Shape invariant modelling pricing kernels and risk aversion

Maria Grith† Wolfgang K. Härdle‡ Juhyun Park§

*The financial support from the Deutsche Forschungsgemeinschaft via SFB 649 "Ökonomisches Risiko", Humboldt-Universität zu Berlin is gratefully acknowledged.

†CASE – Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Spandauer Straße 1, 10178 Berlin, Germany.

‡CASE – Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Spandauer Straße 1, 10178 Berlin, Germany.

§Corresponding author. Department of Mathematics and Statistics, Lancaster University, Lanaster, LA1 4YF, U.K. Phone:+44 1524 593606. Email: juhyun.park@lancaster.ac.uk.
Abstract

Pricing kernels play a major role in quantifying risk aversion and investors’ preferences. Several empirical studies reported that pricing kernels exhibit a common pattern across different markets. Mostly visual inspection and occasionally numerically summarise are used to make comparison. With increasing amount of information updated every day, the empirical pricing kernels can be viewed as an object evolving over time. We propose a systematic modelling approach to describing the evolution of the empirical pricing kernels. The approach is based on shape invariant models. It captures the common features contained in the shape of the functions and at the same time characterise the variability between the pricing kernels based on a few interpretable parameters. The method is demonstrated with the European options and returns values of DAX index.

Keywords: pricing kernels, risk aversion, risk neutral density.
JEL classification: C14, C32, G12.
AMS classification: 62G08, 62M10, 62P05.
1 Introduction

Risk analysis and management drew much attention in quantitative finance recently. Understanding the basic principles of financial economics is a challenging task in particular in a dynamic context. With the formulation of utility maximisation theory, individuals’ preferences are explained through the shape of the underlying utility functions. The basic utility functions that undermine individuals’ preferences are thought to be characterized by a concave, convex or linear utility function, describing risk averse, risk seeking or risk neutral behaviour. An economic consideration states that one unit gain and loss does not carry the same values for every individual. This view of state dependent risk behaviour extends the possible shape of utility functions. The comparison is often made through the Arrow-Pratt measure of absolute risk aversion (ARA), as a summary of investor’s risk-averseness. The quantity is originated from the expected utility theory and is defined by

\[ ARA(u) = -\frac{U''(u)}{U'(u)}, \]

where \( U \) is the individual utility as a function of wealth.

The fundamental problem is that individual agents are not directly observable but the dynamics of the market is assumed to reflect in the prices of goods traded in the market. Several efforts are made to relate price processes of assets traded in a market to risk behaviour of investors, by examining stock prices and option prices, since options are securities guarding against losses in risky stocks.

A standard option pricing model in a complete market assumes a risk neutral distribution of returns, which is the fair price, multiplied by risk-free interest, under a no arbitrage assumption. The assumption of no arbitrage assures the existence of risk neutral distributions. If markets are not complete, there are more risk neutral distributions and the fair price depends on the hedging problem. In this paper, we assume that markets are complete. The subjective or historical distribution of observed returns reflects a risk-adaptive behaviour of investors based on subjective assessment of the future market. Then the equilibrium price is the arbitrage free price and the transition from risk neutral pricing to subjective rule is achieved through the pricing kernel. Assuming those densities exist, the pricing kernel \( K \) is defined by the ratio of those densities:

\[ K(u) = \frac{q(u)}{p(u)}. \]

where \( q \) is the risk neutral density and \( p \) is the historical density. When considering several markets simultaneously, the movements of markets over time make this quantity naturally time varying, for which we introduce the time index \( t \) in the pricing kernel as \( K_t \).

A plot of estimates of pricing kernels in 2006 is shown in Figures 1 and 2. Figure 1
Figure 1: Examples of intertemporal pricing kernels with various maturities: 0.097 (blue/dash-dotted) 0.0694 (red/circle) 0.0611 (magenta/dashed) 0.0472 (cyan/solid) respectively on 12-May-2006 (blue/dash-dotted), 17-May-2006 (red/circle), 22-May-2006 (magenta/dashed), 25-May-2006 (cyan/solid). Expiration date is 02-Jun-2006 for all pricing kernels.

depicts inter-temporal pricing kernels with various maturities in May-June 2006, while Figure 2 depicts monthly pricing kernels with fixed maturity one month. To make these comparable, they are shown on a returns scale, as explained in Section 2.4. The sample of curves appear to have a bump around 1 and have convexity followed by concavity in all cases. The location as well as the magnitude of the bump vary among curves, which reflects individual variability on different dates or under different market conditions. Some features that are of particular economic interest include the maximum of the bump, the spread or duration of the bump and the location of the bump. These characteristics are also reported in the literature and shown to be consistent across different markets.

