Risk Neutral Density Estimation with a Functional Linear Model

Marine Carrasco*

Idriss Tsafack†

October 18, 2022

Abstract

This paper proposes a nonparametric estimator of the risk neutral density (RND) based on cross-sectional European option prices. We recast the arbitrage-free equation for option pricing as a functional linear regression model where the regressor is a curve and the independent variable is a scalar corresponding to the option price. Then, we show that the RND can be viewed as the solution of an ill-posed integral equation. To estimate the RND, we use an iterative method called Landweber-Fridman. Then, we establish the consistency and asymptotic normality of the estimated RND. These results can be used to construct a confidence interval around the curve. Finally, some Monte Carlo simulations and application to the S& P 500 options show that this method performs well compared to alternative methods.

Keywords : Risk neutral density, option pricing, regularization, functional regression, Landweber-Fridman, Nonparametric.

*Economics departement, University of Montreal, 3150 rue Jean-Brillant, Montreal, (QC) H3T 1N8, Canada (marine.carrasco@umontreal.ca). Carrasco thanks NSERC for partial financial support.

†Economic Departement, University of Montreal, 3150 rue Jean-Brillant, Montreal, (QC) H3T 1N8, Canada (idriss.tsafack.teufack@umontreal.ca)

1 Introduction

Estimating the risk neutral density (RND) has been an important topic for financial market participants and monetary policymakers. This tool is used for derivatives pricing, hedging and market sentiment analysis. Additionally, it is used to analyze the trader's reaction to a potential shock in the financial market and predict the extreme shocks probabilities. For the policymaker, this tool is used to evaluate the effectiveness of monetary policies through direct observation of changes in investor's expectations and beliefs to future maturities (see Souissi (2017)).

This concept is also fundamental in the arbitrage-free asset pricing theory (see the textbooks by Campbell et al. (1997) and Cochrane (2005)). Indeed, the RND is the density measure under which the price of each security in the market is equal to the expected value of its future payoff discounted back to the present given a risk-free interest rate. This means that for most of the securities in the market, the number of states of the economy could be very large, which in turn leads to situations where the number of potential future payoffs is very dense. Then the future payoffs can be considered as a continuous function of the potential states of the economy. On the same line, since the set of different states of the economy is very large, the RND can be viewed as a continuous function of the future payoffs which form is unknown. Therefore, these properties should be taken into account in the estimation procedure of the RND.

To address the estimation of the functional form of the RND, two main approaches have been suggested in the literature. The first approach relies on parametric modeling (see Black & Scholes (1973), Bahra (1997), Figlewski (2010)), which focus on considering a specific form for the RND and then estimate the related parameters. The most used distribution in this context is the log-normal density (Jarrow & Rudd (1982)) or a linear mixture of the log-normal distributions (Bahra (1997)). Unfortunately, this approach fails to capture all the features of the data (see Ait-Sahalia & Lo (2000)).

The second approach is the nonparametric technique. Several papers use the fact

that the RND is the rescaled second derivative of the put (or call) price. Indeed, for a put option with strike price κ and time to maturity $T-t$, the payoff is $Z(s) = \max(\kappa - s, 0)$ and the put option price P satisfies

$$P(\kappa) = e^{-r(T-t)} \int_0^\kappa (\kappa - s) f(s) \, ds \tag{1}$$

where f is the risk neutral density and r is the risk-free rate. So differentiating (1) twice with respect to κ yields

$$f(S_T) = e^{r(T-t)} \left. \frac{\partial^2 P}{\partial \kappa^2} \right|_{\kappa = S_T}.$$

To recover the RND, one needs to estimate the second derivative of the put (or call) price. Ait-Sahalia & Lo (2000) suggest to estimate the function P by nonparametric kernel regression and then differentiate it twice to recover the RND. However, the resulting estimator is very volatile. Hence, the authors use a semiparametric approach based on Black and Scholes formula and realized volatility. To be fully nonparametric, Ait-Sahalia & Duarte (2003) propose to use a constrained locally polynomial kernel smoothing (see also Grith et al. (2012)) which has the advantage to reduce the variance. Additionally, Jackwerth & Rubinstein (1996) propose a method to extract RND by minimizing a criterion which penalizes unsmoothness. Garcia & Gençay (2000) propose to fit the option pricing function using neural network. While our paper focuses on the estimation of the RND at a specific time t, Panigirtzoglou & Skiadopoulos (2004) and Bliss & Panigirtzoglou (2004) tackle the estimation of the dynamics of RND over time.

Bondarenko (2003) suggests to estimate the density with a nonparametric method called the positive convolution approximation (PosConv). Let $D^{-2}f$ be the second integral of f. For a cross-section of put prices P_i and assuming that the risk free rate r is 0, the density is selected to satisfy the following criterion

$$\min_{\hat{f}} \sum_{i=1}^{n} \left(P_i - D^{-2} \widehat{f}(\kappa_i) \right)^2$$

subject to $\widehat{f}(\kappa) = \sum_j a_j \phi_h(\kappa - z_j)$ for a fine grid of values z_j and $\phi_h(x) = \frac{e^{-\left(x^2/h^2\right)}}{\sqrt{2\pi}h}$, $a_j \geq 0$, $\sum_j a_j = 1$. In other words, the density is approximated by a weighted sum of normal densities, i.e. a linear mixture of normal densities. One of the challenges is to select the number of normal densities to be included and the bandwidth h driving the variance of the normal distributions. The author uses a two-step data driven method to select these tuning parameters.

An alternative method consists in projecting on a polynomial basis. This idea was suggested by Shimko et al. (1993) to estimate the RND and predict the observed implied volatility. Rosenberg (1998) suggests to use a sigma-shaped polynomial technique for the same problem. Also, Yatchew & Härdle (2006) use a nonparametric least-squares with smoothness penalty and are able to impose various shape restrictions. Fengler (2009) uses smoothing spline. More recently, Vogt (2014) propose to approximate the RND with the squared of a series expansion based on Hermite polynomials and Kundu et al. (2018) use Bernstein polynomial basis. For an overview of the parametric and nonparametric methods used to estimate RND, we refer the reader to the book by Jondeau et al. (2010) and the recent review by Figlewski (2018).

To avoid imposing as much restrictions as possible, this paper proposes to estimate nonparametrically the risk neutral density for European option pricing by exploiting the functional data analysis framework. The advantage of this approach is to connect the fundamental theory of asset pricing and the functional feature of the RND while realizing a good fitting performance. Additionally, the estimation does not rely on any latent form of the distribution for approximation. We deal with a functional linear regression model where the predictor is a function representing the future payoffs at the maturity and the response is a scalar representing the call and put price of the considered security. Then, the call and put prices are treated as a weighted sum of all the potential pay-offs of the considered option at the time-to-maturity, with the weights represented by the density values.

The contribution of this paper is to use the functional data analysis framework in order to estimate the risk neutral density estimation. This approach has not been explored in the extant literature. The estimation of the density function in this context leads to an ill-posed inverse problem. Taking the naive inverse would lead to an estimator which would not be consistent. To overcome this issue, we propose to use a regularization technique called the Landweber-Fridman (LF) method. The LF technique is an iterative method used to solve an equation related to option pricing problem, in such a way that the estimation is made without inverting any operator in the procedure. The advantage of this method is that it relies neither on a basis projection nor on a kernel smoothing and helps to reduce the variability of the estimated density coming from the ill-posed problem. Another advantage is the possibility to derive directly the asymptotic normality results and confidence sets for the estimated density and the predictions of option prices. Compared to functional principal components commonly used in functional regression (see Bosq (2000)), Landweber Fridman method requires weaker conditions on the eigenvalues of the covariance operator. In particular, they do not need to be disjoint and multiple eigenvalues are allowed. As the resulting estimator is not a density, we then apply a density correction procedure in order to obtain a nonnegative function which integrates to one. We establish the consistency and the asymptotic normality of the estimated RND. Finally, we analyze its performance using some Monte Carlo simulations and real data of S&P 500. Based on our empirical analysis, we find that the proposed estimation method yields better out-of-sample results compared to the approach by Bondarenko (2003).

This paper is related to the literature on functional data analysis. For a general treatment, see the books by Bosq (2000), Ferraty & Vieu (2006). Theoretical results on the model where the predictor is a function and the response is a scalar are developed by Cardot et al. (1999), T. T. Cai & Hall (2006), Delaigle & Hall (2012), Tsafack (2020) among others. The more general case where both predictor and response variables

are functions is discussed by Kargin & Onatski (2008), Z. Cai et al. (2009), Kokoszka & Zhang (2010), Park & Qian (2012), Chang et al. (2016), Benatia et al. (2017) among others. Park & Qian (2012) consider a functional regression model where the predictor and response variables are densities. In our paper, the density is the functional parameter to be estimated. Another difference is that Park & Qian (2012) tackle the ill-posed problem in their model by estimating the operator via the functional principal components (FPCA) while we use LF method. Compared to FPCA, LF has the advantage to require weaker assumptions on the eigenvalues of the covariance operator of the predictor function. In a follow-up paper, Chang et al. (2016) analyze the nonstationarity in the time-series of densities via a unit root test.

