Efficient Estimation of Pricing Kernels and Market-Implied Densities

Jeroen Dalderop*

May 25, 2021

Abstract

This paper studies the nonparametric identification and estimation of projected pricing kernels implicit in European option prices and underlying asset returns using conditional moment restrictions. The proposed series estimator avoids computing ratios of estimated risk-neutral and physical densities. Instead, we consider efficient estimation based on the conditional Euclidean empirical likelihood or continuously-updated GMM criterion, which takes into account the informativeness of option prices of varying strike prices beyond observed conditioning variables. In a second step, we convert the implied probabilities into predictive densities by matching the informative part of cross-sections of option prices. Empirically, pricing kernels tend to be U-shaped in the S&P 500 index return given high levels of the VIX, and call and ATM options are more informative about their payoff than put and OTM options.

Keywords: Option Prices, Risk Aversion, Density Forecasting, Empirical Likelihood
JEL Codes: C14, G13

1 Introduction

In equilibrium asset pricing models, the risk premium on any traded asset return is determined by its covariance with the stochastic discount factor (Cochrane, 2009). When derivative products are traded whose payoffs span the outcomes of an asset return, the ratio of the latter's 'risk-neutral' and physical density identifies the conditional expectation of the stochastic discount factor given the return. This 'projected' pricing kernel reveals how investor's relative risk aversion varies with the return, provides testable restrictions on utility functions in consumption-based models, and allows computing bias-corrected risk-neutral densities for forecasting and risk management applications.

*Department of Economics, University of Notre Dame, 3060 Jenkins-Nanovic Halls, Notre Dame, IN 46556, USA. Email address: jdalderop@nd.edu. I am grateful to Yacine Aït-Sahalia, Oliver Linton, Richard Smith, and seminar participants at Cambridge, Notre Dame, and Wisconsin Business School for helpful comments.

The pricing kernel's relation to preferences motivates the use of nonparametric estimation methods that avoid parametric functional form restrictions. Nonparametric pricing kernel estimates are typically computed as the ratio of estimated risk-neutral and physical densities (Aït-Sahalia and Lo, 2000; Härdle et al., 2014; Beare and Schmidt, 2014). The former are identified from the cross-section of European option prices at varying strike prices, while the latter are based on the historical distribution of returns. Alternatively, only the physical density is estimated directly, and the pricing kernel is chosen to fit a cross-section of option prices (Rosenberg and Engle, 2000; Audrino and Meier, 2012; Grith et al., 2012). Critically, either approach assumes the estimated physical density reflects the information set of forward-looking investors. Any omitted state variables will erroneously be attributed to the pricing kernel. Indeed, Linn et al. (2017) suggest that the empirical puzzle of non-monotonically decreasing pricing kernels results from the restricted information set in computing the physical density, rather than a violation of standard preference assumptions. Instead, Linn et al. (2017) and Cuesdeanu and Jackwerth (2018a) propose estimating the pricing kernel nonparametrically based on probability integral transforms and log scores, respectively, of the adjusted risk-neutral densities, following Bliss and Panigirtzoglou (2004) who used parametric utility-based adjustments. However, methods based on predictive ability require the pricing kernel to be a function of the return and observed conditioning variables only. If the pricing kernel depends on omitted state variables, we show that its maximum likelihood estimator does not necessarily identify the ratio of the risk-neutral and physical densities.

To avoid any informational mismatch between the physical and risk-neutral densities, this paper studies the nonparametric estimation of the pricing kernel from conditional moment restrictions on the discounted gains or losses on options with varying strike prices. By conditioning the pricing kernel, option prices, and their payoffs on the same set of observed variables, these moment restrictions hold regardless of whether unexplained variation in option prices reflects predictive information on the underlying return or omitted variables in the pricing kernel. The conditional moments can be consistently estimated by local sample averages, and do not require estimating either the risk-neutral or physical distribution directly. Moreover, our method allows for finite and unbalanced cross-sections of strike prices, as long as the number of time periods grows to infinity. Finally, pricing restrictions on the risk-free rate and underlying return itself can be naturally incorporated by augmenting the vector of moments.

The estimation approximates the pricing kernel by a series expansion, and combines the conditional moment restrictions for option prices into a sieve minimum-distance criterion (Ai and Chen, 2003). Its minimizer yields a consistent estimator when the number of time periods and number of basis functions increase appropriately. The series estimator can avoid some of the instability of density ratio estimates by bounding its variation under a suitable norm. Moreover, shape constraints such as non-negativity and monotonicity can be directly imposed. However, as there are few liquid option prices that are far out-of-the-money, the tails of the pricing kernel may be poorly identified. After observing that option prices identify the first two partial moments of the risk-neutral distribution outside their liquidly traded range of strike prices, we 'paste' parametric tails to the pricing kernel in order to stabilize the estimates.

To achieve asymptotic efficiency, we weigh the conditional moments inversely proportional to the sample variance of discounted option gains and losses. These depend not only on the heteroskedasticity of option pricing errors and option payoffs for varying strike prices, but also on their contemporaneous dependence due to information available to investors beyond the conditioning variables. In a simulation study, we show that both two-stage and continuously-updated variants of the efficiently weighted estimator outperform equal weighting for reasonable sample sizes. The continuously-updated variant is motivated by the conditional Euclidian empirical likelihood problem studied in Antoine et al. (2007), which yields implied probabilities for the joint distribution of option prices and subsequent returns as well as Lagrange multipliers that are interpreted as local measures of mispricing.

In a second step, we extract market-based predictive densities by minimizing the Kullback-Leibler divergence to the estimated physical density, subject to matching the informative part of contemporaneous option prices. This approach encompasses the extreme cases where either all or none of the variation in option prices is due to variation in the physical density, conditional on the observed variables. Compared to the XMM estimator by Gagliardini et al. (2011), which combines current option prices with historical stock returns, we observe both series over the entire sampling period in order to measure option price informativeness. Specifically, we define an informativeness parameter as the weight assigned to the bias-corrected risk-neutral density in a linear combination with the historical density when maximizing the predictive likelihood. Alternatively, we consider an informativeness function for varying strike price based on local regressions of discounted option payoffs on their prices, which can be efficiently estimated using the first-stage implied probabilities.

Empirically, we find pricing kernels that are generally decreasing but hump-shaped in the S&P 500 index return conditional on low levels of the VIX, and U-shaped pricing kernels when the VIX is moderately high. These non-monotonic shapes are found in many empirical studies, as surveyed by Cuesdeanu and Jackwerth (2018b), yet are obtained without any distributional assumptions on option prices and underlying returns, avoiding their informational mismatch, and using efficient weighting. Moreover, we find considerable heterogeneity in informativeness across moneyness levels, as call and ATM options are more informative about their payoff than put and OTM options, respectively, after conditioning on the VIX.

The remainder of this paper is organized as follows. Section 2 introduces the projected pricing kernel, including its identification from conditional moment restrictions, consistent and efficient sieve minimum-distance estimation, simulation performance, and empirical results. Section 3 studies the relative entropy problem that yields implied conditional densities which can be used for forecasting. Section 4 concludes.

2 Nonparametric pricing kernel estimation

Standard no-arbitrage theory implies the existence of a stochastic discount factor $M_{t,t+1}$, such that the price of any traded asset equals the conditional expectation of its discounted payoff X_{t+1}:

$$P_t = E_t\left(M_{t,t+1}X_{t+1}\right). \tag{1}$$

Equilibrium-based asset pricing models typically specify $M_{t,t+1}$ as a fixed discount factor multiplied with the marginal rate of substitution over consumption of a representative agent. Since traditional time-separable utility specifications struggle to match observed equity risk premia, numerous extensions have been studied, such as habit formation and recursive preferences; for an overview see Ludvigson (2011). Furthermore, while linear stochastic discount factor models are commonly used for explaining cross-sectional returns, concerns about functional form misspecification have led Bansal and Viswanathan (1993), Chapman (1997), and others to investigate more general nonlinear functions of priced risk factors.

Due to their payoff structure, option prices are particularly useful for recovering any nonlinear dependence of the stochastic discount factor on the underlying asset return. Moreover, the dynamics of option prices are informative about how this shape varies over time. In this paper, we therefore project the stochastic discount factor on both the asset return R_{t+1} as well as some observed conditioning variables Z_t that may affect its distribution:

$$m(R_{t+1}, Z_t) := E(M_{t,t+1} \mid R_{t+1}, Z_t). \tag{2}$$

This projected pricing kernel has the same pricing properties as its original for payoffs that depend on R_{t+1}, given Z_t, and can be determined model-free from option prices and returns. As such, in the absence of agreement on the specification of the stochastic discount factor $M_{t,t+1}$, the projected pricing kernel provides a conditional moment that any valid specification should satisfy.

Specifically, the projected pricing kernel is identified as the ratio of the conditional densities of the 'risk-neutral' distribution $\mathbb{Q}$ and the physical distribution $\mathbb{P}$:

$$\frac{d\mathbb{Q}(R_{t+1} \mid Z_t)}{d\mathbb{P}(R_{t+1} \mid Z_t)} = m(R_{t+1}, Z_t). \tag{3}$$

The risk-neutral density can be recovered from option prices with varying strike price (Breeden and Litzenberger, 1978). Equation (3) therefore suggests estimating the projected pricing kernel as the ratio of estimated risk-neutral and physical densities conditioned on Z_t. This method is challenged by small values in the tails of the estimated physical density, which may lead the ratio estimate to behave erratically and require wide confidence bands (Härdle et al., 2014). Moreover, ratio estimates may suffer from small-sample biases, which in itself might explain non-monotonic empirical pricing kernels (Leisen, 2017). In the following we avoid separate estimation of the densities, and instead formulate conditional moment restrictions which identify the projected pricing kernel and can be consistently estimated using local sample averages.

2.1 Conditional moment restrictions

For a traded futures price F_t with return $R_{t+1} = \frac{F_{t+1}}{F_t}$ and a one-period bond with risk-free return R_t^f, the general pricing equation (1) implies:

$$1 = E_t \left(M_{t,t+1} R_{t+1} \right) \tag{4}$$

$$1 = E_t \left(M_{t,t+1} R_t^f \right). \tag{5}$$

A useful assumption is that the projected pricing kernel is separable in the risk free return R_t^f. In this case, in terms of the risk free rate $r_t^f = \log R_t^f$, we write

$$m(R_{t+1}, R_t^f, Z_t) = e^{-r_t^f} \mu(R_{t+1}, Z_t), \tag{6}$$

After conditioning down on Z_t, the stock and bond pricing equations (4) and (5) yield

$$0 = E(\mu(R_{t+1}, Z_t) R_{t+1} - R_t^f \mid Z_t)$$
$$0 = E(\mu(R_{t+1}, Z_t) - 1 \mid Z_t).$$

Similarly, the no-arbitrage price of a European call option with strike price K and expiration date $t+1$ is given by the conditional expectation

$$C_t = E_t \left(M_{t,t+1} (F_{t+1} - K)^+ \right)$$
$$\Leftrightarrow \tilde{C}_t = E_t \left(e^{r_t^f} M_{t,t+1} (R_{t+1} - \kappa_t)^+ \right), \tag{7}$$

where $\tilde{C}_t \equiv e^{r_t^f} \frac{C_t}{F_t}$ is the normalized option price, and $\kappa_t \equiv \frac{K}{F_t}$ is the 'moneyness' of the option. In most models, the price level F_t does not affect the distribution of the return R_{t+1}, so that the option price is homogeneous in K and F_t and the normalized option prices are stationary.

Conditioning the expected discounted payoff in (7) on Z_t introduces an option pricing error:

$$\tilde{C}_t = E\left(\mu(R_{t+1}, Z_t)(R_{t+1} - \kappa_t)^+ \mid Z_t\right) + \varepsilon_t, \qquad E(\varepsilon_t \mid Z_t) = 0 \tag{8}$$

The pricing error ε_t would disappear if the true model is Markovian in Z_t. Otherwise, it captures omitted state variables induced by pooling option prices from different time periods under a restricted information set. Moreover, the error may account for observational error in option prices, typically due to market microstructural noise or the use of interpolation methods.

The pricing restrictions can be combined into the conditional moment vector

$$\rho(Z_t, \kappa_t, \mu) := E\left(\mu(R_{t+1}, Z_t)h(R_{t+1}, \kappa_t) - P_t \mid Z_t, \kappa_t\right) = 0, \tag{9}$$

in terms of the payoff vector and price vector, respectively, given by

$$h(R_{t+1}, \kappa_t) = \begin{pmatrix} (R_{t+1} - \kappa_t)^+ \\ R_{t+1} \\ 1 \end{pmatrix}, \qquad P_t = \begin{pmatrix} \tilde{C}_t \\ R_t^f \\ 1 \end{pmatrix}. \tag{10}$$

where $\tilde{C}_t = (\tilde{C}_{t1}, ..., \tilde{C}_{tn_t})^T$ is the vector of n_t normalized call option prices traded at time t with normalized strike prices $\kappa_t = (\kappa_{t1}, ..., \kappa_{tn_t})^T$.[1]

The conditional moment restrictions (9) form the basis for identification and estimation of the projected pricing kernel. The restrictions hold for any set of observed conditioning variables $Z_t \in \mathcal{F}_t$. Each choice of Z_t identifies a projected pricing kernel that can be compared against that implied by a stochastic discount factor model. However, the more variation in option prices is captured by Z_t, the smaller the pricing errors ε_t, and the more accurate the projected pricing kernel can be estimated. Typical state variables would be volatility proxies, lagged returns, or proxies related to the business cycle. Alternatively, Z_t could include a time argument to describe long-term trends in risk aversion.