This observation motivated us to consider a common shape modelling approach for the series of pricing kernels with explicit components of location and scale known as shape invariant modelling. Detailed account of this approach is given in Section 2. With an ease of recovering pricing kernels embedded in the market data, the shape invariant modelling
approach provides a systematic treatment in examining discrepancy between empirical findings and theoretical exposition. Furthermore, this modelling approach facilitates characterisation of time varying risk perception. For example, Leland (1980) showed that the ARA can be expressed in terms of risk-neutral and subjective distributions:

\[
\text{ARA}(u) = \frac{p'(u)}{p(u)} - \frac{q'(u)}{q(u)} = -\frac{d \log K(u)}{du},
\]

and thus the corresponding ARA at time \( t \) can be derived from the pricing kernel \( K_t \) as

\[
\text{ARA}_t(u) = -\frac{d \log K_t(u)}{du}.
\]

In view of this relationship, the pricing kernel is the natural intermediary in risk management. In particular, the relationship between properties, especially shape, of utility functions and risk-adaptive behaviours are captured through the pricing kernel. For this modelling approach, we first need to recover pricing kernels empirically.
1.1 Empirical pricing kernels

With increasing availability of large market data, several approaches to recovering pricing kernels from empirical data have been proposed. As many of them estimate $p$ and $q$ separately to recover $K$, potentially relevant are studies focusing on recovering risk neutral density, which are abundant, see e.g. Jackwerth (1999), Bondarenko (2003) for comparison of different approaches. For the estimation of $p$ nonparametric kernel methods or parametric models such as GARCH or Heston models are popular choices.

Most of earlier works adopt a static viewpoint, showing a snap shot of markets on selected dates but report that there is a common pattern across different markets. The first dynamic viewpoint appears in Jackwerth (2000), who recovers a series of pricing kernels in a consecutive time and claims that these do not correspond to the basic assumption of asset pricing theory. In a similar framework Giacomini and Härdle (2008) perform a factor analysis based on the so-called dynamic semiparametric factor models, while Giacomini et al. (2008) introduce time series analysis of daily summary measures of pricing kernels to examine variability between curves.

Due to evolution of markets over time under different circumstances, these quantities are intrinsically time varying. Thus, approaches that do not take into account the changing market make limited use of information available in the current data. On the other hand, changes over time may not be completely arbitrary, as there are common rules and underlying laws that assure the dynamic evolution of the market system. Moreover, variability observed in pricing kernels, as shown in Figures 1 and 2, is not necessarily linear, and thus factors constructed from a linear combination of observations are only meaningful for explaining aggregated effects.

We take a dynamic viewpoint but focus on the common characteristics observed across different studies and markets. We then explain individual variability as a deviation from a reference. This strategy is transferred to characterisation of changes in risk behaviour. The changes in corresponding ARA measures can be examined together and especially the one corresponding to the reference pricing kernel may be viewed as a typical pattern of risk behaviour.

1.2 Common shape modelling

The new message here is an analysis of a sequence of pricing kernels through shape-invariant models. As noted earlier, the multiple curves share a common shape, but allow for individual variability which may be explained by shifts in horizontal as well as vertical directions. This is the assumption that we will consider in our analysis, a functional data analysis viewpoint (Ramsay and Silverman 2002 & 2005, Ferraty and Vieu 2006).
When multiple curves are viewed as realizations of a common function, treating curves individually wastes our resources. By combining all the information across the curves, the common functional form can be obtained efficiently. Assuming no horizontal variability, a principal component analysis (PCA) type linear decomposition procedure will further provide an optimal representation with a few common components. With additional variability in horizontal direction, however, a simple summary statistic such as mean obtained from a type of averaging operation becomes meaningless (Kneip and Gasser, 1992). Consequently, the PCA type of analysis without taking into account the obvious horizontal variability produces many more spurious directions of maximal variability and thus interpretation of components becomes arbitrary. It is possible though to remove the additional variability before applying PCA by estimating the horizontal transformation, nonparametrically or parametrically, through curve alignments. Still interpretability of results may be an issue.

Instead, we directly model the pricing kernels using a semi-parametric approach where the common function is modelled nonparametrically but individual effects are in a parametric functional form. With this method we will be able to use the maximal information of multiple curves with interpretable individual effects. Further generalization is possible by allowing the transformation to be nonlinear or unspecified, see for example Kneip and Gasser (1992), Gasser and Kneip (1995), Ramsay and Li (1998) and Gervini and Gasser (2004) among many others. In the context of functional data analysis, this is dealt with under curve alignment, synchronization or registration. Inclusion of random shift and transformation as well as other covariate effects also can be achieved (Ke and Wang, 2001).

In this article we restrict to parametric specification as the data in consideration suggests a relatively simple structure and parameters are easily related to interpretation of economic interest. The paper is organised as follows. Section 2 describes the shape invariant model in detail, which serves the basis of our analysis. Empirical results of empirical pricing kernels as well as risk aversion measures are provided in Section 3 followed by discussion and conclusion in Section 4.

