

Nonparametric State-Price Density Estimation using High Frequency Data

Jeroen Dalderop¹

Faculty of Economics
University of Cambridge

May 29, 2015

Abstract

The state-price density (SPD) implicit in option prices can be estimated using nonparametric time series regression, as first done by Ait-Sahalia and Lo (1998). This paper uses high frequency data to estimate the SPD using a kernel regression locally around the time point of interest. The call pricing function is allowed to vary smoothly over time which is aligned with time-varying preferences and beliefs. We show consistency and asymptotic normality of our estimator under infill asymptotics, when the traded strike prices follow a mixing, locally stationary time series. To allow for random and possibly endogenous trading times, we generalize the results in Vogt (2012) towards sampling times described by point processes. We investigate a data driven choice of the sampling windows that automatically adapts to the speed of information flow in the market. In an application to S&P 500 E-mini European call and put options we discuss the properties of the local-in-time cross-validated bandwidths.

Keywords: Option Pricing, Kernel Regression, High Frequency Data, Local Stationarity

JEL Codes: C14, G13

1 Introduction

Option prices contain detailed information about the expectations and preferences of market participants. This information is summarized in the state-price density (SPD), which gives the prices of Arrow-Debreu securities that pay out when the underlying asset ends up within the infinitesimal interval $(x, x + dx)$ for varying levels of the outcome x . High values of the SPD in a particular part of the distribution indicate a high marginal rate of substitution for wealth in

¹I wish to thank my supervisor Oliver Linton for his guidance and support, as well as Seokyoung Hong, Ryoko Ito, Alexei Onatski, Eric Renault, Mike Tehranchi, Steve Thiele, Jasmine Xiao, and seminar participants in the Econometrics Workshop at the University of Cambridge for helpful comments. Financial support by CERF is gratefully acknowledged.

these states. For example, a thick left tail indicates a high market price for insurance against large losses.

Estimators of the SPD are useful in at least four contexts. First, they allow practitioners to compute the no-arbitrage price of new or illiquid derivatives simply by integration. They can also be used to filter out the pricing error in observed option prices, which prevents biases in implied volatilities (Chen and Xu, 2014). Second, risk managers can compute market-implied values of risk measures such as Value-at-Risk and Expected Shortfall directly from the SPD for the maturity of interest (Aït-Sahalia and Lo, 2000). Third, event studies measure changes in the SPD to analyze the impact of economic events to investor beliefs and preferences Beber and Brandt (2006). Finally, together with an estimator of the objective density from historical data it allows computing empirical pricing kernels (Rosenberg and Engle, 2000). These can be related to the risk preferences of a representative agent in consumption based equilibrium models. The finding that the implied Arrow-Pratt risk aversion coefficient does not monotonically decrease with wealth has been referred to as the ‘pricing kernel puzzle’ (Jackwerth, 2000).

Both parametric and nonparametric estimation approaches have been proposed. The former requires parametrizing both the dynamics of the underlying asset and the utility function of a representative investor. The resulting theoretical option prices yield the moment conditions which can be estimated by GMM techniques. However the classical Black-Scholes model with a lognormal SPD provides poor fit to the data, as evidenced by the implied volatility smile. Also more flexible parametric models have been reported to miss important features of the data (Bakshi et al., 1997). As a result, estimated risk premia are notably sensitive to the precise model specification (Broadie et al., 2007).

In the nonparametric approach usually no specific model is assumed for the asset return dynamics or the SPD. Instead, Aït-Sahalia and Lo (1998) proposed to estimate the call pricing function via nonparametric regression of the observed call prices on variables as stock price, strike price, and time-to-maturity. The SPD is then obtained from the second derivative of call pricing function with respect to the strike price using the result by Breeden and Litzenberger (1978). Proposed estimation methods include kernel smoothers (Aït-Sahalia and Lo, 1998; Aït-Sahalia and Duarte, 2003), smoothing splines (Yatchew and Härdle, 2006), and methods based on the shape constraints required by the absence of arbitrage.

An important difference in the literature is the horizon over which option prices are pooled together. Aït-Sahalia and Lo (1998) and other early studies are based on sampling periods that may span multiple years, during which the SPD is treated as a constant function of time-to-maturity. Their estimate of the SPD can therefore be interpreted as some average SPD over the sample period, rather than a point estimate for a given point in time. More recently, the increasing availability of high frequency data has allowed large enough sample sizes to estimate

SPDs using only one or a few trading days (e.g. Härdle and Hlavka (2009), Chen and Xu (2014)). This motivates the use of infill asymptotics, where consistent estimators are obtained by letting the number of observations within a fixed interval go to infinity. Still, it is typically assumed that the SPD for given time-to-maturity is constant over some interval chosen by the econometrician.