2.2 Identification

The identification of the projected pricing kernel $\mu(R_{t+1}, Z_t)$ relies on observing the return, option prices, and conditioning variables, over many time periods. The number of options with different moneyness levels does not need to become large each period.[2] Instead, only the mean option price given any moneyness level κ and conditioning value z is needed for identification.

In practice, there are few liquid option prices whose strike prices are in the tails of the return

[1] For notational simplicity, we assume put option prices are converted into call option prices by put-call parity.

[2] Large cross-sections of option prices would be required for the consistent estimation of the risk-neutral density (e.g. Lu and Qu (2021)), or the dynamic recovery of latent states (Andersen et al., 2015).

distribution. This makes it hard to identify the tails of the projected pricing kernel, which could explain the different shapes found across empirical studies (e.g. Rosenberg and Engle, 2000; Beare and Schmidt, 2014).

The stability of the estimators can be increased by adopting a semiparametric approach that imposes some functional form in regions which are poorly identified. Our identification strategy treats the normalized strike prices κ_t as a continuous but truncated random variable, and assumes a parametric functional form for the tails of the projected pricing kernel. The randomness of κ_t implies that the projected pricing kernel is point identified in the actively traded range. The parametric functional form implies that the tail risk parameters are point identified by adding the pricing equations for stock and risk free returns. These parameters can still be unspecified functions of the exogenous variables Z_t.

Suppose that the moneyness κ_t of traded option prices is a continuous random variable, bounded in the range (κ^l, κ^u). Although each trading period only a fixed set of strike prices is traded, their normalization by the stock price leads to a continuous range around one. The conditional moments in (9) associated with call option prices are

$$\rho^C(Z_t, \kappa_t, \mu) := E\left(\mu(R_{t+1}, Z_t)(R_{t+1} - \kappa_t)^+ - \tilde{C}_t \mid Z_t, \kappa_t\right) = 0, \tag{11}$$

Provided R_{t+1} is conditionally independent of κ_t given Z_t, the normalized strike prices only affects the moment via the known payoff function.[3] Hence

$$E\left(\mu(R_{t+1}, Z_t)(R_{t+1} - \kappa_t)^+ | Z_t = z, \kappa_t = \kappa\right) = \int_\kappa^\infty \mu(y, z)(y - \kappa) f(y|z) dy,$$

in terms of the conditional density $f_{R_{t+1}|Z_t}(y|z)$. Differentiating twice with respect to κ yields

$$\frac{\partial^2}{\partial^2 \kappa} \int_\kappa^\infty \mu(y, z)(y - \kappa) f(y|z) dy = \mu(\kappa, z) f(\kappa|z), \qquad \kappa \in [\kappa^l, \kappa^u],\ z \in \mathcal{Z},$$

so that (11) identifies the projected pricing kernel over the optionable range as

$$\mu_0(\kappa, z) = \frac{\partial^2 E\left(\tilde{C}_t \mid Z_t = z, \kappa_t = \kappa\right) / \partial^2 \kappa}{f(\kappa|z)}, \qquad \kappa \in [\kappa^l, \kappa^u],\ z \in \mathcal{Z}. \tag{12}$$

The projected pricing kernel in the lower and upper tails of the return is partially identified from the stock, the risk-free return, put options at the lowest optionable moneyness κ^l with payoff $(\kappa^l - R_{t+1})^+$, and call options at the highest optionable moneyness κ^u:

Proposition 1. *Suppose that $\mu_0(y, z)$ is point identified for $(y, z) \in [\kappa^l, \kappa^u] \times \mathcal{Z}$, where $\kappa^l < \kappa^u$,*

[3] In an efficient and liquid market the strike prices that are traded are unlikely to have any predictive power for the market return, as asset prices would adjust immediately.

that the conditional expectations of put options $E(P_t \mid z, \kappa^l)$ and call options $E(C_t \mid z, \kappa^u)$ are known for all $z \in \mathcal{Z}$, and the conditional moment restrictions for the stock (4) and risk free rate (5) hold. Then the following four partial conditional expectations are identified:

$$\int_0^{\kappa^l} \mu(y,z) f(y|z) dy, \quad \int_0^{\kappa^l} \mu(y,z) y f(y|z) dy, \quad \int_{\kappa^u}^{\infty} \mu(y,z) f(y|z) dy, \quad \int_{\kappa^u}^{\infty} \mu(y,z) y f(y|z) dy.$$

The four quantities provide Type I (partial) integral restrictions for $\mu(y,z)$ of the type common in the literature on nonparametric IV estimation (e.g. Newey and Powell, 2003; Darolles et al., 2011). Without the argument z in the pricing kernel, a sufficient condition for identification is bounded completeness of the conditional density $f(y|z)$ (Blundell et al., 2007). However, the presence of z prevents full nonparametric identification of $\mu(y,z)$ in the tail regions, in the absence of at least one conditioning variable $Z_{1t} \subsetneq Z_t$ that does not affect the pricing kernel.[4] Unfortunately, variables such as volatility factors, lagged returns, or business cycle proxies that affect the conditional return distribution, are likely to influence attitudes towards risk taking and thus the pricing kernel as well.

Instead, by assuming a parametric form of the pricing kernel in the tails, the four identified partial moments can be used to identify up to four parameters that can be functions of z. For example, we could specify the pricing kernel tails as exponentially-affine:

$$\mu(R_{t+1}, Z_t) = \begin{cases} e^{-\beta_0^l(Z_t) - \beta_1^l(Z_t) R_{t+1}}, & 0 \leq R_{t+1} < \kappa^l \\ \alpha(R_{t+1}, Z_t), & \kappa^l \leq R_{t+1} \leq \kappa^u \\ e^{-\beta_0^u(Z_t) - \beta_1^u(Z_t) R_{t+1}}, & \kappa^u < R_{t+1} < \infty \end{cases} \tag{13}$$

The middle part α is nonparametrically identified by the option prices. For given $Z_t = z$, the two parameters β_0 and β_1 for each tail are identified by the partial moments from Proposition 1. To avoid kinks in the pricing kernel, we restrict $(\beta_0, \beta_1, \alpha)$ to satisfy smooth pasting conditions which match the levels and slopes on both sides of the boundary points κ^l and κ^u.

Parametric models for the tails of the pricing kernel automatically control their asymptotic behavior when R goes to either 0 or ∞. The nonparametric function α in the middle has a bounded domain and can be reasonable assumed to lie in a compact set. This overcomes the ill-posed inverse problem and simplifies the asymptotic theory, as the solution becomes a continuous mapping of the distribution of the data (Newey and Powell, 2003). Alternatively, this problem can be avoided by writing the pricing equation as a Type II integral equation in the unknown marginal utility function (Escanciano et al., 2015) or discrete state marginal utilities (Ross, 2015).

[4] Chen and Ludvigson (2009) use instrumental variables to identify the consumption habit function under the assumed completeness of an expected return weighted-conditional density.

2.3 Estimation

Sample analogues of the conditional moments (9) can be used to estimate the projected pricing kernel using a minimum-distance criterion. Asymptotic theory for estimating an unknown function identified by conditional moments has been developed in Newey and Powell (2003), Ai and Chen (2003), and others, and is adapted here to the option pricing setting.

Define the infinite-dimensional parameter vector $\theta = (\beta_0, \beta_1, \alpha) \in \Theta_\beta^4 \times \Theta_\alpha \equiv \Theta$, where Θ_β and Θ_α are compact spaces. The previous section implies that there is a unique $\theta_0 \in \Theta$ which minimizes the population criterion

$$Q(\theta) = E\left(\rho(Z_t, \kappa_t, \theta)^T \Sigma_t^{-1} \rho(Z_t, \kappa_t, \theta)\right), \tag{14}$$

for any series of positive definite weighting matrix Σ_t. Now define the theoretical pricing function

$$P(Z_t, \kappa_t, \theta) = E\left(\mu(R_{t+1}, Z_t, \theta) h(R_{t+1}, \kappa_t) \mid Z_t\right). \tag{15}$$

so that $\rho(Z_t, \kappa_t, \theta) = P(Z_t, \kappa_t, \theta) - E(P_t \mid Z_t, \kappa_t)$. A closely related objective is

$$Q'(\theta) = E\left((P(Z_t, \kappa_t, \theta) - P_t)^T \Sigma_t^{-1} (P(Z_t, \kappa_t, \theta) - P_t)\right), \tag{16}$$

which minimizes the fit between theoretical and observed prices according to a weighted nonlinear least-squares criterion. Since the difference

$$Q'(\theta) - Q(\theta) = E\left(\varepsilon_t^T \Sigma_t^{-1} \varepsilon_t\right) \tag{17}$$

with $\varepsilon_t = P_t - E(P_t | Z_t, \kappa_t)$ does not depend on the parameter vector θ, $Q'(\theta)$ is also uniquely minimized at $\theta_0 \in \Theta$. The sample analogue of $Q(\theta)$ based on first stage nonparametric estimators $\hat{\rho}(Z_t, \kappa_t, \theta)$ and $\hat{\Sigma}_t$, yields the sample criterion

$$Q_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} \hat{\rho}(Z_t, \kappa_t, \theta)^T \hat{\Sigma}_t \hat{\rho}(Z_t, \kappa_t, \theta). \tag{18}$$

The sample analogue of $Q'(\theta)$ requires a first stage nonparametric estimator $\hat{P}(Z_t, \kappa_t, \theta)$ of the pricing function:

$$Q'_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} (\hat{P}(Z_t, \kappa_t, \theta) - P_t)^T \hat{\Sigma}_t (\hat{P}(Z_t, \kappa_t, \theta) - P_t). \tag{19}$$

This criterion measures the discrepancy between observed prices P_t and estimated theoretical prices $\hat{P}$ over the periods in the sample.

The minimizers of the sample analogues $Q_T(\theta)$ and $Q'_T(\theta)$ are numerically identical when the first-stage estimators for $\hat{\rho}(Z_t, \kappa_t, \theta)$ and $\hat{P}(Z_t, \kappa_t, \theta)$ are series estimators and θ appears linearly in ρ. This follows from the idempotency of the projection matrix of basis functions of (Z_t, κ_t) (Newey and Powell, 2003). In this case $Q'_T(\theta)$ is a two-stage least-squares objective. When the first-stage estimators are based on kernel smoothing, as defined below, the difference between minimizers of the two criteria vanishes in large samples for shrinking bandwidths. Conditions for consistency of the minimizer of $Q'_T(\theta)$ can be stated in terms of the pricing errors alone instead of the discounted option gains and losses. Since the latter are kinked functions of the market return, estimation based on $Q_T(\theta)$ may require additional regularity conditions. Another advantage of the pricing error-based criterion (19) is that the last observed price P_T can still be included, despite its return R_{T+1} not being observed. The first-stage pricing function estimator can simply leave out the last period.

The benefits of distinguishing (Z_t, κ_t) become apparent when estimating the pricing function $\hat{P}(Z_t, \kappa_t, \theta)$. Provided R_{t+1} is conditionally independent of κ_t given Z_t, it follows that

$$P(z, \kappa, \theta) = \int_0^\infty \mu(y, z, \theta) h(y, \kappa) f(y|z) dy, \tag{20}$$

where the conditional density $f(R_{t+1}|Z_t)$ does not depend on the option moneyness κ_t. This suggests the plug-in estimator

$$\hat{P}(z, \kappa, \theta) = \int_0^\infty \mu(y, z, \theta) h(y, \kappa) \hat{f}(y|z) dy, \tag{21}$$

where $\hat{f}(y|z)$ is a nonparametric conditional density estimator, such as the kernel-based density

$$\hat{f}_h(y|z) = \frac{\sum_{t=1}^{T-1} K_{h_y}(R_{t+1} - y) K_{h_z}(Z_t - z)}{\sum_{t=1}^{T-1} K_{h_z}(Z_t - z)} \tag{22}$$

where $K_h(\cdot) = \frac{1}{h} K(\frac{\cdot}{h})$ for some kernel $K(\cdot)$ and bandwidths h_y and h_z. Smoothing with respect to R_{t+1} is optional, and setting $h_y = 0$ reduces the integral to a sum. However, setting $h_y > 0$ helps ensure strictly positive theoretical prices even for far out-of-the-money options.

