Journal of Financial Econometrics, 2013, Vol. 11, No. 2, 370–399

Shape Invariant Modeling of Pricing Kernels and Risk Aversion

Maria Grith
CASE–Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin

Wolfgang Härdle
CASE–Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin

Juhyun Park
Department of Mathematics and Statistics, Lancaster University

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 characterize 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 macroeconomic variables. Implied risk aversion function and utility function can also be derived. The method is demonstrated with the European options and returns values of the German stock index DAX. (*JEL*: C14, C32, G12)

KEYWORDS: pricing kernels, risk aversion, risk neutral density

The financial support from the Deutsche Forschungsgemeinschaft via SFB 649 "Ökonomisches Risiko", Humboldt-Universität zu Berlin is gratefully acknowledged. Address correspondence to Juhyun Park, Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, UK, or email: juhyun.park@lancaster.ac.uk

doi:10.1093/jjfinec/nbs019 Advanced Access publication November 18, 2012

1 METHODOLOGY

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 maximization 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 behavior, 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 behavior becomes an increasingly important issue. The fundamental problem is that individual agents are not directly observable but it is assumed that the prices of goods traded in the market reflect the dynamics of their risk behavior. Several efforts have been made to relate the price processes of assets and options traded in a market to risk behavior of investors, since options are securities guarding against losses in risky assets.

A standard option pricing model in a complete market assumes a *risk neutral* distribution of returns, which gives the fair price under no arbitrage assumptions. 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 behavior 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)

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

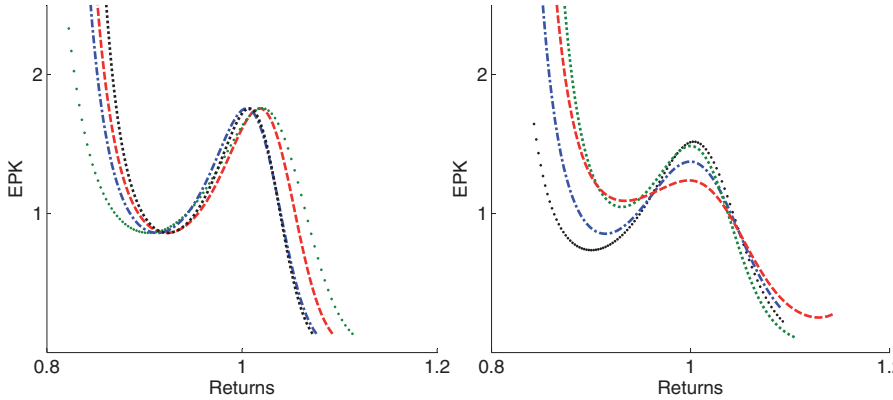

Figure 1 Examples of inter-temporal pricing kernels for various maturities in January–February 2006 (left) and monthly pricing kernels from the first six months in 2006 for maturity one month (right).

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.

1.2 Dynamics of Empirical Pricing Kernels (EPKs)

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 studies focusing on recovering risk neutral density, see e.g. Jackwerth (1999), and 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 obtained from European options data on the German stock index DAX (Deutscher Aktien index) in 2006 are shown in Figure 1, based on separate estimation of p and q. A detailed account of estimation is given in Section 3.4. To make these comparable, they are shown on a continuously compounded returns scale. Throughout the article, the pricing kernel is considered as a function of this common scale of returns. 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). 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 investment horizons. 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, properties of the pricing kernel are intrinsically related to 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 behavior. 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 to as *the pricing kernel puzzle* in the literature. There have been many attempts to reconcile the underlying economic theory with the empirical findings. A recent solution suggested by Hens and Reichlin (2012) relates the puzzle to the violation of the fundamental assumptions in the equilibrium model framework.

Most of earlier works adopt a static viewpoint, showing a snapshot 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 consecutive times and claims that these do not correspond to the basic assumptions 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, Härdle and Handel (2008) introduce time series analysis of daily summary measures of pricing kernels to examine variability between curves.

Chabi-Yo, Garcia, and Renault (2008) 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 are 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 some consistency across different 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.

Considering the pricing kernels as an object of curves, we approach the problem of estimating the pricing kernels and implied risk aversion functions from a functional data analysis viewpoint (Ramsay and Silverman, 2002). 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 amenable 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 or a reference.

Our contribution is three-fold. Firstly, we analyze the phenomenon of pricing kernel puzzle from a dynamic viewpoint using shape invariant modeling approach. The starting question was how to compare the empirical evidence. By taking into account variability among curves, we quantify a *trend of the puzzle* in the series of the pricing kernels by a few interpretable parameters. Secondly, we provide a unified framework for estimation and interpretation of ARA and utility functions consistent with the underlying pricing kernels with the same set of parameters. The ARA corresponding to the reference pricing kernel could be viewed as a typical pattern of risk behavior for the period under consideration. Due to nonlinear transformation involved in deriving ARA from the pricing kernel function, this common ARA function does not necessarily coincide with the simple average ARA functions. Thirdly, the output of the analysis provides a summary measure to study the relationship with macroeconomic variables. Through real data example we have related the changes in risk behavior to some macroeconomic variables of interest and found that local risk loving behavior is procyclical. We acknowledge that we do not provide an economic explanation to the puzzle but rather try to understand the nature of the phenomenon by means of statistical analysis.

