Shape invariant modelling of pricing kernels and risk aversion*

Maria Grith† Wolfgang Härdle† Juhyun Park‡

Abstract

Several empirical studies reported that pricing kernels exhibit a common pattern across different markets. The main interest in pricing kernels lies in validating the presence of the peaks and their variability in location among curves. Motivated by this observation we investigate the problem of estimating pricing kernels based on the shape invariant model, a semi-parametric approach used for multiple curves with shape-related nonlinear variation.

This approach allows us to capture the common features contained in the shape of the functions and at the same time characterise the nonlinear variability with a few interpretable parameters. These parameters provide an informative summary of the curves and can be used to make a further analysis with macro economic variables. Implied risk aversion function and utility function also can be derived. 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.

*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.

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

1 Introduction

1.1 Pricing kernel and risk aversion

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. Namely a concave, convex or linear utility function is associated with risk averse, risk seeking or risk neutral behaviour respectively. The comparison is often made through the Arrow-Pratt measure of absolute risk aversion (ARA), as a summary of aggregate 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.

With an economic consideration that one unit gain and loss does not carry the same value for every individual, understanding state dependent risk behaviour becomes an increasing issue. The fundamental problem is that individual agents are directly observable but it is assumed that the prices of goods traded in the market reflect the dynamics of risk behaviour. Several efforts have been made to relate the price processes of assets of stocks and options traded in a market to risk behaviour of investors, 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 gives the fair priceunder a no arbitrage assumption. If markets are not complete, there are more risk neutral distributions and the fair price depends on the hedging problem. 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, write q for the risk neutral density and p for the historical density. The pricing kernel $\mathcal{K}$ is defined by the ratio of those densities:

$$\mathcal{K}(u) = \frac{q(u)}{p(u)}\,,$$

Through the intermediation of these densities, there exists a link between the pricing kernel and ARA, see for example Leland (1980)

$$ARA(u) = \frac{p'(u)}{p(u)} - \frac{q'(u)}{q(u)} = -\frac{d\log\mathcal{K}(u)}{du}\,.$$

In this way, rather than specifying a priori preferences of agents (risk neutral, averse or risk seeking) and implicitly the monotonicity of the pricing kernel, we can infer the risk patterns from the shape of the pricing kernel.

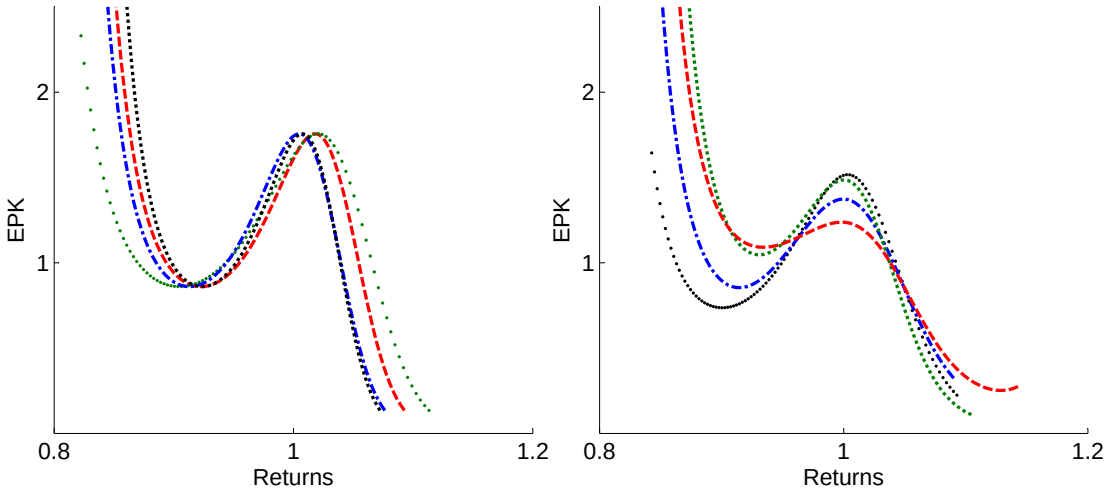

Figure 1: Examples of intertemporal pricing kernels for various maturities (left) and monthly pricing kernels from the first 6 months in 2006 for maturity one month (right).

1.2 Dynamics of 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 $\mathcal{K}$, potentially relevant are many studies focusing on recovering risk neutral density, 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.

Examples of empirical pricing kernels are shown in Figure 1. These are estimated based on European options data on DAX index in 2006. The detailed estimation method is explained in Section 3.5. Figure 1 depicts inter-temporal pricing kernels with various maturities in January-February 2006 (left), and monthly pricing kernels with fixed maturity one month in 2006 (right). To make these comparable, they are shown on a returns scale. Throught the article, the pricing kernel is considered as a function of this common scale of returns.

The sample of curves appears to have a bump around 1 and has 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.

From a statistical perspective, the properties of the pricing kernel are intrinsically related to the assumptions about the data generation process. A very restrictive model, with normal marginal distributions, is the Black-Scholes model. This results in an overall decreasing pricing kernel in wealth, which is consistent with overall risk-averse behaviour. These preferences are often assumed in the classical economic theory of utility maximizing agent and correspond to a concave indirect von Neumann and Morgenstern utility function. However, under richer parametric specifications or nonparametric models monotonicity of the pricing kernel has been rejected in practice (Rosenberg and Engle, 2002; Giacomini and Härdle, 2008). The phenomenon of locally nondecreasing pricing kernel is referred *pricing kernel puzzle* in the literature. There have been many attempts to reconcile the underlying economic theory with the empirical findings. A recent solution is suggested in Hens and Reichlin (2010), relating the puzzle to the violation of the fundamental assumptions in the equilibrium model framework.

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.

Chabi-Yo et al. (2007) explain the observed dynamics or the puzzles by means of latent variables in the asset pricing models. Effectively they propose to build conditional models of the pricing kernels given the state variables reflecting preferences, economic fundamentals or beliefs. Within this framework they were able to reproduce the puzzles, in conjunction with some joint parametric specifications for the pricing kernel and the asset return processes.

Due to evolution of markets over time under different circumstances, the pricing kernels 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, is not necessarily linear, and thus factors constructed from a linear combination of observations are only meaningful for explaining aggregated effects.

We approach the problem of estimating the pricing kernels and their risk aversion function from a functional data analysis viewpoint (Ramsay and Silverman, 2002), consider-

ing the pricing kernels as an object of curves. The main interest in pricing kernels lies in validating the presence of the peaks and their variability in location among curves. Motivated by this observation we investigate the estimation method based on the *shape invariant model*, which will be formally introduced in Section 2. This is chosen over the commonly adopted functional principal component analysis to accommodate the nonlinear features such as variation of peak locations, which encapsulate quantities related to economic interpretation. The shape invariant model allows us to capture the common characteristics, reported across different studies on different markets. We then explain individual variability as a deviation from the common curve viewing it as a reference. This framework also enables us to reproduce the observed pricing kernel puzzle. Our approach can be viewed as an alternative way of introducing state dependence in pricing kernels (Chabi-Yo et al., 2007), as illustrated in Section 2. In addition, based on the pricing kernel estimates we derive implied market risk behaviour based on the ARA measures and its corresponding utility function. The ARA corresponding to the reference pricing kernel may be viewed as a typical pattern of risk behaviour for the period under consideration. Through real data example we have related the changes in risk behaviour to some macro economic variables of interest and found that local risk loving behaviour is procyclical.

The paper is organised as follows. Section 2 motivates common shape modelling approach and Section 3 reviews the shape invariant model and describes it in detail in the context of pricing kernel estimation. This section serves the basis of our analysis. Numerical studies based on simulation are found in Section 4. Application to real data example is summmarised in Section 5. Note that all figures of pricing kernels are shown on the returns scale used in Figure 1.

2 Common shape modelling

2.1 Shape invariant model (SIM) for pricing kernel

When considering several markets simultaneously, we introduce the time index t in the pricing kernel as $\mathcal{K}_t$. We consider a common shape modelling approach for the series of pricing kernels with explicit components of location and scale known as shape invariant modelling. Then we formulate the relationship among K_ts as

$$\mathcal{K}_t(u) = \theta_{t1}\mathcal{K}_0\Big(\frac{u-\theta_{t3}}{\theta_{t2}}\Big) + \theta_{t4}\,. \tag{1}$$

We assume that the functional form of $\mathcal{K}_0$ is unknown. The common shape function $\mathcal{K}_0$ can be interpreted as a reference curve and deviation from the reference curve is described by four parameters $\boldsymbol{\theta}_t = (\theta_{t1}, \theta_{t2}, \theta_{t3}, \theta_{t4})$ that represent a scale change and a shift in horizontal and vertical direction.