Instead of minimizing $Q'_T(\theta)$ over the infinite dimensional functional space Θ, the method of sieves controls the complexity of the model in relation to the sample size by minimizing over an approximating finite-dimensional sieve spaces $\Theta_T \subseteq \Theta_{T+1} \subseteq \cdots \subseteq \Theta$ which become dense in Θ. This gives rise to the Sieve-Minimum Distance estimator (Ai and Chen, 2003)

$$\hat{\theta}_T = \min_{\theta \in \Theta_T} Q'_T(\theta). \tag{23}$$

Each Θ_T depends on a finite number of parameters only. When $\hat{\theta}_T$ is a finite linear combination of a known set of functions it is a series estimator. The compactness assumption on Θ relies on the norm that is specified. Our specification follows that of Newey and Powell (2003), who use the Sobolev norm which sums the integrals of partial derivatives. Consider a d-dimensional function g and let λ denote a $d \times 1$ vector of nonnegative integers, with $|\lambda| = \sum_{l=1}^{d} \lambda_1^d$, and $D^{\lambda} g(y) = \frac{\partial^{|\lambda|} g(y)}{\partial y_1^{\lambda_1} \ldots \partial y_d^{\lambda_d}}$. Then for some positive integers m and p, define the unweighted Sobolev norm

$$\|g\|_{m,p} = \left\{ \sum_{|\lambda| \leq m} \int \left(D^{\lambda} g(z) \right)^{p} dz \right\}^{1/p}. \tag{24}$$

It is implicitly assumed that the random variables that enter g are standardized to have mean zero and variance one. For $p = 2$ and some positive constant B, define the functional space

$$\Theta = \left\{ g(\cdot) : \|g\|_{m,2}^2 \leq B \right\} \tag{25}$$

Now consider the finite-dimensional series approximator $g_L(z) = \sum_{l=1}^{L} \gamma_j p_j(z)$, in terms of a set of basis functions $(p_1(z), \ldots, p_L(z))$. Define

$$\Lambda_L = \sum_{|\lambda| \leq m} \int D^{\lambda} p_L(z) D^{\lambda} p_L(z)^T dz \tag{26}$$

Then $g_L(\cdot) \in \Theta$ if and only if $\gamma^T \Lambda_J \gamma \leq B$. Thus we can define the optimization in (23) to be over the compact finite-dimensional subspace $\Theta_T \subseteq \Theta_{T+1} \subseteq \ldots \subseteq \Theta$, where

$$\Theta_T = \left\{ g(z) = \sum_{l=1}^{L(T)} \gamma_j p_j(z) : \gamma^T \Lambda_{L(T)} \gamma \leq B \right\}.$$

Also define the Sobolev sup-norm

$$\|g\|_{m,\infty} = \max_{|\lambda| \leq m} \sup_{z} \left| D^{\lambda} g(z) \right|. \tag{27}$$

Then the closure $\bar{\Theta}$ of Θ with respect to the norm $\|g\|_{m,\infty}$ is compact for that norm (Gallant and Nychka, 1987; Newey and Powell, 2003). For the case of vector-valued functions, such as $\theta = (\beta_0, \beta_1, \alpha)$ in model (13), define for $p \geq 1$

$$\|\theta\|_{m,p} = \|\beta_0\|_{m,p} + \|\beta_1\|_{m,p} + \|\alpha\|_{m,p}. \tag{28}$$

For each of these functions we can construct multivariate basis functions using tensor products of univariate basis functions $p_L(w)$ (Ai and Chen, 2003). If we use L basis functions for each variable, the series estimator $\hat{\alpha}$ has L^{d_z+1} coefficients in total, and the estimators $(\hat{\beta}_0, \hat{\beta}_1)$ have L^{d_z} coefficients each.

When the pricing kernel is of the partially exponential-affine form (13), we re-define the infinite-dimensional parameter space with smooth pasting conditions as

$$\Theta = \left\{ (\beta_0, \beta_1, \alpha) \in \Theta_\beta^4 \times \Theta_\alpha : \frac{\partial^j}{\partial y^j} \alpha(y, z) \Big|_{y=\kappa^l, \kappa^u} = \frac{\partial^j}{\partial y^j} e^{-\beta_0^l(z) - \beta_1^l(z) y} \Big|_{y=\kappa^l, \kappa^u}, \quad j = 0, 1, z \in \mathcal{Z} \right\},$$

in terms of the individual parameter's subspaces

$$\Theta_\beta = \{\beta(z) : \mathcal{Z} \to \mathbb{R}; \|\beta\|_{m,2}^2 \leq B_0\}$$
$$\Theta_\alpha = \left\{\alpha(y, z) : [\kappa^l, \kappa^u] \times \mathcal{Z} \to \mathbb{R}^+; \|\alpha\|_{m,2}^2 \leq B_\alpha \right\}.$$

The space Θ_α restricts the function α to be positive on its domain $[\kappa^l, \kappa^u]$ for any z. Such a restriction is not required for the functional parameters (β_0, β_1) due to the exponential link function. The smooth pasting conditions in Θ are automatically satisfied when profiling (β_0, β_1) over the level and slope of α at the boundary points:

$$\beta_0^j(z) = \log(\alpha(\kappa^j, z)) - \frac{\alpha_y(\kappa^j, z)}{\alpha(\kappa^j, z)} \kappa^j, \quad j \in \{l, u\} \tag{29}$$
$$\beta_1^j(z) = \frac{\alpha_y(\kappa^j, z)}{\alpha(\kappa^j, z)}$$

Profiling the tail parameters allows for unrestricted optimization over the coefficients of $\hat{\alpha}$, apart from the positivity constraints. Alternatively, a nonlinear constraint could be used in the optimization. We enforce the positivity constraint on a grid of values $(y_i)_i \in [\kappa^l, \kappa^u]$ using a linear system of inequality constraints on the coefficients of the series approximator $\hat{\alpha}$.[5]

2.4 Consistency

We demonstrate consistency of the estimator $\hat{\theta}$ given by (23) of the unknown functions θ that characterize the projected pricing kernel $\mu(R_{t+1}, Z_t, \theta)$. Consider the following set of assumptions:

A1 For every $\theta \in \Theta$ with $\theta \neq \theta_0$, $\rho(Z_t, \kappa_t, \theta) \neq 0$ for some (Z_t, κ_t), while $\rho(Z_t, \kappa_t, \theta_0) = 0$ a.s. for some $\theta_0 \in \Theta$

A2 $\hat{f}_h(R_{t+1}|Z_t = z) \xrightarrow{p} f(R_{t+1}|Z_t = z)$, the joint distribution function $f(R_{t+1}, Z_t)$ is continuous and bounded, $f(R_{t+1}|Z_t, \kappa_t) = f(R_{t+1}|Z_t)$, R_{t+1} is supported on $\mathbb{R}^+$, and (Z_t, κ_t) has compact and convex support $\mathcal{Z} \times [\kappa_l, \kappa_u] \subsetneq \mathbb{R}^{d_z+1}$

A3 $\hat{\Sigma}_t$ and Σ_t are positive definite matrices, such that Σ_t is bounded and $\hat{\Sigma}_t \xrightarrow{a.s.} \Sigma_t$ for every t

A4 $|\mu(R_{t+1}, Z_t, \theta) - \mu(R_{t+1}, Z_t, \tilde{\theta})| \leq b(R_{t+1}, Z_t)\|\theta - \tilde{\theta}\|^v$ for some $v > 0$ with $E\left(b(R_{t+1}, Z_t)^2 R_{t+1}^2 | Z_t\right) < \infty$, and $\text{Var}\left(\mu(R_{t+1}, Z_t, \theta_0)|Z_t\right) < \infty$

[5] All optimizations in the simulation and estimation results are performed using Matlab's `fmincon`.

A5 $(R_{t+1}, Z_t, \kappa_t, \varepsilon_t)$ is a strong mixing stationary process, with $E\left(\varepsilon_t^2\right) < \infty$

Theorem 1. *Under assumptions A1-A5, the estimator defined by (23) satisfies*

$$\|\hat{\theta} - \theta_0\|_{m,\infty} \xrightarrow{p} 0 \tag{30}$$

when $T \to \infty$, $L \to \infty$, $L^{d_z+1}/T \to 0$, $h_y \to 0$, $h_z \to 0$, *and* $Th_z^{d_z} \to \infty$.

We can specialize this result to the case of the partially exponential-affine pricing kernel (13). Consider the additional conditional moment bounds:

A6 $E\left(R_{t+1}^6 | Z_t\right) < \infty$ and $E\left(e^{4\beta(Z_t)R_{t+1}} | Z_t\right) < \infty$ for all $\beta \in \Theta_\beta$

Corollary 1.1. *Under assumptions A2-A3 and A5-A6, the estimator defined by (23) with pricing kernel (13) satisfies for any* $|\lambda| \le m$*:*

$$\max_{(y,z)\in[\kappa^l,\kappa^u]\times\mathcal{Z}} \left| D^\lambda \left(\hat{\alpha}(y,z) - \alpha(y,z)\right) \right| \xrightarrow{p} 0$$

$$\max_{z\in\mathcal{Z}} \left| D^\lambda \left(\hat{\beta}_i^j(z) - \beta_i^j(z)\right) \right| \xrightarrow{p} 0 \qquad i \in \{0,1\},\ j \in \{l,u\}$$

when $T \to \infty$, $L \to \infty$, $L^{d_z+1}/T \to 0$, $h_y \to 0$, $h_z \to 0$, *and* $Th_z^{d_z} \to \infty$.

2.5 Efficiency

The efficiency of the estimator is determined by the weighting matrices Σ_t. In semiparametric conditional moment models, the parametric component is efficiently estimated when the weighting matrices are set as $\Sigma_t = \Omega_t^{-1}$ (Ai and Chen (2003)), where

$$\Omega_t = \text{Var}(\mu(R_{t+1}, Z_t, \theta_0)h(R_{t+1}, \kappa_t) - P_t \mid Z_t, \kappa_t).$$

Our simulation results in subsection 2.7 suggest that using estimates of this weighting matrix also improves the efficiency of the nonparametric pricing kernel estimator. If option prices contain no additional information about the future return beyond Z_t, then this conditional covariance matrix sums up the variances of the discounted payoffs and the option prices. Otherwise, option prices that are informative in the sense of positively correlating with their own payoffs given Z_t, will have less volatile discounted gains and losses, and receive extra weight through Σ_t.

Define the option discounted gain or loss $u_t(\kappa, \theta) = \mu(R_{t+1}, Z_t, \theta)h(R_{t+1}, \kappa) - P_t(\kappa)$. The kernel-smoothed centered estimator of the conditional variance-covariance matrix is

$$\widehat{\Omega}_t(\theta) = \sum_{s=1}^{T} w_{ts}(u_s(\kappa_t, \theta) - \bar{u}_t(\kappa_t, \theta))(u_s(\kappa_t, \theta) - \bar{u}_t(\kappa_t, \theta))^T,$$

where $\bar{u}_t(\kappa_t, \theta) = \sum_{s=1}^T w_{ts} u_s(\kappa_t, \theta)$, and the weights $w_{ts} = \frac{K_h(Z_s - Z_t)}{\sum_r K_h(Z_r - Z_t)}$ measure the distance between observations according to their conditioning variables in terms of the kernel $K(\cdot)$. Let $\tilde{\theta}$ be an initial consistent estimator of θ, such as the one obtained using $\Sigma_t = I_{n_t}$. We then obtain a feasible two-stage estimator by setting $\widehat{\Sigma}_t = \widehat{\Omega}_t^{-1}(\tilde{\theta})$.

Since κ_t are treated as random, the option prices form repeated cross-sections with different sets of moneyness levels in the range $[\kappa^l, \kappa^u]$. While the counterfactual $P_s(\kappa_t)$ is not directly observed for $s \neq t$, we can form a cross-sectional estimate $\hat{P}_s(\kappa_t)$ by interpolating option prices with nearby moneyness levels in κ_s. Many local or global interpolation methods have been developed and shown to be accurate even with relatively small cross-sections of option prices. Moreover, the estimation error will be reflected in the efficient weighting matrix based on $\widehat{\Omega}_t$.

2.6 Conditional Euclidian empirical likelihood formulation

Information-theoretic estimation involves choosing model parameters by maximizing the non-parametric likelihood of the observations subject to theoretical moment conditions. Treating the option prices as random given the conditioning variables, we define the Conditional Euclidian Empirical Likelihood (CEEL) problem based on the conditional moment restrictions (9) as

$$\begin{aligned} \min_{(p_{ts}),\theta} \quad & \sum_{t=1}^T \sum_{s=1}^T \left(\frac{p_{ts} - w_{ts}}{w_{ts}} \right)^2 w_{ts} \qquad (31) \\ \text{s.t.} \quad & \sum_{s=1}^T p_{ts} \left(\mu(R_{s+1}, Z_s, \theta) h(R_{s+1}, \kappa_t) - P_s(\kappa_t) \right) = 0, \quad t = 1, \ldots, T \\ & \sum_{s=1}^T p_{ts} = 1. \end{aligned}$$

For each period t, implied probabilities p_{ts} are chosen as close as possible to the empirical probabilities w_{ts} subject to the pricing restrictions locally around Z_t. Compared to the conditional empirical likelihood problem in Kitamura et al. (2004), this problem uses a quadratic or Euclidian distance criterion which is computationally attractive (Antoine et al., 2007).[6] In particular, the implied probabilities are given in closed form as a function of θ by

$$\begin{aligned} p_{ts}(\theta) &= w_{ts}(1 - \lambda_t(\theta)^T (u_t(\kappa_t, \theta) - \bar{u}_t(\kappa_t, \theta))), \\ \lambda_t(\theta) &= \widehat{\Omega}_t(\theta)^{-1} \bar{u}_t(\kappa_t, \theta). \end{aligned}$$

[6] Both are special cases of the generalized conditional empirical likelihood problems considered in Smith (2007).

The expected discounted payoffs under the discrete probabilities p_{ts} are close to that under the continuous kernel density estimator with $h_y \approx 0$, as

$$\sum_{s=1}^{T} p_{ts}\mu(R_{s+1}, Z_s, \theta)h(R_{s+1}, \kappa_t) \approx \int \mu(y, Z_t, \theta)h(y, \kappa_t)\hat{f}_{h,p}(y|Z_t)dy, \tag{32}$$

where

$$\hat{f}_{h,p}(y|Z_t) = \sum_{s=1}^{T} p_{ts}K_{h_y}(y - R_{s+1}). \tag{33}$$

The criterion (31) when optimized over the implied probabilities can be shown to be equal to

$$\mathcal{L}_T(\theta) = \sum_{t=1}^{T} \bar{u}_t(\kappa_t, \theta)^T \widehat{\Omega}_t(\theta)^{-1}\bar{u}_t(\kappa_t, \theta). \tag{34}$$

This is a continuously-updated variant of the minimum-distance criterion (18). The probability limit of $\mathcal{L}_T(\theta)$ identifies the same θ as the minimizers of $Q(\theta)$ and $Q'(\theta)$ in Section 2.2, as the weighting matrix only matters for efficiency. The diagonals of the weighting matrix indicate that an efficient estimator gives high weight to those strike prices for which the discounted option trade gain or loss have a low variance. In particular, this weight is increasing in the covariance between the option price and its discounted payoff.