Our paper aims to relax this assumption by allowing the regression function to vary smoothly over time. This is achieved by using a kernel estimator that fits a regression curve around each time point of interest. Such an estimator is able to capture changes in the option pricing function due to information becoming available during trading time. This allows a direct way to measure updates in investor beliefs and risk aversion in response to market events. Smoothing in time requires a sampling window to be chosen for each time point of interest. We investigate a data driven choice of the sampling windows that automatically adapts to the speed of the information flow in the market. This makes it possible to find an appropriate bias-variance trade-off which would be neglected when using a fixed sampling period such as a day or a week.

The theoretical treatment of the time-varying kernel regressor is set up generally, as it may be of interest beyond our application to option pricing. As a theoretical contribution we generalize the asymptotic normality result in Vogt (2012) on locally stationary time series towards random sampling times. Since these may endogenously determined with the covariate, we wish to avoid conditioning on them and instead consider the sampling variability due to the sampling times themselves. This is motivated by recent work on the informational content of endogenous trading times (e.g. Li et al. (2009), Renault and Werker (2011)). In particular we model the sampling times as a point process with a conditional intensity function that may depend on both its own history as well as that of the covariates. This framework can incorporate the main empirical features of trading times, such as clustering behaviour, interaction with covariates, and time-of-the-day effects which cause non-stationary trading volume. We discuss appropriate mixing conditions for both the point process and the covariate such that, when the mean intensity goes to infinity, the time-varying kernel estimator is asymptotically normal with the same variance as when the observations were independent. We illustrate these conditions by means of Hawkes processes, both self-exciting and driven by the covariate.

In the application to state-price density estimation we apply the time-varying kernel estimator to data on transaction prices and bid and ask quotes on S&P 500 E-mini European call and put options. We discuss dimension reductions such that the option pricing function can be represented as an unknown function of time and moneyness of the option only. We show how our estimator can be used to study the dynamic behaviour of the state-price density, over both interday and intraday periods. The bandwidths are chosen by a local-in-time cross validation method, which leads to sampling windows during trading time typically to be a couple hours for transactions data, and five minutes to half an hour for bid and ask quotes.

The remainder of this research proposal is organized as follows. Section 2 provides the theoretical treatment of our time-varying kernel regressor under random sampling times. Section 3 discusses the application of this estimator in state-price density estimation, and provides empirical results. Section 4 concludes.

2 Time-varying kernel regression with random sampling times

Nonparametric time series regression typically assumes that the functional form of the regression relationship between two time series is unknown, yet constant throughout the sample. Estimators are then consistent when the time span becomes infinitely large under general mixing conditions (see e.g. Robinson (1983), Bosq (1996)). However, for some applications it may be not be realistic to treat the regression function as time-invariant. In such cases the conditional expectation $E(Y_t|X_t = x) =: m(t, x)$ is a function of both t and x . An increasing time span now no longer guarantees consistent estimators, as this requires a growing amount of observations locally around the time point of interest.

When the data is of the type that there is a large number of observations even within small time windows, such as with high frequency data in finance, such local information on the dependency between the time series is in fact available. This motivates the use of asymptotic analysis based on infill asymptotics. In particular, we study the asymptotic properties of a kernel smoother which smoothes in both the state and the time domain, when the number of observations within a fixed time interval goes to infinity. There are many financial applications that use smoothing in the time domain in the infill setting, for example to estimate spot volatility (e.g. Barndorff-Nielsen et al. (2008), Kristensen (2010)), spot covariation (e.g. Barndorff-Nielsen and Shephard (2004), Zhang (2011)), or time-varying betas for diffusions Mykland et al. (2006). Typically, the data are treated as discrete observations from a continuous semimartingale, such as an Itô process. When the mesh of the grid of observations goes to zero, the complete sample path of the process becomes known and can be used to estimate continuous time models (for an overview see Aït-Sahalia (2007)). However, for data constructed in this way simultaneously smoothing in the state dimension is generally not possible. To see this, note that continuity implies that the range of observed value of the semimartingale can be chosen arbitrary small by letting the time band become small enough. Hence in the limit only the value at that particular time point is observed, from which clearly no consistent estimator can be created.