This parametrisation in (1) is commonly known as shape invariant models (SIM), originally introduced by Lawton et al. (1972), and includes as a special case complete parametric models with known $\mathcal{K}_0$. Detailed account of this approach is given in Section 3.

The new message here is an analysis of a sequence of pricing kernels through shape-invariant models. Although we start with different motivation, our approach is in line with that of Chabi-Yo et al. (2007). In constrast to their approach, we impose a structural constraint that is related to the shape of the function. This way we strike a balance between flexibility much desired in parametric model specification and interpretability of the results lacking in full nonparametric models.

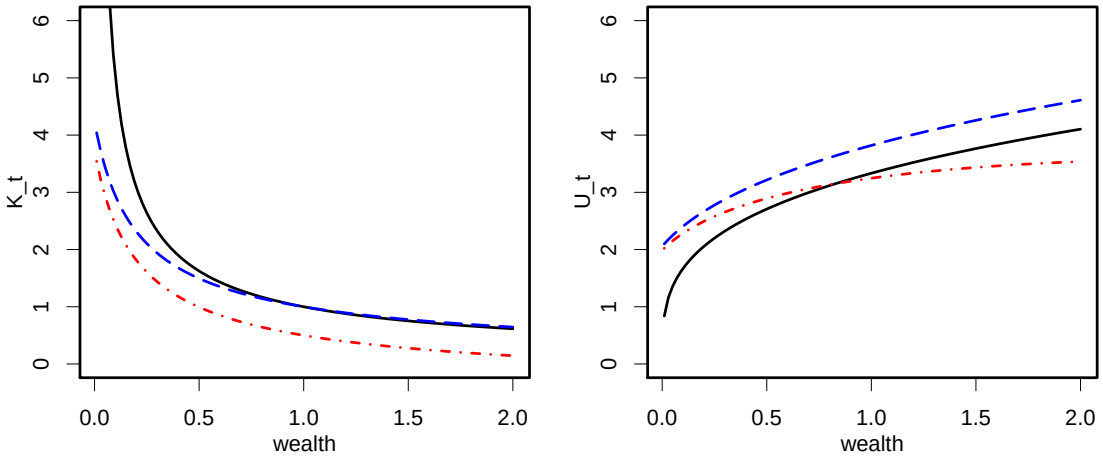

Figure 2: Example of location and scale shift pricing kernels (left) and corresponding utility functions (right) of a power utility. Solid line in each plot represents reference curves of $\mathcal{K}_0(u) = u^{-\gamma}$ and $U_0(u) = u^{1-\gamma}/(1-\gamma)$ with $\gamma = 0.7$ respectively. Parameters are $\theta_{t1} = 1.1, \theta_{t2} = 1, , \theta_{t3} = 1 - \theta_{t1}^{(1/\gamma)}$, and $\theta_{t4} = 0$ for dot-dashed (red) and $\theta_{t4} = -0.5$ for dashed (blue) lines.

2.2 SIM and Black-Sholes model

To appreciate the model formulation, it is instructive to consider utility functions implied by this family of pricing kernels together. The utility function can be derived from

$$U_t(u) = \alpha \int_0^u \mathcal{K}_t(x)\, dx\,,$$

for a constant α. Figure 2 shows an example of transformation based on a power utility function, which corresponds to risk averse behaviour, marked as solid line. Pricing kernels $\mathcal{K}_t$ are shown on the left and the corresponding utility functions U_t are on the

right. The dashed and dot-dashed lines represent $\mathcal{K}_t$ and U_t with appropriate parameters $\boldsymbol{\theta}_t$ in the equation (1). Depending on the choice of parameters, 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 μ 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

$$\mathcal{K}(u) = \lambda u^{-\gamma}\,, 0 < \gamma < 1\,,$$

with appropriate constants λ and γ. The corresponding utility function is a power utility

$$U(u) = \lambda \frac{u^{1-\gamma}}{1-\gamma}\,.$$

Assume that $\lambda = 1$ and suppose that $\mathcal{K}_0$ is the Black-Scholes power function $u^{-\gamma}$. Then the class of pricing kernel implied in (1) is given by

$$\begin{aligned}\mathcal{K}_t(u) &= \theta_{t1}\Big(\frac{u-\theta_{t3}}{\theta_{t2}}\Big)^{-\gamma} + \theta_{t4} \\ &= \theta^*_{t1}(u-\theta_{t3})^{-\gamma} + \theta_{t4}\,,\end{aligned}$$

where $\theta^*_{t1} = \theta_{t1}\theta^{\gamma}_{t2}$. Notice that with this family of functions θ_{t1} and θ_{t2} are not identifiable and $\mathcal{K}_t$ is defined for $u > \theta_{t3}$. For the sake of argument we set $\theta_{t2} = 1$ for the moment. The corresponding utility function is

$$\begin{aligned}U_t(u) &= \int_{\theta_{t3}}^{u} \mathcal{K}_t(x)\,dx \\ &= \frac{\theta_{t1}}{1-\gamma}(u-\theta_{t3})^{(1-\gamma)} + \theta_{t4}(u-\theta_{t3}) \\ &\stackrel{\text{def}}{=} \theta^{**}_{t1}(u-\theta_{t3})^{(1-\gamma)} + \theta_{t4}(u-\theta_{t3})\,.\end{aligned}$$

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 2 for comparison.

2.3 Identifiability condition for SIM

The previous section illustrates two aspects of applicability of the shape invariant models. The class of functions that can be generated by the relation (1) is rich, but in order to uniquely identify the model parameters, some restriction is necessary. For example, we

have seen that the two scale parameters in the pricing kernel functions corresponding to the Black-Scholes model are not separable. Basically unless there exist some qualitatively distinct common characteristics for each curve, the model is not identifiable (Kneip and Gasser, 1988). In the case of no prior structural information available as in the case of pricing kernels, it is sufficient to consider a few landmarks such as peaks and inflection points.

Even with unique $\mathcal{K}_0$, some translation and scaling of parameters lead to multiple representations of the models. For uniqueness of parameters, we will impose normalizing conditions suggested in Kneip and Engel (1995):

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

in the sense that there exists an *average curve*. This is not restriction at all and can be replaced by any appropriate combination of parameters. Alternatively, we could consider the first curve as a reference, as done in Härdle and Marron (1990), which implies the restriction $\boldsymbol{\theta}_1 = (1,1,0,0)$. Generally an application-driven normalisation scheme can be devised and examples are found in Lawton et al. (1972).

2.4 SIM implied risk aversion and utility function

In general the utility function corresponding to $\mathcal{K}_t$ is given by

$$\begin{aligned} U_t(u) &= \theta_{t1}\theta_{t2}\Big\{U_0\Big(\frac{u-\theta_{t3}}{\theta_{t2}}\Big)-U_0\Big(-\frac{\theta_{t3}}{\theta_{t2}}\Big)\Big\}+\theta_{t4}u \\ &\equiv \theta^*_{t1}U_0\Big(\frac{u-\theta_{t3}}{\theta_{t2}}\Big)+\theta^*_{t4}+\theta_{t4}u\,. \end{aligned}$$

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

$$ARA_t(u)=\frac{-\frac{\theta_{t1}}{\theta_{t2}}\mathcal{K}_0'\Big(\frac{u-\theta_{t3}}{\theta_{t2}}\Big)}{\theta_{t1}\mathcal{K}_0\Big(\frac{u-\theta_{t3}}{\theta_{t2}}\Big)+\theta_{t4}}\,. \tag{2}$$

For example, assuming $\mathcal{K}_0(u)=u^{-\gamma}$ with $\theta_{t2}=1$ gives

$$ARA_t(u)=\gamma\big\{(u-\theta_{t3})+(\theta_{t4}/\theta_{t1})(u-\theta_{t3})^{\gamma+1}\big\}^{-1}\,.$$

When $\theta_{t4}=0$, this function is monotonically decreasing but in general this is not the case.