The rest of the paper is organized as follows. Section 2 introduces the theoretical model and the estimation method. Section 3 establishes the rate of convergence and the asymptotic normality results of the estimator. It also suggests a data-driven method to select the optimal tuning parameter. Section 4 presents the results of the simulations. Section 5 is dedicated to the empirical analysis. Section 6 concludes. An online appendix describes the implementation and contains all the proofs. An online appendix $https://www.dropbox.com/s/bk13cppnqnqz4j0/RND\_Appendix.pdf?dl=0$ describes the implementation and contains all the proofs.

2 Estimation of risk-neutral density

This section shows that the risk-neutral density can be viewed as the solution to an integral equation and explains how to estimate it using Landweber-Fridman estimation technique.

2.1 Option pricing formula

In the intertemporal equilibrium models, the current price of a security can be expressed as the expected net present value of its future payoffs discounted back to the present. The expectation is obtained with respect to the risk-neutral density (Cox & Ross (1976)), also called state-price density or equivalent-martingale measure (see Harrison & Kreps (1979)). More specifically, in the derivative market, an option is defined as a contract giving the right (and not the obligation) to buy or sell a risky asset with price s at a predetermined value called strike price κ at (or within) a given maturity date of the contract. There exist many kinds of options in the market. The main ones are the American and the European options. In this paper, we focus on European options characterized by the fact that the exercise of the contract is possible only at the given maturity date.

Then, in the context of a complete market, the price of a European put option P_t with a maturity $T-t$, an underlying price at maturity S_T and a strike price κ, is equal to the expected pay-offs $Z(S_T)$ discounted back to the present. In other words, it is given by

$$P_t = e^{-r_{t,T}(T-t)} \int_0^\infty Z(S_T) f(S_T|(T-t)) dS_T \tag{2}$$

where $f(S_T|(T-t))$ is the unobserved risk-free density (RND) of S_T conditional to the maturity T, $r_{t,T}$ is the riskless interest rate between date t and T, and $Z(S_T) = max(\kappa - S_T, 0)$ the pay-off. To simplify notation, we will denote $\tau = T-t$ and $r_{t,T} = r$. It is important to mention that we use a cross-section of option prices all observed at the same time t.

The previous equation holds when it is assumed that the market is complete, this means that market participants have all the informations about the risky assets. Because of illiquidity in the market, transaction cost, taxes, measurement errors, the market is usually incomplete (see Gourieroux & Jasiak (2001)). Then, it may exist an error term

capturing all those uncontrolled informations and this uncertainty may vary according to the time to maturity of the option. The longer the time-to-maturity is, the bigger the variability (see Ait-Sahalia et al. (2018), Driessen et al. (2009)). Then, for each option i (corresponding to a strike price κ_i) at a same fixed time t and the same maturity τ, we have the following equation :

$$Y_i = \int_0^{\infty} Z_i(s) f(s|\tau) ds + \varepsilon_i \tag{3}$$

where $Y_i = e^{r\tau} P_i$, ε_i is a conditionally zero-mean, homoskedastic error term. For the sake of this model we assume that there is an infinite possibilities of pay-offs at the maturity date, which means that the set of potential payoffs is very dense and the conditional density is a function taking its values on the real line $\mathbb{R}$. This leads to a functional linear regression with the functional predictor represented by the future payoff $Z_i(s) = max(\kappa_i - s, 0)$ and a scalar response (Y_i). The call options are also considered by using the appropriate payoff $Z_i(s) = max(s - \kappa_i, 0)$.

2.2 RND as the solution of an integral equation

Let us define $\mathbb{H} = L^2([0, +\infty))$ the space of square integrable functions mapping from the interval $[0, +\infty)$ to the set of real numbers $\mathbb{R}$. $\mathbb{H}$ is a Hilbert-space endowed with an inner product $< ., . >$ and a norm $||.||$, which are respectively defined as follows: $< f, g >= \int_0^{+\infty} f(t)g(t)dt$ and $||f|| = \left(\int_0^{+\infty} f^2 \right)^{1/2}$.

Let us consider the sample $((\kappa_1, Y_1), ..., (\kappa_n, Y_n))$ of independent pairs following the same distribution as the population version (κ, Y). We consider the functional linear model where $(Z_i)_{i=1...n}$ is the sample of functional predictor variables of the regression representing the set of possible pay-offs at maturity for each option and $(Y_i)_{i=1...n}$ is the scalar response.

For each $i \in \{1, ..., n\}$, $Z_i(s) = max(\kappa_i - s, 0)$, hence Z_i is random only through κ_i. Additionally, the predictor function $(Z_i(s))_{i=1...n}$ is such that $Z_i(s) \geq 0$, this means

that $E(Z_i(s)) \geq 0$ for each $s \in [0, +\infty)$ and $E(Y_i) > 0$. We assume that $E[\kappa_i^3] < +\infty$. Indeed, this assumption guarantees that $\int_0^{+\infty} E(Z^2(s))ds < \infty$, which means that the predictor function is square integrable (see Lemma 1 in Appendix). Then, for each time t the model is a cross-sectional regression presented as follows :

$$Y_i = \int_0^{+\infty} Z_i(s)f(s)ds + \varepsilon_i \tag{4}$$

where $(f(s))_{s\in[0,+\infty)}$ is a function that belongs to the space $\mathbb{H}$ and ε_i, $i = 1, ..., n$, are independent and homoskedastic[1] such that $\mathbb{E}(\varepsilon_i|\kappa_1, ..., \kappa_n) = 0$ and $\mathbb{E}(\varepsilon_i^2|\kappa_1, ..., \kappa_n) = \sigma^2 < \infty$ for each $i \in \{1, ..., n\}$.

By premultiplying both sides of Equation (4) by $Z_i(u)$ and taking the expectation, we obtain

$$\mathbb{E}[Z_i(u)Y_i] = \int_0^{+\infty} \mathbb{E}[Z_i(u)Z_i(s)]f(s)ds + \mathbb{E}[Z_i(u)\varepsilon_i].$$

Since $\mathbb{E}[Z_i(u)\varepsilon_i] = 0$, then

$$\mathbb{E}[Z_i(u)Y_i] = \int_0^{+\infty} \mathbb{E}[Z_i(u)Z_i(s)]f(s)ds.$$

In a compact form we can write

$$C_{zy} = Kf, \tag{5}$$

where $C_{zy}(u) = \mathbb{E}[Z_i(u)Y_i]$ is the cross-covariance function between the predictor variable Z and the response variable Y and K is the covariance operator from $\mathbb{H}$ to $\mathbb{H}$ defined as

$$Kf = \int_0^{\infty} \mathbb{E}[Z_i(u)Z_i(s)]f(s)ds.$$

Our main goal is to estimate the function f solution of (5). Equation (5) is a Fredholm

[1]Heteroskedasticity could be introduced if one considers options with different time-to-maturity. The same estimation procedure could be applied but the standard errors would need to be adjusted to account for heteroskedasticity.

equation of the first kind. Solving this equation is an ill-posed problem as K is a bounded operator mapping from an infinite dimensional space $\mathbb{H}$ to $\mathbb{H}$. This means that the direct inverse of K is not continuous and K is not invertible in $\mathbb{H}$ but only on a subset of $\mathbb{H}$. If the operator K were invertible, we could estimate f using $f(s) = K^{-1}C_{zy}(s)$ for each $s \in [0, +\infty)$. However, in our context, estimating f by $\hat{K}^{-1}\hat{C}_{zy}$ would lead to an unstable estimator of the functional parameter, which would not be consistent (see Carrasco, Florens, and Renault, 2007). To overcome this issue, we propose to use a regularization technique called Landweber-Fridman (LF) method. This method will stabilize the inverse of K and permits to obtain a consistent estimator of f. In the next section, we will present the LF method.

2.3 The Landweber-Fridman method

Recall that we want to estimate f solution of the equation $C_{zy} = Kf$. The main idea of the Landweber-Fridman method is to approach the solution to this equation by an iterative algorithm similar to the fixed point procedure with the goal of minimizing $\|C_{zy} - Kf\|$. Instead of iterating all the way to convergence, the algorithm stops after a finite number of iterations. The early termination stabilizes the solution. It is a regularization technique used in the inverse problem literature (see Engl et al. (1996) and Carrasco et al. (2007)). The algorithm is presented below. Let ω be a constant such that $0 \leq \omega \leq 1/\lambda_1$, where λ_1 is the largest eigenvalue of K.