Instead, we treat our covariate as a discrete time series, which, within any given time interval, visits all points in its range infinitely often when the sample size goes to infinity. This ensures there is sufficient variation in the covariate around each point in time. To achieve this, we adapt strong mixing conditions to the setting where the observation become close to each other in

calendar time. In combination with a local stationarity condition on the covariate, this suffices to achieve consistent estimators of the regression function at each point in (calendar) time. Our setting is very similar to that of Vogt (2012), who shows the consistency of the time-varying kernel regressor for locally stationary covariates. Although Vogt (2012) considers long span asymptotics, consistency is derived in rescaled time. Hence his results can be shown to carry over to our setting up to a rescaling of the time dimension to the unit interval.

Besides adapting the results in Vogt (2012) to the infill asymptotic setting, we also generalize them to allow for random sampling times which are possibly dependent on the covariates. Therefore we derive the unconditional distribution of our estimator, when the sampling times are modelled as a point process. To allow interaction with the covariate we allow the conditional intensity function to depend on both its own history as well as that of the covariates. Similar issues are studied in the literature on marked point processes, see for example Ellis (1991) and Pawlas (2009) for results on density estimation for a stationary regressor observed at random time points. The effect of random sampling times on estimators using high frequency data has been emphasized in Aït-Sahalia and Mykland (2003), and Duffie and Glynn (2004), among others.

2.1 Model

Let $\{(Y_{i,n}, X_{i,n})\}$ be a bivariate stochastic process, observed at the random times $\{t_{i,n}\}$. We restrict attention to the case where $Y_{i,n}$ and $X_{i,n}$ are observed synchronously. The sampling times $\{t_{i,n}\}$ are modelled as a point process and associated with the counting measure $N_n(a, b) = \#\{i : a \leq t_{i,n} \leq b\}$.² Define also the counting process $N_n(t) := N_n(0, t)$ of the number of events up to time $t \in \mathbb{R}$. Throughout we use double-index notation, where the second index n creates a sequence of point processes with increasing mean arrival rate. This is made more precise in Section 2.3. The information flow for each n is described by the natural filtration $\mathcal{F}_{t,n}^N$ and $\mathcal{F}_{t,n}^x$ of the counting process $\{N_n(t)\}$ and the covariates $\{X_{i,n}\}$ observed before time t , respectively, and their joint filtration $\mathcal{F}_{t,n} = \mathcal{F}_{t,n}^N \cup \mathcal{F}_{t,n}^x$.

Consider the time-varying nonparametric regression model

$$Y_{i,n} = m(t_{i,n}, X_{i,n}) + \epsilon_{i,n}, \quad (1)$$

with $E(\epsilon_{i,n}|t_{i,n}, X_{i,n}) = 0$ and $\text{Var}(\epsilon_{i,n}|t_{i,n} = t, X_{i,n} = x) = \sigma^2(t, x)$. Note $m(t, x)$ is the conditional expectation of $Y_{i,n}$ given its observation time $t_{i,n} = t$ and the covariate $X_{i,n} = x$. We are interested in estimating $m(\cdot)$ on a fixed time interval $(0, T)$, hence the number of observations in our sample is $N_n(0, T)$.

²Formally, $\{(t_{i,n}, Y_{i,n}, X_{i,n})\}$ is a univariate marked point process, with bivariate vector of marks $(Y_{i,n}, X_{i,n})$. In the case of asynchronous sampling times, this would be a bivariate marked point process, with univariate marks.

2.2 Estimator

From now on without loss of generality we set $T = 1$. A natural estimator for $m(\cdot)$ at the design point (t, x) is the Nadaraya-Watson (or locally constant) kernel estimator

$$\hat{m}_h(t, x; 0) = \frac{\sum_{i=N_n(0)+1}^{N_n(1)} K_{h_t}(t - t_{i,n}) K_{h_x}(x - X_{i,n}) Y_{i,n}}{\sum_{i=N_n(0)+1}^{N_n(1)} K_{h_t}(t - t_{i,n}) K_{h_x}(x - X_{i,n})}, \quad (2)$$

with $K_h(\cdot) = \frac{1}{h} K(\frac{\cdot}{h})$ for some kernel function K . This estimator takes a weighted average over points close to (t, x) , where the weights are controlled by the bandwidths h_t and h_x for the time and state dimension, respectively. The kernel smoothing method can be generalized to fitting a local polynomial around (t, x) , yet for ease of exposition we stick to the locally constant estimator. Note that if the kernel has a bounded support, say, $[-1, 1]$, then for any $t \in (h_t, 1 - h_t)$ the sum in (2) has only $N_n(t - h_t, t + h_t)$ non-zero terms. In this case the sum needs only to be taken from $i = N_n(t - h_t) + 1$ to $i = N_n(t + h_t)$. We aim to study the limiting distribution of this estimator when n goes to infinity, that is, when the expected number of observations within the interval becomes infinitely large.