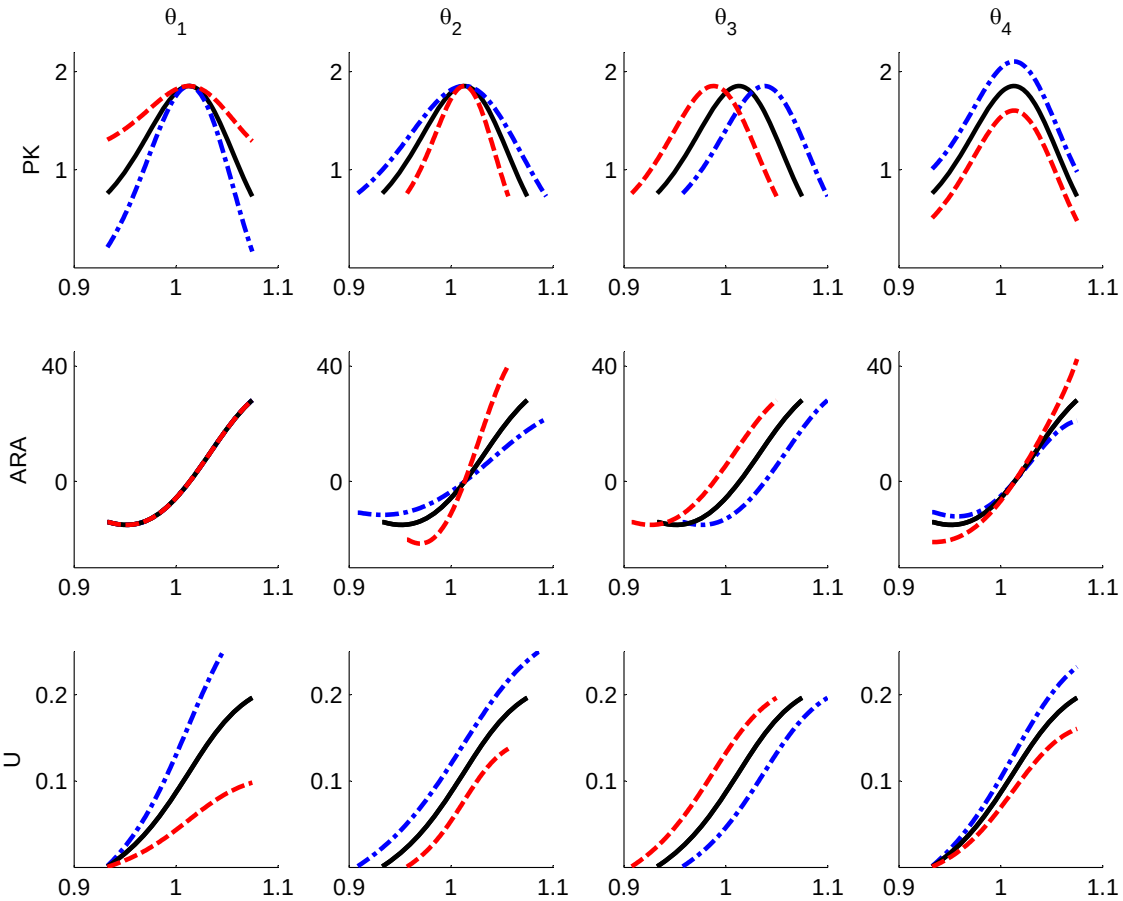

Figure 3: Effect of parameters on pricing kernel (top), ARA (middle) and utility functions (bottom) compared to the baseline model $\theta_0 = (1, 1, 0, 0)$ (black). Dot-dashed lines are used for increasing direction and dashed lines for decreasing direction.

In order to gain some insights, we take a closer look at the changes that individual effects in the family of scale and shift parameters. These effects are demonstrated in Figure 3. We vary each θ_i with respect to a baseline model and then we show how these modifications translate into changes of the risk attitudes and the corresponding utility functions. The parameters used in Figure 3 are $\boldsymbol{\theta} = (0.5, 0.7, -0.025, -0.25)$ in dashed red line and $\boldsymbol{\theta} = (1.5, 1.3, 0.025, 0.25)$ in dot-dashed blue line.

For this exercise we first standardise the common curve that we have estimated via the shape invariant model so that the peak occurs at the value 0 on the abscissa and the effect of the scale and shift parameters is separately captured. But we added the peak coordinates back for visualisation to be in line with other figures shown on returns scale. We observe that an increase in θ_1 marks the bump of the pricing kernel more distinctive

while the shape of ARA remains unchanged compared to the baseline model because, as we can see from (2), ARA does not depend on θ_1 when $\theta_4 = 0$. Yet, the effect of θ_1 on ARA can be analysed by considering two distinct cases: $\theta_4 > 0$ and $\theta_4 < 0$. These specifications are important because the direction of change in the slope of ARA is dictated by the sign of θ_4. In the present case - after normalisation - θ_1 varies around 0 and its effect on ARA is almost nil.

A larger value in the parameter θ_2 as compared to a benchmark value of stretches the x-axis, which implies larger spread of the bump. When we vary θ_2 alone the slope of $\text{ARA}(\theta_2 u)$ is $1/\theta_2^2 \left[\left\{ \mathcal{K}_0'^2(u) - \mathcal{K}_0''(u)/\mathcal{K}_0(u) \right\} / \mathcal{K}_0^2(u) \right]$. The term in brackets does not depend on θ_2; it is equal to the slope of $ARA(u)$. Therefore, there is an inverse relationship between the direction of change in the parameter and that of the absolute value of the slope. These changes in slope occur around an inflection point that corresponds to the peak of the pricing kernel.

A positive increment in θ_3 shifts both curves to the left without any modification in the shape. θ_4 simply translates pricing kernel curves above or below the reference curve following a sign rule. Similarly to θ_2, the shape of ARA modifies around the fixed inflection point that marks the change from risk proclivity (negative ARA) to risk aversion (positive ARA). The effect of θ_4 on the values of ARA is straightforward: since θ_4 adds to the $\mathcal{K}_0$ in the denominator its increase will diminish the absolute ARA level and the other way around. Insulating the effects of a change in θ_4 on the slope of $ARA(u)$ analytically proves to be a more complicated task than in the case of θ_2 because the change in the slope depends jointly on the change in θ_4 and on the pricing kernel values and its first two derivatives. In our case, the slope around the inflection point increases when θ_4 decreases.

As for the utility function, positive changes in θ_1 and θ_4 increases its absolute slope. In the horizontal direction, θ_3 translates the curve to the left or right similarly to the pricing kernel and ARA while θ_2 shrinks or expands its domain.

With this information at hand we can characterise the changes in risk patterns in relation with economic variables of interest, see Section 5.4.

3 Fitting Shape Invariant Models

3.1 Model formulation

Strictly speaking, there is no realisation of the pricing kernels available, however, their estimates are readily available from market data. Our main interest lies in quantifying the variation among the pricing kernels given those estimates. For the purpose of analysis, we treat the estimates as something observable and denote by Y_t, similar to the

regression formulation with direct measurements Y_t. A particular choice of estimates of individual pricing kernels is not part of the model formulation but affects the starting values for the estimation of shape invariant model. Our choice of initial estimates will be explained in Section 3.5.

Suppose that these are evaluated at fine grid points u_j. Let $\{Y_{tj}, t = 1, 2, \cdots, T; j = 1, 2, \cdots, n\}$ be the estimates evaulated at $\{u_j\}$ in an interval J satisfying the relation

$$Y_{tj} = \mathcal{K}_t(u_j) + \varepsilon_{tj}\,,$$

where K_t satisfies the shape invariant model relation 1 and ε_{tj} are independent errors with mean zero and standard deviation σ_t^2.

It is possible to elaborate our approach to incorporate simultaneous estimation with a two-step state dependent dynamic model formulation whereby the dynamics of the observed return processes and the unobserved pricing kernel processes as a state variable are specified separately. However, with current advancement in the methodology, this is only possible with limited parametric model choices, see for example Chabi-Yo et al. (2007), and extension to a flexible shape invariant model is left for future work.

3.2 Estimation of SIM

The model in (1) is equivalently written as

$$\mathcal{K}_t(\theta_{t2}u + \theta_{t3}) = \theta_{t1}\mathcal{K}_0(u) + \theta_{t4}\,, \qquad \theta_{t1} > 0\,, \quad \theta_{t2} > 0\,. \tag{3}$$

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

$$\int \{\hat{\mathcal{K}}_2(\theta_2 u + \theta_3) - \theta_1\hat{\mathcal{K}}_1(u) - \theta_4\}^2 w(u)\,du\,, \tag{4}$$

where $\hat{\mathcal{K}}_i$ 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

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

The parameter estimates are compared only in the common region defined by w but the individual curve estimates are defined on the whole interval. Weights can be extended to account for additional variability.

The normalisation leads to:

$$T^{-1}\sum_{t=1}^{T}\mathcal{K}_t(\theta_{t2}u+\theta_{t3})=\mathcal{K}_0(u)\,. \tag{5}$$

Formula (5) was exploited also in the algorithm proposed by Kneip and Engel (1995). We adopt a similar strategy here.