The paper is organized as follows. Section 2 motivates common shape modeling 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. An application to real data example is summarized in Section 5.

2 COMMON SHAPE MODELING

2.1 Shape Invariant Model for Pricing Kernel

We consider a common shape modeling approach for the series of pricing kernels with explicit components of location and scale. To represent varying pricing kernels, we introduce the time index t in the pricing kernel as $\mathcal{K}_t$ and consider a general regression model:

$$Y_t = \mathcal{K}_t + \varepsilon_t,$$

where ε_t represents an error with mean 0 and variance σ_t^2. We begin with a working assumption of independent error as in Kneip and Engel (1995). The effect of

dependent error is investigated in simulation studies in Section 4.2. The relationship among K_ts is specified as

$$\mathcal{K}_t(u) = \theta_{t1} g\left(\frac{u-\theta_{t3}}{\theta_{t2}}\right) + \theta_{t4}, \tag{1}$$

with some unknown constants $\boldsymbol{\theta}_t = (\theta_{t1}, \theta_{t2}, \theta_{t3}, \theta_{t4})$ and an unknown function g. The common shape function g can be interpreted as a reference curve. Deviation from the reference curve is described by four parameters $\boldsymbol{\theta}_t = (\theta_{t1}, \theta_{t2}, \theta_{t3}, \theta_{t4})$ that capture scale changes and a shift in horizontal and vertical direction. This parametrization in (1) is commonly known as a shape invariant model (SIM), originally introduced by Lawton, Sylvestre and Maggio (1972) and studied by Kneip and Engel (1995). Note that the model includes as a special case complete parametric models with known g.

In contrast to standard applications of SIM as a regression model, the SIM application to pricing kernel estimation does not, strictly speaking, satisfy the model assumption. There is no realization of the pricing kernels available and thus our formulation of regression model should be viewed as an approximation. The original data used would be intraday options data and daily returns data, which are collected from separate sources with sample sizes of different orders of magnitude but estimation of p and q can be effectively done independently of each other. It may be 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 are specified and the unobserved pricing kernel processes enter as a state variable. However, with current advancement in the methodology, this is only possible with limited parametric model choices, see for example Chabi-Yo *et al*. (2008), and extension to a flexible shape invariant model is left for future work.

Instead we exploit the fact that preliminary estimates of pricing kernels based on separate estimation of p and q are readily available from market data and this can easily substitute Y. From now on, we treat the estimates as something observable and denote by Y_t, similar to the regression formulation with direct measurements Y_t and state the asymptotic result without further complication of pre-processing steps. After all, these estimates of curves are available from the beginning and the SIM aims to characterize a structural relationship among these curves. This however may impact the parametric rate of convergence attainable (Kneip and Engel, 1995) because our observations are already contaminated by a nonparametric error of estimation. As is shown in Section 3.6, the dominating error comes from the estimation of q, which involves second derivative estimation. The optimal rate of convergence for estimating second derivative is known to be $\mathcal{O}(N^{-2/9})$, where N is the sample size used (Stone, 1982). This implies that $\sigma_t^2 = \alpha_{N,t} v^2$ where $\alpha_{N,t}$ is a constant of order $\mathcal{O}(N^{-2/9})$, which should be understood as the multiplication factor for the parametric rate of convergence.

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

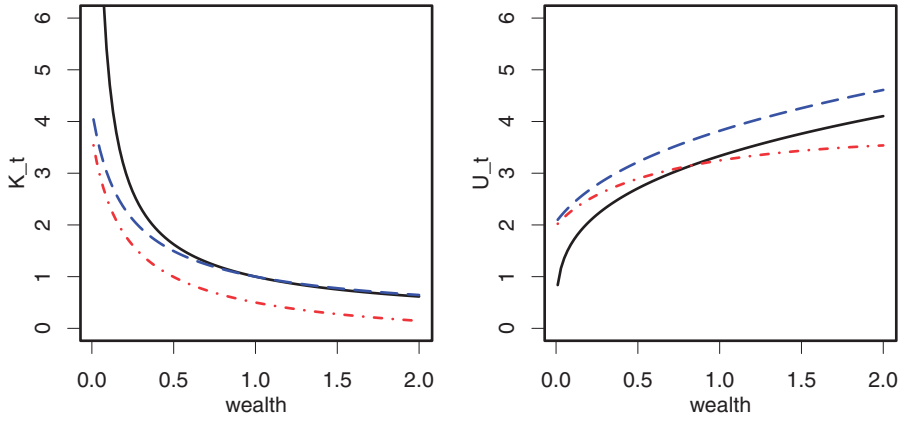

Figure 2 Example of location and scale shift family of pricing kernels (left) and corresponding utility functions (right). Solid line in each plot represents reference curves of $g(u)=u^{-\gamma}$ and $U_0(u)=u^{1-\gamma}/(1-\gamma)$ with $\gamma=0.7$, respectively. Parameters are $\theta_{t1}=1.1, \theta_{t2}=1,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

invariant model. Our choice of initial estimates will be explained in Section 3.4. Our main interest lies in quantifying the variation among the pricing kernels given those estimates.