2 Shape invariant modelling

Let \( \{Y_{ij}, t = 1, \ldots, T; j = 1, \ldots, n\} \) be a noisy sample of curves measured at \( \{u_j\} \) in an interval \( J \) satisfying the model

\[
Y_{ij} = K_t(u_j) + \varepsilon_{ij},
\]
where $\varepsilon_{tj}$ are independent errors with mean zero and standard deviation $\sigma^2$. To take into account the underlying shape relationship, we consider the form

$$K_t(u) = \theta_{t1}K_0\left(\frac{u - \theta_{t3}}{\theta_{t2}}\right) + \theta_{t4}.$$  

(2)

![Figure 3](image.png)

Figure 3: Example of location and scale shift pricing kernels (left) and corresponding utility functions (right) of a power utility. Solid line in each plot represents reference curves of $K_0(u) = u^{-\gamma}$ and $U_0(u) = x^{1-\gamma}/(1 - \gamma)$ with $\gamma = 0.7$ respectively. Parameters are $\theta_{t2} = 0, \theta_{t1} = 1.1, \theta_{t2} = 1 - \theta_{t1}^{(1/\gamma)}$, and $\theta_{t4} = 0$ for dot-dashed (red) and $\theta_{t4} = 0.5$ for dashed (blue) lines.

The common shape function $K_0$ is a reference curve and deviation from the reference curve is described by four parameters ($\theta_{t1}, \theta_{t2}, \theta_{t3}, \theta_{t4}$) that represent a scale change and a shift in horizontal and vertical direction. It is instructive to consider utility functions implied by this family of pricing kernels together. The utility function can be derived from

$$U(u) = \alpha \int_0^u K(x) \, dx,$$

for a constant $\alpha$. Write $U_0$ for the utility function corresponding to $K_0$. Figure 3 shows an example of transformation based on a power utility function, shown as solid line. Pricing kernels $K_t$ are shown on the left and the corresponding utility functions $U_t$ are on the right. The dashed and dot-dashed lines represent $K_t$ and $U_t$ with appropriate parameters $\theta$ in the equation (2). Depending on the parameter choices, the utility function can be made increase quickly or slowly. As an illustration, we consider the Black-Scholes model with power utility function. The Black-Scholes model assumes that
the stock price follows a geometric Brownian motion
\[ dS_t/S_t = \mu dt + \sigma dW_t, \]
which gives rise to a log normal density for the historical density \( p \). Under the risk neutral measure, the drift \( \mu \) is replaced by the riskless rate \( r \) and the density \( q \) is also log normal. The pricing kernel can be written as a power function
\[ K(u) = \lambda u^{-\gamma}, 0 < \gamma < 1, \]
with appropriate constants \( \lambda \) and \( \gamma \). The corresponding utility function is a power utility
\[ U(u) = \lambda u^{1-\gamma}. \]
Assume that \( \lambda = 1 \) and suppose that \( K_0 \) is the Black-Scholes power function \( u^{-\gamma} \). Then the class of pricing kernel implied in (2) is given by
\[ K_t(u) = \theta_1 (u - \theta_3)^{-\gamma} + \theta_4 \]
where \( \theta_{t1}^* = \theta_{t1} \theta_{t2} \). Notice that with this family of functions \( \theta_{t1} \) and \( \theta_{t2} \) are not identifiable and \( K_t \) is defined for \( u > \theta_{t3} \). For the sake of argument we set \( \theta_{t2} = 1 \) for the moment. The corresponding utility function is
\[ U_t(u) = \int_{\theta_{t3}}^u K_t(x) \, dx \]
\[ = \frac{\theta_{t1}}{1-\gamma} (u - \theta_{t3})^{(1-\gamma)} + \theta_{t4}(u - \theta_{t3}) \]
When \( \theta_{t4} = 0 \), this produces again a transformed power utility. When \( \theta_{t4} \neq 0 \), there is additional linear term in the function. See Figure 3 for comparison.

The parametrisation in (2) is commonly referred as shape invariant models (SIM), originally introduced by Lawton et al. (1972), and includes as a special case complete parametric models with known \( K_0 \). When the common function is unspecified, this type of approach needs self-modelling regression.

Model (2) is very general and needs further restriction. For example, the Black-Scholes family of power utility functions shown in Figure 3 has ambiguity between \( \theta_{t1} \) and \( \theta_{t2} \). Issues of identifiability of the common function as well as parameters are studied in Kneip and Gasser (1988). Basically unless there exist some qualitatively distinct common characteristics for each curve, the model is not identifiable. In the case of no prior structural information is available, it is enough to consider peaks and inflection
points for smooth curves as reference points. At least two of such points called landmarks would be sufficient. Even with unique \( K_0 \), some translation and scaling of parameters lead to multiple representation of the models. For uniqueness of parameters, we will impose normalizing conditions:

\[
T^{-1} \sum_{t=1}^{T} \theta_{t1} = 1, \quad T^{-1} \sum_{t=1}^{T} \theta_{t2} = 1, \quad T^{-1} \sum_{t=1}^{T} \theta_{t3} = 0, \quad T^{-1} \sum_{t=1}^{T} \theta_{t4} = 0.
\]

in the sense that there exists an average curve, see Kneip and Engel (1995). This is not restriction at all and can be replaced by any appropriate combination of parameters. Alternatively the restriction \( \theta_1 = (1, 1, 0, 0) \) takes the first curve as a reference, as in Härdle and Marron (1990) or an application-driven normalisation scheme can be introduced (Lawton et al., 1972).

