Universität St.Gallen

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

Francesco Audrino and Pirmin Meier

April 2012 Discussion paper no. 2012-10

School of Economics and Political Science,
Department of Economics

University of St. Gallen

Editor:	Martina Flockerzi University of St. Gallen School of Economics and Political Science Department of Economics Varnbüelstrasse 19 CH-9000 St. Gallen Phone +41 71 224 23 25 Fax +41 71 224 31 35 Email seps@unisg.ch
Publisher:	School of Economics and Political Science Department of Economics University of St. Gallen Varnbüelstrasse 19 CH-9000 St. Gallen Phone +41 71 224 23 25
Electronic Publication:	Fax +41 71 224 31 35 http://www.seps.unisg.ch

Empirical pricing kernel estimation using a functional gradient descent algorithm based on splines[1]

Francesco Audrino and Pirmin Meier

Author's address: Prof. Dr. Francesco Audrino
University of St. Gallen
Bodanstrasse 6
CH-9000 St. Gallen
Phone +41 71 2242431
Fax +41 71 2242894
Email francesco.audrino@unisg.ch

Pirmin Meier
University of St. Gallen
Bodanstrasse 6
CH-9000 St. Gallen
Phone +41 71 2242433
Fax +41 71 2242894
Email pirmin.meier@unisg.ch

[1]Financial support by the Swiss National Science Foundation (Grant nr. 105214-130078) is gratefully acknowledged. We thank Enrico De Giorgi, Matthias Fengler, Kameliya Filipova, Anselm Ivanovas, Simon Knaus and the participants at the CSDA International Conference on Computational and Financial Econometrics (CFE) 2011 in London for helpful comments.

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.

JEL Classification

C13, C14, C51, C53, C58, C63.

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. (2009), Detlefsen et al. (2010), Barone-Adesi and Dallo (2010) and Fengler and Hin (2011). In addition, Yang (2009) and Grith et al. (2011a) 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. (2011)).

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. (2008), 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. (2011b) for example consider state-dependent preferences as possible explanations, whereas others such as Shefrin (2005), Ziegler (2007), De Giorgi and Post (2008), Polkovnichenko and Zhao (2010) and Hens and Reichlin (2011) 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 2 contains a review of basic asset pricing theory (mainly to set up notations) and formally introduces the pricing kernel puzzle. Section 3 provides a detailed description of our new estimation method using a FGD algorithm based on B-splines. Section 4 presents the empirical results and Section 5 concludes the paper.

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

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

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

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), 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

$$M^*_{t,t+1} = E_t[M_{t,t+1}|X_{t+1}]. \tag{2}$$

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 $\mathcal{X}_{t+1}$. Conversely, no arbitrage implies that the SDF in (2) 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 positive SDF which is a function of the primitive payoff X_{t+1}. 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

$$M^*_{t,t+1} = E_t[M^*_{t,t+1}]\frac{q}{p} = e^{-r_{t,t+1}}\frac{q}{p}, \tag{3}$$

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.

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.

$$M_{t,t+1} = \frac{u'(C_{t+1})}{u'(C_t)}, \tag{4}$$

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 (4) should be a decreasing function of aggregate consumption. 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. (2011b) for example consider state-dependent preferences as possible explanations whereas others like Shefrin (2005), Ziegler (2007), De Giorgi and Post (2008), Polkovnichenko and Zhao (2010) and Hens and Reichlin (2011) focus on results coming from the behavioral finance literature to try to understand this puzzling behavior of the pricing kernel.

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.

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.

3.1.1 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.

Using an estimate of the projected kernel $\hat{M}^*_{t,t+1}(r_{t+1})$ together with an adequately estimated return density $\hat{f}_t(r_{t+1})$, we get the fitted model price $\hat{P}_{i,t}$ as

$$\hat{P}_{i,t} = E_t[\hat{M}^*_{t,t+1}(r_{t+1})g_i(r_{t+1})] = \int \hat{M}^*_{t,t+1}(r_{t+1})g_i(r_{t+1})\hat{f}_t(r_{t+1})\, dr_{t+1}. \quad (5)$$

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 conditional density f_t. This question will be addressed in Subsection 3.1.2 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. (2011) among others for more details on applications of vega-weighted option valuation errors.