- At the first iteration, take $f_0(.) = \omega C_{zy}(s)$.
- For $h = 1, ..., \frac{1}{\alpha} - 1$, calculate

$$f_h(s) = f_{h-1}(s) + \omega(C_{zy}(s) - Kf_{h-1}(s)) \tag{6}$$

  where α is a regularization parameter chosen so that $\frac{1}{\alpha} - 1$ is an integer, hence $0 < \alpha < 1$.

- For convenience, the resulting estimator f_h is denoted f_α with $f_\alpha(s) = K_\alpha^{-1} C_{zy}(s)$, for each $s \in [0, +\infty)$ and K_α^{-1} denotes the regularized inverse of K defined below.

After $\frac{1}{\alpha} - 1$ iterations, the regularized inverse of K is given by

$$K_\alpha^{-1}\phi(s) = \omega \sum_{l=0}^{\frac{1}{\alpha}-1} (I - \omega K)^l \phi(s)$$

where $s \in [0, +\infty)$. The regularization parameter α will be chosen via a data driven method described later. Let us denote $(\lambda_j, v_j)_{j\geq 1}$ the eigensystem of the covariance operator K, then we can also write K_α^{-1} in terms of the eigensystem of K as follows

$$K_\alpha^{-1}\phi = \sum_{j=1}^{\infty} \frac{q(\alpha, \lambda_j)}{\lambda_j} < v_j, \phi > v_j$$

for each function ϕ and $q(\alpha, \lambda_j) = \left[1 - (1 - \omega\lambda_j)^{1/\alpha}\right]$.

The true operators K_α^{-1}, K and C_{zy} are unobservable. In practice, they are replaced by their empirical counterparts. Then, the estimated density function is given by $\hat{f}_\alpha(s) = \hat{K}_\alpha^{-1}\hat{C}_{zy}(s)$. In other words, we have

$$\hat{f}_\alpha(s) = \omega \sum_{l=0}^{\frac{1}{\alpha}-1} (I - \omega\hat{K})^l \hat{C}_{zy}(s) \tag{7}$$

where $\hat{K}$ is the integral operator from $\mathbb{H}$ to $\mathbb{H}$ with kernel

$$\hat{k}(u, s) = \frac{1}{n} \sum_{i=1}^{n} Z_i(u) Z_i(s)$$

and

$$\hat{C}_{zy}(s) = \frac{1}{n} \sum_{i=1}^{n} Z_i(s) Y_i.$$

In Section 2 of the online appendix, we explain how to compute $\hat{f}_\alpha$ using only products of matrices instead of operators. This is how we implement the method in the simulations and application.

Other regularization techniques could be used to solve (5). For instance, Tikhonov regularization is a popular method which has been used by Benatia et al. (2017) for the estimation of a functional regression with functional response. It is easy to implement but suffers from saturation (its rate of convergence can not improve beyond a certain level) while LF does not suffer of saturation. Another popular regularization technique is principal components, but analyzing its properties requires some assumptions on the decay rate of the eigenvalues λ_j which are not needed here, see for instance Hall & Horowitz (2007) for an application of principal components to a functional regression with scalar response.

2.4 Density correction

Our estimator of f, $\hat{f}_\alpha$, is not necessarily positive and does not integrate to one. As the true function f is a density, we propose to transform our estimator $\hat{f}_\alpha$ into a density using the methods proposed by Glad et al. (2003). The correction is different depending on whether $\int_0^\infty \max\left\{0, \hat{f}_\alpha(s)\right\} ds \geq 1$ or $\int_0^\infty \max\left\{0, \hat{f}_\alpha(s)\right\} ds < 1$.

Case 1: Case where $\int_0^\infty \max\left\{0, \hat{f}_\alpha(s)\right\} ds \geq 1$.

The corrected estimator is given by

$$\tilde{f}_\alpha(s) = \max\left\{0, \hat{f}_\alpha(s) - \xi\right\}$$

where ξ is a positive constant chosen so that $\int_0^\infty \tilde{f}_\alpha(s)\, ds = 1$.

Note that such a ξ necessarily exists because $\int_0^\infty \tilde{f}_\alpha(s)\, ds \geq 1$ for $\xi = 0$ and $\int_0^\infty \tilde{f}_\alpha(s)\, ds = 0$ for $\xi = \infty$.

Case 2: Case where $\int_0^\infty \max\left\{0, \hat{f}_\alpha(s)\right\} ds < 1$.

The corrected estimator $\check{f}_\alpha$ is computed as follows

$$\check{f}_\alpha(s) = \begin{cases} \max\left\{0, \hat{f}_\alpha(s)\right\} + \eta_M & \text{for } |s| \leq M, \\ \max\left\{0, \hat{f}_\alpha(s)\right\} & \text{for } |s| > M, \end{cases}$$

where

$$\eta_M = \frac{1}{2M}\left[1 - \int_0^\infty \max\left\{0, \hat{f}_\alpha(s)\right\} ds\right].$$

Remarks.

1. In Case 1, Theorem 1 of Glad et al. (2003) shows that $\tilde{f}_\alpha$ is always better than $\hat{f}_\alpha(s)$ in the sense that $\left\|\tilde{f}_\alpha - f\right\|^2 \leq \left\|\hat{f}_\alpha - f\right\|^2$ for all n, almost surely, hence $E\left\|\tilde{f}_\alpha - f\right\|^2 \leq E\left\|\hat{f}_\alpha - f\right\|^2$ so that the mean integrated squared error (MISE) of $\tilde{f}_\alpha$ is always smaller than that of $\hat{f}_\alpha$.

2. In Case 2, Theorem 2 of Glad et al. (2003) establishes that

$$E\left\|\check{f}_\alpha - f\right\|^2 \leq E\left\|\hat{f}_\alpha - f\right\|^2 + \frac{3}{2M}.$$

Hence, one can make the MISE of $\check{f}_\alpha$ arbitrary close to that of $\hat{f}_\alpha$ by choosing M dependent of n and large, for instance $M = n$.

3. An algorithm to select ξ in practice is presented in Section 1 of the online appendix.

4. An alternative correction of $\hat{f}_\alpha$ could have relied on a normalization

$$\frac{\max\left\{0, \hat{f}_\alpha\right\}}{\int_0^\infty \max\left\{0, \hat{f}_\alpha(s)\right\} ds}. \tag{8}$$

However, there is no garantee that this normalization improves the accuracy of the estimator. The MISE of the normalized estimator may actually be worse than that of the original estimator as discussed in Glad et al. (2003).

3 Asymptotic properties of the Landweber-Fridman estimator

3.1 Convergence rate

In this section, we derive the convergence rate of the conditional mean square error (MSE) of $\hat{f}_\alpha$. For this purpose, the following assumptions are imposed.

A1. (κ_i, Y_i) are i.i.d. κ_i has a continuous density on $\mathbb{R}^+$with $E[\kappa_i^6] < \infty$.

A2. $\int_0^{+\infty} f^2(t)dt < \infty$, $E[\varepsilon_i|\kappa_1, ..., \kappa_n] = 0$, $E[\varepsilon_i^2|\kappa_1, ..., \kappa_n] = \sigma^2 < +\infty$, $E[\varepsilon_i^4|\kappa_1, ..., \kappa_n] < \infty$.

A3. The eigenvalues of the covariance operator K are positive and ordered in decreasing order, that is $\lambda_1 \geq \lambda_2 \geq ... > 0$. Similarly, the eigenvalues of $\widehat{K}$ are ordered in decreasing order, $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq ... \geq \hat{\lambda}_n$.

A4. We assume that for some $\mu \geq 0$, f satisfies

$$\sum_{j=1}^{\infty} \frac{< f, v_j >^2}{\lambda_j^\mu} < \infty.$$

Assumption **A1** imposes that $(Z_i, Y_i)_{i=1,...n}$ are independent, identically distributed as (Z, Y). It is useful in order to derive the consistency of the covariance operators $\hat{K}$ and $\hat{C}_{zy}$, and to prove the central limit properties of the estimated functions. The fact that $E[\kappa_i^3] < +\infty$ guarantees that the predictor functions $(Z_i)_{i=1,...,n}$ are square integrable. Moreover, it also guarantees that the covariance operator K is nuclear, which in turn implies that it is Hilbert-Schmidt[2] (see Lemma 1 in Appendix). However, we need a stronger condition ($E[\kappa_i^6] < +\infty$) to guarantee that $E\left(\|Z_i\|^4\right) < \infty$ and hence show the consistency of $\hat{K}$ to K (see Lemma 1 in Appendix and comment below).