- Initialize
  - Let $\hat{K}_t = Y_t$ and set starting values $(\theta_{t2}^{(0)}, \theta_{t3}^{(0)})$ for $t = 1, 2, \cdots, T$.
  - Construct an initial estimate $\mathcal{K}_0^{(0)}$ by

$$\mathcal{K}_0^{(0)}(u) = T^{-1}\sum_{t=1}^{T}\hat{\mathcal{K}}_t(\theta_{t2}^{(0)}u+\theta_{t3}^{(0)})\,.$$

- For rth step, $r = 1, 2, \cdots, R$,
  - Determine parameters $\boldsymbol{\theta}^{(r)}$ separately for $t = 1, 2, \cdots, T$ by minimizing

$$\int\{\hat{\mathcal{K}}_t(\theta_{t2}u+\theta_{t3})-\theta_{t1}\mathcal{K}_0^{(r-1)}(u)-\theta_{t4}\}^2 w(u)\,du\,.$$

  - Normalise parameters: for $j = (1, 2)$ and $k = (3, 4)$

$$\theta_{tj}^{(r)} \leftarrow \frac{\theta_{tj}^{(r)}}{\sum_t \theta_{tj}^{(r)}}\,, \qquad \theta_{tk}^{(r)} \leftarrow \theta_{tk}^{(r)} - T^{-1}\sum_t \theta_{tk}^{(r)}\,.$$

  - Update $\mathcal{K}_0^{(r-1)}$ to

$$\mathcal{K}_0^{(r)}(u) = T^{-1}\sum_{t=1}^{T}\hat{\mathcal{K}}_t(\theta_{t2}^{(r)}u+\theta_{t3}^{(r)})\,.$$

- Determine final estimates:

$$\begin{aligned}\tilde{\boldsymbol{\theta}}_t &= \boldsymbol{\theta}_t^{(R)}\,,\\ \tilde{\mathcal{K}}_0(u) &= T^{-1}\sum_{t=1}^{T}\hat{\mathcal{K}}_t(\tilde{\theta}_{t2}u+\tilde{\theta}_{t3})\,.\end{aligned}$$

Kneip and Engel (1995) proved consistency of the estimator. In particular despite non-parametric initial curve estimates, the parameters are shown to be $\sqrt{T}$ consistent. In their analysis it is noted that the initial estimates of the curves are of minor importance

compared to the final estimate of K_0. So the original algorithm includes the final updating of each curve. This improves precision of the estimates because the pooled sample estimate reduces variance $\tilde{\mathcal{K}}_0$, which allows undersmoothing at the final stage to reduce bias. However, this final updating step is not practical for our situation with indirect measurements and is not implemented here for pricing kernel estimation. On the other hand we can take advantage of having smooth curves evaluated at finite grid points as data. It is easier to improve the initialisation step, explained in Section 3.3. 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.

As a working model we have assumed an independent error. If there is a reasonable dependence structure available, this could be incorporated easily in the estimation algorithm with weighted least squares estimation in (4). The effect of independence assumption mainly appears in the standard error estimation and a correction can be made with a sandwich variance-covariance estimator. To assess the effect of model misspecification, we also carried out some simulation studies with dependent errors and reported the results in Section 4.

3.3 Starting values

If there is no scale change in horizontal direction, due to prominent peaks in each curve, the parameter θ_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 θ_2. Suppose that the peak is identified by u satisfying $K_t'(u) = 0$. Then we have

$$0 = \mathcal{K}_t'(u) = \frac{\theta_{t1}}{\theta_{t2}} \mathcal{K}_0'\Big(\frac{u - \theta_{t3}}{\theta_{t2}}\Big)\,.$$

Writing u_t^* for $\mathcal{K}_t'$ and u_0^* for $\mathcal{K}_0'$ leads to a simple linear relation:

$$u_t^* = \theta_{t2} u_0^* + \theta_{t3}\,. \tag{6}$$

If an inflection point is used, we would have

$$0 = \mathcal{K}_t''(u) = \frac{\theta_{t1}}{\theta_{t2}^2} \mathcal{K}_0''\Big(\frac{u - \theta_{t3}}{\theta_{t2}}\Big)\,,$$

which gives rise to the same relation as (6), with the corresponding u_t^{**} and u_0^{**} substituted. The coefficients of intercept and slope estimates are used for starting values of θ_{t3} and θ_{t2} respectively.

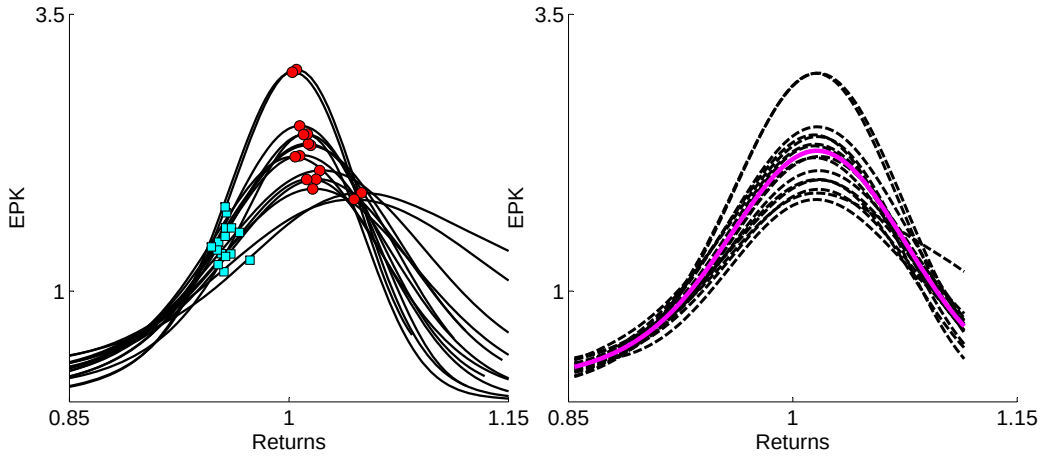

Figure 4: Initial estimates $\mathcal{K}_t(u)$ (left) and final estimates $\mathcal{K}_t(\theta_{t2}u+\theta_{t3})$ from SIM (right) with $\mathcal{K}_0$ overlayed. Marked in the left plot are two landmarks identified for estimation of the starting values of $(\theta_{t2}, \theta_{t3})$.

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 $\mathcal{K}_t$ 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 nonlinearity detected, the model needs to be extended to include a nonlinear transformation. With our example, this was not the case.

3.4 Nonlinear optimisation

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

$$\sum_j \left\{\hat{\mathcal{K}}_t(\theta_{t2}u_j + \theta_{t3}) - \theta_{t1}\hat{\mathcal{K}}_0(u_j) - \theta_{t4}\right\}^2 w(u_j)\,. \tag{7}$$

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

3.5 Initial estimates of $\mathcal{K}$

To start the algorithm the initial estiamtes of $\mathcal{K}$ should be supplied. An example of initial estimates of $\mathcal{K}$ is shown in Figure 4 on the scale of continuously compounded returns. These are obtained from separate estimation of p and q, which are described in detail in Section 5.

3.6 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. This is due to the difference between the daily observations available for estimating p and the intraday observations available for estimating q. Thus it might be 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.

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

$$\begin{aligned} \hat{\mathcal{K}}(u) - \mathcal{K}(u) &= \frac{\hat{q}(u)}{\hat{p}(u)} - \frac{q(u)}{p(u)} \\ &\simeq \frac{\hat{q}(u) - q(u)}{p(u)} - \frac{q(u)}{p(u)} \frac{\hat{p}(u) - p(u)}{p(u)} \,. \end{aligned}$$

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 ϵ. Provided that $p(u) > \epsilon$ and $q(u) > \epsilon$, the asymptotic approximation is straightforward and asymptotic bias and variance can be

computed from

$$\begin{aligned}
\mathsf{E}[\hat{\mathcal{K}}(u) - \mathcal{K}(u)] &\simeq \frac{\mathsf{E}[\hat{q}(u) - q(u)]}{p(u)} - \frac{q(u)}{p(u)}\frac{\mathsf{E}[\hat{p}(u) - p(u)]}{p(u)} \\
&= \mathcal{O}(h_q^4) + \mathcal{O}(h_p^2) + o(h_p^2 + h_q^4)\,, \\
\mathsf{Var}[\hat{\mathcal{K}}(u) - \mathcal{K}(u)] &\simeq \mathcal{K}^2(u)\Big\{\frac{\mathsf{Var}[\hat{q}(u)]}{q^2(u)} + \frac{\mathsf{Var}[\hat{p}(u)]}{p^2(u)}\Big\} \\
&= \mathcal{O}\big\{(Nh_q)^{-1}\big\} + \mathcal{O}\big\{(nh_p)^{-1}\big\} + o\big\{(Nh_q)^{-1} + (nh_p)^{-1}\big\}\,.
\end{aligned}$$

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 dominated by the estimation of q and for the purpose of asymptotics p can be regarded as a fixed quantity. For this reason we actually implement a semiparametric estimator for q to stablise the estimator.