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*. (2008). In contrast 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.

2.2 SIM and Black–Scholes Model

To appreciate the model formulation, given in the Equation (1), 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 based on a power utility function, which corresponds to risk averse behavior. Pricing kernels $\mathcal{K}_t$ are shown on the left and the corresponding utility functions U_t are on the right. The solid lines represent reference curves and 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 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 distribution for the historical density p. Under the risk neutral measure, the drift μ is replaced by the riskless rate r but the density q is still 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 g is a power function, say $u^{-\gamma}$. Then the class of pricing kernels implied by (1) is given by

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

where $\theta_{t1}^* = \theta_{t1}\theta_{t2}^{\gamma}$. 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 a unique g, 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*. These conditions are 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 normalization scheme can be devised and examples are found in Lawton, Sylvestre and Maggio (1972).

2.4 SIM Implied Risk Aversion and Utility Function

The utility function corresponding to $\mathcal{K}_t$ is given by

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

where $G(t)=\int_0^t g(u)du$. 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

$$ARA_t(u)=\frac{-\frac{\theta_{t1}}{\theta_{t2}}g'\left(\frac{u-\theta_{t3}}{\theta_{t2}}\right)}{\theta_{t1}g\left(\frac{u-\theta_{t3}}{\theta_{t2}}\right)+\theta_{t4}}. \tag{2}$$

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

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

When $\theta_{t4}=0$, this function is monotonically decreasing but in general this is not the case. Note the common ARA function corresponding to g is γu^{-1} compared to the mean ARA function computed by taking the sample average $T^{-1}\sum_{t=1}^{T}ARA_t(u)$.

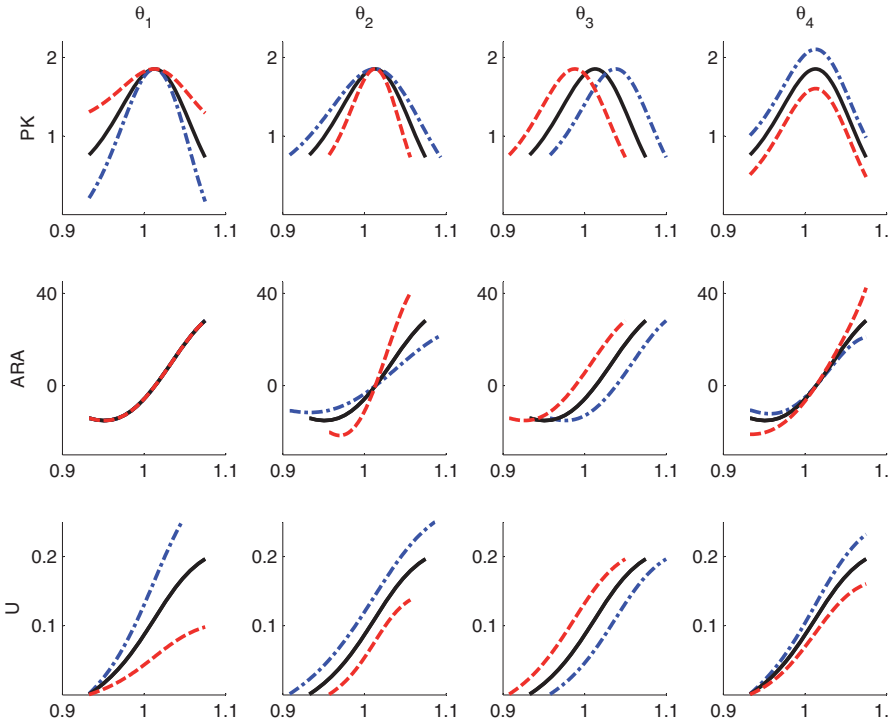

Figure 3 Effect of parameters on pricing kernel (top), ARA (middle), and utility function (bottom) compared to the baseline model $\theta_0 = (1,1,0,0)$ (black). Dot-dashed lines are used for increasing direction and dashed lines for decreasing direction.

In order to gain some insights, we take a closer look at the changes in relation to individual scale and shift parameters. These individual 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 line and $\boldsymbol{\theta} = (1.5, 1.3, 0.025, 0.25)$ in dot-dashed line.

