\}$, from left to right, colored according to their size.

Chapter 4

On working with option data: implications of data choice

Anselm Ivanovas
Pirmin Meier

Abstract

Option prices are helpful for various applications other than option pricing itself, given that they contain information on the risk factors driving the underlying price process. Interestingly, the step of data preparation is often marginalized in the empirical literature. This paper aims to close this gap. We discuss why filtering option prices might be sensible and present the most common approaches. In addition, and more importantly, we introduce new filtering methods based on the idea of "market activity". Using a large data set of intra-day S&P 500 option prices, we analyze the effect of the different filters. We provide evidence that they yield to statistically different strike/moneyness ranges and show that these differences can have a non-negligible impact on tail sensitive applications such as moment estimation.

Keywords: Option data, filtering, market activity, moment estimation.

4.1 Introduction

Exchange traded option data is a valuable tool in the financial literature, and its usage extends beyond the already wide field of option pricing itself. Option prices, especially when considered over a range of strikes, contain information about investors' risk attitudes and their expectations regarding the future evolution of the underlying. Thus they are helpful for various applications such as for example estimating the implied risk-neutral density (see (Figlewski, 2010) for a review of methods), managing risk (Aït-Sahalia and Lo, 2000), or studying the empirical pricing kernel and investors' risk preferences (Jackwerth, 2000; Rosenberg and Engle, 2002).

A very popular data choice for such research topics are options written on the S&P 500. This index is considered a good indicator of the real economy (because of its diverse constituents), and the corresponding options are therefore expected to contain information regarding investors' expectations on this economy. However, with up to 2300 - 2400 different liquid options available at a time for trading on the Chicago Board of options exchange, there is an enormous possible data set that researchers can exploit for their specific applications. It is therefore well-accepted practice to focus on some subset of this full data set for empirical work. While there exist some basic conventions in the literature as to which subset to choose, there does not seem to exist a final consensus on this topic. Moreover, the main focus of research papers is usually on the theoretical contribution and the final results, whereas the step of data preparation, known as "filtering" of option prices, is often marginalized.

In this paper, however, we focus specifically on data choice and cover in more detail this simple but important topic in empirical research. Departing from the full set of intraday S&P 500 option data, we discuss the reasoning to consider specific subsets and show how to specify them. To this end, we list and discuss a list of popular data filters (especially for the tails) and suggest a further tail filtering concept. Our approach builds on the idea that the resulting data should be recent and economically representative. More precisely, we develop three methods that rely on proxies for market activity, based on intraday data. For ease of application, we condense the resulting filter to a simple linear model that defines the data cut-off points as a function of time to maturity. We find that our new filter is generally more restrictive in the tails than other common filters applied in the literature.

In a second part of the paper, we compare the effect of different tail filters and provide evidence that they yield statistically different data ranges. In addition, we investigate whether these differences can have an impact on the outcomes of empirical applications. More precisely, we calculate option-implied moments of the risk-neutral distribution using the method of Bakshi et al. (2003). Not surprisingly, we find that in particular estimates for higher moments like skewness and kurtosis show a substantial dependence on the used filter for option data in the tails. However, we do not find that other (non tail-sensitive) applications depend as crucially on the filter.

Given that many applications are not particularly tail-sensitive, and that there exist appropriate extrapolation tools for those that are (see Chapter 3), we believe that our new activity-based filters offer an interesting alternative to the existing filtering techniques, because of their economic justification.

The rest of this paper is organized as follows: Section 4.2 contains a description of the option terminology used in this paper and introduces the data that is subsequently employed for our analysis. In Section, 4.3 we discuss why filtering might be sensible and provide a summary of the most common techniques. Further, we introduce new filtering methods that are based on the idea of market activity. In Section 4.4, we compare whether different filters yield to statistically different data sets and investigate the impact of these differences on empirical applications. Section 4.5 concludes the paper.

4.2 Terminology and data description

In this section we introduce the option data terminology used in this paper. We also provide a description of our data set, since it is needed to subsequently describe the filtering techniques.

Option data terminology

Some of the terms related to option data are used ambiguously in the literature. Therefore we summarize the most important ones and how they are defined in this paper. While we aim to provide a filter that can be applied to end-of-day data, we will primarily use intraday data to define our filters. Consequently, this list also contains terms related to intraday/tick data.

Trade (tick) An actual transaction of a specific option recorded by the exchange. Consists of a trading price and a trading volume.

Trade price The (unique) price recorded in a transaction of an option. Also known as “transaction price”. Trade prices have the advantage that they are consensus prices since two market participants actually agreed to a transaction at this price.

Quote (tick) We understand quotes as “level I quotes”, i.e. the combination of best bid and best ask offer in the order book at a given time. A quote therefore consists of a bid price (price at which the buyer offers to buy), a corresponding bid volume (number of contracts the buyer wants to buy), and analogously ask price and volume for the seller. A new level I quote is recorded whenever any of these four elements changes.

Quote price A quote has the disadvantage that it consists of two price indications and therefore no actual consensus price. When referring to a quote price we use the *mid-price* as an estimate for the latent true price, which is a common convention in the literature.

Closing price The closing price generally refers to the last price of an option observed at or before market closing. But the term is used somewhat ambiguously in the literature. On the one hand, it does not answer whether this last price is taken from quote prices only, trade prices only or even both (depending on what occurred more recently). On the other hand, it is not clear whether a closing price is necessarily the actual market price. This is because data providers might apply some sort of filtering or smoothing/fitting to obtain the finally reported price. We use the *mid-price of the last available quote tick at or before 15:15 CST*, which is the closing time of the market for S&P 500 options on the Chicago Board of Options Exchange (CBOE).

Option data

We use data from market data express, a subsidiary of CBOE. Specifically, their MDR-SPX product, which covers intraday level I quotes and trades of all (regular) S&P 500 options.

The market for SPX options is one of the most active index options markets in the world and it operates every trading day between 08:30 and 15:15 CST. At any time, option contracts with up to 12 times to maturity are traded, with the exact expiration date being the Saturday following the third Friday of the month. The exercise-settlement value is equal to the opening value of the index on the last business day (typically a Friday) before the expiry date, and trading usually ends one day earlier (typically a Thursday). Options on the S&P 500 are European options, exercise takes place on the last business day before expiration, which results in a delivery of cash on the next trading day after expiration.[1]

Our data period covers all trading days from 2008 to 2011. For our analysis, we exclude "quarterlies" and "weeklies"[2] and focus on the regular March, June, September and December expiries, plus three "near month" expiries. We further exclude options with expiries longer than one year or shorter than ten calendar days (see also next section about maturity filtering). Unless stated differently, all prices correspond to closing prices as defined in the previous section. We further assume that the risk-free interest rate r is deterministic and does not change intraday. We further consider dividends in the form of a dividend yield δ, for which we make the same assumptions. We obtain both on a daily basis by linear regression of several (at least 4) at-the-money (ATM) put-call option pairs, separately for each maturity as follows:

$$C_i - P_i = \alpha + \beta K_i,$$

with C_i and P_i the prices of a call and a put respectively, with common strike K_i. We then obtain estimates of annualized r and δ through the put-call parity, i.e. $r = \frac{\log(-\beta)}{-\tau}$ and $\delta = \frac{\log(\alpha/S)}{-\tau}$, where S denotes the underlying price and τ the time to maturity in fractional years.

4.3 Option data filters

When data filters are applied in empirical research, the discussion and motivation is often limited. We therefore review some reasoning why filtering might be sensible.

Bluntly speaking, the core idea in empirical economic research is to infer results or to make statements about a (theoretical) representative agent. By using financial market data, one tries to observe and use consensus prices between several market participants. Through the argument of price formation, information in such data can then be interpreted as "average attitude" of several market participants.

---

[1] The S&P 500 index option specifications can be found at `http://www.cboe.com/products/indexopts/spx_spec.aspx`

[2] Next to the regular option class (with expiry on Saturdays), the CBOE also has special one with expiries at the end of each quarter (end of March, June, September and December respectively) and one with very short contract duration of one week, expiring each Friday.

While there might be reasons to investigate specific parts of markets or single agents, the argument in favor of S&P 500 options in asset pricing is almost exclusively to make a statement about a "broad market". While rarely mentioned explicitly, we further make the assumption (and impose it for this paper) that this intention also implies the aforementioned idea of observing a consensus or equilibrium state between a broad variety of market participants.

Additionally to this economic representativeness of the data, the concern arises whether the data itself is technically accurate, i.e. if the exchange-recorded prices are actually a good measure for the economic information content. Specifically for option prices, this includes the issue of microstructure noise (minimum tick sizes, bid-ask bounce, non-synchronicity) of the option prices themselves and potentially the corresponding underlying asset.

We understand data filtering as a tool to ensure or at least improve these two main features (representativeness and accuracy) of the filtered data.

A note on liquidity A notion that often appears in the context of data filtering is liquidity. For example, data is filtered because some portion of the data has "higher liquidity" or to remove "illiquid options". Having the two aforementioned goals in mind, we briefly discuss the role and meaning of market liquidity.

Strictly speaking, liquidity describes the (sufficient) availability of securities or contracts for trading by a market participant, i.e. that there is enough volume offered by one or several counterparties to trade at a specific price without large changes in the price. While there are several possible measures for liquidity, the typical one for option contracts in the literature is the ex post trading volume. However, if one uses closing prices, a desired property is that the option prices are recent (and fairly synchronous). Both cannot necessarily be judged by aggregated daily trading volume. Consequently, we consider a different measure and focus on price update frequency rather than trading volume as measure of liquidity. We believe that the frequency is a better proxy to judge how recent prices are (e.g. at market closing).[3] We specifically use the *number* of quote or trade ticks during the day.

---

[3] This is of course only possible with intraday data. With classical end-of-day prices, the aggregated