3.1.2 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 autoregres-

[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.

sive term for estimating a conditional mean, i.e.,

$$r_t = \ln(S_t/S_{t-1}) = \mu r_{t-1} + \epsilon_t, \quad \epsilon_t \sim h(0, \sigma_t^2) \tag{6}$$

and

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha\epsilon_{t-1}^2 + \gamma \max(0, -\epsilon_{t-1})^2. \tag{7}$$

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.

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_{t+1}^2, drawing a second innovation $z_{[2]}$, updating the conditional variance σ_{t+2}^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 (5).

[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).

We have

$$\hat{P}_{i,t}(\theta_t) = \frac{\Delta r_{t+1}}{n} \sum_{j=1}^{n} \hat{M}^*_{t,t+1}(\tilde{r}_{j,t+1};\theta) g_i(\tilde{r}_{j,t+1}) \hat{f}_t(\tilde{r}_{j,t+1}), \tag{8}$$

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 midpoints 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.

3.1.3 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.

$$M^*_{t,t+1}(r_{t+1};\theta_t) = \theta_{0,t}(r_{t+1})^{-\theta_{1,t}}. \tag{9}$$

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

$$M^*_{t,t+1}(r_{t+1};\theta_t) = \theta_{0,t} T_0(r_{t+1}) \exp(\theta_{1,t} T_1(r_{t+1}) + \cdots + \theta_{N,t} T_N(r_{t+1})) \tag{10}$$

with $N+1$ parameters $(\theta_{0,t}, \ldots \theta_{N,t})$ and where $T_0, \ldots, T_N$ 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. (2011a)).

3.2 Our FGD approach based on splines

Let us now focus on our direct estimation approach using 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.[4] In order to be able to apply this boosting technique, we restrict our pricing kernel to be an additive expansion of the form[5]

$$M^*(r) = M_0^*(r) + \sum_{m=1}^{M} f_m(r), \tag{11}$$

where M_0^* designates the starting model and each f_m denotes a general, arbitrary statistical procedure (function) called base learner in the machine-learning context. Possible choices of the functions f_m are restricted in the following way: f_m 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. In our study, we will focus on B-

[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.

[4]We refer to the Appendix for a short introduction of the FGD method.

[5]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.

spline basis functions as base learners since B-splines in connection with FGD have already proven to yield good results in volatility forecasting (see Audrino and Bühlmann (2009)). More precisely, we have

$$f_m(r) = \beta_{d_m} B_{d_m}(r),$$

where B_{d_m} designs a B-spline basis function and β_{d_m} is the corresponding multiplicative coefficient.[6]

B-splines are piecewise polynomial functions and can therefore be used to approximate a general continuous function.[7] 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 $mesh \in \mathbb{N}$.[8]

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.

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

[6]No-arbitrage conditions on the coefficients of the B-spline basis functions have been recently derived by Fengler and Hin (2011). Although in our algorithm we do not formally restrict the parameters to satisfy those restrictions, in our empirical analysis the estimated coefficients always lead to the absence of arbitrage and fulfill the conditions.

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

[8]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).

with

$$\hat{P} = \frac{\Delta r}{n} \sum_{j=1}^{n} M^*(\tilde{r}_j) g(\tilde{r}_j) \hat{f}(\tilde{r}_j). \tag{12}$$

This fitted model price corresponds to the approximation that has been described in equation (8) and in which M^* denotes the projected pricing kernel as defined in (11).

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 an adequate initial estimate to obtain a satisfactory performance. In our application, we take the power pricing kernel specification introduced in (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).

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

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

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

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

where d is a basis index, $\hat{\beta}_d$ denotes the least-squares estimated coefficient when regressing the residuals U_i versus the new model price components $\frac{\Delta r}{n} \sum_{j=1}^{n} \hat{\beta}_d B_d(\tilde{r}_j) g_i(\tilde{r}_j) \hat{f}(\tilde{r}_j)$ 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 step 2, 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 2 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.

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.

4 Empirical analysis

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 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 focus on out-of-the money put and call options and also 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 describes some characteristics of the one-month option contracts that we use for pricing kernel estimation.