### 2.1 Estimation of SIM

The model in (2) is equivalently written as

\[
K_i(\theta_{t2}u + \theta_{t3}) = \theta_{t1}K_0(u) + \theta_{t4}, \quad \theta_{t1} > 0, \quad \theta_{t2} > 0.
\]

The estimation procedure is developed using the least squares criterion based on nonparametric estimates of individual curves. If there are only two curves, parameter estimates are obtained by minimizing

\[
\int \left\{ \hat{K}_2(\theta_{t2}u + \theta_{t3}) - \theta_{t1}\hat{K}_1(u) - \theta_{t4} \right\}^2 w(u) \, du,
\]

where \( \hat{K}_i \) are nonparametric estimates of the curves. Härdle and Marron (1990) studied comparison of two curves and Kneip and Engel (1995) extended to multiple curves with an iterative algorithm. We consider an adaption of such algorithm here.

The weight function \( w \) is introduced to ensure that the functions are compared in a domain where the common features are defined. We assume that there is an interval \([a, b] \in J\) where boundary effects are eliminated and then define

\[
w(u) = \prod_t 1_{[a,b]} \left\{ (u - \theta_{t3})/\theta_{t2} \right\}.
\]

The parameter estimates are compared only in the common region defined by \( w \) but the individual curve estimates are defined on the whole interval.

The normalization leads to:

\[
T^{-1} \sum_{t=1}^{T} K_i(\theta_{t2}u + \theta_{t3}) = K_0(u).
\]

Formula (5) was exploited also in the algorithm proposed by Kneip and Engel (1995). We adopt a similar strategy here.
• Initialize
  – Estimate individual regression functions \( \hat{K}_t \) by a nonparametric smoother.
  – Set starting values \((\theta_t^{(0)}_{12}, \theta_t^{(0)}_{13})\) for each \( t = 1, \cdots, T \).
  – Construct an initial estimate \( K_0^{(0)} \) by
    \[
    K_0^{(0)}(u) = T^{-1} \sum_{t=1}^{T} \hat{K}_t(\theta_t^{(0)}_{12}u + \theta_t^{(0)}_{13}).
    \]
• For \( r \)th step, \( r = 1, \cdots, R \),
  – Determine parameters \( \theta^{(r)} \) separately for each \( t = 1, \cdots, T \) by minimizing
    \[
    \int \{ \hat{K}_t(\theta_{t12}u + \theta_{t13}) - \theta_{t11}K_0^{(r-1)}(u) - \theta_{t14} \}^2 w(u) \, du.
    \]
  – Normalise parameters: for \( j = (1, 2) \) and \( k = (3, 4) \)
    \[
    \theta_{tj}^{(r)} \leftarrow \frac{\theta_{tj}^{(r)}}{\sum_t \theta_{tj}^{(r)}}, \quad \theta_{tk}^{(r)} \leftarrow \theta_{tk}^{(r)} - T^{-1} \sum_t \theta_{tk}^{(r)}.
    \]
  – Update \( K_0^{(r-1)} \) to
    \[
    K_0^{(r)}(u) = T^{-1} \sum_{t=1}^{T} \hat{K}_t(\theta_t^{(r)}_{12}u + \theta_t^{(r)}_{13}).
    \]
• Determine final estimates:
  \[
  \tilde{\theta}_t = \theta_t^{(R)}, \quad \tilde{K}_0(u) = T^{-1} \sum_{t=1}^{T} \hat{K}_t(\tilde{\theta}_{t12}u + \tilde{\theta}_{t13}).
  \]
This procedure is shown to provide consistent estimates. In particular despite nonparac-
metric pilot curve estimates, the parameters are \( \sqrt{n} \) consistent. Furthermore, Kneip
and Engel (1995) showed in their analysis that the initial estimates of the curves are
of minor importance compared to the final estimate of \( K_0 \) and suggested the final esti-
mates \( \tilde{K}_0 \) be re-estimated with new bandwidths. This improves precision of estimates
because the pooled sample allows reduction in variance in \( \tilde{K}_0 \) and thus can accommodate
undersmoothing at the final stage to reduce bias.

A notable difference in applying the model to the pricing kernels is that the original
data available are not noisy realisations of the underlying function itself, but ratio of
functions estimated with possibly non-equal bandwidths, and thus the updating of each
estimator as a whole may not be practical. However we can take advantage of having
smooth curves evaluated at finite points as data. It is easier to improve the initialisation
step, explained in Section 2.2. This leads to simplification of the estimating procedure
with little compromise of the quality of the fit. In fact, the number of iterations required
is very small and often 3 or 4 is sufficient in practical terms. We found that when the
initial estimates are determined sufficiently accurate, the iteration is not necessary.

2.2 Starting values

If there is no scale change in horizontal direction, due to prominent peaks in each curve,
the parameter $\theta_3$ can be identified easily by the location of the individual peak. If the
models hold true, and there are two unique landmarks identifiable for each curve, simple
linear regression between the individual mark and the average mark provides an estimate
of the slope parameter $\theta_2$. Suppose that the peak is identified by $u$ satisfying $K'_t(u) = 0$.
Then we have

$$0 = K'_t(u) = \frac{\theta_{t1}}{\theta_{t2}} K_0' \left( \frac{u - \theta_{t3}}{\theta_{t2}} \right).$$

Writing $u_\ast^t$ for $K'_t$ and $u^0_\ast$ for $K'_0$ leads to a simple linear relation:

$$u_\ast^t = \theta_{t2} u_0^\ast + \theta_{t3}. \quad (6)$$

If an inflection point is used, we would have

$$0 = K''_t(u) = \frac{\theta_{t1}}{\theta_{t2}^2} K_0'' \left( \frac{u - \theta_{t3}}{\theta_{t2}} \right),$$

which gives rise to the same relation as (6), with the corresponding $u_\ast^{**}$ and $u_0^{**}$ substi-
tuted. The coefficients of intercept and slope estimates are used for starting values of
$\theta_{t3}$ and $\theta_{t2}$ respectively.