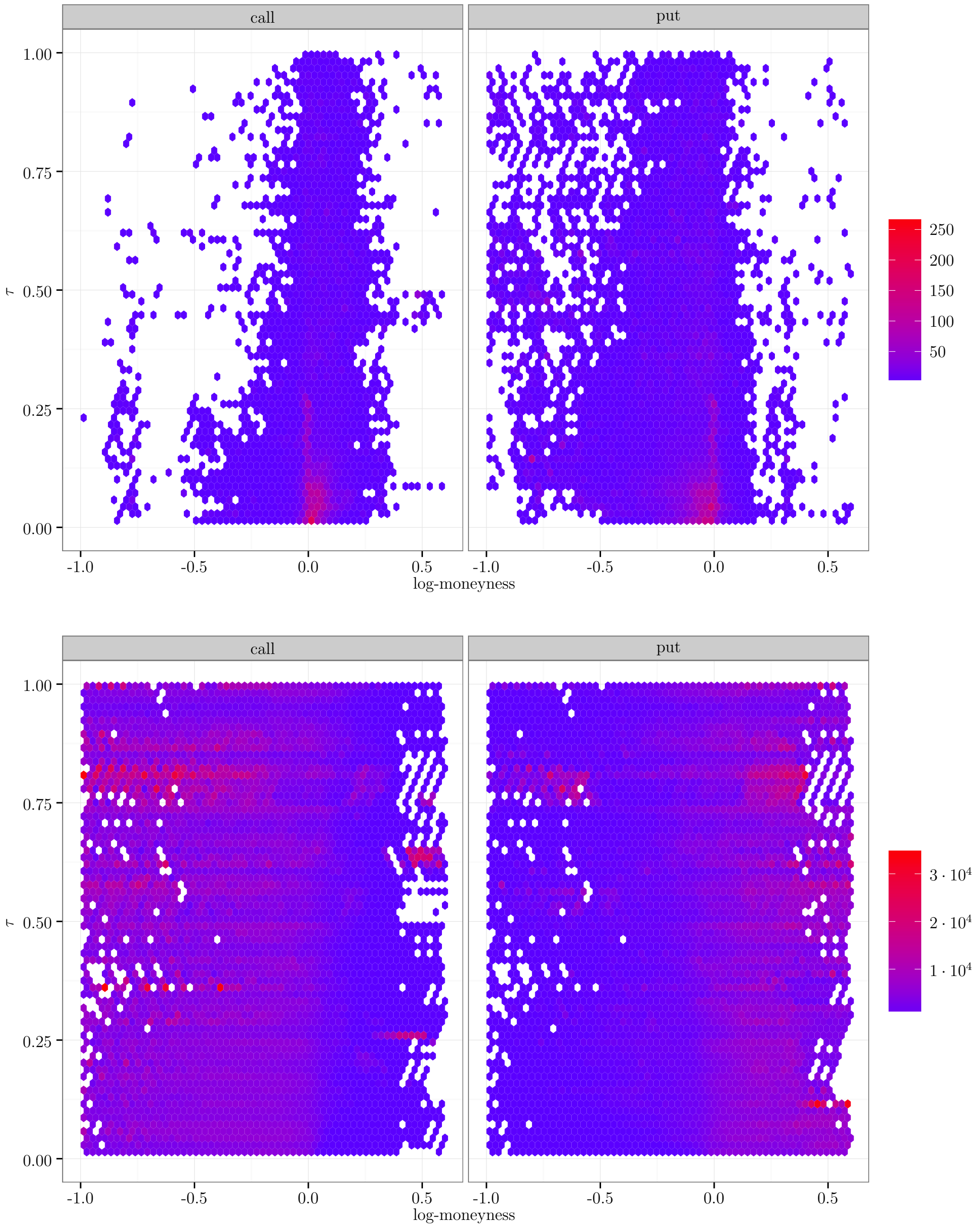


Figure 4.1: Typical distribution of trade (top panel) and quote ticks (bottom panel), depending on log-moneyness and time to maturity τ (in fractional years). The plot shows the average number of quotes and trades per day for the first 200 days of 2010. Data is truncated to the log-moneyness range $[-1, 0.6]$, to focus on the most active part of the market.

Figure 4.1 shows the typical distribution of ticks for both trades (top panel) and quotes (bottom panel), for the first 200 days of 2010. We observe that generally quotes occur

---

volume would be still the best available proxy for frequent pricing activity.

on a larger moneyness domain than trades. And obviously, there are more quotes over the day than actual transactions. If we compare where most of the ticks occur, we see that quotes are more frequent in-the-money (ITM) (independently of time to maturity) while trades are most frequent ATM and for short times to maturity. Put trades occur more out-of-the-money (OTM), while for calls the picture is relatively symmetric with a slight tendency towards ITM options for short expiries. We will come back to these patterns in the following section.

4.3.1 Common filters

Often applied option pricing filters can be categorized into two groups: moneyness/strike and maturity filters. In this section we provide a description of the most frequently used techniques and their reasoning.

Moneyness/strike filters The most common filter in this category is to remove ITM options. The argument for this filter is the higher liquidity (and therefore typically the number of transactions) for OTM options. This is also what we observe in our data for trades. But interestingly, this argument is typically used for end of day data, i.e. closing prices, which are very likely to be constituted of quotes. However, as illustrated in Figure 4.1, one could expect to get, on average, more recent prices (because of an overall higher number of ticks over the day) using ITM data.

But there is an option pricing argument, which still comes in favor of OTM data, even with the goal of having recent prices. Option prices consist of two components: the so called "intrinsic value" ($\max(S-K,0)$ for calls and $\max(K-S,0)$ for puts) which is deterministic, and the "time value", representing the risk premium that a rational investor is willing to pay for a potential upside (option payoff higher than the intrinsic value). Out of these two components, only the latter carries information about investor's expectations. For ITM options, this portion becomes smaller the further it is ITM (as opposed to the intrinsic value). On the other hand, OTM are purely valued by their time value. Therefore, OTM options are considered to have a better "information ratio".

Having excluded ITM data, other filters typically restrict data in the OTM tails, i.e. they specify some moneyness/strike interval, and data outside this interval is discarded. It is common to use one or multiple of the following criteria to specify this interval (by means of options to exclude):

- zero bid price (assumes quote prices, e.g. (Bakshi et al., 2003))
- (mid) prices $< 3/8$ (Jiang and Tian, 2005; Polkovnichenko and Zhao, 2013), there also exist other "threshold values"
- zero trade volume (while it is typically not specified whether this is daily or overall volume, e.g (Markose and Alentorn, 2011), (Christoffersen et al., 2013))
- moneyness outside some fixed interval, e.g. $[0.8, 1.1]$ (Alexander and Nogueira, 2007; Rosenberg and Engle, 2002)

With a few possible exceptions (e.g. an ATM option that was not traded during the day), all of these rules exclude options that are far OTM. Two reasons can be brought forward for this practice, relating back to the two desired data characteristics for filtered data. On the one hand, liquidity (in terms of the number or volume of trades but at some point also the number of quotes) decreases for far OTM options. This means that observed prices might be stale and therefore not completely representative of the current market. On the other hand, far OTM options have very low prices. Hence, the relative impact of (microstructure) noise is larger than for more expensive ATM options. The most extreme example are options with zero bid price, where the bid-ask spread (and therefore the bounds for the unobserved true price of interest) is as large or even larger than the price itself.

Maturity filters There exists a relatively common practice which filters to use regarding time to maturity τ. Typically, options with expiry $\tau > 1$ years are excluded. One argument in favor of this filter is again less liquidity for longer time horizons, but also reduced economic interest. Further, most researchers discard options with very short times to maturity. Such lower limits range from one week up to 20 days. Besides again reduced economic interest for such short maturities, we have made the experience that there can also arise unwanted technical effects.[4]

---

[4] For example negative implied dividend yields or unrealistically large implied risk-free interest rates.

4.3.2 Activity-based filters

To consolidate the filtering techniques, we propose a new approach to filter data in the tails on the basis of economic reasoning and empirical observations. While we finally aim to condense the results to a recommendation that can easily be applied to end-of-day data, our approach is based on intraday data.

Our filtering approach primarily addresses the goal of representative data. But as we will see later, it is also more conservative than any of the filters typically used in the tails. Therefore, the sensitivity to noise is also smaller for the resulting data set. All three methods presented below employ the idea of “market activity”. More precisely, by identifying options for which sufficient pricing/trading activity took place during the day, we want to ensure that prices can actually be considered as a consensus of multiple parties. Further, with higher trading activity also the probability increases that a quote was observed fairly recently (e.g. at market closing).

We use the notation K_i with $i \in \{c, p\}$ for call and put “limit strikes” respectively, i.e. calls with $K > K_c$ and puts with $K < K_p$ are discarded. Unless differently specified, all the following methods are applied for each trading day separately and always expiry by expiry. Therefore, we drop indices for date and expiry for notational simplicity. Further, C_K and P_K denote call and put options with strike K.

Method based on trading activity The first approach is almost identical to the rule “exclude options with zero trading volume”. But we specifically consider the trading volume for each day separately and only discard data outside the limits, identified by the maximum/minimum traded strike. We do not exclude any further options with strikes within these limits, even if they have zero trading volume. This choice makes this filter a pure “tail filter”. Formally, we have

$$K_c = \max(\{K : T(C_K) > 0\})$$

for calls and

$$K_p = \min(\{K : T(P_K) > 0\})$$

for puts, where $T(\cdot)$ counts the number of trade ticks on a specific day for one option.

Method based on intraday quoting frequency Our second filtering method builds on the idea to include only options for which new quotes are observed sufficiently often. Recall that a new level I quote represents an update of either best bid or best ask price (or their corresponding volumes). We assume that frequent updates are needed to ensure accurate prices and to exclude arbitrage opportunities. As a consequence, we use pricing frequency as a proxy for the quality of prices and therefore choose a strike range in which a certain quoting activity can be observed intraday.

Here again we do not consider every option separately, but are only interested in a maximum/minimum strike for each expiry. To determine these limits, we split each trading day in regular time intervals $I_1, I_2, \ldots, I_n$ $(n \geq 1)$ and take the minimum (for OTM puts) and maximum (for OTM calls) strike quoted during each interval, i.e.

$$K_c^j = \max(\{K : Q_{I_j}(C_K) > 0\})$$

for calls and

$$K_p^j = \min(\{K : Q_{I_j}(P_K) > 0\})$$

for puts, where $Q_{I_j}(\cdot)$ counts the number of quote ticks for a specific option in the interval I_j $(1 \leq j \leq n)$. To obtain a daily activity limit we then take the median of these n maximum/minimum strikes. In other words, we choose our limit strikes such that at least during half of the intraday intervals some prices for options up to this strike were quoted. For the empirical application in Section 4.4 we specifically use $n = 81$, which corresponds to 5 minute intervals.

Method based on market maker activity To motivate our last method, we present two stylized facts observed for intraday option data. One concerns the update frequency of option prices and the other the volumes that are quoted.

First we consider ticks per second, a rather simple measure representing the number of all ticks recorded by the CBOE every second of a trading day. Using this measure, we perform a frequency analysis by means of a Fourier transform.

Figure 4.2 shows a plot of the resulting frequencies and their amplitudes for September 27, 2010 (separately for quotes/trades and calls/puts). We truncate the plot for scaling purposes at 0.05 Hz (i.e. periods longer than 20 seconds) to remove the large amplitudes for very small frequencies (indicating almost constant values). We observe a distinct peak for quotes at 0.33 Hz, which indicates a regular effect in the number of ticks every three seconds. Interestingly, the peak does not appear with trade ticks. Furthermore, when repeating this analysis for all days in our data set, such peaks are not always observable, but clustered in time periods before or between certain expiry dates (see Figure 4.10 in the appendix).

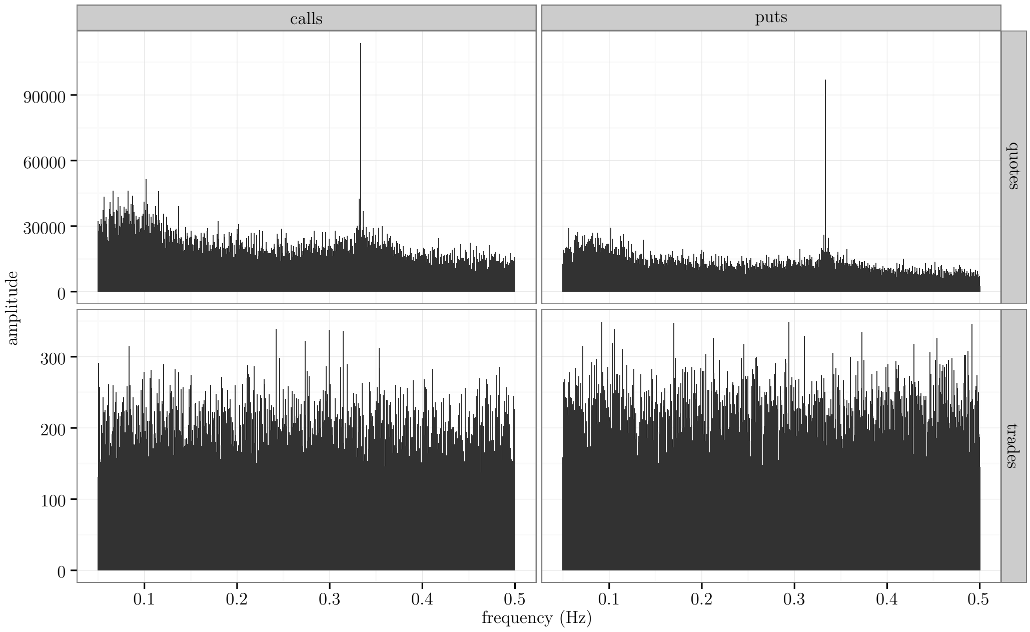


Figure 4.2: Frequency decomposition and corresponding amplitudes for different types of ticks for September 27, 2010. We do not show frequencies lower than 0.05 (20-second periods and longer) for better scaling.