[Table 1 about here.]

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.[9] 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 $\pm 10\%$, which is reported in the table.
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.

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 1980 until December 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 adequately estimated.

[Figure 1 about here.]

[Table 2 about here.]

Figure 1 shows the S&P 500 daily log-returns, the estimated volatility σ_t and the corresponding standardized innovations z_t. Table 2 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.

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

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 $n = 5000$ mid-points 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 parameters and determine the optimal complexity using the same sample-splitting technique as in our FGD procedure.

4.2.1 Estimated pricing kernels

Let us start with an inspection of the pricing kernel shape. Figure 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 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. (2011)).

[Figure 2 about here.]

An additional point visible in Figure 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 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 2 given the lower time frequency.

[Figure 3 about here.]

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 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.

[Figure 4 about here.]

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-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 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.

[Figure 5 about here.]

4.2.2 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 the rolling-window estimation procedure that has been described above. The corresponding results are summarized in Table 3.

[Table 3 about here.]

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 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.

[Figure 6 about here.]

We provide further insights into the pricing performance by means of Figure 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.

[Table 4 about here.]

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 IVRMSE losses 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. We concentrate again on the IV RMSE loss function but qualitatively similar results still hold for the Price RMSE as well. The corresponding outcomes are summarized in Table 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 4 confirms the higher predictive ability of our method in comparison with the chosen benchmarks.

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 (2011) 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.

Appendices

A A short introduction to FGD

Let us present here the main idea of functional gradient descent (FGD) in the framework of general regression. For this purpose, consider data $(X_1, Y_1), \ldots, (X_n, Y_n)$, 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 then 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

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

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

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

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

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

References

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.

Audrino, F. and P. Bühlmann (2001): "Tree-structured GARCH models," *Journal of the Royal Statistical Society, Series B*, 63(4), 727–744.

——— (2003): "Volatility estimation with functional gradient descent for very high-dimensional financial time series," *Journal of Computational Finance*, 6(3), 65–89.

——— (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., 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.

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.

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

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

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

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.

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

Chabi-Yo, F., R. Garcia, and E. Renault (2008): "Recovering Risk Aversion from Option Prices and Realized Returns," *Review of Financial Studies*, 21(2), 973–1011.

Christoffersen, P., S. Heston, and K. Jacobs (2011): "A GARCH Option Model with Variance-Dependent Pricing Kernel," Working Paper.

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

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

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.

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.

Fengler, M. R. and L.-Y. Hin (2011): "Semi-nonparametric estimation of the call price surface," Working Paper.

FRIEDMAN, J. H. (2001): "Greedy function approximation: A gradient boosting machine," *Annals of Statistics*, 29(5), 1189–1232.

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.

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.

GOLUBEV, Y., W. K. HÄRDLE, AND R. TIMONFEEV (2008): "Testing the Monotonicity of Pricing Kernels," SFB 649 Discussion Paper 001, Humboldt-Universität Berlin.

GRITH, M., W. K. HÄRDLE, AND J. PARK (2009): "Shape invariant modelling pricing kernels and risk aversion," Resubmitted to Journal of Financial Econometrics.

GRITH, M., W. K. HÄRDLE, AND M. SCHIENLE (2011a): "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 (2011b): "A Microeconomic Explanation of the EPK Paradox," Invited for resubmission to Journal of Financial Econometrics.

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.

HENS, T. AND C. REICHLIN (2011): "Three Solutions to the Pricing Kernel Puzzle," Working Paper.

HÄRDLE, W. K., Y. OKHRIN, AND W. WANG (2010): "Uniform confidence bands for pricing kernels," SFB 649 Discussion Papper 003, Humboldt-Universität Berlin.

JACKWERTH, J. C. (2000): "Recovering Risk Aversion from Option Prices and Realized Returns," *Review of Financial Studies*, 13, 433–451.

LIU, X., M. B. SHACKLETON, S. J. TAYLOR, AND X. XU (2009): "Empirical pricing kernels obtained from the UK index options market," *Applied Economics Letters*, 16, 1989–993.

LUCAS, R. (1978): "Asset Prices in an Exchange Economy," *Econometrica*, 46(6), 1429–1445.