We used the peak and the inflection points around 1 as landmarks, marked in Figure 4.
The location of the landmarks is defined by the zero crossings of the first and second
derivatives. Because the initial observations $K_t$ are a smoothed curve, we find that
additional smoothing procedure is not required at this stage: a finite difference operation
is sufficient to apply mean value theorem with linear interpolation.

The slope between any two points did not vary much, which is consistent with the model
specification. This step is also used as an informal check and should there be any nonlin-
earity detected, the model needs to be extended to include a nonlinear transformation.
With our example, this was not the case.
Figure 4: Initial estimates of monthly pricing kernels with two landmarks identified. These represent pricing kernels between April 2003 and June 2006 with fixed maturity $\tau = 1$.

2.3 Nonlinear optimisation

Given the estimates of $(\theta_{t2}, \theta_{t3})$, the nonlinear least squares optimisation uses (4), which is approximated by

$$\sum_j \{ \hat{K}_t(\theta_{t2} u_j + \theta_{t3}) - \theta_{t1} \hat{K}_0(u_j) - \theta_{t4} \}^2 w(u_j).$$

(7)

When the initial values of $(\theta_{t2}, \theta_{t3})$ are sufficiently accurate, this produce is simplified to a linear regression. Conditional on $\theta_{t2}, \theta_{t3}$ and $\hat{K}_0$, the solutions to the least square regression with response variable $\hat{K}_t(\theta_{t2} u_j + \theta_{t3})$ and explanatory variable $\hat{K}_0(u_j)$ provide $(\theta_{t1}, \theta_{t4})$. When a further optimisation routine is employed to improve the estimates, these numbers serve as initial values for $(\theta_{t1}, \theta_{t4})$. 
### 2.4 Initial estimates of $K$

The initial estimates of $K$ are obtained from separate estimates for $p$ and $q$. We use intraday European options data on DAX index, provided by European Exchange EUREX and maintained by the CASE, RDC SFB 649 (http://sfb649.wiwi.hu-berlin.de) in Berlin. The data contains the actually traded call prices, the implied index price corrected for the dividends from the futures derivatives on the DAX, the strike prices, the interest rates (linearly interpolated to approximate a \textit{riskless} interest rate for the specific option’s time to maturity, the maturity, the type of the options, calculated future moneyness, calculated Black and Scholes implied volatility, the volume and the date. The extracted observations for our analysis cover the period between April 2003 and June 2006.

Figure 4 shows the resulting estimates on the scale of (compounded) log returns using the estimation methods summarised below. All curves show prominent peak but the intensity as well as the location of the peak varies among the curves, which is further explained by SIM.

**Estimation of the risk neutral density $q$:** We begin with the call price option formula that links the call prices to the risk neutral density estimation. The European call price option formula is given by

$$C(K, \tau, r_{t, \tau}, \delta_{t, \tau}, S_t) = e^{-r_{t, \tau} \tau} \int_0^\infty \max(S_T - K, 0) q(S_T | \tau, r_{t, \tau}, \delta_{t, \tau}, S_t) dS_T$$

where

- $S_t$: the underlying asset price at time $t$,
- $K$: the strike price,
- $\tau$: the time to maturity,
- $T = t + \tau$: the expiration date,
- $r_{t, \tau}$: the deterministic risk free interest rate for that maturity,
- $\delta_{t, \tau}$: the corresponding dividend yield of the asset.

Write $q(S_T)$ for $q(S_T | \tau, r_{t, \tau}, \delta_{t, \tau}, S_t)$. For fixed $t$ and $\tau$, assume $r_{t, \tau} = r$ and $\delta_{t, \tau} = \delta$, the risk neutral density is expressed as

$$q(K) = e^{rt} \frac{\partial^2 C}{\partial K^2}.$$
The relation is due to Breeden and Litzenberger (1978) and serves the basis of many current semi-parametric and nonparametric approaches. We employ the semiparametric estimates of Rookley (1997), where the parametric Black-Scholes formula is assumed except that the volatility parameter $\sigma$ is a function of the option’s moneyness and the time to maturity $\tau$. In this work, we fix the maturity and consider one dimensional regression problems, where the local polynomial smoothing with order 2 is applied to the observations of implied volatility on moneyness scale. A detailed account of the method is found in Huynh et al. (2002).

We focus on options with maturity one month (31 working days/ 23 trading days) across several years, from DAX 30 Index European options traded on Eurex Exchange. Since daily trade is organized in such a manner that we will not find options with maturity one month in every trading day, we select only that days when such securities are traded. This procedure enables us to identify one string of options every month from April 2003 to June 2006; this adds up to 38 days.