This clustered observation of frequency peaks is clearly too systematic to be a random observation. Consequently, we assume that they are due to one or multiple market makers for this specific option class, based on the market maker's obligation to update market quotations intraday in combination with the "market maker rotation" (see Chicago Board of Options Exchange (2014), Rule 8.15d).[5]

The second stylized fact in the data, which we also assume to be a consequence of market maker activity, is summarized in Figure 4.3. It shows a histogram of the volumes quoted on a single trading day (September 27, 2010). We observe distinct spikes for certain volumes (typically round numbers such as 10, 50, 100). While the exact volumes can differ, such peaks are observable on any trading day. Hence, we assume that these are standard volumes offered by market makers to provide liquidity.[6]

---

[5] According to information from the CBOE, there are four "Lead Market Makers" assigned to the SPX class, traded on the "Hybrid 3.0" platform, for which rule 8.15 is applicable.

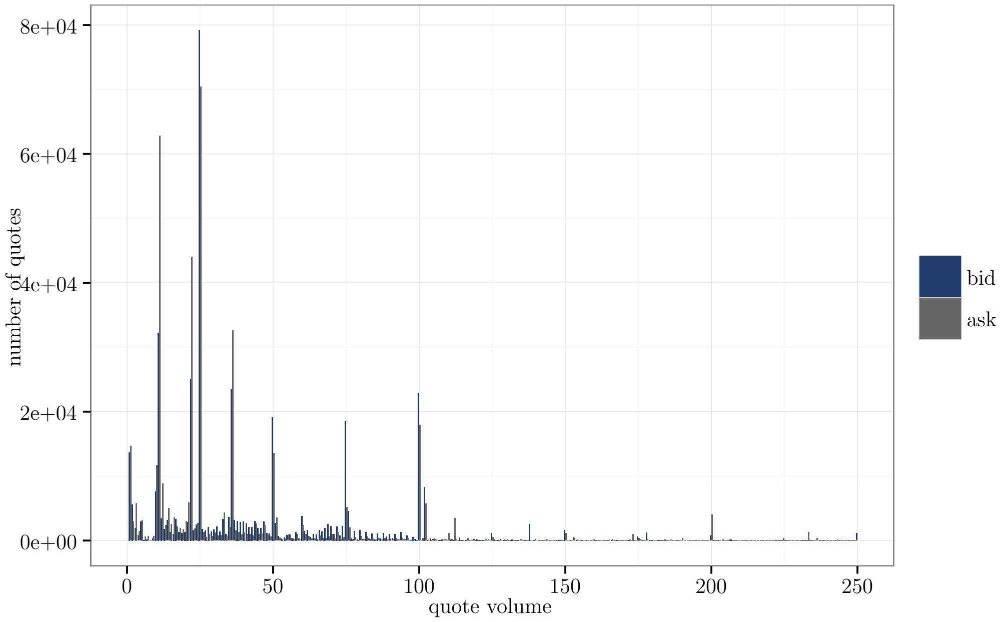


Figure 4.3: Count of quotes (over all options) with specific bid and ask volumes for September 27, 2010. Volumes larger than 250 are not shown, but there are no further peaks observable.

From both stylized facts, we want to identify the part of the market where the market makers are active. Because of their obligation to provide liquidity to other market participants, this region can in consequence be regarded as one where also sufficient other market activity takes place. By arbitrage arguments, this should further imply that the prices can be considered a consensus between the market maker and other market participants.

In order define a filter based on these stylized facts, we decide to rely only on quote volumes, because frequency peaks cannot be observed every day. However, we provide a qualitative argument that both could be used to determine a similar activity range. Since the two stylized facts are also observable for single options, we compare the number of "market maker quotes" to the intensity of the frequency peak for each option. We define market maker quotes (MMQ) as follows:

1. Each trading day, we determine volume peaks separately for all calls and puts and their respective bid and ask prices, using the following iterative procedure:

---

[6] The CBOE does to our best knowledge not publish any specific information on the volumes that market makers are obligated to quote. However, the European EUREX for example publishes this information and typically offers 50 or 200 contracts of any option that does not receive other (or better) quotes from other market participants.

in each step we identify a peak if the number of quotes with this volume exceeds 10% of total quotes, and then remove quotes with the identified volumes before starting the next iteration (i.e. in the next step the 10% are a percentage of the remaining number of quotes). We stop the procedure if there is either no more peak or if no further peak makes up at least 2% of the total daily quotes.

2. We refer to a quote as a MMQ when both bid and ask volume are among the previously identified volume peaks. We count the number of such quotes for each single option.

Figure 4.4 shows in the bottom panel the number of counted MMQ for all options on the example day September 27, 2010, and the top panel a standardized amplitude of the $0.33Hz$ frequency peak for the same options.[7] We show this for multiple expiries, indicated by the color of the dots.

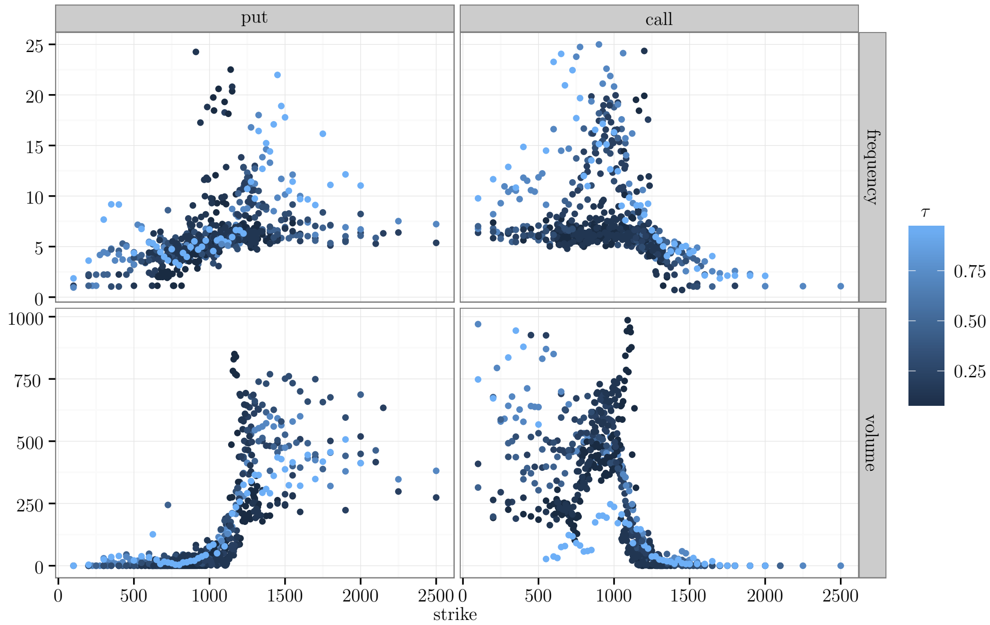


Figure 4.4: Illustration of market maker quotes and frequency peak amplitudes for September 27, 2010. Time to maturity τ in fractional years.

In both panels we observe the same picture. The two measures are high for ITM and ATM options, and decrease OTM. We verified that the decreasing pattern for MMQ remains after standardization (i.e. MMQ divided by total quotes per option). But since the range of the standardized MMQ differs from day to day and with time to maturity, we consider the absolute number of MMQ to be the better measure for the definition of a filter. Defining a strict (OTM) market maker activity limit is difficult for either measure, as there is no clear point beyond which the magnitude could be considered sufficiently small. Therefore, we define the limits by means of an appropriate quantile. In fact, the use of a quantile is more sensible for the MMQ measure, given that the number of quotes can actually be aggregated for all OTM options as opposed to the standardized frequency amplitudes. This further justifies our choice to use only volume peaks for the definition of a filter.

---

[7] We use the amplitude of the $0.33Hz$ peak minus the median amplitude of all considered frequencies, standardized by the standard deviation of the amplitudes of all considered frequencies.

Therefore, we finally specify the filter for each expiry, such that 99% of the MMQ counted OTM are within the strike limits (for each calls and puts). We qualitatively checked that the limits defined in this manner also provide reasonable cut-off points for the measure based on frequency peaks.

4.4 Comparison and implication in practice

In this section we compare the different moneyness/strike filters, as introduced in the previous section. On the one hand, we investigate whether the filtered option data differs, depending the applied filter. On the other hand, we analyze whether the differences in the final data can have an impact on the results of an empirical application.

We compare five of the filters described in Section 4.3. The first two are popular choices in the literature, namely, "non-zero bid price" (NZB) and "price $> 3/8$" (GTE). The other three are the activity-based filters proposed in this paper, i.e. "limits defined by trades" (T), "limits defined by intraday activity" (ITD) and "limits defined by market maker quotes" (MM). Recall that our method based on trade ticks also roughly corresponds to the rule to exclude options with zero trading volume. In all cases we, only use OTM options, given that it is the most widely accepted and an economically reasonable filter.

We apply all filters on each trading day in our data set, expiry by expiry, and pool the obtained limits with time to maturity and log-moneyness as shared variable. Figure 4.5 presents an example plot of these limits for the different methods, including all trading days in 2009. Generally, the limits increase in log-moneyness with increasing time to maturity, and exhibit some variation in the moneyness dimension when pooled over time. Therefore, choosing data in a fixed moneyness range (e.g. $[0.8, 1.1]$ or respectively $[-0.22, 0.1]$ in log-moneyness) seems relatively arbitrary (for any time to maturity). Additionally, we observe that the limits obtained from the five filters cluster in different log-moneyness ranges, which is a first indication that the resulting data sets are in fact different.

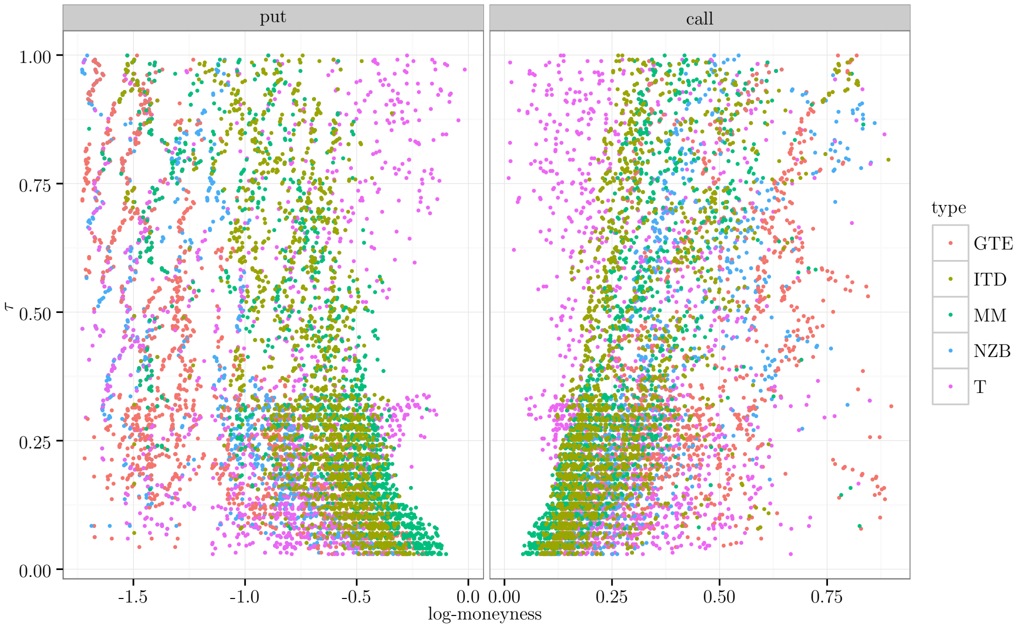


Figure 4.5: Example distribution of data limits (expressed in log-moneyness) for the different filtering methods, for all trading days of 2009. Time to maturity τ in fractional years.