Assumption **A2** imposes that the error term ε_i is homokedastic and Z_i is exogenous. The homoskedasticity assumption is reasonable because we use cross-sectional data.

[2] A covariance operator is nuclear if and only if its eigenvalues are summable, i.e. $\sum_{j=1}^{\infty} \lambda_j < \infty$. It is Hilbert-Schmidt if and only if $\sum_{j=1}^{\infty} \lambda_j^2 < \infty$.

It could be relaxed at the cost of a more complicated expression for the asymptotic variance. **A1** and **A2** are sufficient conditions to ensure that $||\hat{K} - K||^2_{HS} = O_p\left(\frac{1}{n}\right)$, see the proof of Theorem 4 of Carrasco & Florens (2000), where $||.||_{HS}$ is the Hilbert-Schmidt norm of operators. Moreover, assumptions A1 and A2 imply that the fourth moment of Y_i exists, this rules out possible fat-tails of the distribution of Y_i. To relax this assumption, one would need to use appropriate statistical tools (see Ibragimov et al. (2015)) In that case, we expect that the rate of convergence and asymptotic distribution would be different.

Assumption **A3** ensures that the null space of K reduces to 0, $\mathcal{N}(K) = \{0\}$. Hence, f is the unique solution of $C_{xy} = Kf$ and therefore it is identified. Note that the eigenvalues do not have to be distinct and multiple eigenfunctions may be associated with the same eigenvalue. This is quite a bit more general than assumptions usually imposed for principal component method where the eigenvalues need to be distinct and sufficiently spaced from each other (see for instance Hall & Horowitz (2007)).

Assumption **A4** is a source condition important to derive how the bias and estimation error terms behave. To satisfy this condition, the Fourier coefficients of f, $< f, v_j >$, need to decline to zero fast compared to the eigenvalues of K. The parameter μ characterizes the severity of the ill-posed problem. Larger values of μ are associated with less severe ill-posed problem. This source condition is discussed in Engl et al. (1996) and Carrasco et al. (2007) and was used in econometric papers by Chen & Reiss (2011), Darolles et al. (2011), and Gagliardini & Scaillet (2012) among others.

Proposition 1. *Let $\ddot{f}_\alpha$ be the estimated density function corrected either with Case 1 or 2. Under assumptions A1 - A4, if $\alpha^2 n \to \infty$, then*

$$\mathbb{E}\left[\left|\left|\ddot{f}_\alpha - f\right|\right|^2 |\kappa_1, ..., \kappa_n\right] = O_p\left(\alpha^\mu\right) + O_p\left(\frac{1}{\alpha^2 n}\right) \tag{9}$$

where μ is the nonnegative constant defined in Assumption A4. Hence the conditional MISE converges to zero as the sample size increases.

Remarks.

- The first term of the conditional MISE is the squared bias and the second term is the estimation error.
- As α goes to zero, the squared bias term goes to zero, while the estimation error term increases. So, we have the usual trade-off between bias and variance. The optimal parameter α is selected in such a way that the squared bias and the variance term are of the same order.
- If $\alpha \sim n^{-1/(2+\mu)}$, then $MISE \sim n^{-\frac{\mu}{2+\mu}}$.

3.2 Asymptotic normality

Carrasco et al. (2014) derive results on the asymptotic normality of $\hat{f}_\alpha - f_\alpha$ for a fixed α. While this approach can be useful to perform hypothesis testing, it cannot be used to construct confidence intervals of f. Here, we are going to study the pointwise asymptotic normality assuming α goes to zero instead. For related results, see Horowitz (2007) in the context of nonparametric instrumental variables and Carrasco & Florens (2011) who study the asymptotic normality of a deconvolution estimator.

Replacing $\hat{f}_\alpha$ and f_α by their expressions, we obtain

$$\hat{f}_\alpha - f_\alpha = K_\alpha^{-1}\widehat{C}_{z\varepsilon} + \left(\hat{K}_\alpha^{-1} - K_\alpha^{-1}\right)\widehat{C}_{z\varepsilon} + \left(\hat{K}_\alpha^{-1}\widehat{K} - K_\alpha^{-1}K\right) f.$$

The distribution of $\hat{f}_\alpha - f_\alpha$ will be driven by that of $K_\alpha^{-1}\widehat{C}_{z\varepsilon}(s)$. As α depends on n and $K_\alpha^{-1}\widehat{C}_{z\varepsilon}(s) = \frac{1}{n}\sum_{i=1}^n \varepsilon_i \left(K_\alpha^{-1}Z_i\right)(s)$ with $E\left(\varepsilon_i \left(K_\alpha^{-1}Z_i\right)(s)\right) = 0$, it follows that $K_\alpha^{-1}\widehat{C}_{z\varepsilon}(s)$ is a triangular array. A sufficient condition for

$$\frac{\left(K_\alpha^{-1}\widehat{C}_{z\varepsilon}\right)(s)}{\sqrt{V\left(\left(K_\alpha^{-1}\widehat{C}_{z\varepsilon}\right)(s)\right)}} \xrightarrow{d} \mathcal{N}(0,1)$$

is that the Lyapunov's condition holds (Billingsley (1995), Theorem 27.3), i.e., for some $\delta > 0$,

$$\frac{E\left[\left|\varepsilon_i \left(K_\alpha^{-1} Z_i\right)(s)\right|^{2+\delta}\right]}{n^{\delta/2}\left[E\left(\left(\varepsilon_i \left(K_\alpha^{-1} Z_i\right)(s)\right)^2\right)\right]^{1+\delta/2}} \to 0. \tag{10}$$

A sufficient condition for (10) is given in the next assumption.

Assumption A5. $\frac{E\left[\left|\varepsilon_i\left(K_\alpha^{-1} Z_i\right)(s)\right|^3\right]}{n^{1/2}\left[E\left(\left(\varepsilon_i\left(K_\alpha^{-1} Z_i\right)(s)\right)^2\right)\right]^{3/2}} \to 0.$

To obtain

$$\frac{\hat{f}_\alpha(s) - f_\alpha(s)}{\sqrt{\frac{1}{n} E\left(\left(\varepsilon_i \left(K_\alpha^{-1} Z_i\right)(s)\right)^2\right)}} \xrightarrow{d} \mathcal{N}(0,1), \tag{11}$$

it is sufficient that the following condition is satisfied.

Assumption A6.

$$\frac{\left|\left(\hat{K}_\alpha^{-1} - K_\alpha^{-1}\right)\widehat{C}_{z\varepsilon}(s) + \left(\hat{K}_\alpha^{-1}\widehat{K} - K_\alpha^{-1}K\right) f(s)\right|}{\sqrt{\frac{1}{n} E\left(\left(\varepsilon_i \left(K_\alpha^{-1} Z_i\right)(s)\right)^2\right)}} \xrightarrow{P} 0.$$

Finally, to be able to replace f_α by f in (11), we need an extra condition guaranteeing that the bias is negligible.

Assumption A7.

$$\frac{\left|\sum_{j=1}^{\infty}\left(q(\alpha, \lambda_j) - 1\right)\langle v_j, f\rangle v_j(s)\right|}{\sqrt{\frac{1}{n} E\left(\left(\varepsilon_i \left(K_\alpha^{-1} Z_i\right)(s)\right)^2\right)}} \xrightarrow{P} 0.$$

Remark that Assumptions A5 to A7 restrict the rate of convergence of α and possibly the admissible range of values of s. As a result, the rate of convergence of $\hat{f}_\alpha$ depends on s.

An extra assumption is needed to be able to consistently estimate the variance.

Assumption A8. $E(\varepsilon_i^4) < \infty$, $\frac{1}{n} E\left[\left(\left(K_\alpha^{-1} Z_i\right)(s)\right)^4\right] \to 0$, and $n\alpha^2 \to \infty$.

Proposition 2. *Under Assumptions A1 to A7,*

$$\frac{\hat{f}_\alpha(s) - f(s)}{\sqrt{\frac{1}{n}E\left(\left(\varepsilon_i\left(K_\alpha^{-1}Z_i\right)(s)\right)^2\right)}} \xrightarrow{d} \mathcal{N}(0,1).$$

Moreover, under Assumptions A1 to A8,

$$\frac{\hat{f}_\alpha(s) - f(s)}{\sqrt{\widehat{V}_n(s)}} \xrightarrow{d} \mathcal{N}(0,1)$$

where $\widehat{V}_n(s) = \widehat{\sigma}^2 \frac{1}{n^2}\sum_{i=1}^n \left(\left(\hat{K}_\alpha^{-1}Z_i\right)(s)\right)^2$ *and* $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n \left(Y_i - \int z_i(s)\,\hat{f}_\alpha(s)\,ds\right)^2$.