NELSON, D. B. (1991): "Conditional heteroskedasticity in asset returns: A new approach," *Econometrica*, 59, 347–370.

POLKOVNICHENKO, V. AND F. ZHAO (2010): "Probability weighting functions implied by options prices," Working Paper.

ROSENBERG, J. V. AND R. F. ENGLE (2002): "Empirical pricing kernels," *Journal of Financial Economics*, 64, 341–372.

SHEFRIN, H. (2005): *A Behavioral Approach to Asset Pricing*, Academic Press Advanced Finance Series.

SINGLETON, K. J. (2006): *Empirical Dynamic Asset Pricing: Model Specification and Econometric Assessment*, Princeton University Press.

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.

Table 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

Table 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 1980 until December 2010. Panel B reports some characteristics of the standardized innovations z_t.

Panel A: Parameter estimates for AR-GJR-GARCH model					
	$\mu \times 10^2$	$\omega \times 10^6$	β	$\alpha \times 10^2$	γ
Coefficient	2.39	1.78	0.91	1.88	0.11

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	3.48	< 0.001	0.67

Table 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	121.64	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	162.30	1.96	1.61	3.48	3.23
Chebyshev kernel		4.18	2.00	1.69	3.68	3.38

Table 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 IVRMSE losses 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	t-type	sign-type
Power kernel vs. FGD	-0.571	4.345
	(0.284)	($\approx$ 0***)
Chebyshev vs. FGD	-2.397	3.795
	(0.008***)	($\approx$ 0***)

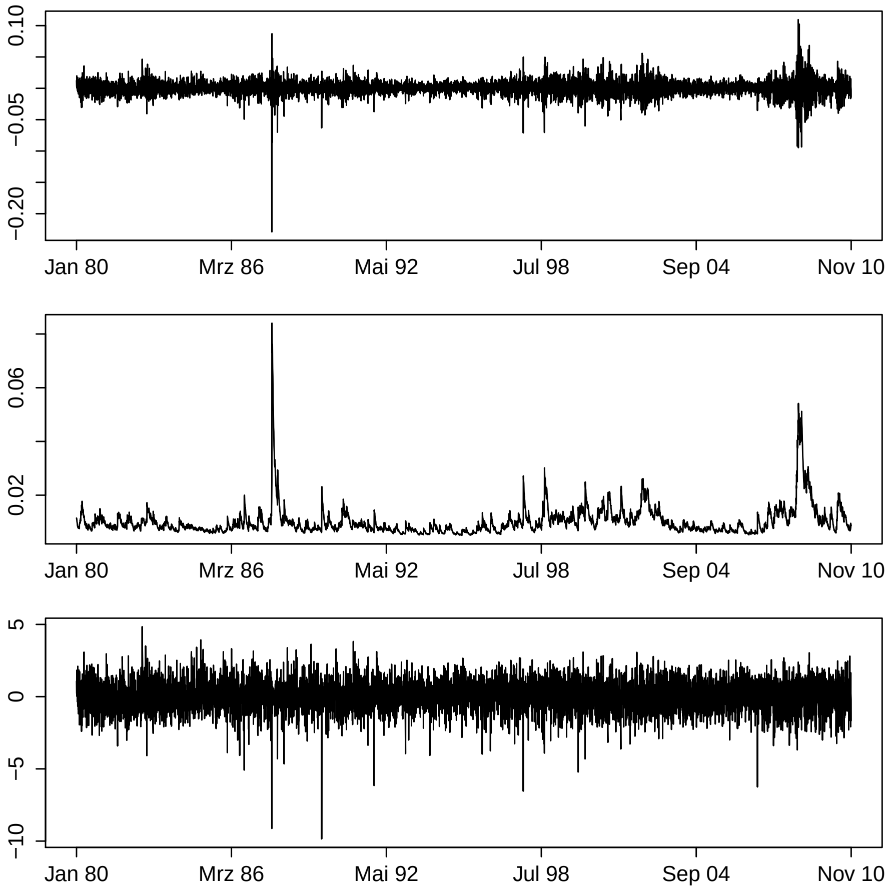

Figure 1: The figure shows daily log-returns of the S&P 500 index from January 1980 until December 2010 (top), the estimated conditional volatility σ_t using the AR-GJR-GARCH model (middle) and the corresponding standardized innovations z_t (bottom). We fit the model using pseudo maximum likelihood based on a normal innovation density.