By expressing the limits in log-moneyness instead of strikes, it is possible to compare them independently of the underlying price. However, log-moneyness still depends on the range of market activity, and is therefore not necessarily comparable under different market conditions. Therefore, we perform the following analysis also with limits expressed in quantiles, which are better comparable even under different market conditions. To obtain the quantiles, we compute the cumulative distribution function (CDF) implied by OTM options for each expiry, using the well-known result of Breeden and Litzenberger (1978). More precisely, we calculate implied volatilities of closing prices and fit a spline of fourth degree with one inner knot ATM to the resulting implied volatility smile (as suggested by Figlewski (2010)). Using the interpolated implied volatilities, we compute option prices at a fine strike interval (we use $\Delta K = 1$ index points) and finally obtain the CDF as the (discounted) numerical first derivative of the pricing function with respect to the strike. A plot in quantile dimension corresponding to Figure 4.5 can be found in the appendix. We observe that the clustering of different filters does also exist in quantile dimension. In the next section, we will test whether these clusters are actually statistically different.

4.4.1 Do different filters lead to different data sets?

To check whether the considered filters lead to different data sets, we compare the averages of the pooled limit strikes, obtained with the different methods. Given that especially the log-moneyness range depends on the time to maturity, we divide the strike limits into several groups, according to their time to expiry. We group the limits for short expiries (the first three months) into periods of 30 days each. After 90 days, we increase the grouping interval to 90 days each. Hence, we obtain a total of 6 groups, with expiries between 10 and 365 days.[8] Using the limits in each group (in moneyness/quantile respectively), we perform pairwise one-sided t-tests to analyze whether the different methods lead to statistically different average strike limits. As mentioned earlier, we do this in log-moneyness dimension as well as for quantiles.

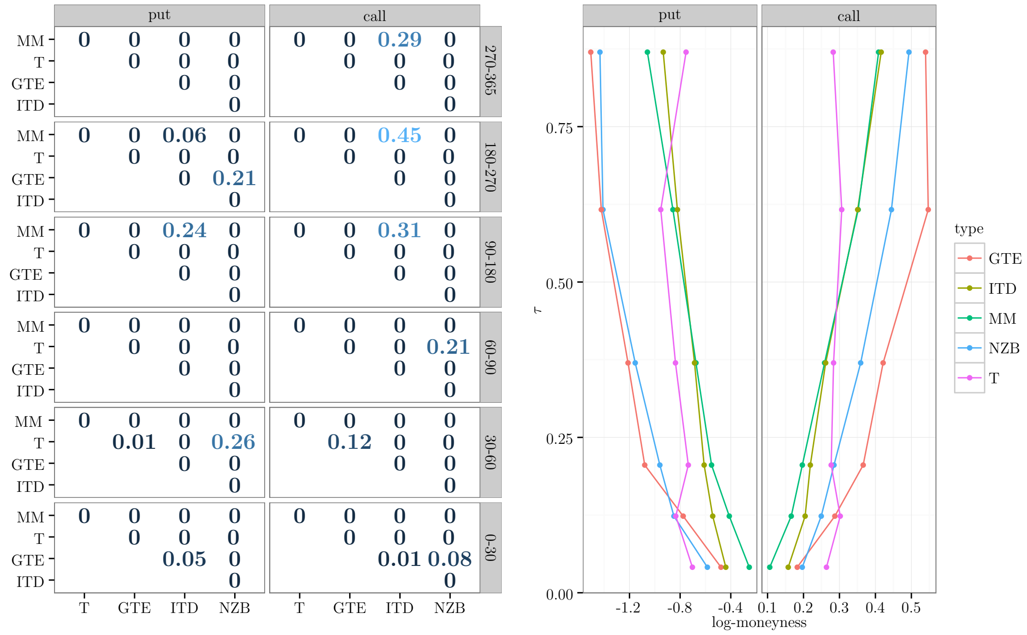


Figure 4.6: Example of pairwise t-tests for the different filters in 2009. The left panel displays the p-values of the t-tests (alternative hypothesis: means are different) for the different time intervals. The right panel displays these means with the time to maturity indicated in fractional years and the reference point set in the middle of each interval. All limits are expressed and tested in log-moneyness.

In Figure 4.6, we present the outcomes of these tests, using all log-moneyness limits from 2009. The corresponding results for quantile limits are shown in Figure 4.7.

We observe that in most cases, the limits are in fact statistically different. This holds true either for any of the four years separately, or all four years grouped together (see

---

[8] Note that the last period contains 95 days.

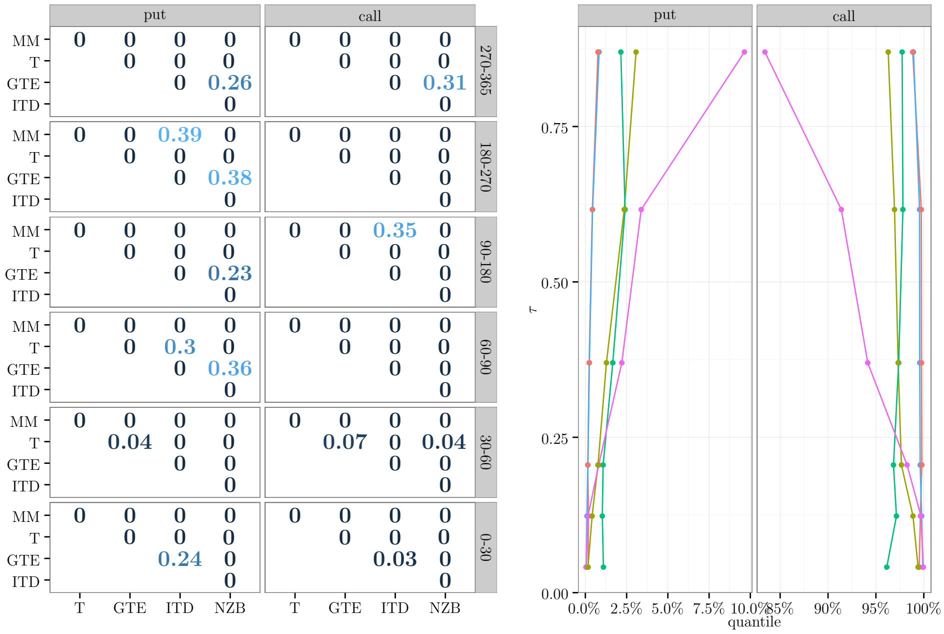


Figure 4.7: Example of pairwise t-tests for the different filters in 2009. The left panel displays the p-values of the t-tests (alternative hypothesis: means are different) for the different time intervals. The right panel displays these means with the time to maturity indicated in fractional years and the reference point set in the middle of each interval. All limits are expressed and tested in log-moneyness.

Figures 4.15 and 4.16 in the appendix). However, the specific test results are somewhat different, depending on whether we perform the tests in moneyness or quantile dimension. This is very likely due to the non-linear mapping of log-moneyness into quantiles.

Qualitatively, we can roughly distinguish two groups of filters in the right panel of Figures 4.6 and 4.7: on the one hand, the filters NZB and GTE and on the other hand, the filters MM and ITD appear to be similar with respect to the limits they define. This is also partially supported by the statistical tests, which are most often insignificant between filters of each group, respectively. We find that filters in the first group are somewhat more "generous" (i.e. they remove less data in the tails), while the filters in the second group are more restrictive. Additionally, the filter T leads to a relatively wide data range for small expiries and is rather restrictive (similar to filters in the second group) for higher maturities. Unexpectedly, for short expiries we observe wider limits for trades than for NZB. One possible explanation for this is, that market participants can *buy* options with zero bid price at the ask price, and because we derive the trade limit from the highest/lowest traded strike, even a single extreme trade of this type then constitutes the limit of the day. On the other hand, this can also be due to pooling (over 30 days) and the fact that the T measure is based on intraday data, while NZB only on closing prices.

Building on the observation that the relation between average strike limits per group and time to maturity is close to linear, we decide to represent the filters by linear functions. This is particularly helpful for the activity-based filters proposed in this paper since it condenses them to a form that can be applied to daily data. We estimate these functions by filter-wise regression of all daily limits (as shown e.g. in Figures 4.5 and 4.12). More precisely, we regress the quantile or moneyness limits on the time to maturity τ (in years), separately for calls and puts.

The results for the data aggregated over 2008-2011 can be found in Table 4.1. We observe again similarities for the two previously described groups of filters. For trades, we obtain quantiles $< 0\%$ or $> 100\%$ as constant parameters. This is due to the special data ranges given by trades, as described above, and a consequence of performing unconstrained linear regressions. The full results (including the coefficients for every year and every filter separately) can be found in Table 4.2, contained in the appendix.

Table 4.1: Limits of different data filtering methods, by means of a linear model. Data taken over all four years (2008 - 2011). Note that quantiles smaller than 0% or larger than 100% are a pure numerical artifact due to unconstrained linear regressions.

type	**log-moneyness**		**Quantile (%)**	
	const	**slope**	**const**	**slope**
		Calls		
NZB	0.18	0.29	99.79	-0.84
ITD	0.13	0.23	98.42	-3.65
GTE	0.22	0.37	99.75	-0.49
T	0.24	0.03	101.42	-18.15
MM	0.11	0.31	95.93	2.88
		Puts		
NZB	-0.59	-1.06	-0.06	1.62
ITD	-0.41	-0.61	0.13	3.24
GTE	-0.62	-1.00	0.01	1.53
T	-0.67	-0.24	-1.26	10.04
MM	-0.27	-0.95	0.99	0.83

As an intermediary conclusion from this section, we can state that the different analyzed filters deliver statistically different data cut-off points in a large number of cases. Further, we can distinguish two groups of filters. One the one hand, some of the typical filters used in the literature that only remove little data in the tails. On the other hand, our activity-based filters, which are more restrictive with respect to the usage of (deep) OTM option data.

Before continuing with a practical application, we want to point out that the observed differences within each of these two groups are rather small (even if significantly different) for typical empirical studies. We therefore only focus on the differences between these two groups in the next section. Since the activity-based filters cannot be applied directly to end-of-day data, we continue using them by means of a linear specification. And for the most general applicability, we use the results in terms of quantiles. To avoid pseudo-accuracy (given the range of single filtering limits as seen e.g. in Figure 4.5), we round the coefficients for the linear filter specification in quantiles for calls and puts as follows:

- For calls: 98.5% quantile for $\tau = 0$ and 92% quantile for $\tau = 1$
- For puts: 0.5% quantile for $\tau = 0$ and 5% quantile for $\tau = 1$

We choose this specification as a balancing solution between the three activity filters, not only based on the results over all four years but also taking into account the variability between different years. It would also be possible to choose a more or less restrictive “summary filter” depending on the application.

4.4.2 Does different data lead to different results?

In the previous, section we show that purely in terms of data, different filters result in different data sets when applied to raw end-of-day data. In this section, we address the question whether these differences in the data have an impact on the outcomes of an empirical application. We have previously pointed out that the two filter groups deliver data sets that differ only in the tails. Hence, we consider an application that is tail-dependent, namely, the estimation of moments of the risk-neutral density (RND) implied by option data. Specifically, we use the “model-free” method introduced by Bakshi et al. (2003) to avoid any issues due to model misspecification. This method allows the estimation of moments up to fourth order, in particular skewness and excess kurtosis. Especially for these higher moments, we expect differences due to more or less data in the tails.

Note that because of data truncation in the tails, several researchers use an extrapolation technique to correct for this effect (see Chapter 3 for a review). The most popular correction is to extrapolate option prices by assuming constant implied volatility beyond the observable option prices. Consequently, we compare the option-implied moments (after having applied the different filters) with and without this correction.

As representative of the more generous group of filters, we use the rule “exclude options with zero bid price” (NZB) since it is easily applicable in practice. On the other hand, we apply the activity-based filter by means of the linear model specified at the end of the previous section.