When many strike prices are available per cross-section, the covariance matrix $\widehat{\Omega}_t(\theta)$ risks becoming close to singularity due to the dependence of gains and losses of option prices with nearby moneyness level. A simple way of regularizing the covariance matrix would be to restrict the number of included strike prices per period.[7] Alternatively, the following local CEEL problem finds a set of implied probabilities for each available strike price:

$$\begin{aligned} &\min_{(p_{its}),\theta} \sum_{t=1}^{T}\sum_{i=1}^{n_t+2}\sum_{s=1}^{T}\left(\frac{p_{its}-w_{ts}}{w_{ts}}\right)^2 w_{ts} \\ &\text{s.t.} \sum_{s=1}^{T} p_{its}\left(\mu(R_{s+1}, Z_s, \theta)h(R_{s+1}, \kappa_{it}) - P_s(\kappa_{it})\right) = 0, \quad t = 1, \dots, T, \quad i = 1, \dots, n_t + 2 \\ &\sum_{s=1}^{T} p_{its} = 1. \end{aligned} \tag{35}$$

Its implied probabilities as a function of θ are given by

$$\begin{aligned} p_{its}(\theta) &= w_{ts}(1 - \lambda_{it}(\theta)^T(u_{it}(\kappa_{it}, \theta) - \bar{u}_{it}(\kappa_{it}, \theta))), \\ \lambda_{it}(\theta) &= \widehat{\Omega}_{t,ii}^{-1}\bar{u}_{it}(\kappa_{it}, \theta). \end{aligned} \tag{36}$$

[7] In our simulations, we find using about 6 out-of-the money options per period, split evenly among calls and puts, leads to stable estimates. The same number is shown to work well in Buchen and Kelly (1996) for extracting risk-neutral densities from finite sets of options based on the relative entropy criterion.

The option-specific Lagrange multipliers $\lambda_{it}(\theta)$ equal the average discounted gain or loss per unit of variance. For a flat pricing kernel ($K = 1$), these resemble Sharpe ratios which measure the expected excess return per unit of standard deviation. For nonlinear pricing kernels, they can be interpreted as preference-adjusted performance metrics. Profiling over the local implied probabilities, the criterion (35) equals

$$\mathcal{L}_T(\theta) = \sum_{t=1}^{T} \sum_{i=1}^{n_t+2} \widehat{\Omega}_{t,ii}(\theta)^{-1} \bar{u}_{it}(\kappa_{it}, \theta)^2. \tag{37}$$

This criterion ignores the cross-sectional dependence of similar options at the cost of some efficiency. However, it allows all available strike prices to be included, with weights depending on their informativeness, without the risk of near-singularity. Hybrid, block-wise approaches between the local and joint CEEL problems can be constructed by creating portfolios of options, the underlying return, and the risk-free rate, so that their payoffs are close to orthogonal.

2.7 Simulation study

We investigate the performance of different versions of the sieve-minimum distance estimator (23) in a simulation experiment. In particular, we consider the estimator with equal weighting and with two-stage and continuously-updated efficient weighting inversely to the local variance $\hat{\Omega}_{t,ii}$ to allow for heteroskedasticity in the strike dimension. Option prices are simulated by the one-factor stochastic volatility jump-diffusion model proposed by Bates (2000), whose return and volatility dynamics under the risk-neutral $\mathbb{Q}$-measure are given by

$$\begin{aligned}
\frac{\mathrm{d}F_t}{F_t} &= r_t \mathrm{d}t + \sqrt{V_t}\mathrm{d}B_t + (J-1)\mathrm{d}N(t), \\
\mathrm{d}V_t &= \kappa(\Theta - V(t))\mathrm{d}t + \gamma\sqrt{V_t}\mathrm{d}W_t,
\end{aligned}$$

where $Corr(\mathrm{d}B_t, \mathrm{d}W_t) = \rho \mathrm{d}t$, $N(t)$ is a Poisson jump process with intensity λ_J that is independent of the Brownian motions, and J is lognormal with location parameter μ_J and volatility σ_J. Under the physical measure $\mathbb{P}$, stochastic volatility follows the same process with different parameters, reflecting the variance risk premium. The model parameters are chosen based on Bates (2000). Simulated data sets $(C_t^{(i)}, R_{t+1}^{(i)})_{i=1}^{S}$ are generated by computing monthly option prices and returns for each of $S = 500$ volatility paths.

The results in Table 1 show that for the relatively small period of 5 years of monthly data, the equally weighted estimator is slightly outperformed in terms of MSE by the two-stage estimator, yet outperforms the continuously-updated estimator. The latter's underperformance is primarily due to its high variance, as both weighted estimators have much lower bias for all three orders K. For the relatively long period of 10 years of monthly data, both weighted estimators outperform

Table 1: Simulated integrated bias, variance, and mean squared error of different sieve-minimum-distance pricing kernel estimators, based on $S = 500$ volatility paths and monthly option prices and returns. The reported MSE and components are based on integrated distances $\int_{\kappa_l}^{\kappa_u} \left(\hat{m}^{(i)}(y) - \bar{m}^{(i)}(y)\right)^2 dy$ from the simulated model-based pricing kernel $\bar{m}^{(i)} = \frac{\bar{f}^{\mathbb{Q},i}(y)}{\bar{f}^{\mathbb{P},i}(y)}$ over moneyness region $(\kappa_l, \kappa_u) = (0.95, 1.05)$, where $\bar{f}^{(i)}(\cdot) = \frac{1}{T}\sum_{t=1}^{T} f_t^{(i)}(\cdot)$ are pathwise-average densities. Series expansion is of varying order $K = 4, 6, 8$, displayed from left to right panels. The Sobolev bound is set as $B = 5$. Rows of the tables refer to the equally weighted (EW), two-stage (2S), and continuously-updated (CU) estimators based on the local CEEL problem (37).

(a) $T = 60$ months

K = 4	Bias2	Var	MSE	K = 6	Bias2	Var	MSE	K = 8	Bias2	Var	MSE
EW	0.33	0.28	0.60	*EW*	0.44	0.32	0.76	*EW*	0.41	0.33	0.75
2*S*	0.21	0.53	0.74	2*S*	0.12	0.57	0.68	2*S*	0.11	0.57	0.68
CU	0.29	0.69	0.98	*CU*	0.14	0.69	0.83	*CU*	0.14	0.69	0.83

(b) $T = 120$ months

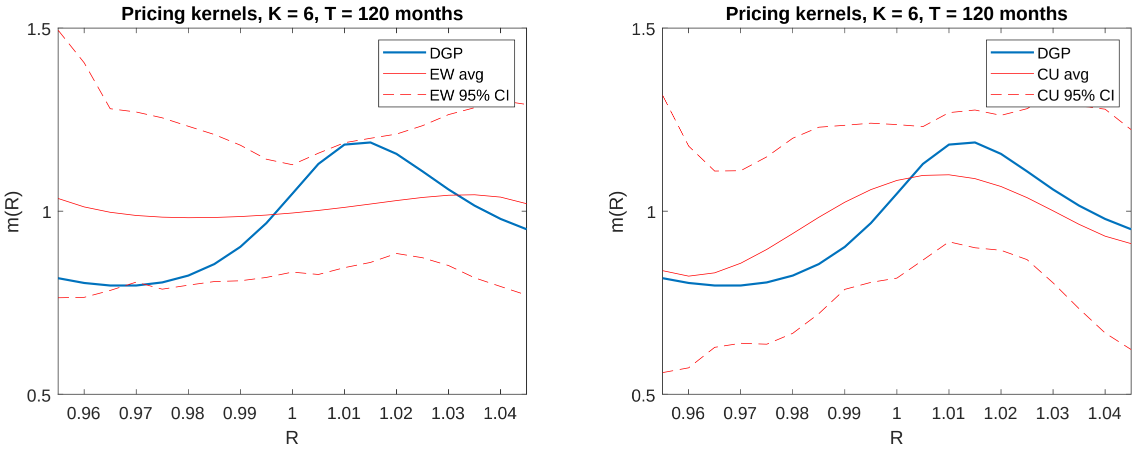
K = 4	Bias2	Var	MSE	K = 6	Bias2	Var	MSE	K = 8	Bias2	Var	MSE
EW	0.15	0.11	0.26	*EW*	0.18	0.13	0.31	*EW*	0.16	0.15	0.30
2*S*	0.09	0.16	0.25	2*S*	0.04	0.16	0.20	2*S*	0.04	0.16	0.20
CU	0.17	0.17	0.34	*CU*	0.05	0.15	0.21	*CU*	0.05	0.15	0.21

the equally weighted estimator in terms of MSE. In particular, this is due to large reductions in the variance of both efficiently weighted estimators. Thus, the simulation results suggests that once the sample size is large enough to accurately estimate the variance of discounted option gains and losses, the feasible efficiently weighted estimators allow for improved estimation of the pricing kernel for various levels of the approximating series order.

Figure 1 shows the simulated average and 95% pointwise quantiles of estimated pricing kernels, as well as the unconditional model-implied pricing kernel (DGP), for sample paths of $T = 120$ months of option prices and returns. The model-implied kernel shows a positive hump for moderately positive returns around 1%, reflecting the higher kurtosis of the risk-neutral versus the objective density. However, the equally weighted pricing kernels appear mostly flat. Meanwhile, the continuously updated estimator yields a similar hump shape on average, with similar levels of pointwise variation.

2.8 Empirical pricing kernels

We apply the sieve minimum-distance estimator (23) for estimating the projecting pricing kernel implicit in option contracts written on the S&P 500 index. The option data are obtained from OptionMetrics over the sampling period January 1996 to June 2016. The contracts have European exercise style with one month maturity. The cum-dividend returns on the S&P 500 index measure the underlying asset return, and the one month Federal Funds rate is used as a proxy for the



(a) Equally weighted.

(b) Continuously updated.

Figure 1: Simulated average and 95% pointwise quantiles of estimated pricing kernels, against the unconditional model-implied pricing kernel (DGP), using the settings from Table 1.

risk-free rate. We include put options within the moneyness range $[0.92, 1]$, and call options within the range $[1, 1.06]$, which excludes around 5% of options of either type due to their low liquidity. Within this range, we compute kernel-smoothed option prices that are used to estimate the covariance matrix of the moments for any set of moneyness levels. The conditioning variable is set as $Z_t = \log VIX_t$, as the VIX is an often-used proxy for market-implied volatility and risk aversion. Periods where the VIX fell outside its 2.5%- and 97.5%-quantiles are trimmed due to the difficulty of estimating their conditional second moments. We use Hermite polynomials as the basis functions of the series estimator, and smoothly paste exponential-affine tails according to (29). The Sobolev bound for α is set as $B_\alpha = 20$, based on values obtained from lower order ($K_R = 4$) unrestricted estimation. The bandwidth h_z is chosen using the nearest-neighbor method with 40 nearby observations.

Figure 2 shows the estimated pricing kernel $\hat{\mu}(R_{t+1}, Z_t)$ for the bivariate expansion with $K_R = 6$ in the return dimension and $K_Z = 3$ in the log VIX dimension, based on local continuously-updated weighting (34). The fitted pricing kernel is not monotonically decreasing in the return, in contrast with standard preference assumptions that imply higher marginal utility for lower wealth levels. Only for low levels of the VIX is the pricing kernel broadly decreasing, while featuring a hump around one. For moderate and large values of the VIX, the fitted pricing kernel appears U-shaped. In particular, the pricing kernel steeply increases for returns on the upside, while modestly sloping upwards on the downside. The different slopes can be explained using the left panel of Figure 3, which shows the kernel-weighted average of option prices and their payoffs for different moneyness levels conditional on low and high values of the VIX. The kernel-smoothed payoffs of both put and call options exceed their average prices, and the more so for out-of-the-money options and for calls. Moreover, since returns are negatively skewed, this call option premium has to be explained by high values of the pricing kernel for relatively modest positive returns, whereas the put option premium can be explained with higher values of the pricing kernel for large negative returns. The over-valuation of right tail payoffs suggests increases in the VIX contain a strong 'preference' component and do not necessarily reflect a higher probability of tail events. In particular, the large call option premium could be explained by a desire to insure against high future volatility, even when this coincides with large positive returns.

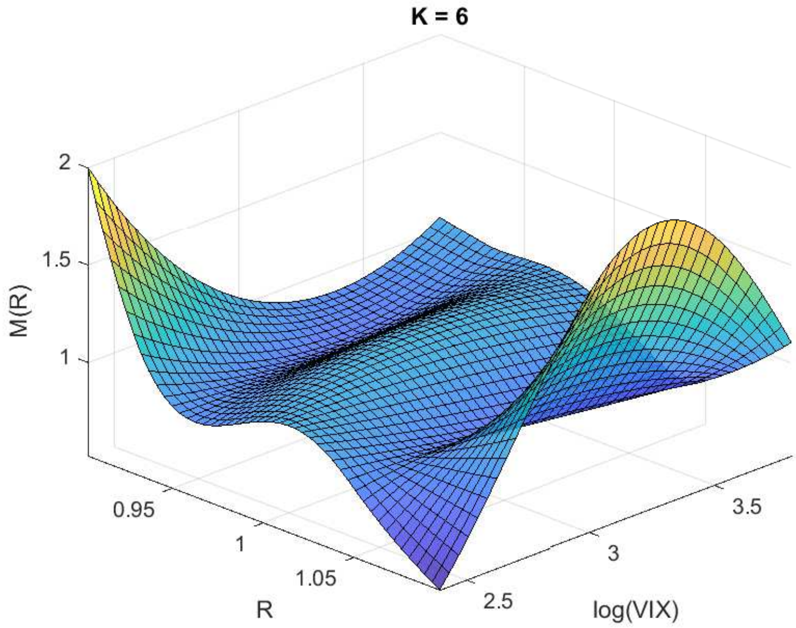


Figure 2: Fitted bivariate pricing kernel expansion with $(K_r, K_z) = 6 * 3$ on option and returns data, from January 1996 until June 2016, based on the local continuous-updating estimator, with log VIX as conditioning variable.