For this exercise we first standardize 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 visualization so that they are comparable to 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 analyzed 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 normalization—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 stretches the x-axis, which implies larger spread of the bump. When we vary θ_2 alone the slope of ARA($\theta_2 u$) is $1/\theta_2^2\left[\left\{g'^2(u)-g''(u)/g(u)\right\}/g^2(u)\right]$. 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 g 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 characterize the changes in risk patterns in relation to economic variables of interest, see Section 5.4.

3 FITTING SHAPE INVARIANT MODELS

3.1 Estimation of SIM

The model in (1) is equivalently written as

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

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

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

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 normalization leads to:

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

Formula (5) was exploited 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 $\left(\theta_{t2}^{(0)},\theta_{t3}^{(0)}\right)$ for $t=1,2,\cdots,T$.
  - Construct an initial estimate $g^{(0)}$ by

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

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

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

  - Normalize 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 $g^{(r-1)}$ to

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

- Determine final estimates:

$$\tilde{\boldsymbol{\theta}}_t = \boldsymbol{\theta}_t^{(R)},$$

$$\tilde{g}(u) = T^{-1}\sum_{t=1}^{T}\hat{\mathcal{K}}_t\left(\tilde{\theta}_{t2}u+\tilde{\theta}_{t3}\right).$$

Kneip and Engel (1995) proved consistency of the estimator. In particular despite nonparametric 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 g. So the original algorithm includes the final updating of each curve. This improves precision of the estimates because the pooled sample estimate reduces the variance of $\tilde{g}$, 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 initialization step, explained in Section 3.2. 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.

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.

3.2 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

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

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

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

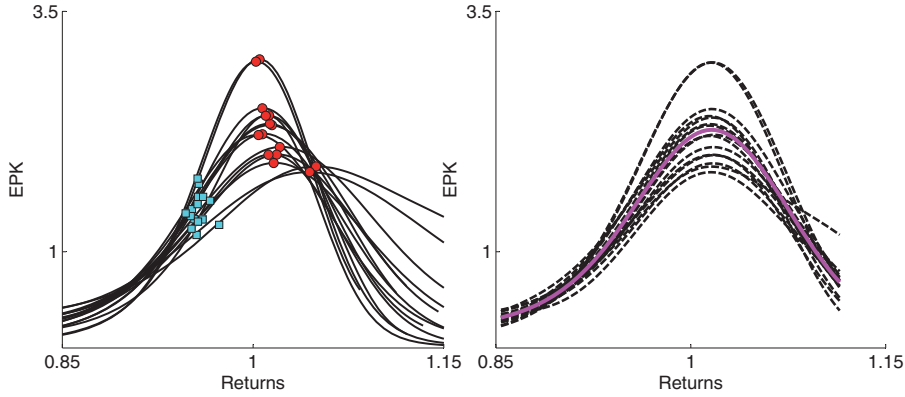

Figure 4 Initial estimates $\mathcal{K}_t(u)$ (left) and final estimates $\mathcal{K}_t(\theta_{t2}u+\theta_{t3})$ from SIM (right) with g overlayed. Marked in the left plot are two landmarks identified for estimation of the starting values of $(\theta_{t2},\theta_{t3})$.

If an inflection point is used, we would have

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

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.

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.

3.3 Nonlinear Optimization

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

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

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 $\hat{g}$, the solutions to the least square regression with response variable $\hat{\mathcal{K}}_t(\theta_{t2}u_j + \theta_{t3})$ and explanatory variable $\hat{g}(u_j)$ provide $(\theta_{t1}, \theta_{t4})$. When a further optimization routine is employed to improve the estimates, these numbers serve as initial values for $(\theta_{t1}, \theta_{t4})$.

3.4 Initial Estimates of $\mathcal{K}$

To start the algorithm the initial estimates 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 below. Individual smoothing parameter choice is discussed in Section 5 with real data example.

3.4.1 Estimation of the historical density p. We use the nonparametric kernel density estimates similar to Ait-Sahalia and Lo (2000) based on the past observations of returns for a fixed maturity τ. 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 modeling assumption is made for the evolution of the stock price such as GARCH, a simulation-based approach could be employed.

At given time t and $T = t + \tau$ we obtain realizations of future return values from a window of historical return values of length J:

$$r_T^k = \log\left(S_{t-(k-1)}/S_{t-\tau-k+1}\right) \quad \text{and} \quad S_T^k = S_t e^{r_T^k}, \quad k = 1, \ldots, J.$$

The probability density of r_T is obtained by the kernel density estimator

$$\hat{p}_{h_p}(r) = \frac{1}{Jh_p}\sum_{k=1}^{J} K\left(\frac{r_T^k - r}{h_p}\right),$$

where K is a kernel weight function and h_p is the bandwidth. Some variations are also explored such as overlapping and nonoverlapping windows with a real data example in Section 5.