Figure 4.8 shows the moments estimated from both filtered data sets which we obtained from closing prices on June 27th, 2010, as an example day.

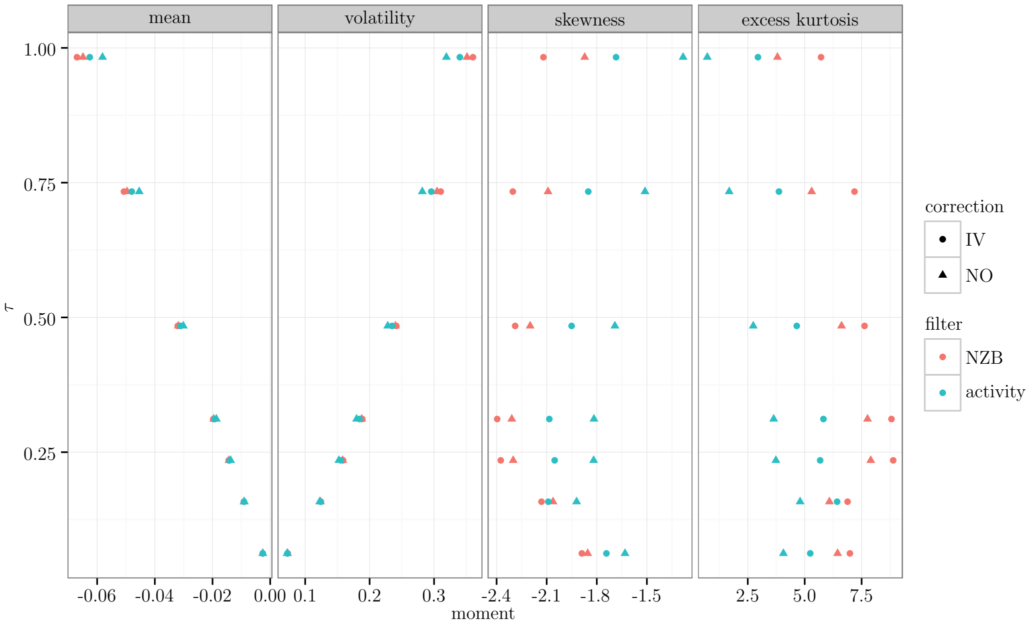


Figure 4.8: Estimated moments for differently filtered data sets on June 27, 2010, using the Bakshi et al. (2003) estimators. Estimates are shown with and without constant IV extrapolation in the tails. Time to maturity τ is shown in fractional years.

We observe substantial differences for skewness and kurtosis, not only between the corrected and un-corrected moments, but more importantly also between the two filter types. The latter can even be larger than what is compensated by the extrapolation technique (i.e. the extrapolation often just shifts results and maintains the difference).[9] Further, the differences increase with higher moments, confirming that more data in the tail is needed for an accurate estimate. One could in fact consider the differences in mean and volatility negligible, but not for skewness and excess kurtosis. Additionally, the differences seem to vary with respect to time to expiry.

---

[9] The effects due to extrapolation can actually differ, depending on the used technique, see also Chapter 3 and Ivanovas (2014).

It remains to be examined whether these differences are generally observable or only depend on the data situation on this specific day. We therefore compute the differences for all trading days in our data set and only report the results for skewness and kurtosis given the small observed differences for the first two moments. All pairwise differences and their conditional medians (depending on time to maturity, obtained via quantile regression) are shown in Figure 4.9 for the corrected and un-corrected case.

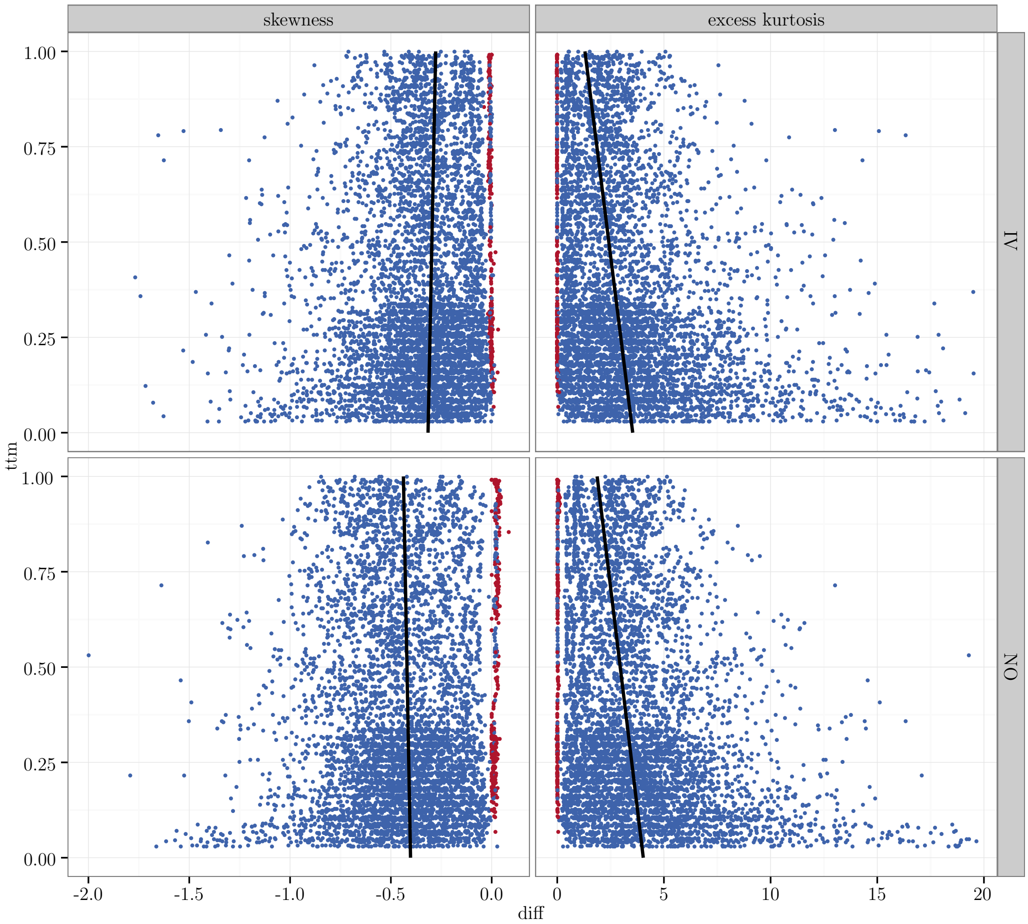


Figure 4.9: Differences of estimated skewness and kurtosis, for all trading days from 2008 - 2011, computed as moment with NZB filter minus the moment with the activity-based filter. Lines show the conditional median (depending on time to maturity). Red dots indicate cases in which both filters lead to the same limit. The top panel displays differences obtained with implied volatility extrapolation, the bottom panel without. Skewness and kurtosis differences are truncated at -2 and 20 respectively for scaling purposes.

Figure 4.9 confirms that these differences are systematic and in the median too large to be ignored. We do not further analyze these differences, since our goal is only to demonstrate that they exist systematically. To verify that these results are not only due to the selected linear specification of the activity-based filter, we repeated the exercise with more generous specifications for this filter (within the range covered by the three original filters). The resulting differences are smaller (as one should expect) but still not negligible.

Next to the results for this method per-se, we want to point out that these observed differences in higher order moment estimates can have further implications, given various applications that build on such estimates. For instance, there are several methods to derive the option-implied RND based on its moments (see Rompolis and Tzavalis (2008), Eriksson et al. (2009) and Polkovnichenko and Zhao (2013), among others).

It is true that the activity-based filter results in less data in the tails, and therefore leads to a larger bias in the estimated moments. However, as shown in Chapter 3 and (Ivanovas, 2014), this bias can be corrected for this specific and other applications, using an appropriate extrapolation technique (especially for the quantile limits considered in this study).

With this correction tool available, researchers have the possibility to choose a filter not based on data quantity alone, but also on economic reasoning. This way one can use the actual data where economically sensible, while imposing own (e.g. parametric) assumptions on the tail. This might be more desirable than relying on data with potentially unclear information content, which we consider as an advantage of our proposed filter.

Other applications We can classify other considered applications into roughly two categories. First, parametric models (for option prices, underlying dynamics, implied volatilities, etc.), which are generally not as sensitive to small differences in the data, compared to the non-parametric moment estimator used previously. Second, other semi-and non-parametric applications. From the first category, we calibrated the model of (Bates, 1996) to the different filtered data sets, but did not obtain significantly different parameters. For the second category (e.g. option-implied asset pricing kernels), the filters considered in this paper only lead to a slightly different support. Specifically in the tails, missing support is often addressed by extrapolation, which can have a bigger impact on final results than the data filter (a topic which we address separately in Chapter 3). We can therefore not confirm that other applications are sensitive to the exact data filter used for option data in the tails.[10] This in turn implies that for such applications a more restrictive, but economically meaningful filter (such as the one suggested in this paper), is a reasonable option.

---

[10] Note that we did not consider any fixed moneyness filters, which remove significantly more data than just the tails, and might therefore have an impact even on these less sensitive applications.

4.5 Conclusion

In this paper we consolidate the approach of data filtering for option data in empirical applications. We provide a summary why filtering (especially in the tails) is meaningful and review some popular filters. Interestingly, although having the same purpose, many of these filters seem relatively arbitrary. As a consequence, we suggest a new tail filtering approach based on the idea that the resulting data should be recent and representative. More precisely, we present three methods that rely on proxies for market activity, which we derive from intraday data. For ease of application, we finally condense the resulting filter to a simple linear model that defines the data cut-off points as a function of time to maturity and provide these linear model specifications in terms of moneyness and quantiles. Our new filter is generally more restrictive than other typical filters applied in the literature, i.e. it removes more data in the tails.

In an empirical analysis, we compare the effect of several filters and show that the filtered data sets are statistically different. In addition, and more importantly, we demonstrate that different filters can have an impact on the outcomes of a tail-sensitive application, using the model-free moment estimators introduced in (Bakshi et al., 2003). For other, but not particularly tail sensitive applications (for example calibration of a parametric model for the underlying asset or estimation of the option-implied risk-neutral density), we do not find noticeable differences depending on the applied filter.

Given that many applications are not particularly tail-sensitive and for the others there exist appropriate extrapolations tools (see Chapter 3), we believe that our new activity-based and economically justified filter represents an interesting alternative to existing (and somewhat arbitrary) filtering techniques.

4.A Additional figures

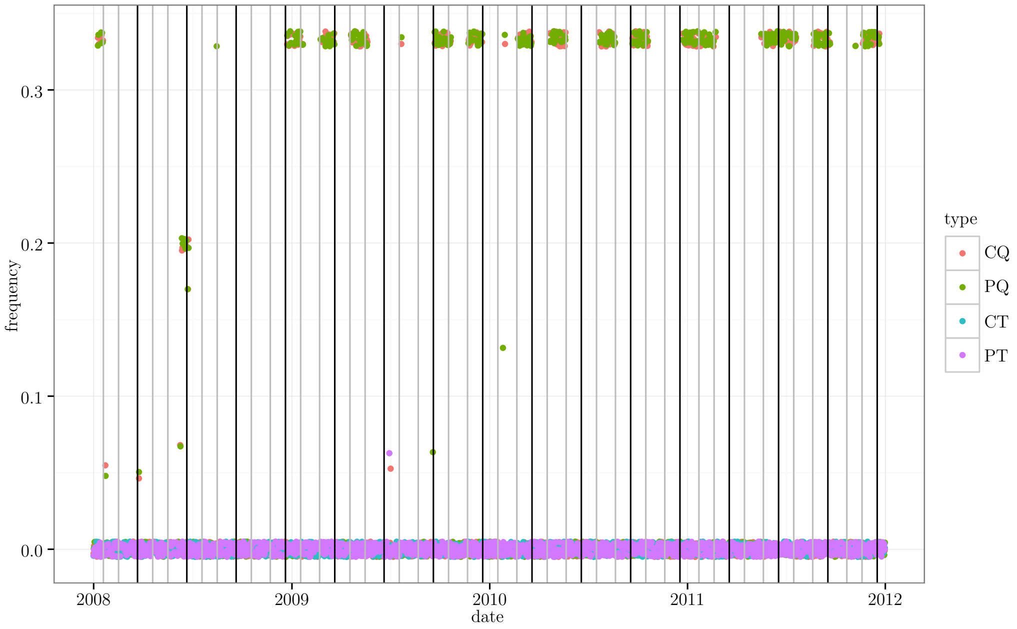


Figure 4.10: The plot shows days on which frequency peaks in quotes/second can be observed, together with the corresponding frequency. Blue dots are for frequency peaks observed in put quotes, orange for call quotes. Black vertical lines show major expiry dates, grey lines the minor expiry dates. Note that the top frequency peak is always observed ad 0.33 Hz and some vertical jitter is added such that put and call dots do not overlay.