The right panel of Figure 3 shows the resulting match between the kernel-weighted average of option prices and their discounted payoffs based on the $K = 6$ order pricing kernel from Figure 2. Compared to the flat pricing kernel ($K = 1$) in the left panel, the fitted pricing kernel is able to explain most of the put and call option premia for both high and low values of the VIX.

Figure 4 shows the estimated Lagrange multipliers (36) for fitting the option price constraints in (35). The left panel shows the multipliers when using a flat pricing kernel, which would apply to risk-neutral investors. In this case, all multipliers are negative, and all options are overpriced. The multipliers are the most negative for out-of-the-money call options, implying that their overpricing is especially large relative to the variation of their gains and losses. The right panel shows the multipliers obtained by fitting a higher-order pricing kernel expansion. Allowing for

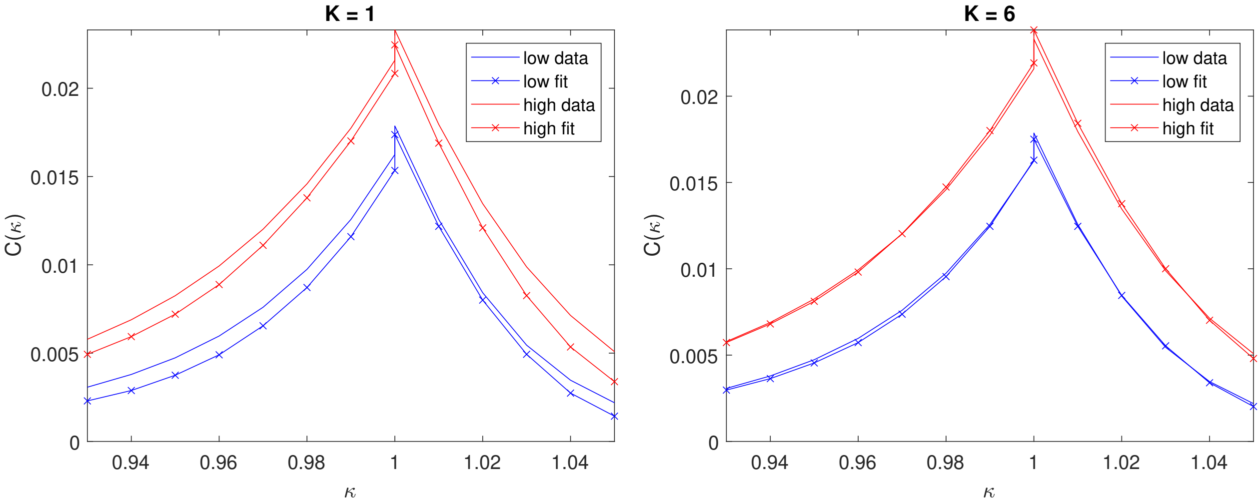
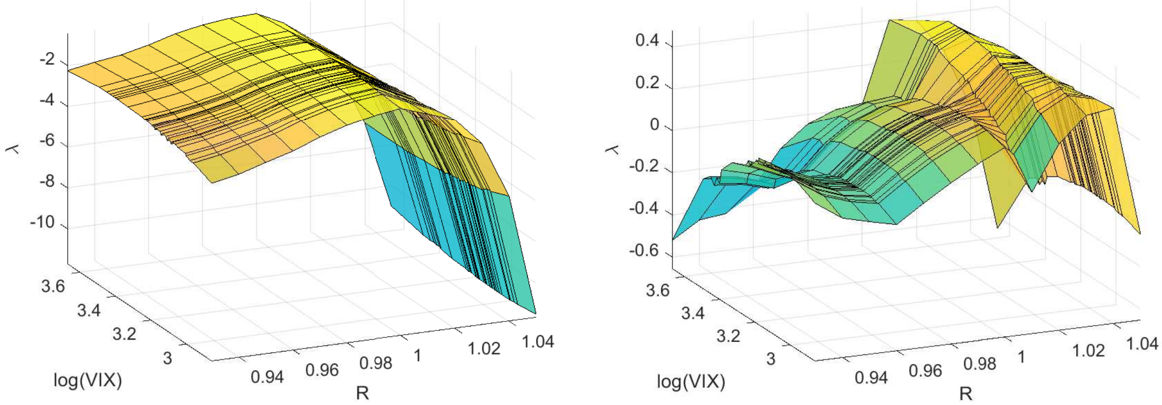


Figure 3: Fitted and smoothed option prices for varying moneyness κ based on a flat pricing kernel (left) and $(K_r, K_z) = 6 * 3$ pricing kernel expansion (right) based on the continuous-updating estimator. 'Low' and 'high' refers to the 10%- and 90%-quantiles, respectively, of the log VIX conditioning variable.

flexible risk aversion much reduces the absolute values of the multipliers, with the largest values occurring in the tails of the return distribution when the VIX is high. Thus the main challenge for the pricing kernel is to rationalize the expensiveness of out-of-the-money call and put options with the historical return distribution. Interestingly, the multipliers corresponding to call prices change signs when moving from at-the-money to out-of-the money strike prices. This points to the difficulty of matching the large negative risk-premium of out-of-the money call options with more moderate overpricing of those at-the-money.



(a) $(K_r, K_z) = 1 * 1$

(b) $(K_r, K_z) = 6 * 3$

Figure 4: Estimated Lagrange multipliers for fitting the option price constraints using a flat (left) versus polynomial expansion (right) pricing kernel, for varying levels of $\log(VIX)$.

2.9 Option-Implied Expected Returns

The estimated pricing kernel also allows computing time-varying expected excess returns in a way that is consistent with option prices. The pricing restrictions for options help to identify the potentially nonlinear shape of the projected pricing kernel, which in turn can be used to compute expected returns from the pricing restriction on the underlying asset. In particular, the stock pricing equation (4) minus the bond pricing equation (5) yields

$$E\left(R_{t+1} - R_t^f \mid Z_t\right) = -\text{Cov}(\mu(R_{t+1}, Z_t, \theta), R_{t+1} \mid Z_t). \tag{38}$$

Since the pricing kernel is generally decreasing with the market return, this formula implies positive risk premia. Time-variation in the risk premium comes from both time-varying risk aversion as well as the time-varying return distribution. An informationally efficient estimate of the risk premia can be computed using the implied conditional probabilities from (31), which automatically satisfy the risk premium formula (38) for any given pricing kernel.

Figure 5 shows the realized returns and annualized expected returns conditional on the VIX. The expected returns shown are those using the initial conditional probabilities and those using their implied counterparts from the CEEL problem (31). Under the initial conditional probabilities, the expected returns suddenly increase with high levels of the VIX, as there are few data points in this area. However, the CEEL implied expected returns are more stable, as they use knowledge of the estimated pricing kernel from option prices to extrapolate through regions with few observations. Moreover, the implied expected returns are slightly decreasing in the VIX, against standard predictions on the risk-return trade-off. This finding is line with the U-shaped pricing kernels given above average values of the VIX from Figure 2, which imply that assets that pay off during large positive returns trade at a premium.

3 Market-based Density Forecasting

Option prices can be used to extract market-implied densities which incorporate information available to investors in addition to historical data observed by econometricians. In particular, the risk-neutral density at a given time t can be estimated based on the Breeden and Litzenberger (1978) result

$$f^{\mathbb{Q}}(\kappa|t) = \frac{\partial^2}{\partial^2 \kappa} E\left(C_{i,t}|\kappa_{i,t} = \kappa\right), \qquad \kappa \in [\kappa^l, \kappa^u] \tag{39}$$

where $(C_{i,t})_i$ is a cross-section or rolling window of option prices observed at time t. The call pricing function and its second derivative can be estimated using nonparametric methods such as kernel smoothing or series approximation, requiring only a suitable choice of tuning parameters.

A fundamental problem with using risk-neutral densities for forecasting is their biases due

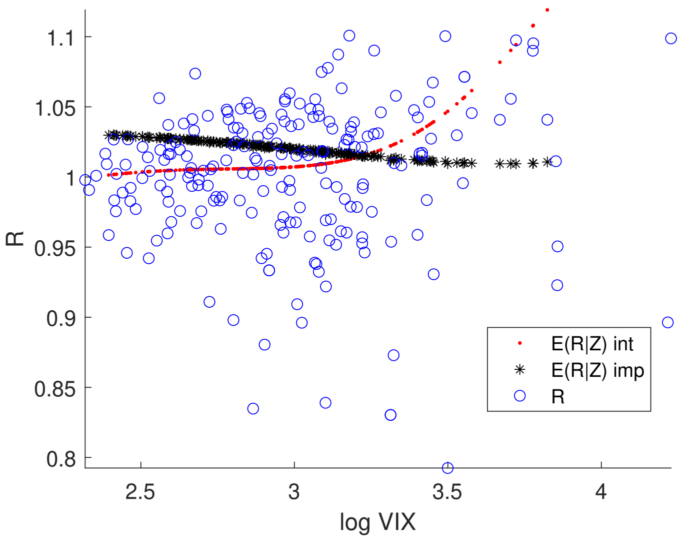


Figure 5: Actual monthly returns and annualized expected returns on S&P 500 futures under initial and implied conditional probabilities, for the $(K_R, K_Z) = (6, 3)$ order bivariate pricing kernel expansion, for varying levels of the VIX, using data from January 1996 until June 2016.

to the risk aversion of investors. If risk aversion is constant, we can 'correct' the risk neutral density into an unbiased estimate of the conditional physical density using a model for the pricing kernel. Several parametric models have been suggested to perform this transformation, such as the Generalized Beta distribution (Liu et al., 2007; Vincent-Humphreys et al., 2012), or exponential-affine models (Hördahl and Vestin, 2005; Gagliardini et al., 2011; León et al., 2012). Using our nonparametrically estimated pricing kernel from Section 2, a simple plug-in estimator of the physical density $f_t^{\mathbb{P}}(R_{t+1})$ based on equation (3) would be

$$\hat{f}_t^{\mathbb{P}}(R_{t+1}) = \frac{\hat{f}_t^{\mathbb{Q}}(R_{t+1})}{\hat{\mu}(R_{t+1}, Z_t)}. \tag{40}$$

This density estimator is sensitive to small values of the estimated pricing kernel in the denominator. Moreover it is not guaranteed to be non-negative and integrate to unity.

Instead, we propose to extract the bias-corrected predictive density by minimizing the Kullback-Leibler divergence or relative entropy to a prior historical density, subject to matching contemporaneous option prices. The relative entropy criterion guarantees a positive density, and has been used to estimate risk-neutral densities in Stutzer (1996) and Buchen and Kelly (1996).[8] For forecasting, Robertson et al. (2005) and Giacomini and Ragusa (2014) use the criterion to exponentially 'tilt' a prior density to incorporate additional moment information motivated by economic theory.

[8] Stutzer (1996) and Buchen and Kelly (1996) do so by minimizing the Kullback-Leibler divergence between the risk-neutral and historical distributions, without modeling the pricing kernel.

More fundamentally, the pricing kernel may have missing conditioning variables due to unobserved changes in risk aversion or subjective beliefs. The density (40) or its entropic estimate would fully attribute such unobserved preference dynamics to the physical density. Instead, we modify the conditional moments of the relative entropy problem to reflect that option prices are only partially informative about future returns. We do so based on an encompassing density that linearly combines the bias-corrected option-implied density and historical conditional density given observed variables. Finally, we consider conditional moments that reflect heterogeneity in the informativeness of option prices across moneyness levels.

3.1 Encompassing densities

In general, both the conditional physical density $f_t^{\mathbb{P}}(y)$ and conditional pricing kernel $\mu_t(y) \equiv \frac{f_t^{\mathbb{Q}}(y)}{f_t^{\mathbb{P}}(y)}$ may depend on information available to market participants, but not to econometricians. Restricting the information set to some observed conditioning variables Z_t therefore risks inducing dynamic misspecification. In particular, by writing

$$\log \frac{f_t^{\mathbb{Q}}(y)}{f^{\mathbb{Q}}(y|Z_t)} = \log \frac{f_t^{\mathbb{P}}(y)}{f^{\mathbb{P}}(y|Z_t)} + \log \frac{\mu_t(y)}{\mu(y|Z_t)},$$

we see that any variation in the risk-neutral density not explained by Z_t implies unexplained variation in the physical density, the pricing kernel, or both.