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. We write the approximate distribution for $\hat{\boldsymbol{\theta}}_t$ as

$$\hat{\boldsymbol{\theta}}_t \approx \mathrm{N}(\boldsymbol{\theta}_t, \Sigma_t)\,.$$

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 $\hat{e}_{tj} = \hat{\mathcal{K}}_t(u_j) - \tilde{\mathcal{K}}_t(u_j)$ where $\hat{\mathcal{K}}$ is the initial estimates and $\tilde{\mathcal{K}}$ is the SIM estimates and let

$$\hat{\sigma}_t^2 = \frac{1}{n}\sum_{j=1}^{n} \hat{e}_{tj}^2\,.$$

The covariance matrix can be estimated as

$$\hat{\Sigma}_t = \hat{\sigma}_t^2\Big[n^{-1}\sum_{j=1}^{n}\big\{\bigtriangledown_{\boldsymbol{\theta}}\,\tilde{\mathcal{K}}_t(u_j;\tilde{\boldsymbol{\theta}})\big\}\big\{\bigtriangledown_{\boldsymbol{\theta}}\,\tilde{\mathcal{K}}_t(u_j;\tilde{\boldsymbol{\theta}})\big\}^{\top}\Big]^{-1}\,,$$

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

$$\begin{aligned}
\frac{\partial\mathcal{K}(u)}{\partial\theta_1} &= \mathcal{K}_0\Big(\frac{u-\theta_3}{\theta_2}\Big)\,, \\
\frac{\partial\mathcal{K}(u)}{\partial\theta_2} &= -\frac{\theta_1}{\theta_2}\Big(\frac{u-\theta_3}{\theta_2}\Big)\mathcal{K}_0'\Big(\frac{u-\theta_3}{\theta_2}\Big)\,, \\
\frac{\partial\mathcal{K}(u)}{\partial\theta_3} &= -\frac{\theta_1}{\theta_2}\mathcal{K}_0'\Big(\frac{u-\theta_3}{\theta_2}\Big)\,, \\
\frac{\partial\mathcal{K}(u)}{\partial\theta_4} &= 1\,.
\end{aligned}$$

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.

4 Numerical studies of SIM estimation

Applying the SIM to pricing kernels involves two rather separate estimation steps, the initial estimation of the pricing kernels and the SIM estimation given the pre-estimates. The former has been studied extensively and in particular the properties of the nonparametric methods that we have used are well established in the literature. This section mainly concerns the latter.

We identify the two main factors that could affect the performance of SIM estimation to be error misspecification and smoothing parameter selection for the individual curves. Their effects are evaluated in a designed simulation studies. Their effects on pricing kernel estimation are separately studied in Section 5.4, in comparison to the standard nonparametric approach used in Jackwerth (2000).

4.1 Generating curves

In each simulation 50 curves are generated at 50 (random) grid points. In order to mimic the common shape of the observed pricing kernel, we generated the common curve by a ratio of two densities

$$\mathcal{K}_0(u) = q_0(u)/p_0(u)\,,$$

where p_0 is density of Gamma(0.8,1) distribution and q_0 is density of mixture $w *$ Gamma$(0.2, 1) + (1 - w) * N(0.91, 0.3^2)$ distribution with $w = 0.3$. In accordance with the normalisation scheme, the $\boldsymbol{\theta}$ values are set as in Table 1. The values of realised the standard deviation were chosen to be similar to the observed ones in the real data example.

	Distribution	Mean	Standard deviation
θ_1	Log-normal	1	0.33
θ_2	Log-normal	1	0.28
θ_3	Normal	0	0.27
θ_4	Normal	0	0.35

Table 1: Parameter values of $\boldsymbol{\theta}$.

4.2 Error specification

For the error specification, we have included dependent errors in time as well as in moneyness as following.

- Case 1: Independent error: $\varepsilon_{t,j} \sim N(0, \sigma^2)$
- Case 2: Dependent error in moneyness:

$$\varepsilon_{t,j} = \phi \varepsilon_{t,j-1} + u_{t,j}\,, \qquad u_{t,j} \sim N(0, \sigma_u^2)$$

- Case 3: Dependent error in time: $\varepsilon_{t,j} \sim N(0, \sigma_t^2)$

$$\log(\sigma_t) = \alpha + \beta \log(\sigma_{t-1}) + v_t\,, \qquad v_t \sim N(0, \sigma_v^2)$$

- Case 4: Dependent error in moneyness and time:

$$\begin{aligned} \varepsilon_{t,j} &= \phi \varepsilon_{t,j-1} + u_{t,j}\,, & u_{t,j} &\sim N(0, \sigma_{ut}^2)\,, \\ \log(\sigma_{ut}) &= \alpha + \beta \log(\sigma_{u,t-1}) + v_t\,, & v_t &\sim N(0, \sigma_v^2) \end{aligned}$$

Case 1 and 2 are commonly assumed but Case 3 and 4 were rarely used in the literature with SIM estimation. Table 2 lists the parameter values set for the simulation. These values are chosen to be comparable in terms of overall variability among cases.

4.3 Smoothing parameter selection

We consider three versions of the least squares cross-validation (CV) based criteria for bandwidth selection:

$$\mathrm{CV}_t(h) = \sum_{i=1}^{n} \left\{ Y_{t,i} - \widehat{\mathcal{K}}_{t,h}^{-(i)}(u_i) \right\}^2 ,$$

where $\widehat{\mathcal{K}}_{t,h}^{-(i)}$ is the local linear fit without using the i-th observation. For each observed curve we find the optimal bandwidth $h_t^* = \arg\min \mathrm{CV}_t(h)$. Due to considerable variability in the x-dimension we standardise the optimal bandwidths ($\tilde{h_t^*} = h_t^*/s_t$), where

		Error 1	Error 2	Error3
Case 1	σ	0.02	0.05	0.10
Case 2	ϕ	0.75	0.75	0.75
	σ_u	0.02	0.03	0.09
Case 3	α	-3.69	-2.99	-2.30
	β	0.75	0.52	0.53
	σ_v	0.01	0.02	0.02
Case 4	α	-2.41	-1.89	-1.39
	β	0.45	0.40	0.42
	ϕ	0.75	0.45	0.45
	σ_v	0.10	0.25	0.25

Table 2: Parameter values for error specification.

s_t is the empirical standard deviation, and we choose the common bandwidth as follows:

$$h_{opt,1} = \max(\tilde{h_t^*}) \qquad h_{opt,2} = \text{average}(\tilde{h_t^*}) \qquad \text{or} \qquad h_{opt,3} = \arg\min \sum_t \text{CV}_t(h)\,.$$

Finally, we multiply h_{opt} by s_t and use these values to perform smoothing of each curve.

4.4 Results of simulation

We considered various simulation scenarios based on the combinations of the case of errors and bandwidth selection methods. Table 3 summarises the results of the goodnees of fit measured by MSE for the case $\sigma = 0.05$. For comparison we added in the last row the MSE for the standard nonparametric estimates based on individual optimal bandwidths to their advantage. For larger error ($\sigma = 0.1$, not shown) we also observed some dramatic deterioration with case 4. Nevertheless the simulation studies suggest that the overall error is in the same order of magnitude and we suspect that the impact of these factors is limited. The fit was however best with smoothing parameters selected by h_1.

5 Real data example

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

$\sigma = 0.05$					
methods	parms.	case 1	case 2	case 3	case 4
h_1	θ_1	31	32	67	65
	θ_2	60	70	84	77
	θ_3	54	62	81	76
	θ_4	32	32	77	75
	K_is	1.2	1.6	1.5	1.5
h_2	θ_1	67	68	80	69
	θ_2	115	115	110	99
	θ_3	111	110	105	103
	θ_4	70	72	99	85
	K_is	1.1	1.6	1.9	1.9
h_3	θ_1	67	71	67	73
	θ_2	115	108	91	82
	θ_3	111	100	88	84
	θ_4	70	74	83	88
	K_is	1.1	1.6	1.8	1.8
	np$K$	3.5	2.0	4.2	3.6

Table 3: Comparison of SIM estimation with respect to error misspecification and smoothing parameter selection. Numbers are MSE multiplied by 10000. K_is computes the average MSE for all curves from SIM and np$K$ without SIM but using individual optimal bandwidths for each curve.

strike prices, the interest rates (linearly interpolated to approximate a *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 June 2003 and June 2006.