2.3 Random sampling times

We assume that the counting processes are such that for every n and $t \in \mathbb{R}$

$$\begin{aligned} \mathbb{P}(N_n(t, t + \Delta) = 1 | \mathcal{F}_t^N, \mathcal{F}_t^x) &= \lambda_n(t) \Delta + o_P(\Delta), \\ \mathbb{P}(N_n(t, t + \Delta) > 1 | \mathcal{F}_t^N, \mathcal{F}_t^x) &= o_P(\Delta), \end{aligned}$$

for $\Delta \rightarrow 0$, which defines the conditional intensity function $\lambda_n(t)$. Note in particular that $\lambda_n(t)$ depends on both past event times as well as past observed covariates. We denote its unconditional expectation by $\bar{\lambda}_n(t) := E(\lambda_n(t))$. The second property is known as orderliness and ensures there are no simultaneous events, see Daley and Vere-Jones (2003) for details. This implies that

$$\text{Var}(dN_n(t)) = E\{(dN_n(t))^2\} = E(dN_n(t)) = \bar{\lambda}_n(t) dt.$$

Furthermore assume that $N_n(t)$ admits a covariance density, defined for $t \neq s$ by

$$\mu_n(t, s) = \frac{\text{Cov}(N_n(t, t + dt), N_n(s, s + ds))}{dtds},$$

while $\mu_n(t, t)$ is defined such that the function is continuous.

The following conditions require the average intensity and covariance density at each time point to be proportional to n and n^2 , respectively:

(C1) $\frac{\bar{\lambda}_n(t)}{n} < C$ for some $C < \infty$ for every t , $\bar{\lambda}_n(t)$ is twice differentiable on $(0, 1)$ for every n , and

$$\frac{\bar{\lambda}_n(t)}{n} \rightarrow \nu(t), \quad (3)$$

for some twice differentiable function $\nu(t)$, satisfying $\inf_{t \in (0,1)} \nu(t) > 0$ and $\sup_{t \in (0,1)} \nu(t) < \infty$.

(C2) For each time $t \in (0, 1)$ there exists a function $\gamma_t : \mathbb{R} \mapsto \mathbb{R}$, with

$$\eta(t) := \int_{-\infty}^{\infty} \gamma_t(\tau) d\tau < \infty,$$

such that

$$\frac{\mu_n(t, t + \frac{\tau}{n})}{n^2} \rightarrow \gamma_t(\tau),$$

and $\left| \frac{\mu_n(t, t + \frac{\tau}{n})}{n^2} \right| \leq M(\tau)$ for some integrable function $M : \mathbb{R} \mapsto \mathbb{R}_+$ independent of t .

Example. A natural specification of the sequence of counting processes $\{N_n(t)\}$ is

$$N_n(t) = N \left(n \int_0^t \pi(s) ds \right), \quad (4)$$

for $t \in (0, 1)$, where $N(\cdot)$ is a stationary point process with intensity ρ and covariance density $\mu(\cdot)$. In this specification the index n ‘speeds up the clock’, whereas the function $\pi : (0, 1) \mapsto \mathbb{R}_+$ captures a deterministic pattern in the mean intensity, satisfying without loss of generality $\int_0^1 \pi(s) ds = 1$. It holds that

$$\begin{aligned} \nu(t) &= \lim_{n \rightarrow \infty} \frac{\bar{\lambda}_n(t)}{n} = \rho \pi(t), \\ \gamma_t(\tau) &= \lim_{n \rightarrow \infty} \frac{\mu_n(t, t + \frac{\tau}{n})}{n^2} = \pi^2(t) \mu(\tau \pi(t)). \end{aligned}$$

□

The point process notation allows us to write the sums in (2) as integrals with respect to the counting measure $N_n(t)$. For example, for $t \in (h_t, 1 - h_t)$,

$$\sum_{i=N_n(0)+1}^{N_n(1)} K_{h_t}(t - t_{i,n}) = \int_{t-h_t}^{t+h_t} K_{h_t}(t - u) N_n(du). \quad (5)$$