A class of encompassing models specifies the conditional physical density as the linear combination

$$f_t^c(y;\pi,\mu) = \pi \frac{f^{\mathbb{Q}}(y|Z_t)/\mu(y|Z_t)}{\int f^{\mathbb{Q}}(y|Z_t)/\mu(y|Z_t)\mathrm{d}y} + (1-\pi) f^{\mathbb{P}}(y|Z_t). \tag{41}$$

This generalizes the parametric encompassing densities in Liu et al. (2007). The first density is a market-implied density corrected for its bias in states with similar Z_t, normalized to integrate to unity. The second density $f^{\mathbb{P}}(y|Z_t)$ is unbiased given Z_t, but is subject to missing conditioning variables. The parameter π can be interpreted as the informativeness of the risk-neutral density, or as a Bayesian posterior probability (Barone-Adesi et al., 2020).[9] When $\pi = 1$, any unexplained variation in option prices is attributed to investor beliefs, and the pricing kernel only varies with Z_t. When $\pi = 0$, unexplained variation in option prices is attributed to the pricing kernel, and the predictive density only varies with Z_t.

We define the parameter π_0 as the solution to the maximum likelihood population problem

$$\pi_0 = \arg\max_{\pi} E\left(\log f_t^c(R_{t+1};\pi,\mu_0)\right),$$

[9] Barone-Adesi et al. (2020) consider a Bayesian approach that allows for time-varying informativeness π_t.

where μ_0 is the projected pricing kernel identified by (12). If instead μ were unrestricted, its maximum likelihood value would generally differ from the projected pricing kernel when the conditioning information is misspecified. To see this, suppose $\pi = 1$, let $Z = \emptyset$, and suppose the physical density depends on an unobserved state variable S. Then optimizing the unconstrained maximum likelihood over the function μ yields

$$\mu^* = \arg\min_{\mu \in L^2} E\left(\log\left(\frac{f^{\mathbb{Q}}(R|S)/\mu(R)}{\int f^{\mathbb{Q}}(y|S)/\mu(y)\mathrm{d}y}\right)\right),$$

which by calculus of variation is characterized implicitly by

$$\mu^*(R) = \int \frac{f^{\mathbb{Q}}(R|s)/f^{\mathbb{P}}(R)}{\int f^{\mathbb{Q}}(y|s)/\mu^*(y)\mathrm{d}y} f_S(s)\mathrm{d}s.$$

When the conditioning variables are correctly specified ($S = \emptyset$), the RHS implies that $\mu^*(R)$ equals the projected pricing kernel. With non-trivial missing conditioning variables ($S \neq \emptyset$), the pricing kernel $\mu(R) = \frac{f^{\mathbb{Q}}(R)}{f^{\mathbb{P}}(R)}$ does not generally satisfy this equation as denominator values unequal to one distort the expectation over S.

The 'informativeness' parameter π_0 is identified from the likelihood provided the averaging densities are not identical, which is the case with missing conditioning variables.

3.2 Partially smoothed option prices

The market-implied bias-corrected density and the physical density in (41) are not directly observed, and have to be estimated from a finite set of option prices and historical returns, respectively. Instead of combining two separately estimated densities, the following population problem defines the encompassing density by matching a combination of current and historical option prices:

$$\begin{aligned} f_t^c(\cdot;\pi,\mu) = {} & \arg\min_{f_t(\cdot)} \int f_t(y)\log\frac{f(y|Z_t)}{f_t(y)}dy \\ \text{s.t.} & \int \mu(y,Z_t)h(y,\kappa_t)f_t(y)\mathrm{d}y = \pi P_t(\kappa_t) + (1-\pi)E(P_t(\kappa_t)\big|Z_t) \\ & \int p_t(y)dy = 1. \end{aligned} \tag{42}$$

The problem uses the relative entropy criterion to choose among the set of densities that match the partially smoothed option prices at currently observed strike prices. When $\pi = 1$, it extracts conditional densities by exactly matching the cross-section of option prices, as proposed for estimating parametric pricing kernels in Gagliardini et al. (2011).[10] This is justified when all

[10] Gagliardini et al. (2011) focus on using the latest cross-section of option prices and the historical time series of stock prices. The estimator described in this paper also uses the time series of option prices to identify the projected pricing kernel.

state variables are included in Z_t and there is no option pricing error ε_t. When $\pi = 0$, current option prices contain no additional conditioning information beyond the variable Z_t. Still, in this case the entropic density $f_t(\cdot)$, unlike the historical density $f(\cdot|Z_t)$, takes into account the constraint that expected option prices for given Z_t should match their expected discounted payoff given the pricing kernel. The higher $\pi \in [0,1]$, the less option prices are smoothed, and the stronger implied densities react to current option prices.

Suppose we observe the data $(R_{t+1}, (C_{it})_i, Z_t)_t$ consisting of returns, cross-sections of option prices, and conditioning variables. In particular, the relative entropy distance criterion leads to the problem

$$\min_{(f_t(\cdot))} \frac{1}{T}\sum_{t=1}^{T} \int f_t(y) \log \frac{\hat{f}_h(y|Z_t)}{f_t(y)} dy \tag{43}$$

$$\text{s.t.} \int \mu(y, Z_t) h(y, \kappa_t) f_t(y) dy = \hat{P}_t(\pi) \tag{44}$$

$$\int p_t(y) dy = 1, \qquad t = 1, \cdots, T$$

where

$$\hat{f}_h(y|Z_t) = \sum_{s=1}^{T-1} w_{ts} K_{h_y}(R_{s+1} - y)$$

$$\hat{P}_t(\pi) = \pi P_t + (1-\pi)\hat{E}_h\left(P_t|Z_t\right)$$

$$\hat{E}_h\left(P_t|Z_t\right) = \sum_{s=1}^{T-1} w_{ts} P_s$$

The densities that solve this problem are of the form

$$f_t(y) = \hat{f}_h(y|Z_t)\exp(\lambda_t^T(\mu(y, Z_t)h(y, \kappa_t) - \hat{P}_t(\pi)) + \mu_t),$$

which depends on the Lagrange multipliers and normalization constant

$$\lambda_t = \arg\min_{\lambda} \int f_h(y|Z_t) \exp\left(\lambda^T(\mu(y, Z_t)h(y, \kappa_t) - \hat{P}_t(\pi))\right) dy \tag{45}$$

$$\mu_t = \log\left(\int \hat{f}_h(y|Z_t)\exp(\lambda_t^T(\mu(y, Z_t)h(y, \kappa_t) - \hat{P}_t(\pi)))dy\right)^{-1}.$$

While yielding positive densities, a downside of exponential tilting is that the multipliers are not analytically available. This makes the relative entropy criterion less computationally attractive than the Euclidian least squares criterion for estimating the pricing kernel, which can require reasonable high-dimensional optimization. Instead, the pricing kernel μ and informativeness parameter π can be chosen in a two-step approach. First, minimize the conditional Euclidian distance criterion (31) over the sieve space to find the estimator $\hat{\mu}$, which is consistent regardless of whether option price have additional predictive return information. Second, maximize the predictive log-likelihood $\mathcal{L}_T(\pi) = \frac{1}{T}\sum_{t=1}^{T} \log f_t^c(R; \pi, \hat{\mu})$ over π, where f_t^c minimizes the relative entropy to the historical density subject to the partially smoothed option prices in (43). The implied probabilities $p_{ts}(\hat{\mu})$ from the first-stage can be used instead of w_{ts} to efficiently estimate the conditional density of returns and means of the option prices. While Euclidian implied probabilties are not necessarily non-negative, they could be made so using a shrinkage method at no asymptotic cost (Antoine et al., 2007).

3.3 Local informativeness

The density combination (41) assumes that the informativeness parameter π does not vary with the region of the return. Similarly, the conditional moment combination (42) assumes that option prices of varying strike prices are equally informative about their payoff. However, Chakravarty et al. (2004) present evidence that the share of informed trading in option markets varies with the strike price, as they offer varying leverage, trading volume, and spreads. To assess this, we define the local informativeness $\pi(\kappa)$ as a function of moneyness κ as the slope of the local regression

$$\mathbb{E}\left(\mu(R_{t+1}, Z_t)h(R_{t+1}, \kappa) \mid Z_t, P_t(\kappa)\right) = \pi(\kappa)P_t(\kappa) + (1 - \pi(\kappa))\mathbb{E}(P_t(\kappa) \mid Z_t).$$

Since option prices are unbiased forecasts of their discounted payoffs given Z_t, the constant in this local regression is restricted to equal $(1 - \pi(\kappa))\mathbb{E}(P_t(\kappa) \mid Z_t)$. For given moneyness, we estimate $\pi(\kappa)$ by a local regression of discounted payoffs $\mu(R_{t+1}, Z_t)h(R_{t+1}, \kappa)$ on current option prices $P_t(\kappa)$. In particular, we estimate the local coefficients as $\hat{\pi}(\kappa) = \frac{1}{T}\sum_{t=1}^{T} \hat{\pi}(\kappa, Z_t)$, where

$$\hat{\pi}^C(\kappa, Z_t) = \frac{\hat{E}_h\left((\hat{\mu}(R_{t+1}, Z_t)h(R_{t+1}, \kappa) - \hat{P}_t(\kappa))P_t(\kappa) \mid Z_t\right)}{\hat{E}_h\left((P_t(\kappa) - \hat{P}_t(\kappa))P_t(\kappa) \mid Z_t\right)}$$

are the locally weighted regression coefficients given Z_t. The conditional expectations in the numerator and denominator can be estimated efficiently using the implied probabilities from the Euclidian empirical likelihood problem, following Antoine et al. (2007). De-centering both payoffs and prices by their implied conditional expectations $\hat{P}_t(\kappa)$ given Z_t automatically removes the intercept when $\hat{\mu}$ is obtained by the CEEL method. The estimated regression coefficients can thus be interpreted as the informative fraction of unexplained variation in option prices of a given moneyness level.

Figure 6 shows the estimated informativeness coefficients $\hat{\pi}(\kappa)$ as a function of moneyness, for a flat pricing kernel and for the sixth order polynomial from Figure 2. The graph reveals that put option prices tend to be less informative about their own payoffs than call option prices, controlling for the VIX as conditioning variable. This could be explained by the VIX already heavily weighting OTM put options in its construction, in line with its interpretation as a 'fear gauge'. Meanwhile, call options appear informative about upside risk that is not fully reflected by the VIX. When taking into account the probability-weighting induced by the pricing kernel ($K = 6$), ATM call options are even more informative, suggesting they are particularly able at predicting large positive returns that are over-weighted by the fitted U-shaped pricing kernel from Figure 2.

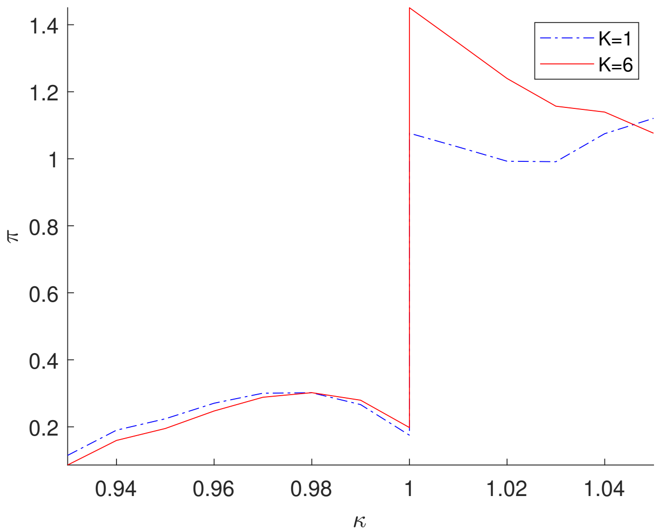


Figure 6: Estimated local informativeness parameter of option prices of varying moneyness, for a flat pricing kernel ($K = 1$) and a sixth order bivariate polynomial ($K = 6$), beyond $Z_t = \log VIX_t$.

3.4 Empirical implied predictive densities

Figure 7 shows the implied conditional density estimates for four week returns on S&P 500 index futures, with the VIX as the conditioning variable, based on the optimization problem (43) with full option price smoothing $\pi = 0$. The implied densities are well-behaved, and feature a left skewness and relatively fat left tail. The scale of the densities increases with the VIX, suggesting the VIX at least partially predicts future uncertainty instead of only affecting the pricing kernel. However, these densities do not reflect any contemporaneous predictive information in option prices.

Figure 8 shows the predictive log likelihood of the minimum relative entropy densities as a function of the informativeness parameter π. Uniformly over π, the likelihood is higher when using the $(K_R, K_Z) = (6, 3)$ instead of the flat $(K_R, K_Z) = (1, 1)$ order bivariate pricing kernel expansions. This confirms the need to correct the local biases in the risk-neutral density due to nonlinear preferences over return outcomes. The predictive likelihood is maximized at $\hat{\pi} = 0.84$

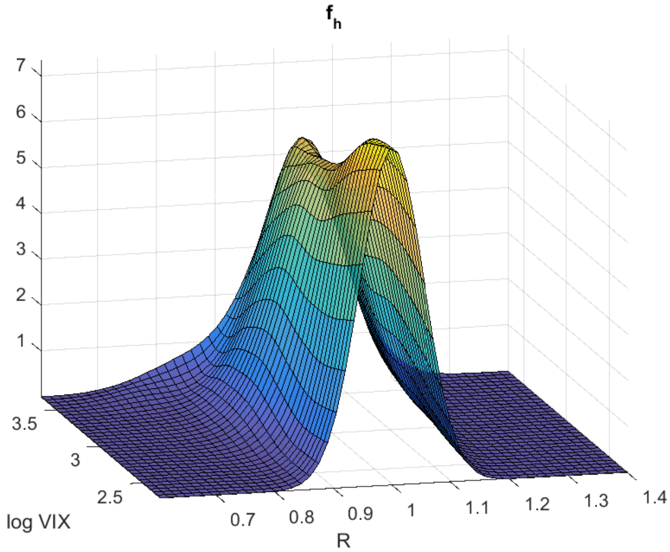


Figure 7: Implied conditional density estimates for four week excess returns on the S&P 500 index given $Z_t = \log VIX_t$, based on $\pi = 0$ and a $(K_R, K_Z) = (6, 3)$ bivariate expansion of the projected pricing kernel.

for the nonlinear dynamic pricing kernel, and at $\hat{\pi} = 0.75$ for the flat pricing kernel. Thus, by controlling for changes in the pricing kernel due to the VIX, option prices become more informative. The figure also shows the predictive log likelihood obtained when using the locally estimated informativeness parameter $\hat{\pi}(\kappa)$, and defining the conditional moments in (44) as

$$\int \mu(y, Z_t)h(y, \kappa_{it})f_t(y)dy = \hat{\pi}(\kappa_{it})P_{it} + (1 - \hat{\pi}(\kappa_{it}))\hat{E}_h\left(P_{it}|Z_t\right), \qquad \forall i.$$

While not targeted, densities obtained using the varying informativeness parameter yield a higher predictive likelihood than using any fixed informativeness parameter. This confirms that once the VIX is conditioned on, call option prices are more informative about upside returns than put options are about downside returns.