3.5 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.

Define $F=S_te^{(r-\delta)\tau}$ and $m=X/F$ is moneyness. Write Φ and ϕ for the cumulative distribution function and its density of standard normal random variable, respectively. The Black–Scholes model assumes

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

In a semiparametric call price function, the volatility parameter σ is expressed as a function of the option's moneynes and the time to maturity τ:

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

To derive the second derivative of C, it is simpler to work with a standardized call price function $c(m,\tau)=e^{r\tau}C(X,\tau,r,\delta,\sigma)/F=\Phi(d_1)-m\Phi(d_2)$. The derivatives of C and c are related as

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

With some manipulation we obtain the following expressions, which are only functions of $(\sigma, \sigma', \sigma'')$:

$$\frac{\partial c}{\partial m} = \phi(d_1)\frac{\partial d_1}{\partial m} - \Phi(d_2) - m\phi(d_2)\frac{\partial d_2}{\partial m}$$

$$\frac{\partial^2 c}{\partial m^2} = -d_1\phi(d_1)\left(\frac{\partial d_1}{\partial m}\right)^2 + \phi(d_1)\frac{\partial^2 d_1}{\partial m^2} - \phi(d_2)\frac{\partial d_2}{\partial m} - \phi(d_2)\frac{\partial d_2}{\partial m} + md_2\phi(d_2)\left(\frac{\partial d_2}{\partial m}\right)^2 - m\phi(d_2)\frac{\partial^2 d_2}{\partial m^2},$$

where

$$\frac{\partial d_1}{\partial m} = -\frac{1}{\sqrt{\tau}}\frac{1}{m\sigma(m,\tau)} + \frac{1}{\sqrt{\tau}}\ln(m)\frac{\sigma'(m,\tau)}{\sigma^2(m,\tau)} + \frac{\sqrt{\tau}}{2}\sigma'(m,\tau)$$

$$\frac{\partial d_2}{\partial m} = \frac{\partial d_1}{\partial m} - \sqrt{\tau}\sigma'(m,\tau)$$

$$\frac{\partial^2 d_1}{\partial m^2} = \frac{1}{m^2\sqrt{\tau}\sigma(m,\tau)} + \frac{2}{\sqrt{\tau}}\frac{\sigma'(m,\tau)}{\sigma^2(m,\tau)}\left\{\frac{1}{m} - \ln(m)\frac{\sigma'(m,\tau)}{\sigma(m,\tau)}\right\} + \sigma''(m,\tau)\left\{\frac{\ln(m)}{\sigma^2(m,\tau)\sqrt{\tau}} + \frac{\sqrt{\tau}}{2}\right\}$$

$$\frac{\partial^2 d_2}{\partial m^2} = \frac{\partial^2 d_1}{\partial m^2} - \sqrt{\tau}\sigma''(m,\tau).$$

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 be the implied volatility corresponding to the call price C_i with moneyness m_i. The local polynomial smoothing estimates are obtained by minimizing

$$\sum_i \left\{\sigma_i - \sum_{j=0}^{3}\beta_j(m)(m_i - m)^j\right\}^2 W\left((m_i - m)/h_q\right),$$

where $W(\cdot)$ is a weight function. The estimates are computed as $\hat{\sigma}(m) = \hat{\beta}_0(m), \hat{\sigma}'(m) = \hat{\beta}_1(m)$ and $\hat{\sigma}''(m) = 2\hat{\beta}_2(m)$. Substituting the estimates to the above expressions gives an estimate of q. The density estimates are defined on the scale of S_T. To define the density on the same returns scale $r_T = \log(S_T/S_t)$ as p, a simple transformation can be applied:

$$q(r_T) = q(S_T)S_T.$$

Notice that all results are shown on a continuously compounded 1-month period returns $R_T = 1 + r_T = 1 + \log(S_T/S_t)$.

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 modeling. The shape invariant modeling 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

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

Since $\hat{q}$ involves estimation of second derivative of a regression function, the error is dominated by the estimation of q. The optimal rate of convergence for q is $\mathcal{O}(N^{-2/9})$

while that for p is $\mathcal{O}(n^{-2/5})$. These will be equivalent when $M = \mathcal{O}(n^{39/15}) > \mathcal{O}(n^2)$. In practice M is of much smaller order and therefore the leading error terms come from 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 stabilize 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) show 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 nonlinear least square methods. There is no suggested estimate for the asymptotic covariance matrix but a consistent estimate can be constructed as in standard nonlinear 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

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

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

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

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

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

	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

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 nonstandard 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 it 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 the following simulation studies. The 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 uniform) 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

$$g(u) = q_0(u)/p_0(u),$$

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 normalization scheme, the $\boldsymbol{\theta}$ values are set as in Table 1. The values of the

Table 2 Parameter values for error specification

		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

standard deviation were chosen to be similar to the observed ones in the real data example.