5.1 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 (Ait-Sahalia and Duarte, 2003)

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

where

- S_t: the underlying asset price at time t,

- X: the strike price,
- τ: 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 τ, assume $r_{t,\tau} = r$ and $\delta_{t,\tau} = \delta$, the risk neutral density is expressed as

$$q(u) = e^{r\tau} \frac{\partial^2 C}{\partial X^2}|_{X=u} .$$

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 σ is a function of the option's moneyness and the time to maturity τ. In this work, we fix the maturity and consider it as one dimensional regression problem, for which the local polynomial smoothing with degree $p = 3$ is applied to the observations of implied volatility on moneyness scale. The detailed derivation is given in Appendix.

We have identified options data with maturity one month (31 working days/ 23 trading days) from June 2003 to June 2006, from DAX 30 Index European options traded on Eurex Exchange, which adds up to 37 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:

$$S^* \geq C_i \geq \max(S^* - X_i e^{-r\tau}, 0) ,$$

where S^* is the adjusted price for intraday movements.

Moneyness is computed for each pair (S_i^*, K_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 4 are defined on a returns scale (continuously compounded 1 month-period returns $R_T = 1 + \log(S_T/S_t)$) different from the moneyness definition used by Rookley (1997)).

5.2 Estimation of the historical density p

We use the nonparametric kernel density estimates similar to Ait-Sahalia and Lo (2000) based on the past two years' observations of returns from the maturity. With this approach 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. Alternatively if additional modelling assumption is made for the evolution of the stock price such as GARCH, a simulation based approach could be employed.

Jackwerth (2000) argues that some discrepancies between the nonparametric 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). For comparison to the earlier works, we also experimented with a choice of time varying equity premium and constant equity premium (we demean the densities and supplant it with the risk free rate on the estimation day plus 8% equity premium per annum as in Jackwerth (2000) adjusted for the corresponding maturity), overlapping and nonoverlapping returns, window length 500, 1000, 1500 trading days (which corresponds to 2, 4 and 6 years respectively). The estimates with the different choice of parameters are then compared subsequently in terms of pricing kernel, implied risk aversion and implied utility function estimation.

5.3 Smoothing parameter selection

In constrast to the simulation studies, the effect of smoothing parameter is less transparent with real data when we estimate p and q separately. At first glance, the bandwidth selection for q seems 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 $\hat{q}$. 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 bandwidth choices.

For a systematic choice, we employed a version of CV criteria ($h_{opt,1}$ defined in Section 4.3) for p and q estimation. For estimation of q, we have used the least squares CV

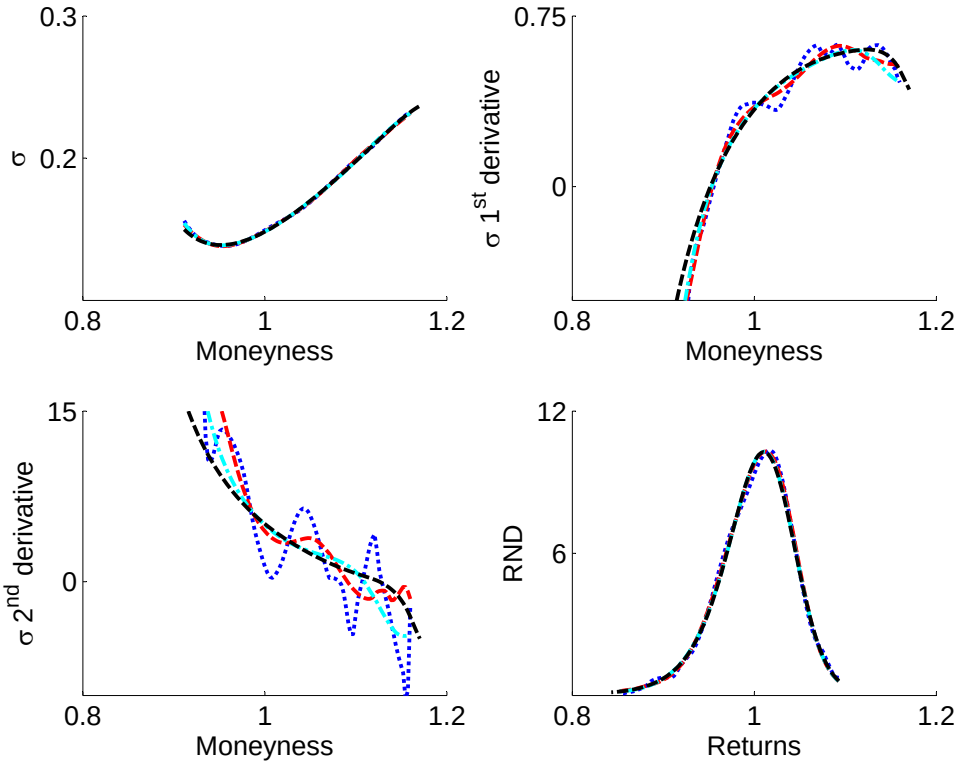

Figure 5: Example of q estimates with varying bandwidths (0.05, 0.1, 0.15, 0.20). The first three panels show estimates of implied volatility, its first and second derivative. The corresponding densities are shown in lower right panel. Estimates are stable for a wide range of bandwidths choices.

using local cubic estimation to include the second derivative of C.

$$\begin{aligned} \mathrm{CV}_t(\tilde{h}) \quad = \quad & \sum_{i=1}^{n}\sum_{j\neq i}^{n}\Big\{Y_{t,i} - \widehat{C}^{(0)}_{\tilde{h},-i}(X_{t,i}) - \widehat{C}^{(1)}_{\tilde{h},-i}(X_{t,i})(X_{t,j}-X_{t,i}) \\ & -\frac{1}{2}\widehat{C}^{(2)}_{\tilde{h},-i}(X_{t,i})(X_{t,j}-X_{t,i})^2\Big\}^2 w(X_{t,i}) \end{aligned} \tag{8}$$

where $\hat{C}^{(k)}_{\tilde{h},-i}$ are kth derivative estimate obtained without ith observation $(X_{t,i}, Y_{t,i})$ and $0 \leq w(X_{t,i}) \leq 1$ is a weight function. See, for example, Li and Racine (2007) for details. The h_1-optimal bandwidth is selected in implied volatility space, which turns out to be

$h_q = 0.2$. For estimation of p, we have used the likelihood CV for each curve:

$$\ln L = \sum_{i}^{n} \ln \hat{p}^{-(i)}(X_i)$$

where $\hat{p}^{-(i)}(X_i)$ is leave-one-out kernel estimator for $p(X_i)$. However we found that the optimal bandwidth selected tends to systematically oversmooth and thus we chose a value close to the maximum of individually optimal bandwidths, which is in our case $h_p = 0.05$.

5.4 Estimation of EPK, ARA, Utility function

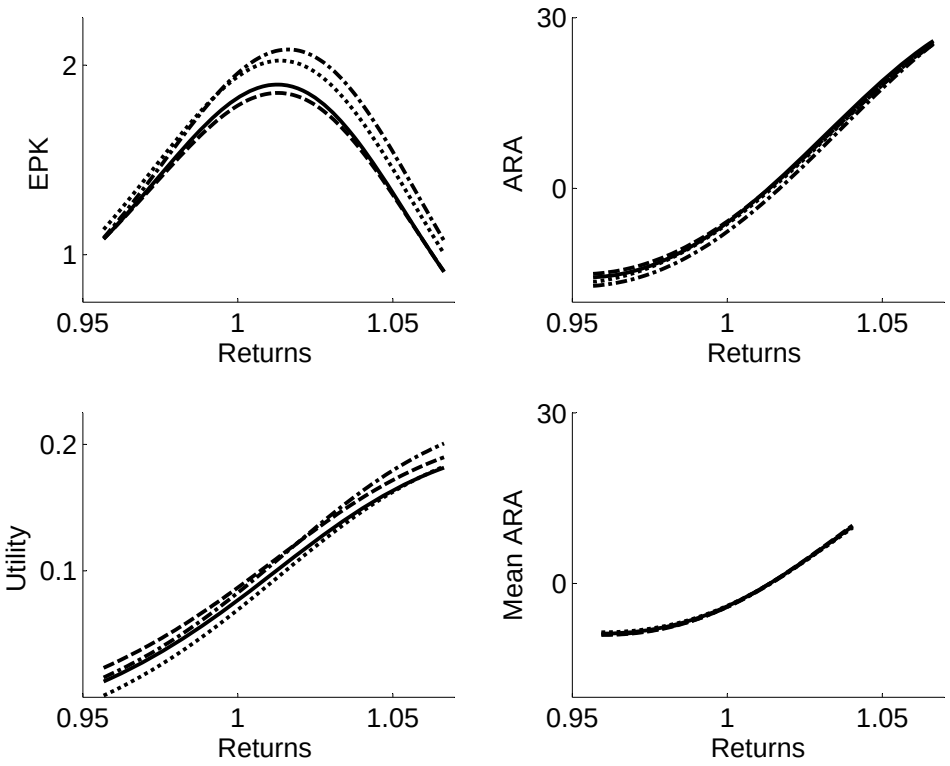

Figure 6: Illustration of SIM with common EPK, ARA, utility function and mean ARA.