The index stock price varies within one day and we would need to identify the price at which a certain transaction has taken place. However, several authors (e.g. Jackwerth (2000)) report that the change of the index price is stale and we use instead the prices of futures contracts closest to the time of the registered pair option strike to derive the corresponding stock price corrected for dividends, following a methodology described in Fengler (2005).

For each day, we use only at-the-money and out-of-the-money call options and in-the-money puts to compute the Black-Scholes implied volatilities. This guarantees that unreliable observations (high volatility) will be removed from our estimation samples. Since, as mentioned before, the intraday stock price varies, we use its median to compute the risk neutral density. For this price, we verify if our observations satisfy the no arbitrage condition:

$$S^* \geq C_i \geq \max(S^* - K_i, 0),$$

where $S^*$ is the adjusted (for dividend) price.

Moneyness is computed for each pair $(S^*_i, X_i)$, where after we assume that the volatility does not depend on the changes in the intraday stock price. (Notice that the results in Figure 6 are defined on a returns scale (continuously compounded 1 month-period returns $V = 1 + \log(S_T/S_t)$) different from the moneyness definition used by Rookley (1997)). We have used the following transformation of probability measure:

$$\mathcal{P}(V \leq v) = \mathcal{P}(S_T \leq Se^{v-1}) = \int_0^{Se^{v-1}} q(S_T) dS_T.$$

The risk neutral density $q$ on the log returns is then

$$\frac{d}{dv} \mathcal{P}(V \leq v) = Se^{v-1}q(S_T)$$

(8)
Estimation of the historical density $p$: The return values provide information on the historical density $p$. In contrast to the option prices data, the number of observations at maturity is limited and additional modelling assumption on the evolution of stock prices is necessary. Jackwerth (2000), among many others, used GARCH models to simulate the return values at maturity, while Ait-Sahalia and Lo (2000) used a nonparametric kernel density estimates based on past return values. Alternatively the returns of the stock prices are assumed to vary slowly and thus the process can be assumed stationary for a short period of time. With this assumption, a nonparametric kernel density estimates is used in Ait-Sahalia and Lo (2000). Jackwerth (2000) argues that some discrepancies between estimates are attributed to overlapping and non-overlapping windows of the past observations selected. Nevertheless with varying degrees of assumptions on the model, common characteristics such as peaks and skewness are reportedly observed in a wide range of estimates (Härdle et al., 2009).

In this work we use the nonparametric kernel density estimates similar to Ait-Sahalia and Lo (2000) based on the past two years’ observations from the maturity. We experimented with varying sizes of the window of the past observations including non-overlapping and overlapping returns. According to our experiences the difference in shapes between them is not considerable but the prominence of peak varies more with the bandwidth choices.

2.5 Smoothing parameter selection

Initially individual estimates of $p$ are obtained based on Gaussian kernel with bandwidth choice from the Silverman’s rule of thumb (Wand and Jones, 1995) and then a common bandwidth is obtained by taking either the average or the maximum. The common bandwidth is introduced to avoid systematic bias driven by different bandwidth choices - difference in peaks can be amplified by different magnitude of the bias due to different bandwidths introduced. However, we find that, due to inherent variability, taking the average of the initial bandwidths sometimes produce multi-modality and taking the maximum tends to oversmooth. So the final bandwidth is selected in the range between the mean and the maximum by visual inspection.

The bandwidth selection for $q$ is expected to be more influential than that of $p$ in gauging performance of the estimates, as it involves derivative estimation. Figure 5 examines the effect of the bandwidth choices on $q$. Top left panel shows the implied volatility estimates overlayed, the top right shows the first derivative estimates and bottom left shows the second derivative estimates respectively, which are used as inputs to create the estimates of $q$ on bottom right panel. The bandwidths used are 0.05, 0.10, 0.15, 0.20. With the apparent undersmoothing at the smallest bandwidth, there is notable variability in terms of smoothness in estimation of implied volatility and its derivatives however the resulting density estimates demonstrate robustness. Similar observations are made to other dates. However by smoothing on implied volatility domain, we find
that the estimates are stable with relatively a wide range of bandwidths. The reported estimates used 0.20 for all estimation.

2.6 Word on asymptotics

There are two layers of estimation involved. The first step deals with individual curve estimation and the second step introduces shape invariant modelling. The shape invariant modelling is largely robust to how the data are prepared before entering the iterative algorithm and the resulting estimates are interpreted as conditional on the individual curves. Therefore, the main estimation error arises in the first stage where $p$
and $q$ are separately estimated with possibly different sample sizes and separately chosen bandwidths.

In practical terms, the sample size used in estimating $p$ is normally of smaller order, say $n$ compared to $N = nM$ for $q$ for a constant $M$, due to the difference between collected daily observations for $p$ and intraday observations for $q$. Thus it is expected that the estimation error will be dominated by the estimation error of $p$. On the other hand, the underlying function $p$ for which simple kernel estimation is used is much simpler and more stable compared to $q$ for which nonparametric second derivative estimation is required. Thus the source of error will partially be cancelled out.