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

$$\varepsilon_{t,j} = \phi \varepsilon_{t,j-1} + u_{t,j}, \qquad \text{the set of the } u_{t,j} \sim N\left(0, \sigma_u^2\right)$$

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

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

- Case 4: Dependent error in moneyness and time:

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

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

Cases 1 and 2 are commonly assumed but Cases 3 and 4 were rarely used in the literature with SIM estimation. Table 2 lists the parameter values used for 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 $h_t^*=\mathrm{argmin}\mathrm{CV}_t(h)$. Due to considerable variability in the x-dimension we standardize the optimal bandwidths ($\tilde{h}_t^*=h_t^*/s_t$), where s_t is the empirical standard deviation of u_i, and we choose the common bandwidth as follows:

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

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 summarizes the results of the goodness 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 the Deutscher Aktien index (DAX), provided by European Exchange EUREX and maintained by the CASE, RDC SFB 649 (http://sfb649.wiwi.hu-berlin.de) in Berlin. 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, which adds up to 37 days.

We obtain the initial estimates for p and q as described in Section 3.4. For the choice of kernel functions, we have used quartic function for both p and q.

Table 3 Comparison of SIM estimation with respect to error misspecification and smoothing parameter selection

		$\sigma = 0.05$			
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
	$\mathcal{K}_i$s	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
	$\mathcal{K}_i$s	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
	$\mathcal{K}_i$s	1.1	1.6	1.8	1.8
	np$\mathcal{K}$	3.5	2.0	4.2	3.6

Numbers are MSE multiplied by 10000. $\mathcal{K}_i$s computes the average MSE for all curves from SIM and np$\mathcal{K}$ without SIM but using individual optimal bandwidths for each curve.

5.1 Estimation of the Risk Neutral Density *q*

The stock index 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 intraday change of the index price is stale and we use instead the prices of futures contracts closest to the time of the registered pair of the option and strike prices to derive the corresponding stock price, corrected for dividends and difference in taxation following a methodology described in Fengler (2005).

The data contains the actually traded call prices, the implied stock index price corrected for the dividends from the futures derivatives on the DAX, the strike prices, the interest rates (linearly interpolated based on EURIBOR to approximate a *riskless* interest rate for the specific option's time to maturity), the maturity, the type of the options, calculated moneyness, calculated Black and Scholes implied volatility, the volume, and the date. 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 the intraday stock price varies, we use its median (S_t) to compute the risk neutral density and correct the strike

price to preserve the ratio relative to the underlying stock price. For this price, we verify if our observations satisfy the no arbitrage condition:

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

where X_i is the adjusted strike price and C_i is the corresponding call price. For the remaining observations (X_i, C_i) we compute the (m_i, σ_i) counterparts for the fixed S_t by implicitly assuming that the volatility does not depend on the changes in the intraday stock price. The estimates are computed based on these pairs (m_i, σ_i).

5.2 Estimation of the Historical Density *p*

We compute the nonparametric kernel density estimates as described earlier. Jackwerth (2000) argues that some discrepancies between the nonparametric estimates are attributed to overlapping and nonoverlapping windows of the past observations selected. 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 lengths of 2, 4, and 6 years, respectively. The estimates for different choices of parameters are then compared subsequently in terms of pricing kernel, implied risk aversion and implied utility function estimation. We find that with varying degrees of assumptions on the model, common characteristics such as peaks and skewness are reportedly observed in a wide range of estimates.

5.3 Smoothing Parameter Selection

In contrast 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.

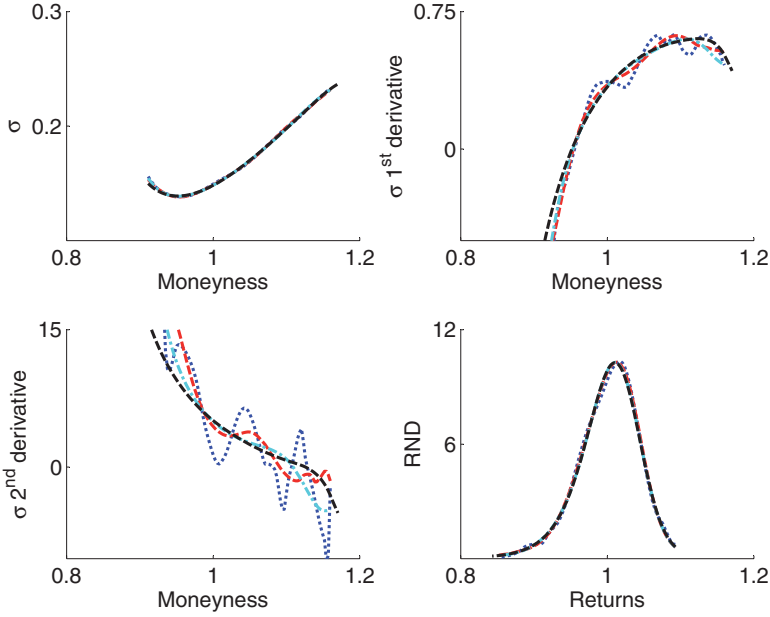

Figure 5 Example of q estimates with varying bandwidths (0.05, 0.1, 0.15, 0.20). The first three panels show estimates of implied volatility, its first and second derivative. The corresponding densities are shown in lower right panel. Estimates are stable for a wide range of bandwidths choices.