It is important to notice that Proposition 2 does not imply that $\hat{f}_\alpha(s)$ converges at a $\sqrt{n}$ rate of convergence because the term $E\left(\left(\varepsilon_i\left(K_\alpha^{-1}Z_i\right)(s)\right)^2\right)$ usually diverges so that the rate is slower. The previous results can be used to construct an asymptotic confidence interval for $f(s)$.

Corollary 1 *Under Assumptions A1 to A8, the asymptotic* $1-a$ *confidence interval for* $f(s)$ *is given by*

$$\hat{f}_\alpha(s) - z_{a/2}\widehat{V}_n(s)^{1/2} \leq f(s) \leq \hat{f}_\alpha(s) + z_{a/2}\widehat{V}_n(s)^{1/2}$$

where $z_{a/2}$ *is the* $1-a/2$ *quantile of the standard normal distribution.*

3.3 Data-driven selection of the tuning parameter

According to the consistency results, it can be noticed that the estimation of the RND depends on the tuning parameter α. Then, this parameter should be selected optimally. Since the main goal is to estimate the RND and therefore predict the call and put prices, we define a prediction criterion to select the optimal parameter α. Then, we choose the regularization parameter in such a way that the mean squared prediction error (MSPE) is minimized. We use the K-fold cross-validation for the selection procedure. Let us

split the initial sample into M subsamples denoted I_1,...,I_M.

$$\alpha_{op} = \underset{\alpha \in \mathcal{I}_\alpha}{\operatorname{argmin}} \frac{1}{M} \sum_{\ell=1}^{M} \frac{1}{card(I_\ell)} \sum_{j \in I_\ell} \left(Y_j - \hat{Y}_j \right)^2. \tag{12}$$

For $\ell \in \{I_1,...,I_M\}$, we estimate the parameter f in the sample $\mathcal{I}_{-\ell}$ representing all the observations not in $\mathcal{I}_\ell$. Then, we predict the response variable in $\mathcal{I}_\ell$ considered as the hold-out sample. $\hat{Y}_j$ is the prediction of the j^{th} observation in $\mathcal{I}_\ell$. Hence, we calculate the MSPE for each candidate α. $\mathcal{I}_\alpha$ is the set of candidate α.

An alternative approach suggested by Engl et al. (1996) is to choose the parameter α such that the following objective function is minimized.

$$\alpha_{op} = \underset{\alpha \in \mathcal{I}_\alpha}{\operatorname{argmin}} \left|\left| \hat{f}_\alpha \right|\right|^2 \left|\left| \hat{C}_{zy} - \hat{K}(\hat{f}_\alpha) \right|\right|^2. \tag{13}$$

4 Simulations

In this section, the Monte-Carlo simulations are used to evaluate the proposed estimation method. For this purpose, we consider a variety of data generating processes. We compare the proposed estimation method with the positive convolution approach (PosConv) suggested by Bondarenko (2003) considered as a benchmark. This simulation follows the same idea as the one by Bondarenko (2003) with a change on the underlying discretization, the error term distribution and the dates. We consider the functional regression model.

$$Y_i = \int_0^{+\infty} Z_i(s) f(s) ds + \varepsilon_i \tag{14}$$

where $(Y_i, Z_i)_{i=1,...n}$ is the sample of generated data. The response variable $Y_i = \exp(r\tau) P_i$ where P_i are call option prices and r is set equal to 0. The predictors are the pay-offs functions Z_i where $Z_i(s) = \max\{s - \kappa_i, 0\}$. To characterize the incompleteness of the market and allow for measurement errors, an error term ε_i is added to the option

price with $\mathbb{E}[\varepsilon_i] = 0$. This error term is assumed to follow a uniform distribution, $\varepsilon_i \sim \mathcal{U}[-0.5, 0.5]$.

The true RND $f(s)$ is specified as a linear mixture of lognormal distributions presented as follows

$$\begin{aligned} f(s) &= \pi_1 \mathcal{LN}(s|\eta_1, \sigma_1) + \pi_2 \mathcal{LN}(s|\eta_2, \sigma_2) + \pi_3 \mathcal{LN}(s|\eta_3, \sigma_3), \\ \pi_1 + \pi_2 + \pi_3 &= 1, \end{aligned}$$

and $\mathcal{LN}(s|\eta_j, \sigma_j)$, $j = 1, 2, 3$ is a lognormal distribution

$$\mathcal{LN}(s|\eta_j, \sigma_j) = \frac{1}{\sqrt{2\pi}\sigma_j s} e^{-\frac{\left(\ln\left(\frac{s}{\eta_j}\right) - \frac{1}{2}\sigma_j^2\right)^2}{2\sigma_j^2}}.$$

Four different models are considered. Model 1 and 2 correspond to option data with maturity between one to two months. Model 3 considers options with maturity between three and six months and Model 4 considers options with maturity exceeding six months.

The set of strike prices is defined as follows $\mathcal{K} = [1500, 1505, 1510, ..., 3000]$, which means that the sample size is $n = 301$ and the underlying follows the aforementioned lognormal mixture distribution. This parametrization is used to match a typical cross-section of the $S\&P$ 500 index options traded at the Chicago Board Options Exchange (CBOE) on June 25, 2017 for Model 1, August 04, 2017 for Model 2 and 4, and June 05, 2017 for Model 3. The parameters of the RND are presented in Table 1.

We simulate the data and evaluate the performance of the estimation method with 2 criteria :

- The Root Mean Squared Prediction Error (RMSPE) between the estimated put prices $\hat{Y}_\alpha$ and the theoretical one Y.

$$RMSPE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_{i,\alpha} - Y_i)^2}.$$

Table 1: Parameters of the lognormal mixture density

Parameters	Model 1	Model 2	Model 3	Model 4
π_1	0.0812	0.0823	0.0562	0.5934
π_2	0.0914	0.8115	0.3347	0.2894
π_3	0.8274	0.1062	0.5791	0.1172
η_1	7.8020	7.7669	7.5628	7.8594
η_2	7.7023	7.8176	7.7690	7.7779
η_3	7.8052	7.8165	7.8340	7.5688
σ_1	0.0285	0.0543	0.2032	0.0369
σ_2	0.0987	0.0214	0.0599	0.0696
σ_3	0.0245	0.0247	0.0323	0.2090
Time to maturity (days)	53	35	148	224
Risk free rate (r in %)	0.95	1.06	0.95	1.06
Current index price (S_0)	2436.10	2476.83	2436.10	2476.83
Trading date	2017-06-05	2017-08-04	2017-06-05	2017-08-04

- The Root Integrated Squared Error (RISE) between the estimated density $\hat{f}_{\alpha,\delta}$ and the theoretical one f.

$$RISE = \frac{1}{||f||}\sqrt{\int_0^\infty (\hat{f}_\alpha(s) - f(s))^2 ds}.$$

Simulations are performed with the following procedure:

1. Select the parameters of the model and fix the risk neutral density $f(s)$.

2. Generate the sample of strike prices and the predictor variable Z_i.

3. Generate the n random variables ε_i from the uniform distribution.

4. Compute the values of the option prices $Y_i = \int_0^\infty Z_i(s)f(s)ds + \varepsilon_i$.

5. Center the variables Y_i and Z_i. Then, apply the LF method on centered data.

6. Run the 10-fold cross-validation procedure in order to select the optimal number of iterations. The whole sample is randomly divided into 10 groups. 9 groups are used as the training sample and the last one is used as the validation sample. This operation is repeated 10 times as we have 10 different groups in the cross-validation

procedure. Then, we average to get an estimate of the mean squared prediction error. This quantity is minimized with respect to α.

7. Estimate the risk neutral density using the optimal tuning parameter α selected in Step 5 and calculate the predicted option price.

8. Calculate the RISE and RMSPE by 10-fold cross-validation.

9. Repeat the steps 2 to 8 with 100 iterations and calculate the average RISE and RMSPE over the simulations.

All the numerical integrations are performed with the trapezoidal rule. It is also possible to use other integration rules such as the Newton-Cotes or adaptative quadrature. For LF, the tuning parameter ω is chosen equal to $1/\sum_j \lambda_j$ where λ_j are the eigenvalues of $\widehat{K}$. To correct the density, we use the formula (8), which is faster than the procedure proposed by Glad et al. (2003). To implement PosConv, we approximate f by $\widehat{f}(x) = \sum a_j \phi_h(x - z_j)$ with $a_j \geq 0$ and $\sum a_j = 1$ using an equispaced grid of 70 points for z_j. The a_j are selected by minimizing[3] $\sum_i \left(Y_i - \int_0^\infty Z_i(s) f(s) ds\right)^2$. Moreover, the bandwidth h is selected by 10-fold cross-validation.