Now suppose we are interested in using $\hat{\nu}_h(t) = \frac{1}{n} \sum_{i=N_n(0)+1}^{N_n(1)} K_{h_t}(t - t_{i,n})$ as an estimator of the average local intensity $\nu(t)$. Provided the bandwidth satisfies $h_t \rightarrow 0$, $nh_t \rightarrow \infty$, and

(C3)

$$\sup_{t \in (0,1)} \left| \frac{\bar{\lambda}_n(t)}{n} - \nu(t) \right| = o(h_t^2),$$

then when $n \rightarrow \infty$, $\hat{\nu}_h(t)$ has expectation

$$\begin{aligned} E(\hat{\nu}_h(t)) &= \frac{1}{n} \int_{t-h_t}^{t+h_t} K_{h_t}(t-u) \bar{\lambda}_n(u) du \\ &= \nu(t) + \frac{1}{2} \mu_2(K) h_t^2 \nu''(t) + o(h_t^2), \end{aligned}$$

and variance (see Appendix)

$$\text{Var}(\hat{\nu}_h(t)) = \frac{R(K)}{nh_t} \{\nu(t) + \eta(t)\} + o\left(\frac{1}{nh_t}\right). \quad (6)$$

Hence combining terms, the asymptotic MSE of $\hat{\nu}_h(t)$ is given by

$$E(\hat{\nu}_h(t) - \nu(t))^2 = \frac{1}{4} \mu_2^2(K) h_t^4 \nu''(t)^2 + \frac{R(K)}{nh_t} (\nu(t) + \eta(t)) + o\left(h_t^4 + \frac{1}{nh_t}\right). \quad (7)$$

This is minimized by the bandwidth

$$h_{AMSE}^*(t) = \left(\frac{\{\nu(t) + \eta(t)\} R(K)}{n \nu''(t)^2 \mu_2^2(K)} \right)^{\frac{1}{5}}. \quad (8)$$

Example. The linear self-exciting process by Hawkes (1971) has conditional intensity function

$$\lambda(t) = \phi + \int_{-\infty}^t g(t-u) N(du), \quad (9)$$

for some parameter ν and non-negative ‘infectivity’ function $g(\cdot)$. A common choice is the exponential model $g(u) = \alpha e^{-\beta u}$, in which case the process is stationary for $\frac{\alpha}{\beta} < 1$, with mean intensity $\rho = \frac{\phi\beta}{\beta-\alpha}$ and covariance density $\mu(t, s) = \frac{\alpha\rho(2\beta-\alpha)}{2(\beta-\alpha)} e^{-(\beta-\alpha)|t-s|}$ for any (t, s) . In the speeding-up setting of (4) with $N(\cdot)$ a Hawkes process, we find that

$$\begin{aligned} \nu(t) &= \frac{\pi(t)\phi\beta}{\beta-\alpha}, \\ \eta(t) &= \int_{-\infty}^{\infty} \pi^2(\tau) \mu(\tau, t) d\tau = \frac{\pi(t)\alpha\phi(2\beta-\alpha)}{\beta-\alpha}. \end{aligned}$$

Hence when the bandwidth for the smoother is chosen based on the Hawkes model, (8) reduces to

$$h_{AMSE}^*(t) = \left(\frac{\pi(t)(1-\alpha/\beta)(1+\alpha(2-\alpha/\beta))R(K)}{n\phi\pi''(t)^2\mu_2^2(K)} \right)^{\frac{1}{5}}.$$

□

2.4 Asymptotic normality

For the asymptotic analysis we need to impose some mixing conditions on both the covariates and error as well as on the point process. Therefore we adapt the usual mixing conditions to the infill setting. For point processes it is common to assume that the number of events within two non-overlapping time intervals become independent once they are sufficiently far apart. Now, since the expected number of observations grows with n , it makes sense to let the distance between two time intervals depend on n too. This is reflected in the mixing condition below. First we introduce the notation

$$X_{\{u\},n} := X_{N_n(u),n},$$

and similarly for $\epsilon_{\{u\},n}$. Note that $X_{\{u\},n}$ is a piecewise deterministic, càglàd function on the real line. It can be seen as the value of the covariate after the last jump before or at time u . Also define

$$\mathcal{F}_{a,n}^b = \sigma\{N_n(A), A \subseteq (a, b), X_{\{t\},n} : a \leq t \leq b\}.$$