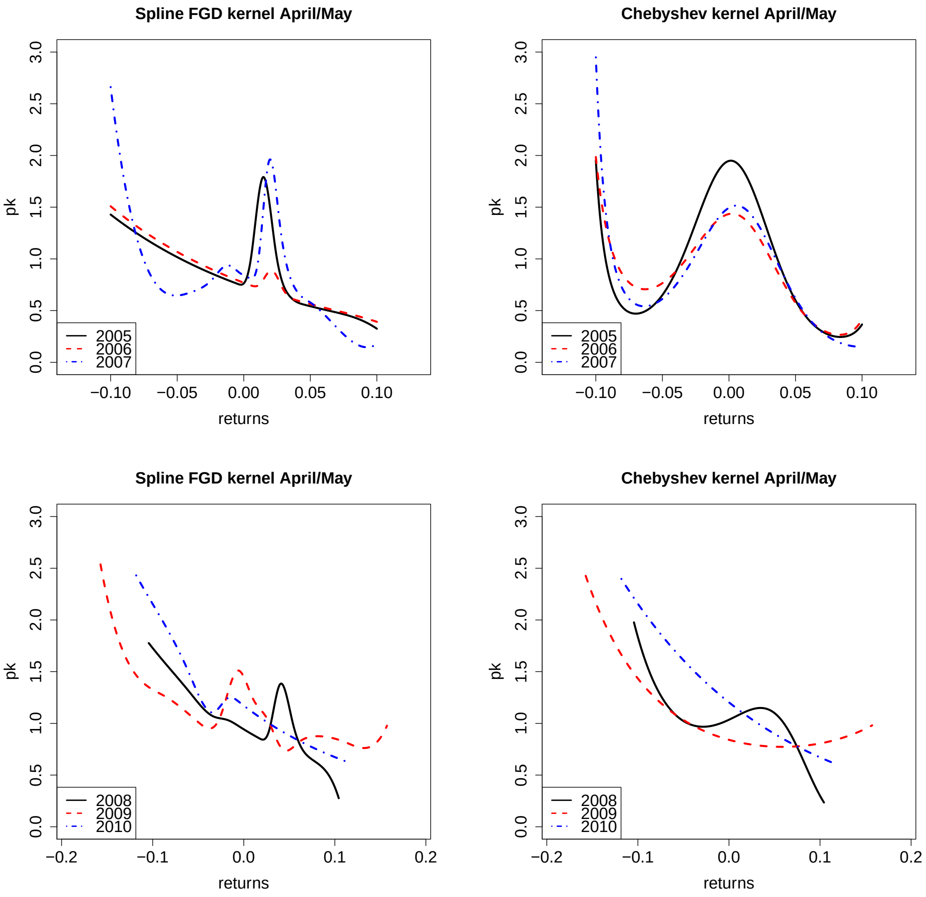

Figure 2: The figure shows a selection of estimated pricing kernels using April/May option data each year from 2005 to 2010. We apply the first option cross-section to fit the model whereas the second one is used to do model evaluation and to determine the optimal model complexity. The left part shows results for the B-spline FGD approach and the right part for the Chebyshev model.