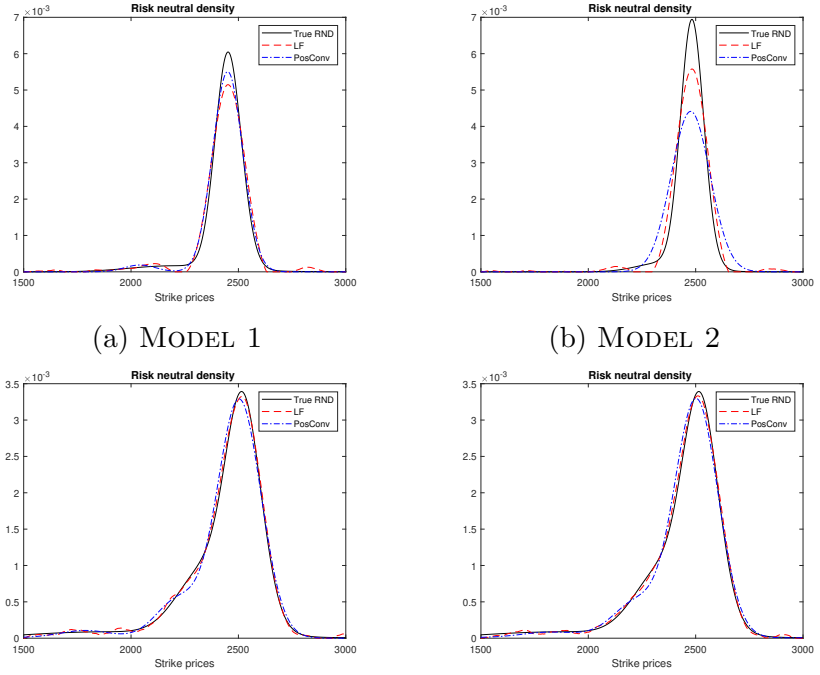
Table 2 shows the results from the simulations. Comparing the LF method with PosConv, we can observe that when considering the option pricing with a time-to-maturity of less than 6 months, PosConv method tends to outperform LF in terms of RMSPE (see results from Model 1). In contrast, for options with a maturity of more than 6 months, LF method tends to outperform PosConv in both RISE and RMSPE (Models 3 and 4). Across all models, LF always displays a smaller RISE than PosConv. Also, we can see that as the time to maturity increases, the RISE improves but the RMSPE increases for both methods. This result can be confirmed by observing Figures 1, 2

[3]We do not use the formula $\sum_i \left(Y_i - D^{-2}\widehat{f}(x_i)\right)^2$ suggested by Bondarenko (2003) because Bondarenko's formula to compute $D^{-2}f$ seems to ignore the fact that f has support $[0, \infty)$ instead of $\mathbb{R}$. Moreover, the correct expression for $D^{-2}f$ is different depending on whether the option is a put or a call.

and 3 which respectively display the risk neutral density, the cumulative distribution, and the predicted option prices. These figures show that the adjustment of the true RND improves with the maturity, but PosConv tends to adjust better thin tails, while LF tends to fit better fat tails distributions. As PosConv approximates the unknown density with a mixture of normal distributions which has relatively thin tails, it makes sense that this method will perform best when the true density has thin tails. Our method on the other hand lets the data speak by themselves.

Table 2: Comparison of the estimation methods

Models	Maturity (days)	Criteria	LF	PosConv
Model 1	53	$RMSPE$	1.513	1.125
		$RISE$	0.132	0.140
Model 2	35	$RMSPE$	2.585	2.640
		$RISE$	0.184	0.340
Model 3	148	$RMSPE$	4.021	5.058
		$RISE$	0.032	0.072
Model 4	224	$RMSPE$	4.850	7.31
		$RISE$	0.030	0.034



(a) Model 1 (b) Model 2

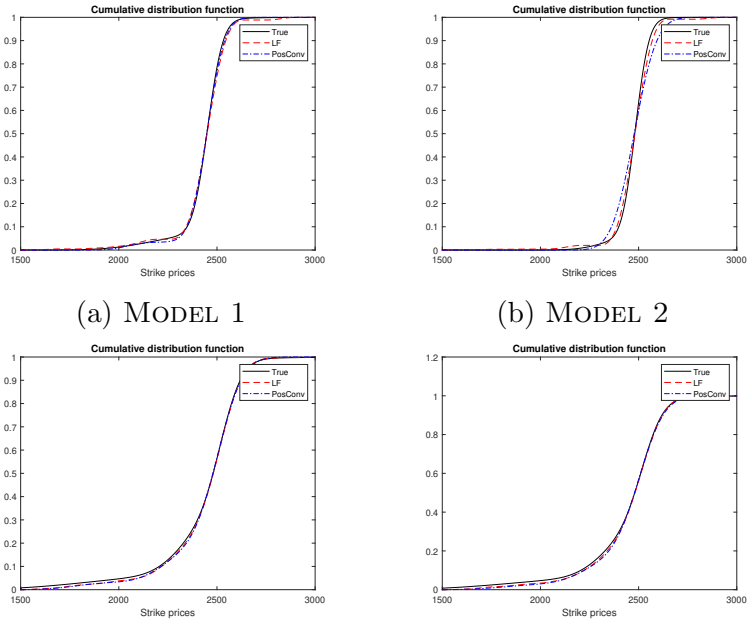
(c) Model 3 (d) Model 4

Figure 1: Estimated risk neutral density

5 Application to S&P500 options

This section focuses on a real data example. For this purpose, we consider the S&P 500 index (SPX) as the data of interest to derive the underlying and the strikes. The S&P 500 option is one of the most liquid and tradable options in the market. It represents the aggregated capitalization of the 500 most important corporations in the U.S. It is also used as a benchmark to see how well the most important companies are behaving. The data are taken from Barratt et al. (2020). They have been extracted from OptionMetric Ivy database and have been made public on the github account of the authors on *https://github.com/cvxgrp/cvx_opt_risk_neutral*.

From the database, the best bid and ask prices of all S&P 500 European options are collected for the date of June 3, 2019 with a maturity of 25 days. The price of the index at the end of the same day is also collected. Indeed, it is equal to 2744.45 dollars. The range of values for s is between 1500 to 3999.50 dollars. Put and call options are



(a) Model 1 (b) Model 2

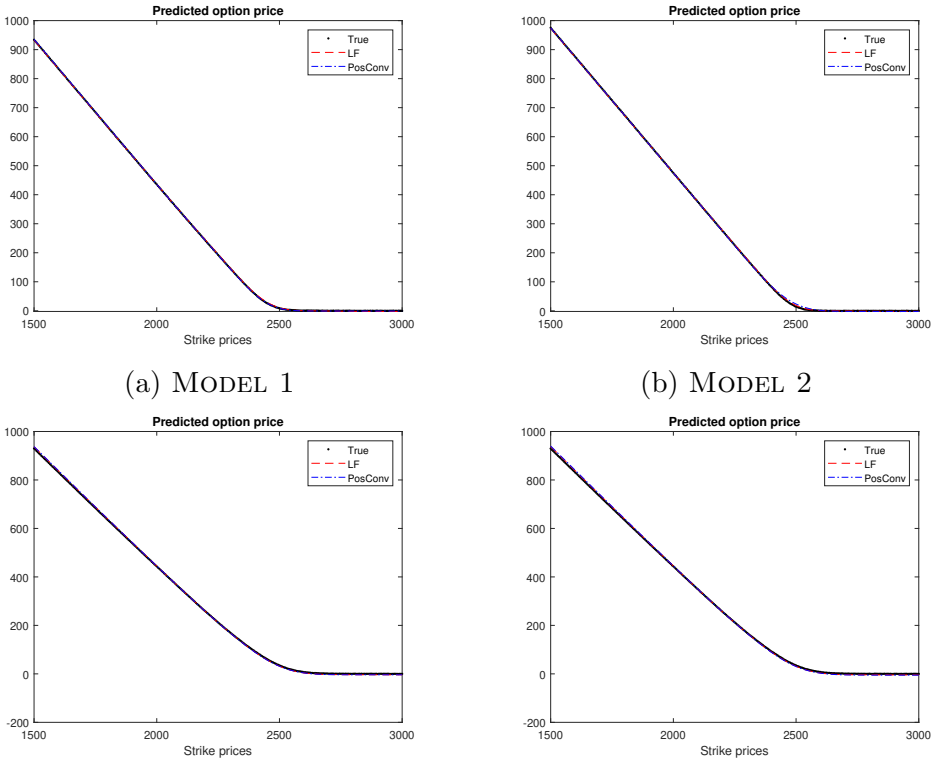
(c) Model 3 (d) Model 4

Figure 2: Estimated cumulative distribution function

both considered in the sample. The option prices are not directly observed. We follow the literature by taking the average of the bid and ask prices as proxy for the actual price. The put and call prices are then stacked together. The resulting sample includes 157 observations (49 calls and 108 puts). Next, we estimate Model (14) where Y_i is the option price and $Z_i(s) = \max(\kappa_i - s, 0)$ for put options and $Z_i(s) = \max(s - \kappa_i, 0)$ for call options.