We have considered in Section 5.2 various options for the parameter choice in estimating p and have ended up with twelve series of pre-estimates of EPK. We are interested in seeing how these choices influence the estimated common curves and θ_t parameters by SIM. Since, as it turns out, the results are very similar among specifications we depict graphically only four of them in Figures 6 and 7: those based on nonoverlapping (solid)

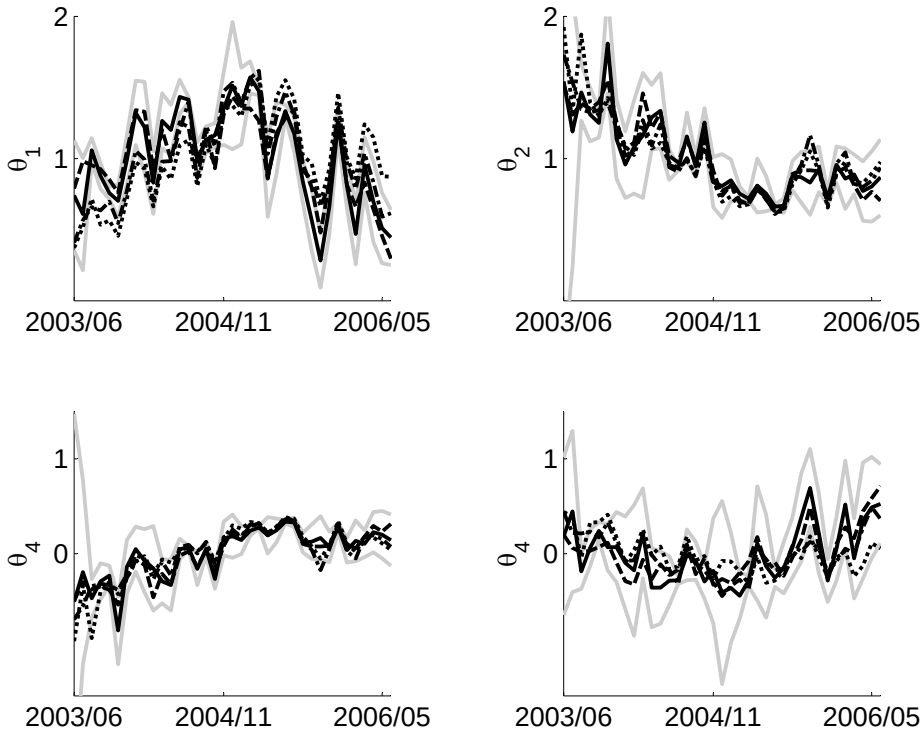

Figure 7: Estimated SIM parameters.

and overlapping (dashed) returns over the last two years, nonoverlapping returns over the last four (dot-dashed) and six (dotted) years respectively with varying equity premium. The added lines in Figure 7 are 95% pointwise confidende band for the first series of pre-estimates.

The common curves are represented in Figure 6. All estimates display a *paradoxal* feature: EPK has a bump, ARA has a region of negative values that correspond to the increasing region in the EPK, utility function has a convex region in the domain around the EPK peak. The variability among curves is expressed by θ_t-s. In Figure 7 we observe that the main difference in the dynamics of different series has to do with the magnitude but less with the direction of change. In addition, we computed the mean of implied ARA corresponding to our estimation period and found that it similar to the the mean ARA for $S\&P500$ appearing in Figure 3 panel C - 19. March 91 to 19 August 1993 in Jackwerth (2000), and to a certain extent to the yearly average from 2003 and 2005 shown in Figure 4 in Chabi-Yo et al. (2007). It is worth noting that the mean ARA and the common ARA curves differ a great deal. This is not surprising since the interpretation of common curve is different from the average curve, in particular the common curve and the mean curves have different scales of the x-domains - by means

of registration.

5.5 Relation to macro economic variables

With an aid of the SIM model for EPK, we wish to characterise changes in risk patterns in relation to economic variables of interest. Before doing this, we should mention that in the case of nonstandard common curves - in our empirical example the peak does not occur at 0 - both θ_1 and θ_2 introduce a shift effect in EPK together with its shape effect. In order to disentangle these effects and improve interpretation we first standardise the EPK curves by the location of the peak before applying SIM. This introduces two more parameters, the horizontal and vertical coordinates of the peaks in the analysis. Since their shift effect is comprised by parameters θ_3 and θ_4 we will not treat them here in more detail.

Previous studies trying to link the parameters describing risk attitudes to the business conditions include Rosenberg and Engle (2002). Based on power pricing kernel specifications they show that risk aversion is counter-cyclical. Other related work investigates the relation between equity premiums (e.g. Fama and French (1989)), smile asymmetry of volatility (Drechsler and Yaron (2010), Bekaert and Wu (2000)) or market efficiency (Marshall et al., 2008). The advantage of our approach over Rosenberg and Engle (2002) is that it allows us to identify how the change in economic variables relates to the shape of a nonparametrically estimated pricing kernel. Due to limited sample size - 37 observations it is impossible to estimate a structural model that correctly deals with the simultaneity of our set of dependent variables. Further research will involve the estimation of a (S)VAR specification, in order to account for the aforementioned endogeneity. We instead evaluate the potential univariate correlations between the estimated θ_t parameters and macro economic variables associated with the business cycle and interpret our results from the perspective of local EPK and risk aversion functions. We use the following variables that have a revealed relation with the state of the economy: credit spread (CS) is the difference between the yield on the corporate bond[1] and the government bond maturing in 5 years; the yield curve slope (YT) refers to the difference between the thirty-year government bond yield and three-months intebank rate; short term interest rate (IR) is the three-months intebank rate; and DAX 30 Performance index as a proxy for consumption. Depending on data availability we collect daily or monthly data. Tests on unit roots failed to reject stationarity in all parameter series and economic variables; we therefore work with their first difference. For conciseness we present only the correlation table for non-overalpping returns over the last two years with varying equity premium and interpret the results below in relation to Figure 3.

In Table 4 we read significant positive correlation between changes in θ_1 and DAX and

[1]Series Euro Area Corporate Bond Yield are based on the German CORPTOP Bond maturing in 3-5 years. Data are sourced from the Commerzbank.

	θ_1	θ_2	θ_3	θ_4	CS	DAX	YT
θ_1	1.00	0.55*	0.02	0.78*	-0.25	0.38**	-0.26
θ_2		1.00	0.38*	-0.04	0.06	-0.12	-0.39**
θ_3			1.00	-0.18	0.07	-0.21	-0.28***
θ_4				1.00	-0.37**	0.62*	-0.04

Table 4: Correlation table for the first difference of SIM parametethers and the selected macro economic variables. (sig. at 1% = *, sig. at 5% = **, sig. at 10% = ***)

negative one with the credit spread, indicating that the EPK becomes more pronounced when the economic indicators suggest an expanding economy; changes in θ_2 and YT are negatively correlated, suggesting that risk aversion slope becomes locally steeper during economic boom. The same interpretation holds for the negative correlation between changes in θ_3 and YT. The hight of the peak varies with the returns on the index, pointing to an increasing local risk proclivity in periods of economic expansion. We have not found any significant correlation between changes in θ_t and in the short term interest rate. Finally, we observe a positive correlation between the increments in θ_1 and θ_2 that suggests that over periods of concerted negative evolution of the economic indicators the EPK bump will shrink in both horizontal and vertical direction, possibly leading to an overall decreasing EPK.