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 for local cubic estimation to include the second derivative of σ:

$$\mathrm{CV}(h_q) = \sum_{i=1}^{n}\sum_{j\neq i}^{n}\left\{\sigma_i - \widehat{\sigma}^{(0)}_{h_q,-i}(m_i) - \widehat{\sigma}^{(1)}_{h_q,-i}(m_i)(m_j - m_i) - \frac{1}{2}\widehat{\sigma}^{(2)}_{h_q,-i}(m_i)(m_j - m_i)^2\right\}^2 w(m_i),$$

where $\widehat{\sigma}^{(k)}_{h_q,-i}$ is the k-th derivative estimate without the i-th observation (m_i, σ_i) and $0 \leq w(m_i) \leq 1$ is a weight function. The h_1-optimal bandwidth in implied volatility space turns out to be $h_q = 0.2$.

For estimation of p, we have used the likelihood CV for each curve on returns scale:

$$\log L(h_p) = \sum_{i=1}^{n} \log \hat{p}^{-(i)}_{h_p}(r_i),$$

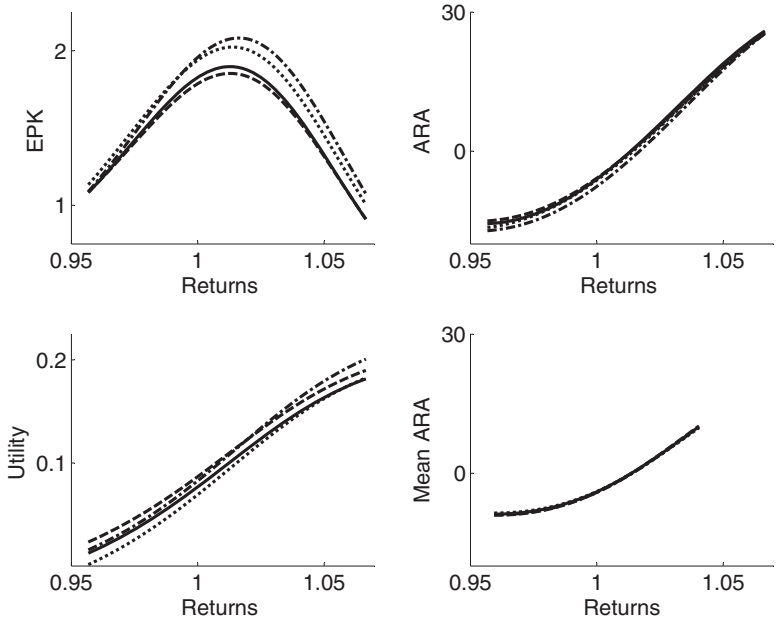

Figure 6 Illustration of SIM with common EPK, ARA, utility function, and mean ARA.

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

5.4 Estimation of Pricing Kernels, ARA and Utility Function

We have considered in Section 5.2 various options for the parameter choice in estimating p and have ended up with 12 series of pre-estimates of pricing kernel. 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) 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 confidence band for the first series of pre-estimates.

The common curves are represented in Figure 6. All estimates display a *paradoxical* feature: pricing kernel has a bump, ARA has a region of negative values that correspond to the increasing region in the pricing kernel, utility function has a convex region in the domain around the peak of the pricing kernel. The variability

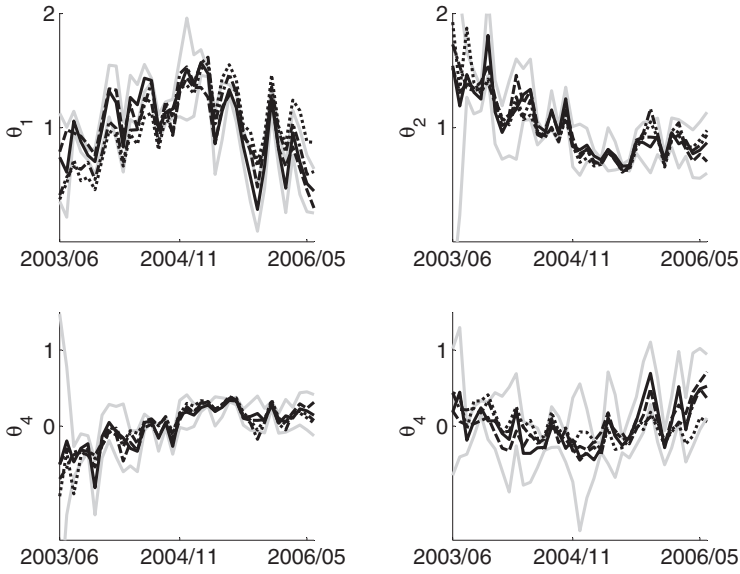

Figure 7 Estimated SIM parameters under variations in the choice of the window lengths of returns values.