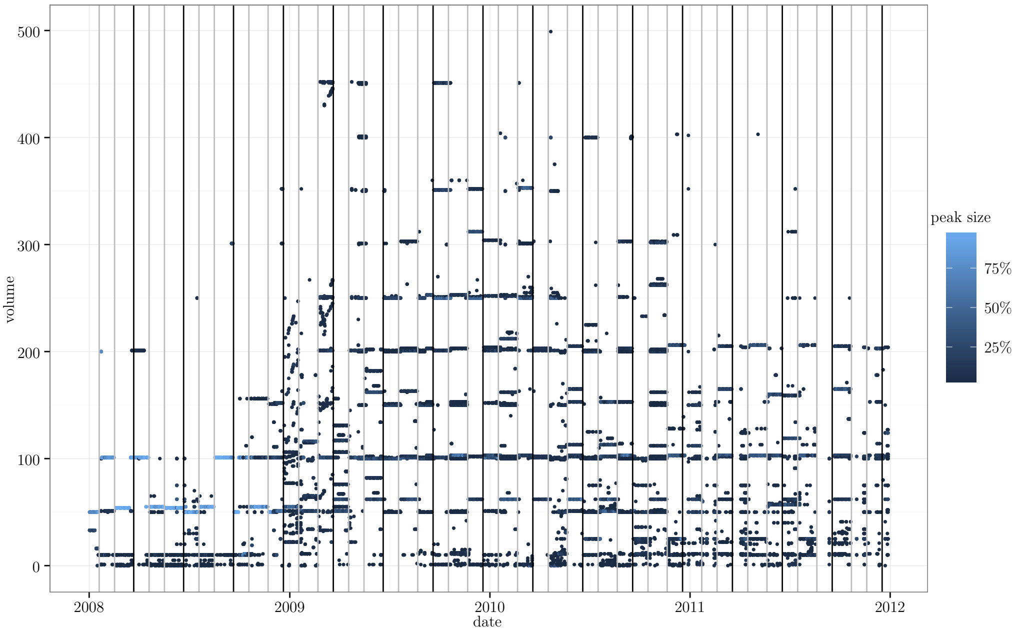


Figure 4.11: Volume peaks identified for all trading days from 2008-2011, pooled for call/put and bid/ask. Black vertical lines show major expiry dates, grey lines the minor expiry dates. Peak sizes (indicated by the dot color) given as percentage of the total quote ticks of the day. Peaks for volumes larger than 500 truncated for scaling reasons.

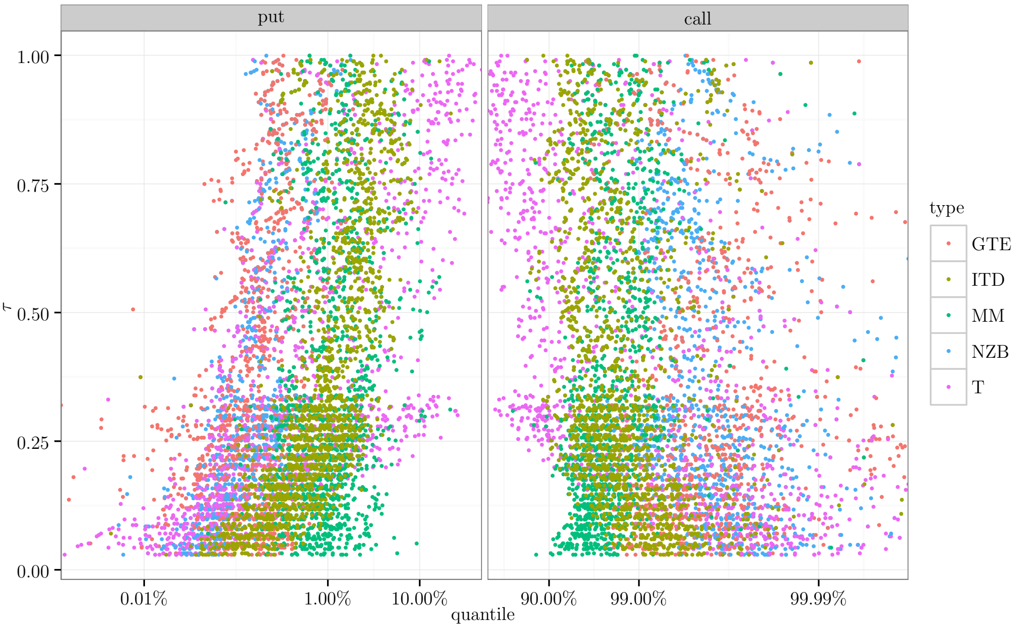


Figure 4.12: Example distribution of data limits (expressed in quantiles) for the different filtering methods, all trading days of 2009. Time to maturity τ in fractional years. Because of the logarithmic x-axis and due to the possible inaccuracies with very small or high CDF values, we truncate the few points with quantile $< 10^{-5}$ or quantile $> 1 - 10^{-5}$ for scaling reasons.

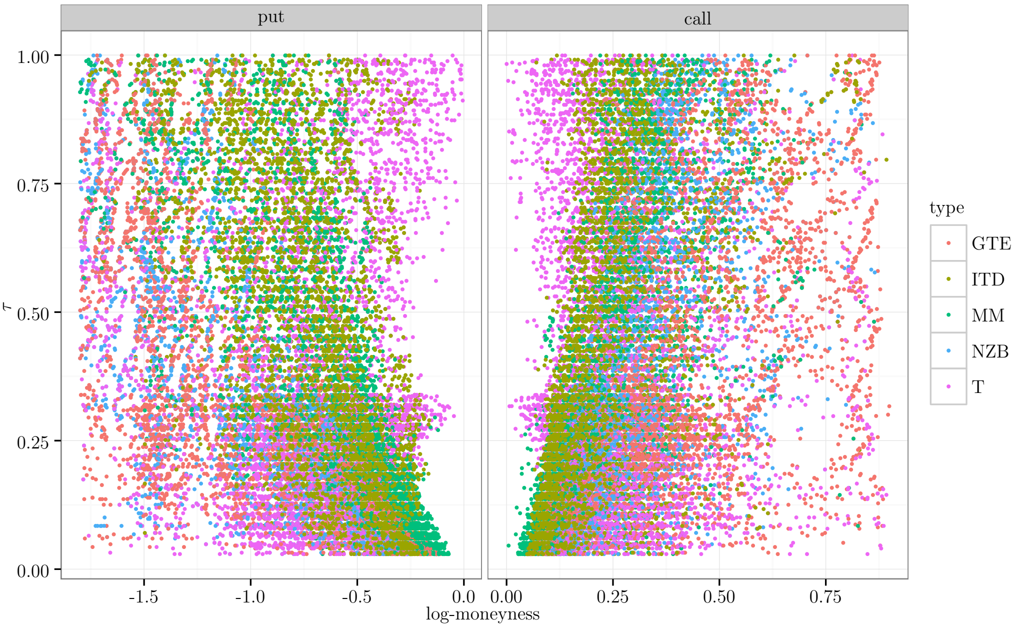


Figure 4.13: Distribution of data limits (expressed in log-moneyness) for the different filtering methods, all trading days between 2008 and 2011. Time to maturity τ in fractional years.

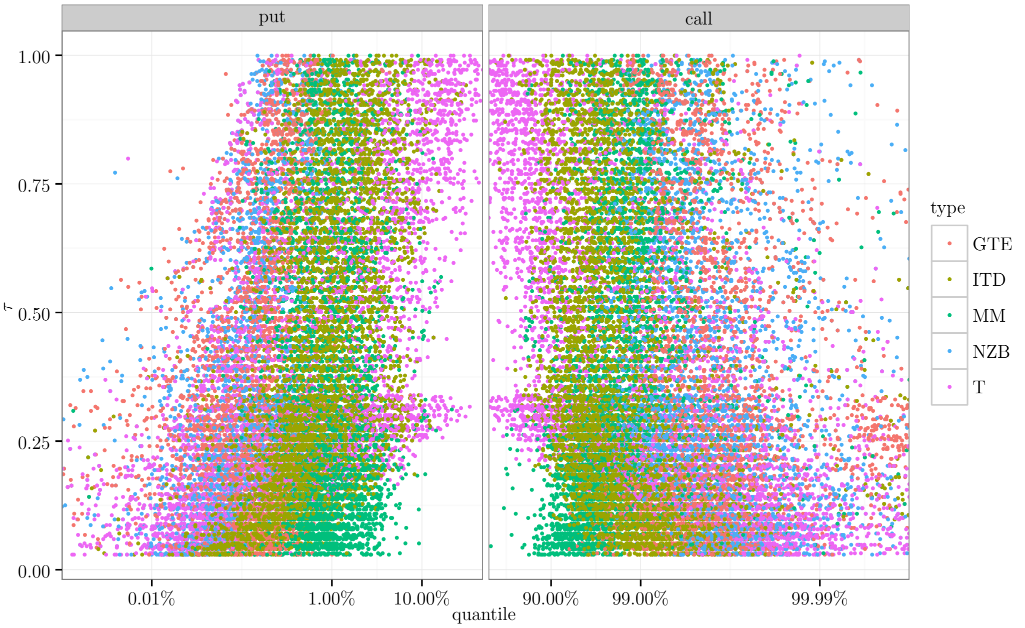
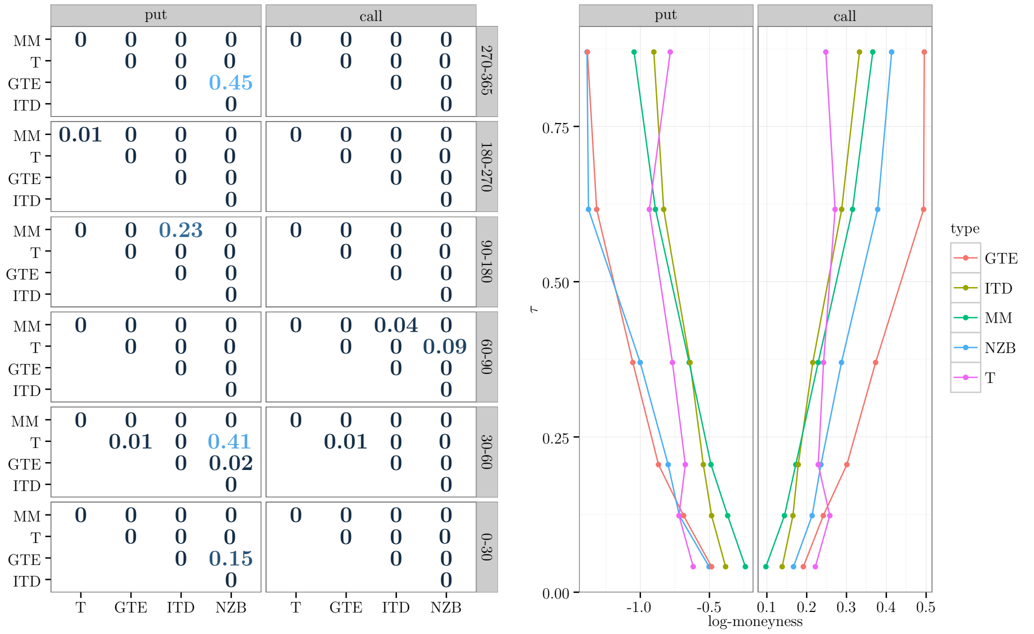