Because the estimates of ratios are constructed from the ratio of the estimates, we can decompose the error as

$$
\hat{\theta}_t \approx \mathcal{N}(0, \Sigma_t).
$$

Numerical instability might occur in the region where $\hat{p} \approx 0$ however this is not of theoretical concern. In fact, the pricing kernel is the Radon-Nikodym derivative of an absolutely continuous measure, and thus $p$ and $q$ are equivalent measures, that is, the null set of $p$ is the same as the null set of $q$. So we can limit our attention to the case where $p(u) > \epsilon$ for some constant $\epsilon$. Provided that $p(u) > \epsilon$ and $q(u) > \epsilon$, the asymptotic approximation is straightforward and asymptotic bias and variance can be computed from

$$
\begin{align*}
\mathbb{E}[\hat{\theta}_t - \mathbb{E}(\theta_t)] & \approx \mathcal{O}(h^4) + \mathcal{O}(h^2) + \mathcal{O}(h^2) + \mathcal{O}(h^4), \\
\text{Var}[\hat{\theta}_t - \mathbb{E}(\theta_t)] & \approx \mathcal{K}^2(u) \left\{ \frac{\text{Var}[\hat{q}(u)]}{q^2(u)} + \frac{\text{Var}[\hat{p}(u)]}{p^2(u)} \right\} \\
& = \mathcal{O}((Nh_q)^{-1}) + \mathcal{O}((nh_p)^{-1}) + \mathcal{O}((Nh_q)^{-1} + (nh_p)^{-1}).
\end{align*}
$$

Since $\hat{q}$ involves estimation of second derivative of a regression function, the error is dominated by the estimation of $q$. Ait-Sahalia and Lo (2000) showed in a similar framework that the error is dominant by the estimation of $q$ and for the purpose of asymptotics $p$ can be regarded as a fixed quantity.

Consistency and asymptotic normality of the parameter estimates are shown in Härdle and Marron (1990) for two curves and in Kneip and Engel (1995) for multiple curves. So we may write that

$$
\hat{\theta}_t \approx \mathcal{N}(\theta_t, \Sigma_t).
$$
Due to the iterative algorithm, the asymptotic covariance matrix is more complicated for multiple curves but Kneip and Engel (1995) shows that, as the number of curves increases, the additional terms arising in the covariance matrix is of lower order than the standard error term due to non-linear least square methods. There is no suggested estimate for the asymptotic covariance matrix but a consistent estimate can be constructed as in standard non-linear least square methods. Define the residual $\hat{e}_{tj} = \hat{K}_t(u_j) - \tilde{K}_t(u_j)$ where $\hat{K}$ is the initial estimates and $\tilde{K}$ is the SIM estimates and let

$$\hat{\sigma}^2 = \frac{1}{nT} \sum_{t=1}^{T} \sum_{j=1}^{n} \hat{e}_{tj}^2.$$ 

The covariance matrix can be estimated as

$$\hat{\Sigma}_t = \hat{\sigma}^2 \left( \frac{1}{n} \sum_{j=1}^{n} [\nabla_{\theta} \hat{K}_t(u_j; \tilde{\theta})][\nabla_{\theta} \tilde{K}_t(u_j; \tilde{\theta})]^T \right)^{-1},$$

where $\nabla_{\theta} K(u; \theta)$ is the first derivative of the function, given by

$$\frac{\partial K(u)}{\partial \theta_1} = K_0(u - \theta_3),$$

$$\frac{\partial K(u)}{\partial \theta_2} = \frac{\theta_1}{\theta_2} K_0(u - \theta_3),$$

$$\frac{\partial K(u)}{\partial \theta_3} = -\frac{\theta_1}{\theta_2} K_0(u - \theta_3),$$

$$\frac{\partial K(u)}{\partial \theta_4} = 1.$$

To see whether the location or scale parameters are different between any pair of curves, we can compute the standard errors of the estimates to make a comparison. A formal hypothesis testing also appears in Härdle and Marron (1990) for kernel-based estimates and in Ke and Wang (2001) for spline-based estimates. For example we might be interested in testing whether a location or a scale parameter can be removed.

Although these results are practically relevant, we note that the methods mentioned all assume direct observations of the underlying function of interest, with one smoothing parameter selection involved. Obtaining comparable rigorous results for our estimator is complicated in the present situation due to the non-standard nature of the estimator being a ratio of two separate nonparametric estimates with independent bandwidths. We consider this out of scope of this paper and leave for separate work.

3 Pricing kernels and risk aversion

The common curve estimate of the pricing kernel, together with shifted curves are shown in Figure 6. The estimate is not defined outside the range where no information is
available on the common shape. Figure 7 shows the parameter estimates from the model (2). Note that the range in y axis of the location parameters ($\theta_3, \theta_4$) and scale parameters ($\theta_1, \theta_2$) differ. $\theta_1$ and $\theta_4$ correspond to the amplitude variability (vertical direction) and $\theta_2$ and $\theta_3$ correspond to phase variability (horizontal direction) across dates.