As in the simulations, the optimal tuning parameter α and optimal bandwidth h are selected by minimizing the RMSPE evaluated by 10-fold cross-validation. Once these tuning parameters are chosen, one can estimate the RND and therefore predict the option prices.

The results are presented in Figure 4. The estimated distribution from the 25-days maturity options displays a bell-shaped curve centerered around the value 2750 dollars. This density also presents long tails. Figure 5 presents the predicted call and put prices with their respective 95% (in-sample) confidence interval. The confidence interval is



(a) Model 1 (b) Model 2

(c) Model 3 (d) Model 4

Figure 3: Predicted option prices

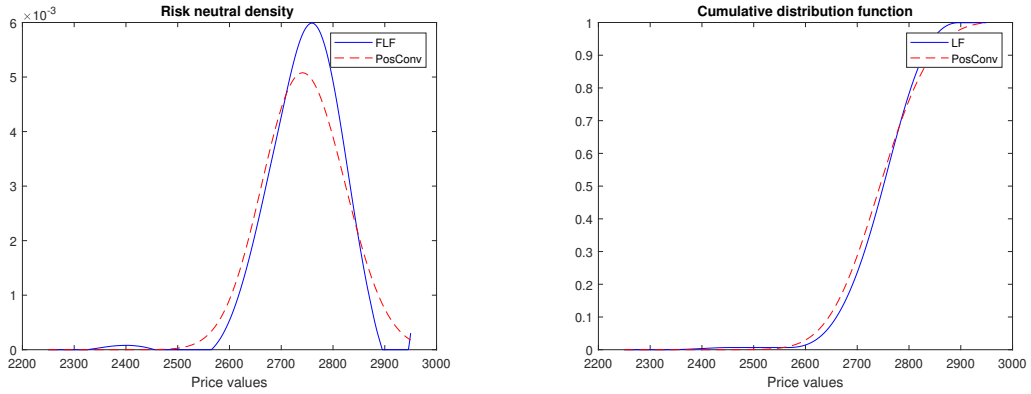
based on LF estimator. It can be noticed that the option prices are well predicted via the suggested estimation approach.

Table 3 reports the LF and PosConv RMSPE evaluated by 10-fold cross-validation. We can see that LF outperforms PosConv method as it displays a smaller RMSPE than PosConv. Additionally, Figure 5 shows that LF tends to predict better than PosConv.

One of the potential reasons of the weaker performance for PosConv is that PosConv uses to estimate f a mixture of normal distributions which have thin tails. In the real data, the RND may have fat tails and it was observed, in the simulations, that LF outperforms PosConv for RND with fat tails.

Table 3: Comparison of the estimation methods on real data

Criterion	LF	PosConv
RMSPE	1.065	2.801

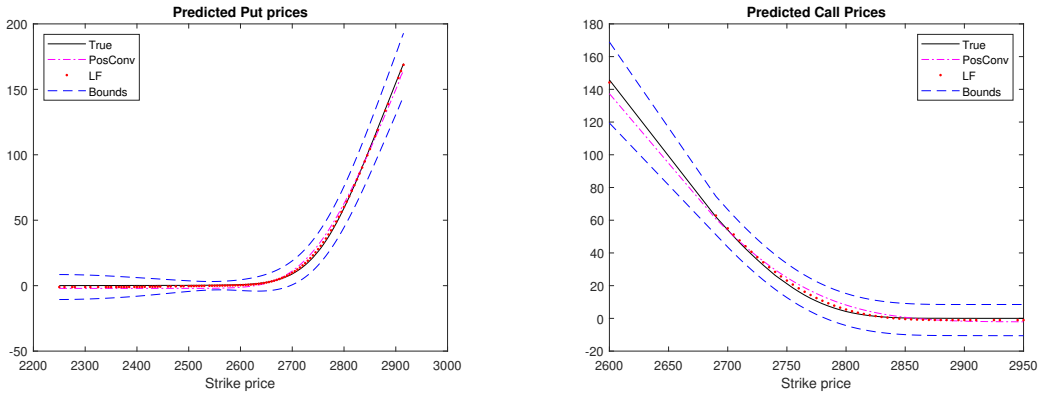


(a) RND (b) PosConv

Figure 4: Estimated RND and CDF with LF and PosConv methods

6 Conclusion

This paper proposes to estimate the risk neutral density for option pricing models with the functional data analysis framework. Indeed, we consider that a European option price of an asset is evaluated as a weighted average of all possible payoffs of the asset, where the weights represents here the risk neutral density of a market participant. To use the functional data analysis framework, we assume that for each asset, one can have an infinity of possible price values at the maturity. This means that at the maturity date, a market participant is exposed to an infinite possibility of payoffs. The set of potential payoffs for each option price is then very dense and is considered as a function. On the same line, the risk neutral density is also considered as a collection of values observed on a very fine grid. Therefore, the model setting considered is a functional linear model where the predictor functions are represented by the set of potential payoffs and the response variable is the price of the option, which is a scalar. The estimation method proposed in this paper is free of any parametric or semi-parametric assumptions



(a) Put prices (b) Call prices

Figure 5: SPX predicted call and put prices with 95% confidence bounds using the LF method

and also takes into account the arbitrage-free theory for option pricing in the model.

One of the main issue of this model is the high dimensionality problem as the inverse of the covariance operator of the predictor variable is not continuous. This problem leads to unstable estimated function. To overcome this issue, we propose to use a regularization technique called Landweber-Fridman. We also control for the positivity and the integration to one of the density by applying a density correction. We derive the consistency and asymptotic normality of the estimated density function. Additionally, we provide confidence intervals for the RND.

Our simulations show that LF estimator captures well the bell-shaped form of the true risk neutral density. Comparing the results of LF method with those obtained by Bondarenko (2003), we observed that the PosConv method of Bondarenko (2003) outperforms our LF method when the tails for the RND are thin and our method outperforms when the tails of the RND are fat. The results from real data on S&P 500 options show that LF outperforms PosConv in terms of RMSPE. Therefore, the proposed approach in this paper can be considered as a promising alternative to the existing ones.

Below, we discuss a few possible extensions.

1. **RND and implied volatility:** one important application of the risk neutral density estimation is the usage of the inversion formula to get the implied volatility. As our model is nonparametric and we derive the estimation results directly on option prices and underlying, this inversion formula cannot be obtained. To estimate implied volatility, one should consider a cross-sectional model of options prices with different maturities for the same asset, which is now a heteroskedastic model. An extension of our results to heteroskedastic errors would be needed to obtain the pricing function for put and call options. The next step would be to calculate the realized volatility, which is the variance of the estimated put and call prices. Then, the implied volatility is nothing else but the mean of the realized volatilities for each time t considered in the data.

2. **RND and positivity of the stochastic discount factor:** the stochastic discount factor is another variable of interest in asset pricing models. Based on our estimation approach, one could worry about the positivity of the discount factor. To address this question, one could estimate the option prices via our method, then consider a utility function and estimate its parameters using the recursive utility framework proposed by Garcia et al. (2003). The last step would be to check if the stochastic discount factor is positive. This is left for future research.

3. **Fat tails:** Many studies have empirically documented the presence of fat-tails in financial returns (see Embrechts et al. (1997) among others). This feature should translate into fat tails of the RND and has motivated some work on RND models allowing for semi-fat and fat tails (see Hartvig et al. (2001) and Figlewski (2010)). Our simulations suggest that our method may work well when the tails of the RND are fat. However, a thorough investigation would be needed to confirm this conjecture and is left for future research.

4. **American options:** This paper focuses on European options only. However, most traded options are American options. Because American options can be exercised before maturity, one cannot express the price of American options as a linear function

of the RND. Some authors have tried to exploit American options to estimate the RND. Melick & Thomas (1997) provide bounds using American options. On the other hand, Tian (2011) develops a method to extract European option prices from American option prices and then use the usual method to estimate the RND. Also, Flamouris & Giamouridis (2002) use the Edgeworth series expansion (ESE) technique to estimate the RND and Borovkova et al. (2012) exploit the partial differential equations approach to derive the option prices. An in-depth analysis of the American options is beyond the scope of the present paper.