Figure 4.14: Distribution of data limits (expressed in quantiles) for the different filtering methods, all trading days between 2008 and 2011. Time to maturity τ in fractional years. Because of the logarithmic x-axis and due to the possible inaccuracies with very small or high CDF values, we truncate the few points with quantile $< 10^{-5}$ or quantile $> 1 - 10^{-5}$ for scaling reasons.



	put T	put GTE	put ITD	put NZB	call T	call GTE	call ITD	call NZB	
MM	0	0	0	0	0	0	0	0	270-365
T		0	0	0		0	0	0	
GTE			0	0.45			0	0	
ITD				0				0	
MM	0.01	0	0	0	0	0	0	0	180-270
T		0	0	0		0	0	0	
GTE			0	0			0	0	
ITD				0				0	
MM	0	0	0.23	0	0	0	0	0	90-180
T		0	0	0		0	0	0	
GTE			0	0			0	0	
ITD				0				0	
MM	0	0	0	0	0	0	0.04	0	60-90
T		0	0	0		0	0	0.09	
GTE			0	0			0	0	
ITD				0				0	
MM	0	0	0	0	0	0	0	0	30-60
T		0.01	0	0.41		0.01	0	0	
GTE			0	0.02			0	0	
ITD				0				0	
MM	0	0	0	0	0	0	0	0	0-30
T		0	0	0		0	0	0	
GTE			0	0.15			0	0	
ITD				0				0	

Figure 4.15: Example of pairwise t-tests for the different filters and all trading days from 2008-2011. The left panel displays the p-values of the t-tests (alternative hypothesis: means are different) for the different time intervals. The right panel displays these means with the maturity indicated in fractional years and the reference point set in the middle of each interval. All limits are expressed and tested in log-moneyness.

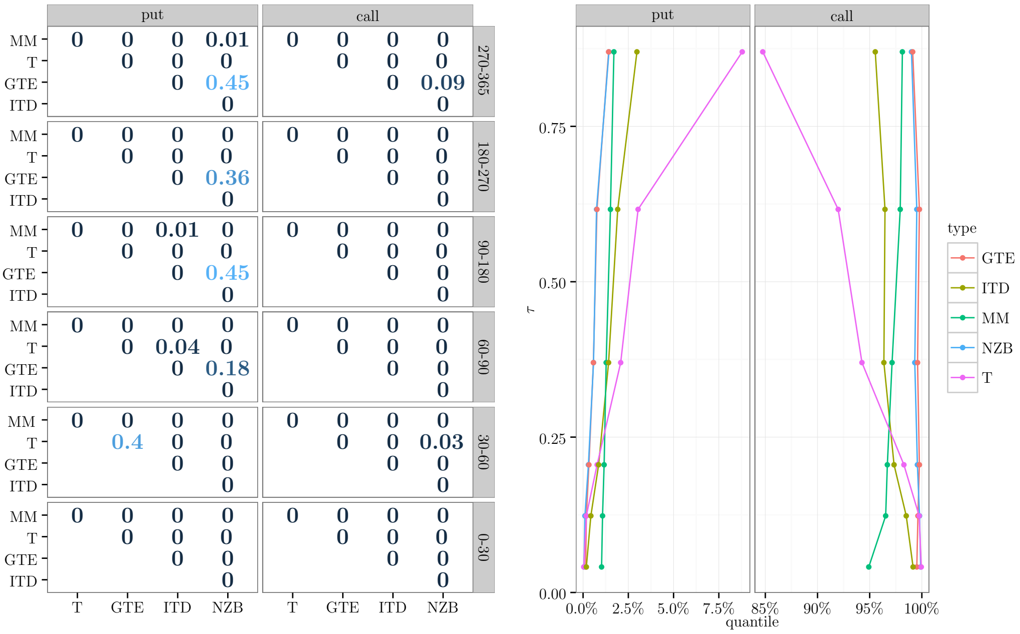


Figure 4.16: Example of pairwise t-tests for the different filters and all trading days from 2008-2011. The left panel displays the p-values of the t-tests (alternative hypothesis: means are different) for the different time intervals. The right panels actually display these means. The reference point of each interval is set in the middle. All limits expressed and tested in quantiles

4.B Additional tables

Table 4.2: Linear model parameters (filtering limit depending on time to maturity) for all filtering methods as considered in Section 4.4.1

	calls				**puts**			
year	**log-moneyness**		**Quantile (%)**		**log-moneyness**		**Quantile (%)**	
	const	**slope**	**const**	**slope**	**const**	**slope**	**const**	**slope**
	type: NZB				type: NZB			
2008	0.19	0.28	99.48	-0.44	-0.60	-0.22	-0.22	4.77
2009	0.21	0.35	99.90	-0.95	-0.74	-0.94	-0.05	0.93
2010	0.15	0.28	99.80	-0.69	-0.46	-1.57	0.04	0.39
2011	0.16	0.26	99.98	-1.28	-0.54	-1.53	0.02	0.35
all	0.18	0.29	99.79	-0.84	-0.59	-1.06	-0.06	1.62
	type: ITD				type: ITD			
2008	0.17	0.21	98.43	-2.73	-0.39	-0.33	0.25	6.26
2009	0.16	0.31	98.77	-3.03	-0.47	-0.56	0.03	3.57
2010	0.11	0.22	98.29	-3.62	-0.35	-0.84	0.25	1.36
2011	0.10	0.19	98.20	-5.25	-0.43	-0.70	0.01	1.72
all	0.13	0.23	98.42	-3.65	-0.41	-0.61	0.13	3.24
	type: GTE				type: GTE			
2008	0.30	0.18	99.50	-0.10	-0.58	-0.24	-0.12	4.64
2009	0.25	0.40	99.87	-0.78	-0.73	-1.03	0.03	0.76
2010	0.15	0.41	99.70	-0.11	-0.46	-1.43	0.08	0.39
2011	0.17	0.51	99.94	-0.99	-0.70	-1.31	0.05	0.29
all	0.22	0.37	99.75	-0.49	-0.62	-1.00	0.01	1.53
	type: T				type: T			
2008	0.24	0.04	100.51	-14.49	-0.56	0.05	-1.15	14.28
2009	0.28	0.01	101.71	-19.79	-0.78	-0.09	-1.45	11.13
2010	0.19	0.03	101.81	-20.02	-0.61	-0.51	-1.26	7.94
2011	0.22	0.04	101.68	-18.37	-0.72	-0.40	-1.18	6.89
all	0.24	0.03	101.42	-18.15	-0.67	-0.24	-1.26	10.04
	type: MM				type: MM			
2008	0.15	0.28	97.04	1.75	-0.31	-0.64	1.53	1.12
2009	0.12	0.35	96.63	1.64	-0.32	-0.88	0.96	1.70
2010	0.08	0.32	95.27	4.17	-0.22	-1.19	0.72	0.24
2011	0.07	0.28	94.40	4.34	-0.24	-1.02	0.70	0.40
all	0.11	0.31	95.93	2.88	-0.27	-0.95	0.99	0.83

Bibliography

Aït-Sahalia, Y. and A. Lo (1998): "Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices," *Journal of Finance*, 53, 499–547.

——— (2000): "Nonparametric Risk Management and Implied Risk Aversion," *Journal of Econometrics*, 94, 9–51.

Alexander, C. and L. Nogueira (2007): "Model-free price hedge ratios for homogeneous claims on tradable assets," *Quantitative Finance*, 7, 473–479.

Audrino, F. and P. Bühlmann (2009): "Splines for Financial Volatility," *Journal of the Royal Statistical Society, Series B*, 71(3), 655–670.

Audrino, F. and D. Colangelo (2010): "Semi-parametric forecasts of the implied volatility surface using regression trees," *Statistics and Computing*, 20(4), 421–434.

Bakshi, G., C. Cao, and Z. Chen (1997): "Empirical performance of alternative option pricing models," *Journal of Finance*, 52, 2003–2049.

Bakshi, G., N. Kapadia, and D. Madan (2003): "Stock Return Characteristics, Skew Laws, and the Differential Pricing of Individual Equity Options," *The Review of Financial Studies*, 16, 101–143.

Bakshi, G., D. Madan, and G. Panayotov (2010): "Returns of Claims on the Upside and the Viability of U-Shaped Pricing Kernels," *Journal of Financial Economics*, 97, 130–154.

Balkema, A. and L. De Haan (1974): "Residual life time at great age," *Annals of Probability*, 2, 792–804.

Banz, R. and M. Miller (1978): "Prices for State- Contingent Claims: Some Estimates and Applications," *Journal of Business*, 51(4), 653–672.

Barone-Adesi, G., F. Bourgoin, and K. Giannopoulos (1998): "Don't look back," *Risk*, 11, 100–103.

Barone-Adesi, G. and H. Dallo (2010): "Is the Pricing Kernel monotone?" Working Paper.

Barone-Adesi, G., R. F. Engle, and L. Mancini (2008): "A GARCH Option Pricing Model with Filtered Historical Simulation," *Review of Financial Studies*, 21, 1223–1258.

Barone-Adesi, G., L. Mancini, and H. Shefrin (2014): "Sentiment, Risk Aversion, and Time Preference," Swiss Finance Institute Research Paper No. 12-21.

Bates, S. S. (1996): "Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options," *The Review of Financial Studies*, 9, 69–107.

Beare, B. K. and L. Schmidt (2011): "An empirical test of pricing kernel monotonicity," Working paper.

Birru, J. and S. Figlewski (2012): "Anatomy of a meltdown: The risk neutral density for the S&P 500 in the fall of 2008," *Journal of Financial Markets*, 15, 151–180.

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

Black, F. and M. Scholes (1973): "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81, 637–654.

Bliss, R. R. and N. Panigirtzoglou (2002): "Testing the Stability of Implied Probability Density Functions," *Journal of Banking and Finance*, 26, 381–422.

——— (2004): "Option-Implied Risk Aversion Estimates," *The Journal of Finance*, 59, 407–446.

Bollerslev, T. and J. M. Wooldridge (1992): "Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances," *Econometric Reviews*, 11, 143–172.

Bondarenko, O. (2003): "Estimation of risk-neutral densities using positive convolution approximation," *Journal of Econometrics*, 116, 85–112.

Borak, S., K. Detlefsen, and W. Härdle (2005): "FFT Based Option Pricing," .

Breeden, D. T. and R. H. Litzenberger (1978): "Prices of State-contingent Claims Implicit in Option Prices," *Journal of Business*, 51(4), 621–651.

Breiman, L., J. H. Friedman, C. J. Stone, and R. A. Olshen (1984): *Classification and Regression Trees*, Chapman & Hall/CRC.

Brown, D. P. and J. C. Jackwerth (2012): "The Pricing Kernel Puzzle: Reconciling Index Option Data and Economic Theory," in *Contemporary Studies in Economics and Financial Analysis: Derivative Securities Pricing and Modelling*, ed. by J. A. Batten and N. Wagner, Bingley, UK: Emerald Group.

Brunner, B. and R. Hafner (2003): "Arbitrage-free estimation of the risk-neutral density from the implied volatility smile," *Journal of Computational Finance*, 7(1), 75–106.

Bu, R. and K. Hadri (2007): "Estimating option implied risk-neutral densities using spline and hypergeometric functions," *Econometrics Journal*, 10(2), 216–244.