Figure 9 shows the resulting implied predictive densities that match, given the fitted pricing kernel, the partially smoothed option prices based on the estimated informativeness function $\hat{\pi}(\kappa)$. The predictive densities reflect the heterogeneity in informativeness across moneyness levels. In particular, the left tail is relatively stable outside of periods of financial markets turbulence during which the VIX was high. Meanwhile, the peaks of the density vary more considerably over time. The densities can be seen to have the largest peaks in the pre-crisis years 2004-2007, reflecting the low market volatility during this period, and vice versa for the financial crisis years 2008-2010.

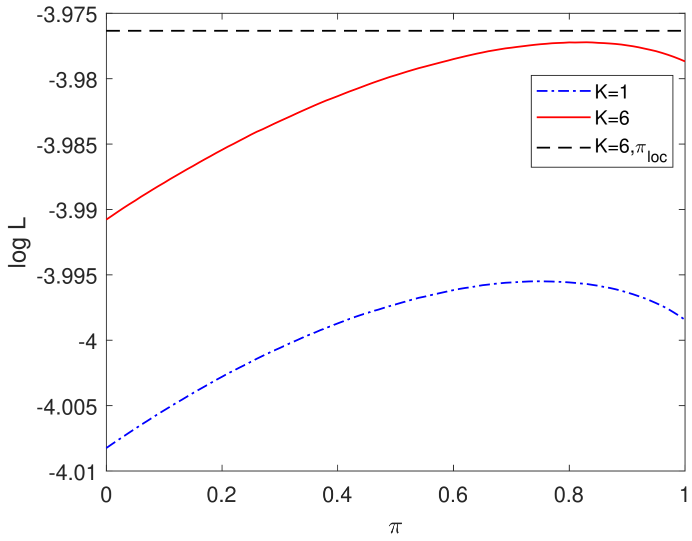


Figure 8: Predictive log likelihoods of four-week S&P 500 index excess returns as a function of the informativeness parameter π, given $Z_t = \log VIX_t$, based on $(K_R, K_Z) = (1, 1)$ and $(K_R, K_Z) = (6, 3)$ order bivariate pricing kernel expansions.

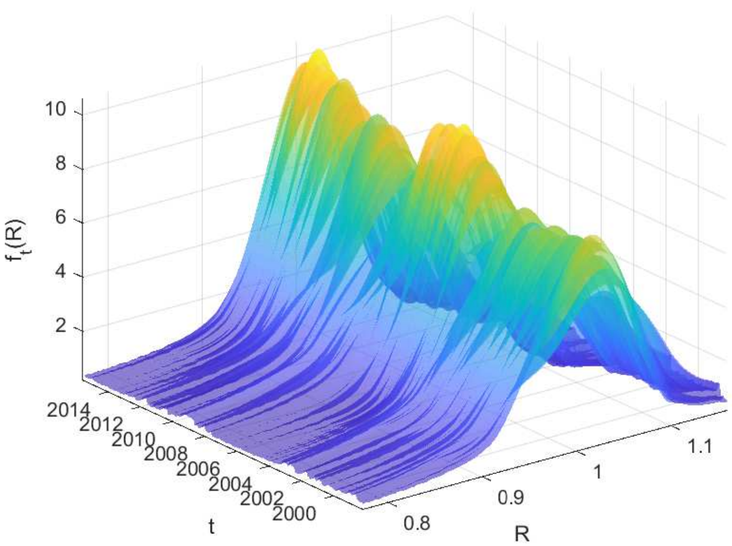


Figure 9: Implied conditional density estimates for four-week S&P 500 index excess returns given $Z_t = \log VIX_t$, based on the estimated informativeness function $\hat{\pi}(\kappa)$ from Figure 6 and a $(K_R, K_Z) = (6, 3)$ order bivariate pricing kernel expansion.

4 Conclusion

This paper studies the nonparametric identification and estimation of projected pricing kernels from conditional moment restrictions on the discounted profits on options, the underlying asset, and a risk-free bond. We avoid the informational mismatch between option-implied and historical return densities by conditioning the pricing kernel on observed variables. The sieve estimator does not rely on ratios of estimated densities, allowing to regularize the estimates in regions with few observations. A conditional Euclidian empirical likelihood formulation yields a continuously-updated efficient estimator, which performs well in a simulation study. Its Lagrange multipliers and implied probabilities are used to characterize option mispricing and option price informativeness, respectively. Empirically, we find pricing kernels that are U-shaped in the S&P 500 index return given high levels of the VIX, and hump-shaped given low levels of the VIX. These confirm non-monotonic shapes in other studies, and provide model- and distributional-free restrictions on structural stochastic discount factor models. In a second stage, we extract option-implied densities that match conditional moments reflecting the informative part of option prices. Predictive densities allowing for heterogeneity in option price informativess across moneyless levels perform better than those based on assigning a fixed weight to the bias-corrected risk-neutral density. This appears primarily due to call and ATM options being more informative about their payoff than put and OTM options, after controlling for the VIX.

A Appendix

A.1 Proofs

Proof of Proposition 1. Denote

$$G(z) = \int_{\kappa^l}^{\kappa^u} \mu(y,z) y f(y|z) dy,$$
$$H(z) = \int_{\kappa^l}^{\kappa^u} \mu(y,z) f(y|z) dy.$$

Both of these are identified by the option prices in the actively traded range. After substituting the identified region, the conditional moment restrictions for the stock (4) and risk free rate (5) can be written as

$$1 = \int_{0}^{\kappa^l} \mu(y,z) y f(y|z) dy + G(z) + \int_{\kappa^u}^{\infty} \mu(y,z) y f(y|z) dy,$$
$$1 = \int_{0}^{\kappa^l} \mu(y,z) f(y|z) dy + H(z) + \int_{\kappa^u}^{\infty} \mu(y,z) f(y|z) dy.$$

Furthermore, for a put option with moneyness κ^l and a call option with moneyness κ^u we have

$$E(P_t \mid Z_t, \kappa^l) = \int_0^{\kappa^l} \mu(y,z)(\kappa^l - y) f(y|z) dy,$$
$$E(C_t \mid Z_t, \kappa^u) = \int_{\kappa^u}^{\infty} \mu(y,z)(y - \kappa^u) f(y|z) dy.$$

Substituting the latter into the stock condition gives

$$1 + E(P_t \mid Z_t, \kappa^l) - E(C_t \mid Z_t, \kappa^u u) - G(z) = \kappa^l \int_0^{\kappa^l} \mu(y,z) f(y|z) dy + \kappa^u \int_{\kappa^u}^{\infty} \mu(y,z) f(y|z) dy,$$

which together with the risk free rate condition yields

$$\begin{pmatrix} \int_0^{\kappa^l} \mu(y,z) f(y|z) dy \\ \int_{\kappa^u}^{\infty} \mu(y,z) f(y|z) dy \end{pmatrix} = \begin{pmatrix} \kappa^l & \kappa^u \\ 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 + E(P_t \mid Z_t, \kappa^l) - E(C_t \mid Z_t, \kappa^u) - G(z) \\ 1 - H(z) \end{pmatrix}, \tag{46}$$

provided that $\kappa^l \neq \kappa^u$, i.e. there is more than one option. Hence the integrated risk-neutral probability masses of the lower and upper tail are identified. The boundary put and call prices then identify the two remaining partial expectations as

$$\int_0^{\kappa^l} \mu(y,z) y f(y|z) dy = \int_0^{\kappa^l} \mu(y,z) f(y|z) dy - E(P_t | Z_t, \kappa^l)$$
$$\int_{\kappa^u}^{\infty} \mu(y,z) y f(y|z) dy = E(C_t | Z_t, \kappa^l) - \int_{\kappa^u}^{\infty} \mu(y,z) f(y|z) dy.$$

Proof of Theorem 1. The proof is based on Lemma A1 in Newey and Powell (2003). This requires checking that (i) there is unique θ_0 that minimizes $Q'(\theta)$ on Θ, (ii) $Q'_T(\theta)$ and $Q'(\theta)$ are continuous, $Q'(\theta)$ is compact, and $\max_{\theta \in \Theta} |Q'_T(\theta) - Q'(\theta)| \xrightarrow{p} 0$, (iii) Θ_T are compact subsets of Θ such that for any $\theta \in \Theta$ there exists a $\tilde{\theta}_T \in \Theta_T$ such that $\tilde{\theta}_T \xrightarrow{p} \theta$ when $T \to \infty$.

The identification condition (i) follows from Section 2.2, together with the positive definiteness of Σ_t.

The compact subset condition in (iii) holds by construction of Θ_T and Θ. Moreover for any $\theta \in \Theta$ we can find a series approximator $\theta_T \in \Theta_T$ that satisfies $\|\theta_T - \theta\| \to 0$.

For (ii), continuity of $Q'_T(\theta)$ follows from the continuity of $\mu(y,z,\theta)$ in θ for all (y,z). The remaining conditions of continuity of $Q'(\theta)$ and uniform convergence can be proved using Lemma A2 in Newey and Powell (2003). This requires pointwise convergence $Q'_T(\theta) - Q'(\theta) \xrightarrow{p} 0$ as well as the stochastic equicontinuity condition that there is a $v > 0$ and $B_T = O_p(1)$ such that for all $\theta, \tilde{\theta} \in \Theta$, $\|Q'_T(\theta) - Q'_T(\tilde{\theta})\| \leq B_T \|\theta - \tilde{\theta}\|^v$. For pointwise convergence, denote

$\epsilon_t(\theta) = P(Z_t, X_t, \theta) - P_t$ and $\hat{\epsilon}_t(\theta) = \hat{P}(Z_t, X_t, \theta) - P_t$, and write

$$\begin{aligned} Q'_T(\theta) - Q'(\theta) &= \frac{1}{T}\sum_{t=1}^{T}\left(\hat{\epsilon}_t(\theta)^T\hat{\Sigma}_t\hat{\epsilon}_t(\theta) - \epsilon_t(\theta)^T\Sigma_t\epsilon_t(\theta)\right) \\ &+ \frac{1}{T}\sum_{t=1}^{T}\epsilon_t(\theta)^T\Sigma_t\epsilon_t(\theta) - E\left(\epsilon_t(\theta)^T\Sigma_t\epsilon_t(\theta)\right). \end{aligned}$$

The first term converges to zero in probability under consistent first stage nonparametric estimation, while the second term does so by the weak law of large numbers. The elements of the summation in the first term can be decomposed as

$$\begin{aligned} \hat{\epsilon}_t(\theta)^T\hat{\Sigma}_t\hat{\epsilon}_t(\theta) - \epsilon_t(\theta)^T\Sigma_t\epsilon_t(\theta) &= \left(\hat{\epsilon}_t(\theta) - \epsilon_t(\theta)\right)^T\hat{\Sigma}_t\hat{\epsilon}_t(\theta) \\ &+ \epsilon_t(\theta)^T\left(\hat{\Sigma}_t - \Sigma_t\right)\hat{\epsilon}_t(\theta) \\ &+ \epsilon_t(\theta)^T\Sigma_t\left(\hat{\epsilon}_t(\theta) - \epsilon_t(\theta)\right) \end{aligned}$$

Since $\hat{\Sigma}_t \xrightarrow{p} \Sigma_t$ with Σ_t bounded,

$$\begin{aligned} E\left(\|\hat{\epsilon}_t(\theta) - \epsilon_t(\theta)\|_2^2\right) &= E\left(\|\hat{P}(Z_t, \kappa_t, \theta) - P(Z_t, \kappa_t, \theta)\|_2^2\right) \\ &= E\left(\left\|\int_0^\infty \mu(y, Z_t, \theta)h(y, \kappa_t)\left(\hat{f}_h(y|Z_t) - f(y|Z_t)\right)dy\right\|_2^2\right) \\ &\to 0 \end{aligned}$$

by consistency of $\hat{f}_h(y|z)$, and

$$\begin{aligned} E\left(\|\epsilon_t(\theta)\|_2^2\right) &\le E\left(\|\epsilon_t(\theta) - \epsilon_t(\theta_0)\|_2^2\right) + E\left(\|\epsilon_t(\theta_0)\|_2^2\right) \\ &= E\left(\|P(Z_t, \kappa_t, \theta) - P(Z_t, \kappa_t, \theta_0)\|_2^2\right) + E\left(\|\epsilon_t\|_2^2\right) \\ &\le C_1\|\theta - \theta_0)\|^{2v} + C_2 = O(1), \end{aligned}$$

using compactness of Θ in the last step, it follows that

$$\frac{1}{T}\sum_{t=1}^{T}\left(\hat{\epsilon}_t(\theta)^T\hat{\Sigma}_t\hat{\epsilon}_t(\theta) - \epsilon_t(\theta)^T\Sigma_t\epsilon_t(\theta)\right) \xrightarrow{p} 0.$$