In summary the sense of the relations between the indicators of the business cycle and the parameters that summarise risk preferences indicates that locally risk loving behavior is procyclical. These findings are also in line with the results found in Rosenberg and Engle (2002).

Appendix

a. Derivation for q estimation

The Black-Scholes model assumes

$$\begin{aligned} C_{BS}(X,\tau) &= S_t e^{-\delta\tau}\Phi(d_1) - e^{-r\tau}X\Phi(d_2) \\ &= e^{-r\tau}F\{\Phi(d_1) - M\Phi(d_2)\}\,, \end{aligned}$$

where $F = S_t e^{(r-\delta)\tau}$ and $M = X/F$. With semiparametric call price function, the volatility parameter σ is expressed as a function of the option's moneyness $M = X/F$ and the time to maturity τ:

$$C(X,\tau,r,\delta,S_t) = C_{BS}(X,\tau,F,\sigma(M,\tau))\,.$$

Define a standardised all price by $c(M,\tau) = e^{r\tau}C(X,\tau,r,\delta,\sigma)/F$. Then

$$\begin{aligned}\frac{\partial C}{\partial X} &= e^{-r\tau}F\frac{\partial c}{\partial M}\frac{\partial M}{\partial X} = e^{-r\tau}\frac{\partial c}{\partial M}\\ \frac{\partial^2 C}{\partial X^2} &= e^{-r\tau}\frac{\partial c^2}{\partial M^2}\frac{\partial M}{\partial X} = e^{-r\tau}\frac{1}{F}\frac{\partial c^2}{\partial M^2}\end{aligned}$$

Recall that

$$c(M,\tau) = \Phi(d_1) - M\Phi(d_2)\,.$$

With some manipulation we have

$$\begin{aligned}\frac{\partial c}{\partial M} &= \phi(d_1)\frac{\partial d_1}{\partial M} - \Phi(d_2) - M\phi(d_2)\frac{\partial d_2}{\partial M}\\ \frac{\partial^2 c}{\partial M^2} &= -d_1\phi(d_1)\Big(\frac{\partial d_1}{\partial M}\Big)^2 + \phi(d_1)\frac{\partial^2 d_1}{\partial M^2} - \phi(d_2)\frac{\partial d_2}{\partial M} - \phi(d_2)\frac{\partial d_2}{\partial M}\\ &\quad + Md_2\phi(d_2)\Big(\frac{\partial d_2}{\partial M}\Big)^2 - M\phi(d_2)\frac{\partial^2 d_2}{\partial M^2}\,,\end{aligned}$$

where

$$\begin{aligned}\frac{\partial d_1}{\partial M} &= -\frac{1}{\sqrt{\tau}}\frac{1}{M\sigma(M,\tau)} + \frac{1}{\sqrt{\tau}}\ln(M)\frac{\sigma'(M,\tau)}{\sigma^2(M,\tau)} + \frac{\sqrt{\tau}}{2}\sigma'(M,\tau)\\ \frac{\partial d_2}{\partial M} &= \frac{\partial d_1}{\partial M} - \sqrt{\tau}\sigma'(M,\tau)\\ \frac{\partial^2 d_1}{\partial M^2} &= \frac{1}{M^2\sqrt{\tau}\sigma(M,\tau)} + \frac{2}{\sqrt{\tau}}\frac{\sigma'(M,\tau)}{\sigma^2(M,\tau)}\Big\{\frac{1}{M} - \ln(M)\frac{\sigma'(M,\tau)}{\sigma(M,\tau)}\Big\}\\ &\quad + \sigma''(M,\tau)\Big\{\frac{\ln(M)}{\sigma^2(M,\tau)\sqrt{\tau}} + \frac{\sqrt{\tau}}{2}\Big\}\\ \frac{\partial^2 d_2}{\partial M^2} &= \frac{\partial^2 d_1}{\partial M^2} - \sqrt{\tau}\sigma''(M,\tau)\,.\end{aligned}$$

Note that this leads to a slightly different derivation from Rookley (1997), albeit using the same principle. In order to compute the derivatives of σ, we used the local polynomial smoothing on implied volatility. Let σ_i to be the implied volatility corresponding to the call price C_i with maturity m_i. The local polynomial smoothing estimates are obtained by minimising

$$\sum_i \{\sigma_i - \sum_{j=0}^{p}\beta_j(m_0)(m_i - m_0)^j\}^2 K_h(m_i - m_0)\,.$$

The estimates are computed as $\hat{\sigma} = \hat{\beta}_0, \hat{\sigma}' = \hat{\beta}_1$ and $\hat{\sigma}'' = 2\hat{\beta}_2$.

References

Ait-Sahalia, Y. and Duarte, J. (2003). Nonparametric option pricing under shape restrictions. *Journal of Econometrics*, 116:9 – 47.

Ait-Sahalia, Y. and Lo, A. W. (2000). Nonparametric risk management and implied risk aversion. *Journal of Econometrics*, 94(1-2):9 – 51.

Bekaert, G. and Wu, G. (2000). Asymmetric volatility and risk in equity markets. *Review of Financial Studies*, 13(1):1–42.

Bondarenko, O. (2003). Estimation of risk-neutral densities using positive convolution approximation. *J. Econometrics*, 116(1-2):85–112. Frontiers of financial econometrics and financial engineering.

Breeden, D. T. and Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *The Journal of Business*, 51(4):621–651.

Chabi-Yo, F., Garcia, R., and Renault, E. (2007). State dependence can explain the risk aversion puzzle. pages 973–1011.

Drechsler, I. and Yaron, A. (2010). What's vol got to do with it. *Review of Financial Studies*, 20(9):1–45.

Fama, E. F. and French, K. R. (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics*, 25:23–49.

Fengler, M. R. (2005). *Semiparametric Modeling of Implied Volatility*. Springer Finance. Springer.

Giacomini, E. and Härdle, W. (2008). Dynamic semiparametric factor models in pricing kernel estimation. In Dabo-Niang, S. and Ferraty, F., editors, *Functional and Operational Statistics*, Contributions to Statistics, pages 181–187. Springer Verlag.

Giacomini, E., Härdle, W., and Handel, M. (2008). Time dependent relative risk aversion. In Bol, G., Rachev, S., and Wrth, R., editors, *Risk Assessment: Decisions in Banking and Finance*, Contributions to Economics, pages 15–46. Physica Verlag.

Härdle, W., Krätschmer, V., and Grith, M. (2009). A microeconomic explanation of the epk paradox. Technical Report 2009-010, SFB 649 Discussion Paper. Submitted to Journal of Financial Econometrics.

Härdle, W. and Marron, J. S. (1990). Semiparametric comparison of regression curves. *Ann. Statist.*, 18(1):63–89.

Hens, T. and Reichlin, C. (2010). Three solutions to the pricing kernel puzzle. Technical report, Swiss Finance Institute.

Jackwerth, J. C. (1999). Option-implied risk-neutral distributions and implied binomial trees: a literature review. *Journal of Derivatives*, 2(7):66–82.

Jackwerth, J. C. (2000). Recovering risk aversion from option prices and realized returns. *Rev. Financ. Stud.*, 13(2):433–451.

Ke, C. and Wang, Y. (2001). Semiparametric nonlinear mixed-effects models and their applications. *Journal of the American Statistical Association*, 96:1272–1298.

Kneip, A. and Engel, J. (1995). Model estimation in nonlinear regression under shape invariance. *Ann. Statist.*, 23(2):551–570.

Kneip, A. and Gasser, T. (1988). Convergence and consistency results for self-modeling nonlinear regression. *Ann. Statist.*, 16(1):82–112.

Lawton, W. H., Sylvestre, E. A., and Maggio, M. S. (1972). Self modeling nonlinear regression. *Technometrics*, 14(3):513–532.

Leland, H. E. (1980). Who should buy portfolio insurance? *Journal of Finance*, 35(2):581–94.

Li, Q. and Racine, R. S. (2007). *Nonparametric econometrics: Theory and Practice*. Princeton University Press.

Marshall, B. R., Cahan, R. H., and Cahan, J. (2008). Technical analysis around the world: Does it ever add value? Technical report, SSRN eLibrary.

Ramsay, J. O. and Silverman, B. W. (2002). *Applied functional data analysis*. Springer Series in Statistics. Springer-Verlag, New York. Methods and case studies.

Rookley, C. (1997). Fully exploiting the information content of intra day option quotes: Applications in option pricing and risk management. Technical report, University of Arizona.

Rosenberg, J. V. and Engle, R. F. (2002). Empirical pricing kernels. *Journal of Financial Economics*, 64:341–372.