Bühlmann, P. and S. van de Geer (2011): *Statistics for High-Dimensional Data: Methods, Theory and Applications*, Springer.

Carr, P. and D. Madan (1999): "Option valuation using the fast Fourier transform," *Journal of Computational Finance*, 2, 61–73.

Chabi-Yo, F. (2012): "Pricing kernels with stochastic skewness and volatility risk," *Management Science*, 58, 624–640.

Chabi-Yo, F., R. Garcia, and E. Renault (2008): "State Dependence Can Explain the Risk Aversion Puzzle," *Review of Financial Studies*, 21(2), 973–1011.

Chicago Board of Options Exchange (2014): "CBOE Rules," .

Christoffersen, P., S. Heston, and K. Jacobs (2013): "Capturing Option Anomalies with a Variance-Dependent Pricing Kernel," *Review of Financial Studies.*, 26, 1962–2006.

Cochrane, J. H. (2005): *Asset Pricing (Revised Edition)*, Princeton University Press.

Conrad, J., R. F. Dittmar, and E. Ghysels (2013): "Ex Ante Skewness and Expected Stock Returns," *Journal of Finance*, 68(1), 85–124.

Cont, R. and P. Tankov (2004): "Non-parametric calibration of jump-diffusion option pricing models," *Journal of Computational Finance*, 7, 1–49.

Cox, J. and S. Ross (1976): "The Valuation of Options for Alternative Stochastic Processes," *Journal of Financial Economics*, 3, 145–166.

David, A. and P. Veronesi (2013): "What Ties Return Volatilities to Price Valuations and Fundamentals?" *Journal of Political Economy*, 121, 682–746.

de Boor, C. (2001): *A Practical Guide to Splines (Revised Edition)*, Springer Series in Applied Mathematical Sciences 27.

De Giorgi, E. and T. Post (2008): "Second-order stochastic dominance, reward-risk portfolio selection, and the CAPM," *Journal of Financial and Quantitative Analysis*, 43(2), 525–546.

Delbean, F. and W. Schachermeyer (1994): "A general version of the fundamental theorem of asset pricing," *Mathematische Annalen*, 300, 463–520.

Detlefsen, K., W. K. Härdle, and R. A. Moro (2010): "Empirical Pricing Kernels and Investor Preference," *Mathematical Methods in Economics and Finance*, 3(1), 19–48.

Diebold, F. X. and R. S. Mariano (1995): "Comparing predictive accuracy," *Journal of Business and Economic Statistics*, 13, 253–263.

Eriksson, A., E. Ghysels, and F. Wang (2009): "The Normal Inverse Gaussian Distribution and the Pricing of Derivatives," *The Journal of Derivatives*, 16, 23–37.

FAN, J. AND L. MANCINI (2009): "Option Pricing with Model-Guided Nonparametric Methods," *Journal of the American Statistical Association*, 104(488), 1351–1372.

FENGLER, M. R. AND L.-Y. HIN (2014): "Semi-nonparametric estimation of the call-option price surface under strike and time-to-expiry no-arbitrage constraints," Working Paper.

FIGLEWSKI, S. (2010): "Estimating the Implied Risk Neutral Density for the U.S. Market Portfolio," in *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*, ed. by T. Bollerslev, J. R. Russell, and M. Watson, Oxford, UK: Oxford University Press.

FISCHER, R. A. AND L. H. C. TIPPETT (1928): "Limiting forms of the frequency distribution of the largest or smallest member of a sample." *Proceedings of the Cambridge Philosophical Society*, 25, 180–190.

FRIEDMAN, J. H. (2001): "Greedy function approximation: A gradient boosting machine," *Annals of Statistics*, 29(5), 1189–1232.

GARCIA, R., E. GHYSELS, AND E. RENAULT (2010): "The Econometrics of Option Pricing," in *Handbook of Financial Econometrics*, ed. by Y. Aït-Sahalia and L. P. Hansen, North-Holland, vol. 1, 479–552.

GATHERAL, J. (2004): "A parsimonious arbitrage-free implied volatility parameterization with application to the valuation of volatility derivatives," Presentation at Global Derivatives.

GATHERAL, J. AND A. JACQUIER (2014): "Arbitrage-free SVI volatility surfaces," *Quantitative Finance*, 14(1), 59–71.

GIACOMINI, E. AND W. HÄRDLE (2008): "Dynamic semiparametric factor models in pricing kernels estimation," N Functional and Operatorial Statistics (Contributions to Statistics), Dabo-Niang, S.; Ferraty, F. (eds.), Springer, Berlin.

GIL-PELAEZ, J. (1951): "Note on the inversion theorem," *Biometrika*, 38, 481–482.

GLOSTEN, L. R., R. JAGANNATHAN, AND D. RUNKLE (1993): "On the relationship between the expected value and the volatility of the nominal excess return on stocks," *Journal of Finance*, 48, 1779–1801.

GNEDENKO, B. (1943): "Sur La Distribution Limite Du Terme Maximum D'Une Série Aléatoire," *Annals of Mathematics*, 44, 423–453.

GOLUBEV, Y., W. K. HÄRDLE, AND R. TIMONFEEV (2011): "Testing the Monotonicity of Pricing Kernels," *AStA - Advances in Statistical Analysis*, to appear.

GRITH, M., W. K. HÄRDLE, AND J. PARK (2013a): "Shape invariant modelling pricing kernels and risk aversion," *Journal of Financial Econometrics*, 11, 370–399.

GRITH, M., W. K. HÄRDLE, AND M. SCHIENLE (2011): "Nonparametric Estimation of Risk-Neutral Densities," To appear in Jin-Chuan Duan, James E. Gentle, and Wolfgang Härdle(eds) Handbook of Computational Finance. Springer Verlag.

Grith, M., V. Krätschmer, and W. K. Härdle (2013b): "Reference Dependent Preferences and the EPK Puzzle," Working Paper.

Hamidieh, K. (2013): "Estimating the Tail Shape Parameter from Option Prices," Working Paper.

Han, B. (2008): "Investor Sentiment and Option Prices," *The Review of Financial Studies*, 21, 387–414.

Hansen, L. and S. Richard (1987): "The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models," *Econometrica*, 55(3), 587–613.

Härdle, W. K., Y. Okhrin, and W. Wang (2010): "Uniform confidence bands for pricing kernels," SFB 649 Discussion Papper 003, Humboldt-Universität Berlin.

Hens, T. and C. Reichlin (2012): "Three Solutions to the Pricing Kernel Puzzle," *Review of Finance*, to appear.

Heston, S. L. (1993): "A closed-form solution for options with stochastic volatility with applications to bonds and currency options," *The Review of Financial Studies*, 6, 327–343.

Ivanovas, A. (2014): "The Bakshi, Kapadia & Madan moment formulas and their intraday application," Working paper.

Jackwerth, J. C. (2000): "Recovering Risk Aversion from Option Prices and Realized Returns," *Review of Financial Studies*, 13, 433–451.

——— (2004): "Option-Implied Risk-Neutral Distributions and Risk Aversion," Research Foundation of AIMR, Charlotteville, USA.

Jiang, G. J. and Y. S. Tian (2005): "The Model-Free Implied Volatility and Its Information Content," *Review of Financial Studies*, 18, 1305–1342.

——— (2007): "Extracting Model-Free Volatility from Option Prices: An Examination of the Vix Index," *Journal of Derivatives*, 14(3), 35–60.

Kostakis, A., N. Panigirtzoglou, and G. Skiadopoulos (2011): "Market Timing with Option-Implied Distributions: A Forward-Looking Approach," *Management Science*, 57(7), 1231–1249.

Lai, W.-N. (2011): "Comparison of methods to estimate option implied risk-neutral densities," *Quantitative Finance*, Advance online publication, doi: 10.1080/14697688.2011.606823.

Lee, R. (2004): "The Moment Formula for Implied Volatility at Extreme Strikes," *Mathematical Finance*, 14(3), 469–480.

Linn, M., S. Shive, and T. Shumway (2014): "Pricing Kernel Monotonicity and Conditional Information," Working paper, University of Michigan.

Lucas, R. (1978): "Asset Prices in an Exchange Economy," *Econometrica*, 46(6), 1429–1445.

Ludwig, M. (2014): "Robust Estimation of Shape Constrained State Price Density Surfaces," Working Paper, University of Zurich.

Markose, S. and A. Alentorn (2008): "Generalized Extreme Value Distribution and Extreme Economic Value at Risk (EE-VaR)," in *Computational Methods in Financial Engineering*, ed. by B. R. E.J. Kontoghiorghes and P. W. in honour of Manfred Gilli, Springer.

——— (2011): "The Generalized Extreme Value Distribution, Implied Tail Index, and Option Pricing," *The Journal of Derivatives*, 18, 35–60.

Metaxoglou, K. and A. Smith (2014): "State Prices of Conditional Quantiles: New Evidence on Time Variation in the Pricing Kernel," Working Paper, University of California, Davis.

Nelson, D. B. (1991): "Conditional heteroskedasticity in asset returns: A new approach," *Econometrica*, 59, 347–370.

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

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

Resnick, S. I. (1987): *Extreme Values, Regular Variation and Point Processes*, Springer, New York.

Rompolis, L. S. (2010): "Retrieving risk neutral densities from European option prices based on the principle of maximum entropy," *Journal of Empirical Finance*, 17, 918–937.

Rompolis, L. S. and E. Tzavalis (2008): "Recovering risk neutral densities from option prices: A new approach," *Journal of Financial and Quantitative Analysis*, 43, 1037–1054.

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

Ross, S. (1976): "Options and Efficiency," *The Quarterly Journal of Economics*, 90, 75–89.

Schmelzle, M. (2010): "Option Pricing Formulae using Fourier Transform: Theory and Application," `http://pfadintegral.com/docs/Schmelzle2010%20Fourier%20Pricing.pdf`.

Shefrin, H. (2008): *A Behavioral Approach to Asset Pricing*, Elsevier Academic Press, Boston, second edition ed.

Shimko, M. (1993): "Bounds of probability," *Risk*, 6(4), 33–37.

von Neumann, J. and O. Morgenstern (1944): *The Theory of Games and Economic Behavior*, Princeton University Press.

Yang, J. (2009): “Semiparametric estimation of asset pricing kernel,” *Applied Financial Economics*, 19, 257–272.

Ziegler, A. (2007): “Why Does Implied Risk Aversion Smile?” *Review of Financial Studies*, 20(3), 973–1011.

Curriculum Vitae

Pirmin Meier, Swiss, born in Lucerne on February 28, 1984.

EDUCATION

2009 - 2014	**Ph.D. in Economics and Finance (PEF)**
	University of St. Gallen
2007 - 2009	**Master of Science in Mathematical Sciences**
	Swiss Federal Institute of Technology Lausanne (EPFL)
2004 - 2007	**Bachelor of Science in Mathematics**
	Swiss Federal Institute of Technology Lausanne (EPFL)
1997 - 2003	**Matura with specialization in mathematics and physics**
	Kantonsschule Reussbühl (Lucerne)

WORK EXPERIENCE

2009 - 2014	**Group of Mathematics and Statistics, University of St. Gallen**
	Research and teaching assistant
2008 - 2009	**Swiss Federal Institute of Technology Lausanne (EPFL)**
	Teaching assistant for courses in probability and statistics
Jan - Mar 2004	**Ruag Aerospace, Emmen**
July 2005	Fatigue Engineering Intern

CONFERENCE PRESENTATIONS

Computational and Financial Econometrics Conference (CFE), 2014, Pisa, Italy.
Humboldt-Copenhagen Conference on Financial Econometrics, 2013, Berlin, Germany.
Computational Statistics Conference (COMPSTAT), 2012, Limassol, Cyprus.
Computational Management Science Conference, 2012, London, UK.
Computational and Financial Econometrics Conference (CFE), 2011, London, UK.