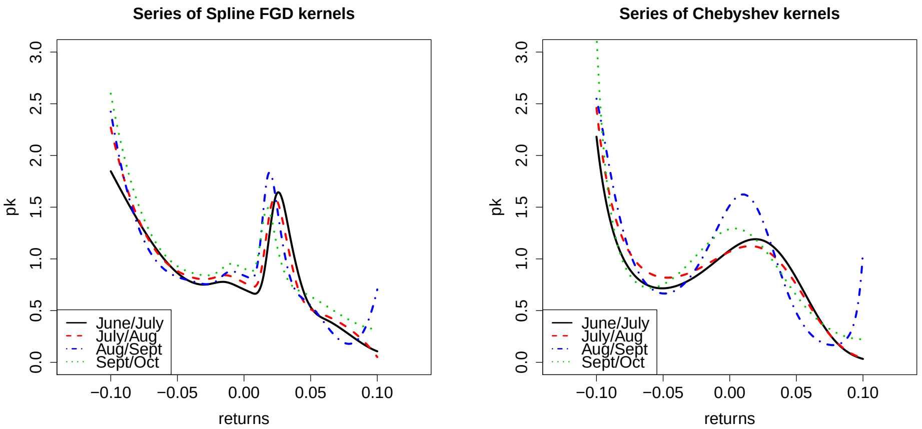

Figure 3: The figure shows a series of four consecutively estimated pricing kernels starting with June/July option data from 2006. We apply a rolling-window procedure and always consider two consecutive cross-sections for estimation. The first is used to fit the model whereas the second one serves as a validation sample to determine the optimal model complexity. Left: Spline FGD approach. Right: Chebyshev specification.

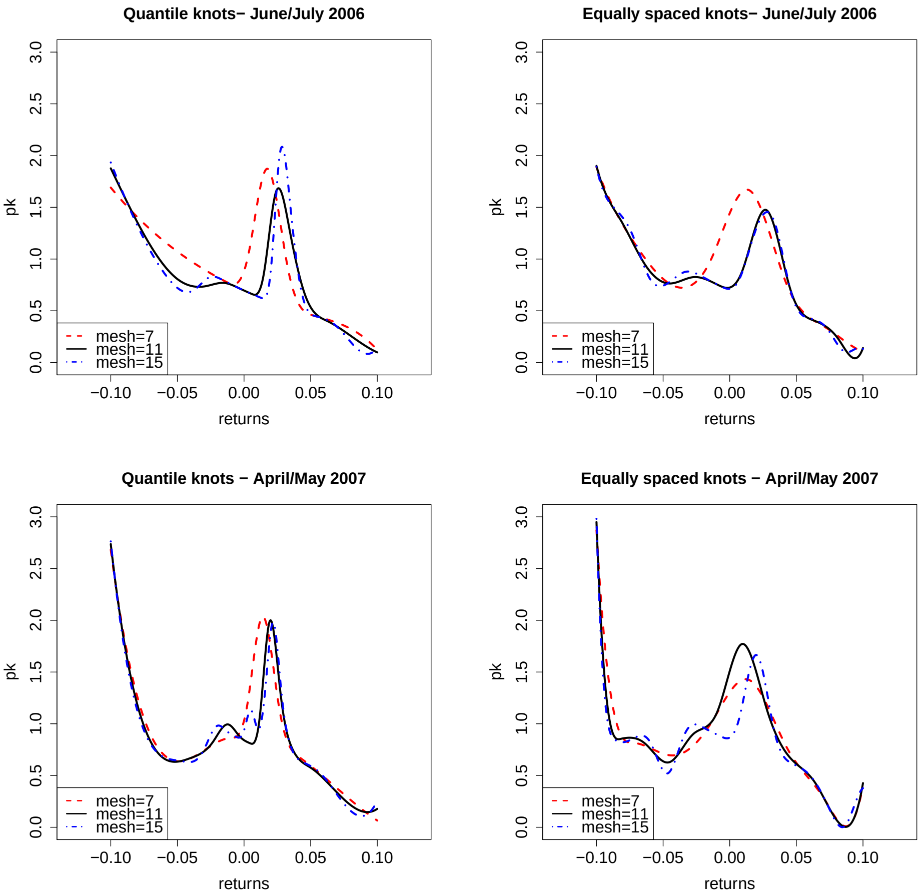

Figure 4: Results of a small robustness check concerning the number of inner knots used in our algorithm and their placement. The left plots are obtained using empirical quantiles of the simulated returns as break points whereas in the right figures we considered equally spaced inner knots. The figures show fitted kernels using June/July option data from 2006 (top) and April/May option data from 2007 (bottom).