References

Ait-Sahalia, Y., & Duarte, J. (2003). Nonparametric option pricing under shape restrictions. *Journal of Econometrics*, *116*(1-2), 9–47.

Ait-Sahalia, Y., Karaman, M., & Mancini, L. (2018). The term structure of variance swaps and risk premia. *Swiss Finance Institute Research Paper*(18-37).

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

Bahra, B. (1997). Implied risk-neutral probability density functions from option prices: theory and application. *working paper, Bank of England*.

Barratt, S., Tuck, J., & Boyd, S. (2020). Convex optimization over risk-neutral probabilities. *arXiv preprint arXiv:2003.02878*.

Benatia, D., Carrasco, M., & Florens, J.-P. (2017). Functional linear regression with functional response. *Journal of Econometrics*, *201*(2), 269–291.

Billingsley, P. (1995). *Probability and measure*. John Wiley & Sons, New York.

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of political economy*, *81*(3), 637–654.

Bliss, R. R., & Panigirtzoglou, N. (2004). Option-implied risk aversion estimates. *The Journal of Finance*, *59*(1), 407–446.

Bondarenko, O. (2003). Estimation of risk-neutral densities using positive convolution approximation. *Journal of Econometrics*, *116*(1-2), 85–112.

Borovkova, S., Permana, F., & Van Der Weide, J. (2012). American basket and spread option pricing by a simple binomial tree. *The Journal of Derivatives*, *19*(4), 29–38.

Bosq, D. (2000). *Linear processes in function spaces: Theory and applications, volume 149 of lecture notes in statistics.* Springer-Verlag New York Inc.

Cai, T. T., & Hall, P. (2006). Prediction in functional linear regression. *The Annals of Statistics*, *34*(5), 2159–2179.

Cai, Z., Li, Q., & Park, J. Y. (2009). Functional-coefficient models for nonstationary time series data. *Journal of Econometrics*, *148*(2), 101–113.

Campbell, J., Lo, A., & MacKinlay, A. (1997). *The econometrics of financial markets.* Princeton University Press.

Cardot, H., Ferraty, F., & Sarda, P. (1999). Functional linear model. *Statistics & Probability Letters*, *45*(1), 11–22.

Carrasco, M., & Florens, J.-P. (2000). Generalization of GMM to a continuum of moment conditions. *Econometric Theory*, *16*, 797–834.

Carrasco, M., & Florens, J.-P. (2011). A spectral method for deconvolving a density. *Econometric Theory*, 546–581.

Carrasco, M., Florens, J.-P., & Renault, E. (2007). Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization. *Handbook of Econometrics*, *6*, 5633–5751.

Carrasco, M., Florens, J.-P., & Renault, E. (2014). Asymptotic normal inference in linear inverse problems. *Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, *73*(74), 140.

Chang, Y., Kim, C. S., & Park, J. Y. (2016). Nonstationarity in time series of state densities. *Journal of Econometrics*, *192*(1), 152–167.

Chen, X., & Reiss, M. (2011). On rate optimality for ill-posed inverse problems in econometrics. *Econometric Theory*, *27*, 497–521.

Cochrane, J. (2005). *Asset pricing: Revised edition*. Princeton University Press.

Cox, J. C., & Ross, S. A. (1976). The valuation of options for alternative stochastic processes. *Journal of Financial Economics*, *3*(1-2), 145–166.

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

Delaigle, A., & Hall, P. (2012). Methodology and theory for partial least squares applied to functional data. *The Annals of Statistics*, *40*(1), 322–352.

Driessen, J., Maenhout, P. J., & Vilkov, G. (2009). The price of correlation risk: Evidence from equity options. *The Journal of Finance*, *64*(3), 1377–1406.

Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling extremal events for insurance and finance*. Springer.

Engl, H. W., Hanke, M., & Neubauer, A. (1996). *Regularization of inverse problems* (Vol. 375). Springer Science & Business Media.

Fengler, M. R. (2009). Arbitrage-free smoothing of the implied volatility surface. *Quantitative Finance*, *9*(4), 417–428.

Ferraty, F., & Vieu, P. (2006). *Nonparametric functional data analysis: theory and practice*. Springer Science & Business Media.

Figlewski, S. (2010). Estimating the implied risk neutral density. *Volatility and Time Series Econometrics: Essay in Honor of Robert F. Engle*.

Figlewski, S. (2018). Risk neutral densities: A review. *Annual Review of Financial Economics*, *10*, 329–359.

Flamouris, D., & Giamouridis, D. (2002). Estimating implied pdfs from american options on futures: A new semiparametric approach. *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, *22*(1), 1–30.

Gagliardini, P., & Scaillet, O. (2012). Nonparametric instrumental variable estimation of structural quantile effects. *Econometrica*, *80*, 1533–1562.

Garcia, R., & Gençay, R. (2000). Pricing and hedging derivative securities with neural networks and a homogeneity hint. *Journal of Econometrics*, *94*(1-2), 93–115.

Garcia, R., Luger, R., & Renault, E. (2003). Empirical assessment of an intertemporal option pricing model with latent variables. *Journal of Econometrics*, *116*(1-2), 49–83.

Glad, I., Hjort, N. L., & Ushakov, N. (2003). Correction of density estimators that are not densities. *Scandinavian Journal of Statistics*, *30*, 415–427.

Gourieroux, C., & Jasiak, J. (2001). *Financial econometrics: Problems, models, and methods*. Princeton University Press.

Grith, M., Härdle, W., & Schienle, M. (2012). Nonparametric estimation of risk-neutral densities. *Handbook of Computational Finance*, 277–304.

Hall, P., & Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics*, *35*(1), 70–91.

Harrison, J. M., & Kreps, D. (1979). Martingales and arbitrage in multiperiod security markets. *Journal of Economic Theory*, *20*, 381–408.

Hartvig, N., Jensen, J., & Pedersen, J. (2001). A class of risk neutral densities with heavy tails. *Finance and Stochastics*, *5*, 115–128.

Horowitz, J. L. (2007). Asymptotic normality of a nonparametric instrumental variables estimator. *International Economic Review*, *48*(4), 1329–1349.

Ibragimov, M., Ibragimov, R., & Walden, J. (2015). *Heavy-tailed distributions and robustness in economics and finance* (Vol. 214). Springer.

Jackwerth, J. C., & Rubinstein, M. (1996). Recovering probability distributions from option prices. *The Journal of Finance*, *51*(5), 1611–1631.

Jarrow, R., & Rudd, A. (1982). Approximate option valuation for arbitrary stochastic processes. *Journal of Financial Economics*, *10*(3), 347–369.

Jondeau, E., Poon, S.-H., & Rockinger, M. (2010). *Financial modeling under non-gaussian distributions*. Springer.

Kargin, V., & Onatski, A. (2008). Curve forecasting by functional autoregression. *Journal of Multivariate Analysis*, *99*(10), 2508–2526.

Kokoszka, P., & Zhang, X. (2010). *Improved estimation of the kernel of the functional autoregressive process* (Tech. Rep.). Technical Report. Utah State University.

Kundu, A., Kumas, S., & Tomar, N. (2018). Option implied risk-neutral density estimation: A robust and flexible method. *Computational Economics*.

Melick, W., & Thomas, C. (1997). Recovering an asset's implied pdf from option prices: An application to crude oil during the gulf crisis. *Journal of Financial and Quantitative Analysis*, *32*, 91–115.

Panigirtzoglou, N., & Skiadopoulos, G. (2004). A new approach to modeling the dynamics of implied distributions: Theory and evidence from the s&p 500 options. *Journal of Banking & Finance*, *28*(7), 1499–1520.

Park, J. Y., & Qian, J. (2012). Functional regression of continuous state distributions. *Journal of Econometrics*, *167*(2), 397–412.

Rosenberg, J. V. (1998). Pricing multivariate contingent claims using estimated risk–neutral density functions. *Journal of International Money and Finance*, *17*(2), 229–247.

Shimko, D., Tejima, N., & Van Deventer, D. R. (1993). The pricing of risky debt when interest rates are stochastic. *Journal of Fixed Income*, *3*(2), 58–65.

Souissi, N. (2017). The implied risk neutral density dynamics: Evidence from the s&p tsx 60 index. *Journal of Applied Mathematics*, *2017*.

Tian, Y. (2011). Extracting risk-neutral density and its moments from american option prices. *Journal of Derivatives*, *18*, 17–34.

Tsafack, I. (2020). Intraday stock market forecasting via functional time series. *Working paper*.

Vogt, E. (2014). Option-implied term structures. *Available at SSRN 2541954*.

Yatchew, A., & Härdle, W. (2006). Nonparametric state price density estimation using constrained least squares and the bootstrap. *Journal of Econometrics*, *133*(2), 579–599.