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 by computing the sample average and found that it was similar to the the mean ARA for S&P500 appearing in Figure 3C—19 March 1991 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.* (2008). It is worth noting that the mean ARA and the common ARA curves differ a great deal due to the nonlinear transformation involved in deriving ARA from the pricing kernel, e.g. see Equation (2) in Section 2.4. 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 Macroeconomic Variables

With an aid of the SIM model for EPK, we wish to characterize 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 standardize the EPK curves by the location of the peak before applying SIM.

Table 4 Correlation table for the first difference of SIM parametethers and the selected macro economic variables

	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

*, **, and *** significant at 1%, 5%, and 10% levels, respectively.

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 separately.

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 (Bekaert and Wu, 2000; Drechsler and Yaron, 2010), or market efficiency (Marshall, Cahan, and Cahan, 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 macroeconomic 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, based on the German CORPTOP Bond maturing in 3–5 years, and the government bond maturing in 5 years; the yield curve slope (YT) refers to the difference between the 30-year government bond yield and three-months interbank rate; short-term interest rate (IR) is the three-months interbank 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 nonoverlapping returns over the past 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 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 height 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 summarize 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).

REFERENCES

Ait-Sahalia, Y., and J. Duarte. 2003. Nonparametric Option Pricing under Shape Restrictions. *Journal of Econometrics* 16: 9–47.

Ait-Sahalia, Y., and A. Lo. 2000. Nonparametric Risk Management and Implied Risk Aversion. *Journal of Econometrics* 94: 9–51.

Bekaert, G., and G. Wu. 2000. Asymmetric Volatility and Risk in Equity Markets. *Review of Financial Studies* 13: 1–42.

Bondarenko, O. 2003. Estimation of Risk-Neutral Densities using Positive Convolution Approximation. *Journal of Econometrics* 116: 85–112.

Breeden, D., and R. Litzenberger. 1978. Prices of State-Contingent Claims Implicit in Option Prices. *The Journal of Business* 51: 621–651.

Chabi-Yo, Y., R. Garcia, and E. Renault. 2008. State Dependence can Explain the Risk Aversion Puzzle. *Review of Financial Studies* 21: 973–1011.

Drechsler, I., and A. Yaron. 2010. What's Vol Got to Do with it. *Review of Financial Studies* 20: 1–45.

Fama, E. F., and K. R. French. 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-Verlag Berlin and Heidelberg.

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

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

Härdle, W., and J. S. Marron. 1990. Semiparametric Comparison of Regression Curves. *The Annals of Statistics* 18: 63–89.

Hens, T., and C. Reichlin. 2012. "Three Solutions to the Pricing Kernel Puzzle." Technical report, Swiss Finance Institute Research Paper No. 10-14. Available at SSRN: http://ssrn.com/abstract=1582888 or http://dx.doi.org/10.2139/ssrn.1582888

Jackwerth, J. C. 1999. Option-Implied Risk-Neutral Distributions and Implied Binomial Trees: a Literature Review. *Journal of Derivatives* 2: 66–82.

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

Ke, C., and Y. Wang. 2001. Semiparametric Nonlinear Mixed-Effects Models and their Applications. *Journal of the American Statistical Association* 96: 1272–1298.

Kneip, A., and J. Engel. 1995. Model Estimation in Nonlinear Regression under Shape Invariance. *The Annals of Statistics* 23: 551–570.

Kneip, A., and T. Gasser. 1988. Convergence and Consistency Results for Self-Modeling Nonlinear Regression. *The Annals of Statistics* 16: 82–112.

Lawton, W. H., E. A. Sylvestre, and M. S. Maggio. 1972. Self Modeling Nonlinear Regression. *Technometrics* 14: 513–532.

Leland, H. E. 1980. Who Should Buy Portfolio Insurance? *Journal of Finance* 35: 581–594.

Marshall, B. R., R. H. Cahan, and J. Cahan. 2008. "Technical Analysis Around the World: Does it Ever Add Value?" Technical report, SSRN eLibrary.

Ramsay, J. O., and B. W. Silverman. 2002. *Applied Functional Data Analysis*. Springer Series in Statistics. New York: Springer. 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 R. F. Engle. 2002. Empirical Pricing Kernels. *Journal of Financial Economics* 64: 341–372.

Stone, C. J. 1982. Optimal Global Rates of Convergence for Nonparametric Regression. *The Annals of Statistics* 10: 1040–1053.