In Section 2, we have discussed utility functions implied by the pricing kernel family. In general the utility function corresponding to $K_t$ is given by

$$U_t(u) = \theta_1 \theta_2 U_0 \left( \frac{u - \theta_3}{\theta_2} \right) - U_0 \left( -\frac{\theta_3}{\theta_2} \right) + \theta_4 u$$

$$\equiv \theta_1^* U_0 \left( \frac{u - \theta_3}{\theta_2} \right) + \theta_4^* + \theta_4 u.$$

The utility function $U_t$ is a combination of a SIM class of the common utility function and a linear utility function. The ARA measure is given by

$$ARA_t(u) = \frac{-\theta_1^* K_0' \left( \frac{u - \theta_3}{\theta_2} \right)}{\theta_1 K_0 \left( \frac{u - \theta_3}{\theta_2} \right) + \theta_4^*}.$$

For example, assuming $K_0(u) = u^{-\gamma}$ with $\theta_1^* = 1$ gives

$$ARA_t(u) = \gamma \left\{ (u - \theta_3) + (\theta_4^*/\theta_{11})(u - \theta_3)^{\gamma+1} \right\}^{-1}.$$

When $\theta_4^* = 0$, this function is monotonically decreasing but in general this is not the case.
The estimated curves between 2003 and 2006 are shown in Figures 8 for 2003 and 2004 and 9 for 2005-2006. Pricing kernels are shown on the left and the corresponding ARA measures are on the right side. We observe gradual shift in location, spread and intensity of the bumps over years in pricing kernels. The corresponding changes in ARA measures are also seen.

4 Conclusion

We have explored a way of combining information contained in a series of empirical pricing kernels. We propose a systematic approach using shape invariant modelling that aims at capturing a common structure as a reference curve and explaining individual variability by the deviations on the horizontal and vertical directions from the reference curve. We have demonstrated the method using the European DAX options and returns data and quantified the variability with four parameters that are easy to interpret in the economic context. Inspection of these summaries would enhance our understanding of the potentially complex objects like pricing kernels. Based on the pricing kernels we derived the risk behaviour based on the ARA measures. It would be interesting to
related the changes in risk behaviour to economic variables of interest.

Several extensions are left for future research. Although the shape invariant modelling framework is independent of the estimation of pricing kernels, this can serve an alternative framework for comparison of empirical pricing kernels and related quantities. Furthermore, time series analysis based on the estimated parameters for longer series will be useful to summarise the dynamics and can be used to develop formal statistical inference methods. We have only considered linear transformation and found it sufficient for our analysis. An extension to nonlinear transformation is relatively straightforward, if necessary. The real issue then would be to incorporate some formal statistical inference methods to distinguish the cases.
Figure 9: Estimated pricing kernels (left) and corresponding ARA measures (right) in 2005 (top) and 2006 (bottom). These are reconstructed based on the model parameters shown in Figure 6 and 7.

References


• Figure 1: Examples of intertemporal pricing kernels with various maturities: 0.097 (blue/dash-dotted) 0.0694 (red/circle) 0.0611 (magenta/dashed) 0.0472 (cyan/solid) respectively on 12-May-2006 (blue/dash-dotted), 17-May-2006 (red/circle), 22-May-2006 (magenta/dashed), 25-May-2006 (cyan/solid). Expiration date is 02-Jun-2006 for all pricing kernels.

• Figure 2: Examples of monthly pricing kernels from the first 6 months in 2006 for maturity one month:Dates are 18-Jan-2006 (blue/dash-dotted), 15-Feb-2006 (red/circle), 22-Mar-2006 (magenta/dashed), 19-Apr-2006 (cyan/solid), 17-May-2006 (black/dotted).

• Figure 3: Example of location and scale shift pricing kernels (left) and corresponding utility functions (right) of a power utility. Solid line in each plot represents reference curves of $K_0(u) = u^{-\gamma}$ and $U_0(u) = x^{1-\gamma}/(1 - \gamma)$ with $\gamma = 0.7$ respectively. Parameters are $\theta_2 = 0, \theta_1 = 1.1, \theta_2 = 1 - \theta_1^{(1/\gamma)}$, and $\theta_4 = 0$ for dot-dashed (red) and $\theta_4 = 0.5$ for dashed (blue) lines.

• Figure 4: Initial estimates of monthly pricing kernels with two landmarks identified. These represent pricing kernels between April 2003 and June 2006 with fixed maturity $\tau = 1$.

• Figure 5: Example of $q$ estimates with varying bandwidths (0.05, 0.1, 0.15, 0.20). The first three panels show estimates of implied volatility, its first and second derivative. The corresponding densities are shown in lower right panel. Estimates are stable for a wide range of bandwidths choices.

• Figure 6: Estimated common shape function $K_0$ (left) and transformed curves $K_t(\theta_2 u + \theta_3)$ of those in Figure 4 on the common domain (right).

• Figure 7: Parameter estimates of the shape invariant models for epk

• Figure 8: Estimated pricing kernels (left) and corresponding ARA measures (right) in 2003 (top) and 2004 (bottom). These are reconstructed based on the model parameters shown in Figure 6 and 7.

• Figure 9: Estimated pricing kernels (left) and corresponding ARA measures (right) in 2005 (top) and 2006 (bottom). These are reconstructed based on the model parameters shown in Figure 6 and 7.