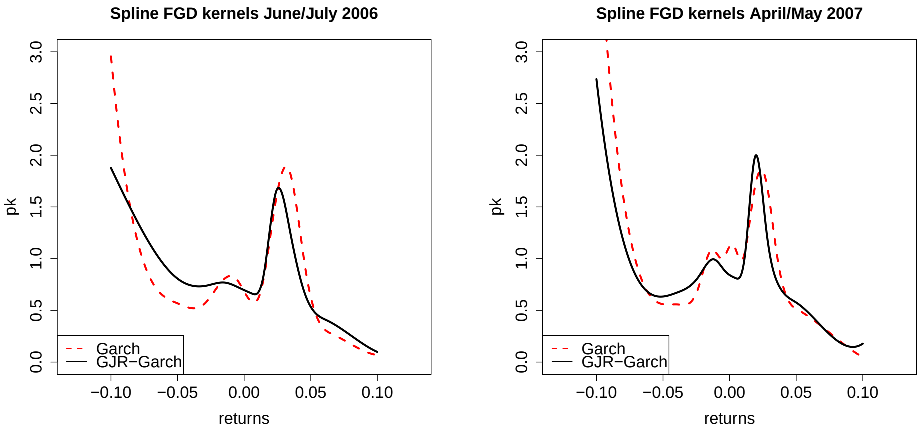

Figure 5: Robustness check to explore the sensitivity of the fitted kernels with respect to the specification of the underlying return process. The plots show estimated pricing kernels using June/July option data from 2006 (left) and April/May option data from 2007 (right).

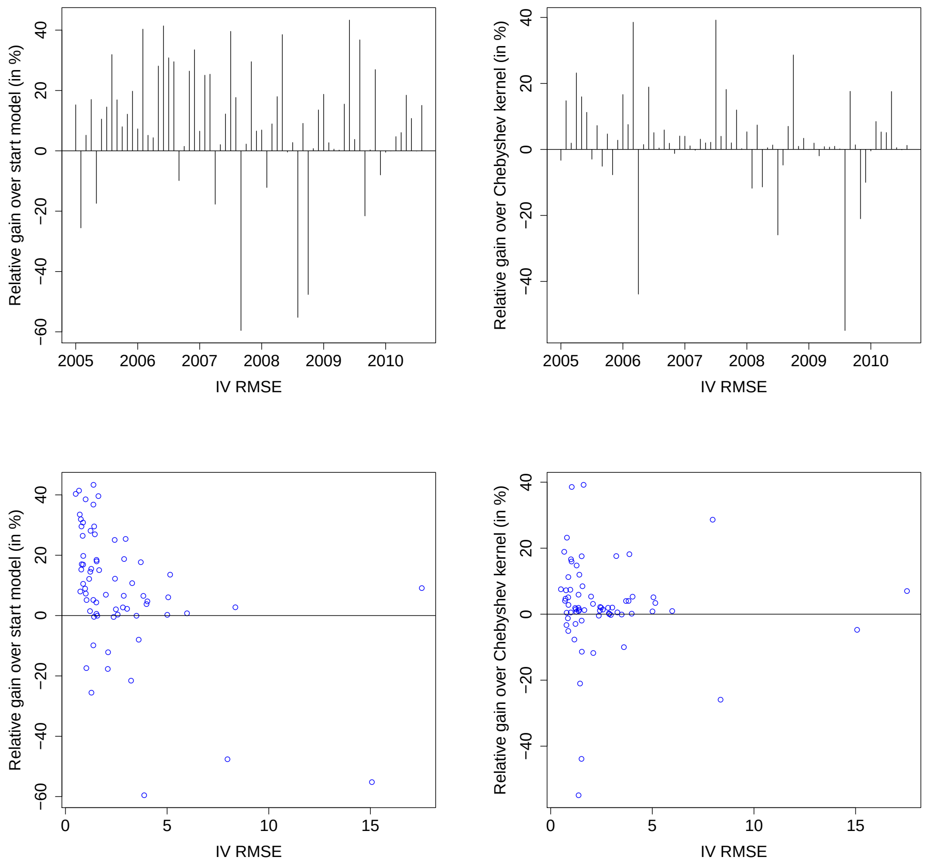

Figure 6: The plots show relative forecasting gains of our new method over the power pricing kernel specification (left) and the Chebyshev model (right) using the IV RMSE loss. The upper part of the figure contains the time-series of such gains whereas the lower part plots these gains versus the observed loss.