We prove the stochastic equicontinuity condition for the case $\Sigma_t = I$, which can be generalized to other p.d. matrices. Let $\epsilon(\theta) = (\epsilon_1(\theta), ..., \epsilon_T(\theta))^T$ and similarly define $\hat{\epsilon}(\theta)$. For any $\theta, \tilde{\theta} \in \Theta$,

$$\begin{aligned} |Q'_T(\theta) - Q'_T(\tilde{\theta})| &= \frac{1}{T}|\hat{\epsilon}(\theta)^T\hat{\epsilon}(\theta) - \hat{\epsilon}(\tilde{\theta})^T\hat{\epsilon}(\tilde{\theta})| \\ &\le \frac{1}{T}\|\hat{\epsilon}(\theta) - \hat{\epsilon}(\tilde{\theta})\|_2^2 + \frac{2}{T}\|\hat{\epsilon}(\theta)\|_2\|\hat{\epsilon}(\theta) - \hat{\epsilon}(\tilde{\theta})\|_2, \end{aligned}$$

where

$$\begin{aligned}
\frac{1}{T}\|\hat{\epsilon}(\theta)-\hat{\epsilon}(\tilde{\theta})\|_2^2 &= \frac{1}{T}\sum_{t=1}^{T}\left(\hat{P}(Z_t,\kappa_t,\theta)-\hat{P}(Z_t,\kappa_t,\tilde{\theta})\right)^2 \\
&= \frac{1}{T}\sum_{t=1}^{T}\left(\int_0^\infty \left(\mu(y,Z_t,\theta)-\mu(y,Z_t,\tilde{\theta})\right) h(y,\kappa_t)\hat{f}_h(y|Z_t)dy\right)^2 \\
&\leq \frac{1}{T}\sum_{t=1}^{T}\int_0^\infty b(y,Z_t)^2 h(y,\kappa_t)^2\hat{f}_h(y|Z_t)dy\|\theta-\tilde{\theta}\|^{2v} \\
&\equiv B_T\|\theta-\tilde{\theta}\|^{2v} \\
&\leq B_T C\|\theta-\tilde{\theta}\|^{v},
\end{aligned}$$

and where the last step follows by compactness of Θ. Also

$$\begin{aligned}
\frac{1}{T}\|\hat{\epsilon}(\theta)\|_2\|\hat{\epsilon}(\theta)-\hat{\epsilon}(\tilde{\theta})\|_2 &\leq \frac{1}{T}\|\hat{\epsilon}(\theta)-\hat{\epsilon}(\theta_0)\|_2\|\hat{\epsilon}(\theta)-\hat{\epsilon}(\tilde{\theta})\|_2+\frac{1}{\sqrt{T}}\|\hat{\epsilon}(\theta_0)\|_2\frac{1}{\sqrt{T}}\|\hat{\epsilon}(\theta)-\hat{\epsilon}(\tilde{\theta})\|_2 \\
&\leq \left(B_T C+\frac{1}{\sqrt{T}}\|\hat{\epsilon}(\theta_0)\|_2\sqrt{B_T}\right)\|\theta-\tilde{\theta}\| \\
&\equiv \tilde{B}_T\|\theta-\tilde{\theta}\|.
\end{aligned}$$

Here

$$\begin{aligned}
\frac{1}{T}\|\hat{\epsilon}(\theta_0)\|_2^2 &= \frac{1}{T}\sum_{t=1}^{T}\left(\hat{P}(Z_t,\kappa_t,\theta_0)-P_t\right)^2 \\
&\leq \frac{2}{T}\sum_{t=1}^{T}\left(\hat{P}(Z_t,\kappa_t,\theta_0)-P(Z_t,\kappa_t,\theta_0)\right)^2+\frac{2}{T}\sum_{t=1}^{T}\epsilon_t^2 \\
&= O_p(1),
\end{aligned}$$

by consistency of $\hat{f}_h(y|z)$, continuity and boundedness of $f(y,z)$, and $E(\epsilon_t^2)<\infty$. Since $h^C(R_{t+1},\kappa_t)=(R_{t+1}-\kappa_t)^+<R_{t+1}$,

$$E\left(b(R_{t+1},Z_t)^2h(R_{t+1},\kappa_t)^2|Z_t\right)\leq E\left(b(R_{t+1},Z_t)^2R_{t+1}^2|Z_t\right)<\infty$$

Similarly it follows that $B_T=O_p(1)$, so that $\tilde{B}_T=O_p(1)$, which proves the stochastic equicontinuity condition with $v=1$. This implies $Q'(\theta)$ is continuous and the pointwise convergence holds uniform over Θ. This proves the three conditions. □

Proof of Corollary 1. The identification condition A1 follows from section 2.2. It remains to check the Hölder condition A4. For any $\theta, \tilde{\theta} \in \Theta$,

$$\begin{aligned}
|\mu(y,z,\theta) - \mu(y,z,\tilde{\theta})| &\leq |e^{-\beta_0(z)-\beta_1(z)y} - e^{-\tilde{\beta}_0(z)-\tilde{\beta}_1(z)y}| + |\alpha(y,z) - \tilde{\alpha}(y,z)| \\
&\leq \left|e^{-\beta_0(z)-\tilde{\beta}_1(z)y} + e^{-\beta_0(z)} + e^{-\tilde{\beta}_1(z)y} + 1\right| \\
&\left(\left|e^{-\beta_0(z)} - e^{-\tilde{\beta}_0(z)}\right| + \left|e^{-\beta_1(z)y} - e^{-\tilde{\beta}_1(z)y}\right| + \left|\alpha(y,z) - \tilde{\alpha}(y,z)\right|\right) \\
&\leq \left|e^{-\beta_0(z)-\tilde{\beta}_1(z)y} + e^{-\beta_0(z)} + e^{-\tilde{\beta}_1(z)y} + 1\right| \\
&\left(1 + e^{-\check{\beta}_0(z)} + ye^{-\check{\beta}_1(z)y}\right)\left(\left|\beta_0(z) - \tilde{\beta}_0(z)\right| + \left|\beta_1(z) - \tilde{\beta}_1(z)\right| + \left|\alpha(y,z) - \tilde{\alpha}(y,z)\right|\right) \\
&\leq Cye^{2B_1 y}\left(\|\beta_0 - \tilde{\beta}_0\|_{m,\infty} + \|\beta_1 - \tilde{\beta}_1\|_{m,\infty} + \|\alpha - \tilde{\alpha}\|_{m,\infty}\right) \\
&= Cye^{2B_1 y}\|\theta - \tilde{\theta}\|,
\end{aligned}$$

where the last inequality uses boundedness of the functions (β_0, β_1) and $(\tilde{\beta}_0, \tilde{\beta}_1)$, and C is an arbitrary large constant. From the Hölder inequality

$$E\left(R_{t+1}^3 e^{2B_1 R_{t+1}}|Z_t\right) \leq E\left(|R_{t+1}|^6|Z_t\right)^{\frac{1}{2}} E\left(e^{4B_1 R_{t+1}}|Z_t\right)^{\frac{1}{2}},$$

condition A4 follows with $b(R_{t+1}, Z_t) = R_{t+1}e^{2B_1 R_{t+1}}$.

□

References

Ai, C. and Chen, X. (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica*, 71(6):1795–1843.

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

Andersen, T. G., Fusari, N., and Todorov, V. (2015). Parametric inference and dynamic state recovery from option panels. *Econometrica*, 83(3):1081–1145.

Antoine, B., Bonnal, H., and Renault, E. (2007). On the efficient use of the informational content of estimating equations: Implied probabilities and euclidean empirical likelihood. *Journal of Econometrics*, 138(2):461–487.

Audrino, F. and Meier, P. (2012). *Empirical pricing kernel estimation using a functional gradient descent algorithm based on splines*. Working paper.

Bansal, R. and Viswanathan, S. (1993). No arbitrage and arbitrage pricing: A new approach. *The Journal of Finance*, 48(4):1231–1262.

Barone-Adesi, G., Fusari, N., Mira, A., and Sala, C. (2020). Option market trading activity and the estimation of the pricing kernel: A bayesian approach. *Journal of Econometrics*, 216(2):430–449.

Bates, D. S. (2000). Post-'87 crash fears in the s&p 500 futures option market. *Journal of Econometrics*, 94(1-2):181–238.

Beare, B. K. and Schmidt, L. D. (2014). An empirical test of pricing kernel monotonicity. *Journal of Applied Econometrics*.

Bliss, R. R. and Panigirtzoglou, N. (2004). Option-implied risk aversion estimates. *The journal of finance*, 59(1):407–446.

Blundell, R., Chen, X., and Kristensen, D. (2007). Semi-nonparametric iv estimation of shape-invariant engel curves. *Econometrica*, 75(6):1613–1669.

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

Buchen, P. W. and Kelly, M. (1996). The maximum entropy distribution of an asset inferred from option prices. *Journal of Financial and Quantitative Analysis*, pages 143–159.

Chakravarty, S., Gulen, H., and Mayhew, S. (2004). Informed trading in stock and option markets. *The Journal of Finance*, 59(3):1235–1257.

Chapman, D. A. (1997). Approximating the asset pricing kernel. *Journal of Finance*, pages 1383–1410.

Chen, X. and Ludvigson, S. C. (2009). Land of addicts? an empirical investigation of habit-based asset pricing models. *Journal of Applied Econometrics*, 24(7):1057–1093.

Cochrane, J. H. (2009). *Asset Pricing:(Revised Edition)*. Princeton university press.

Cuesdeanu, H. and Jackwerth, J. C. (2018a). The pricing kernel puzzle in forward looking data. *Review of Derivatives Research*, 21(3):253–276.

Cuesdeanu, H. and Jackwerth, J. C. (2018b). The pricing kernel puzzle: Survey and outlook. *Annals of Finance*, 14(3):289–329.

Darolles, S., Fan, Y., Florens, J.-P., and Renault, E. (2011). Nonparametric instrumental regression. *Econometrica*, 79(5):1541–1565.

Escanciano, J. C., Hoderlein, S., Lewbel, A., Linton, O., Srisuma, S., et al. (2015). Nonparametric euler equation identification and estimation. Technical report, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Gagliardini, P., Gourieroux, C., and Renault, E. (2011). Efficient derivative pricing by the extended method of moments. *Econometrica*, pages 1181–1232.

Gallant, A. R. and Nychka, D. W. (1987). Semi-nonparametric maximum likelihood estimation. *Econometrica: Journal of the Econometric Society*, pages 363–390.

Giacomini, R. and Ragusa, G. (2014). Theory-coherent forecasting. *Journal of Econometrics*, 182(1):145–155.

Grith, M., Härdle, W. K., and Schienle, M. (2012). Nonparametric estimation of risk-neutral densities. *Handbook of Computational Finance*, pages 277–305.

Härdle, W. K., Okhrin, Y., and Wang, W. (2014). Uniform confidence bands for pricing kernels. *Journal of Financial Econometrics*.

Hördahl, P. and Vestin, D. (2005). Interpreting implied risk-neutral densities: the role of risk premia. *Review of Finance*, 9(1):97–137.

Kitamura, Y., Tripathi, G., and Ahn, H. (2004). Empirical likelihood-based inference in conditional moment restriction models. *Econometrica*, 72(6):1667–1714.

Leisen, D. P. (2017). The shape of small sample biases in pricing kernel estimations. *Quantitative Finance*, 17(6):943–958.

León, Á., Mencía, J., and Sentana, E. (2012). Parametric properties of semi-nonparametric distributions, with applications to option valuation. *Journal of Business & Economic Statistics*.

Linn, M., Shive, S., and Shumway, T. (2017). Pricing kernel monotonicity and conditional information. *The Review of Financial Studies*, 31(2):493–531.

Liu, X., Shackleton, M. B., Taylor, S. J., and Xu, X. (2007). Closed-form transformations from risk-neutral to real-world distributions. *Journal of Banking & Finance*, 31(5):1501–1520.

Lu, J. and Qu, Z. (2021). Sieve estimation of option-implied state price density. *Journal of Econometrics*.

Ludvigson, S. C. (2011). Advances in consumption-based asset pricing: Empirical tests. Technical report, National Bureau of Economic Research.

Newey, W. K. and Powell, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica*, pages 1565–1578.

Robertson, J. C., Tallman, E. W., and Whiteman, C. H. (2005). Forecasting using relative entropy. *Journal of Money, Credit and Banking*, pages 383–401.

Rosenberg, J. and Engle, R. F. (2000). Empirical pricing kernels. *Journal of Financial Economics*, (99-014).

Ross, S. (2015). The recovery theorem. *The Journal of Finance*, 70(2):615–648.

Smith, R. J. (2007). Efficient information theoretic inference for conditional moment restrictions. *Journal of Econometrics*, 138(2):430–460.

Stutzer, M. (1996). A simple nonparametric approach to derivative security valuation. *The Journal of Finance*, 51(5):1633–1652.

Vincent-Humphreys, D., Noss, J., et al. (2012). Estimating probability distributions of future asset prices: empirical transformations from option-implied risk-neutral to real-world density functions. *Bank of England Working Paper*.