Definition. (i) A marked point processes $(N(t), X_{\{t\}})$ with natural filtration \mathcal{F}_t is α -mixing if $\alpha(s) \rightarrow 0$ when $s \rightarrow \infty$, where

$$\alpha(s) = \sup_{\substack{A \in \mathcal{F}_{-\infty}^u, B \in \mathcal{F}_v^\infty \\ |v-u| \geq s}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \quad (10)$$

(ii) A sequence of marked point processes $(N_n(t), X_{\{t\},n})$ with family of natural filtrations $\{\mathcal{F}_{t,n}\}$ is α -mixing *in rescaled time* if $\alpha(s) \rightarrow 0$ when $s \rightarrow \infty$, where

$$\alpha(s) = \sup_{n \geq 1} \sup_{\substack{A \in \mathcal{F}_{-\infty,n}^u, B \in \mathcal{F}_{v,n}^\infty \\ |v-u| \geq \frac{s}{n}}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \quad (11)$$

Part (i) in this definition is a standard mixing conditions for point processes to be found for example in Cranwell and Weiss (1978), or for marked point processes in Pawlas (2009). The mixing condition in rescaled time in part (ii) differs from this one in that for each n , the supremum is taken over intervals whose gap is at least $\frac{s}{n}$. This requires the dependency to vanish proportional to the expected number of observations, or equivalently the integrated intensity, in between intervals, rather than their mere distance in time. This imposes some restrictions on the dependency structure for different n . Informally, the effective distance between two time intervals grows with the amount of information in between them.

Example. (cont.) In the speeding-up setting of (4), it follows almost straight from the definition that α -mixing (in the usual sense) of the baseline process $N(t)$ ensures that the sequence of point processes $\{N_n(t)\}$ is α -mixing in rescaled time. \square

Hence we impose the following condition on the point process:

(C4) $\{(N_n(t), X_{\{t\},n}, \epsilon_{\{t\},n})\}$ is α -mixing in rescaled time, with the mixing coefficients satisfying

$$\int_0^\infty w^\lambda \alpha(w)^{1-2/\delta} dw < \infty, \quad \text{for some } \lambda > 1 - 2/\delta \quad (12)$$

Since we are interested in estimating a regression function locally in time, we do not need to impose global stationarity of the covariates. Instead, we only require the following local stationarity condition, which is slight adaptation of the one in Vogt (2012) to allow for an unequally spaced grid:

(C5) The process $\{X_{i,n}\}$ observed at times $\{t_{i,n}\}$ is locally stationary (see e.g. Vogt (2012)), i.e. for each time point $t \in [0, 1]$ there an exists a stationary process $\{X_{i,n}^*(t)\}$, such that

$$|X_{i,n} - X_{i,n}^*(t)| \leq \left(|t_{i,n} - t| + \frac{1}{n} \right) U_{i,n}(t) \quad \text{a.s.},$$

where $U_{i,n}(t)$ is a positive-valued process satisfying $E(U_{i,n}(t)^\rho | t_{i,n}) < C$ a.s. for some $\rho > 0$ and $C < \infty$ independent of t, i , and n .

Furthermore, we require the following regularity conditions:

(C6) The stationary process $X_{i,n}^*(t)$ has density $f(t, x) > 0$ independent of i and n for all $(t, x) \in (0, 1) \times S$ with S a compact subset of \mathbb{R} . Here $f(t, x)$ is bounded bounded for every t and continuously differentiable w.r.t t for every x . Furthermore the density $f_{X_{i,n}}$, the conditional density $f_{\epsilon_{i,n}|X_{i,n}}$ and the joint conditional density $f_{\epsilon_{i,n}, \epsilon_{j,n}|X_{i,n}, X_{j,n}}$ are bounded for every i, j and n .

(C7) The kernel K is a density, symmetric around zero and with zero value outside $[-1, 1]$. Define $R(K) = \int K^2(x) dx < \infty$ and $\mu_2(K) = \int x^2 K(x) dx < \infty$. Furthermore K is bounded and Lipschitz continuous, i.e. there exists an $L < \infty$ such that $|K(u) - K(v)| \leq L|u - v|$ for every $u, v \in \mathbb{R}$.

(C8) The regression function $m(t, x)$ is twice continuously partially differentiable and Lipschitz continuous w.r.t. t and x . $\sigma(t, x) > 0$ is continuously differentiable in both arguments.

(C9) $E|\epsilon_{i,n}|^{2+\delta} < \infty$ for every i, n for some $\delta > 0$

(C10) The bandwidths satisfy $h_t \rightarrow 0$, $h_x \rightarrow 0$, $nh_t h_x \rightarrow \infty$, $nh_t^5 h_x = O(1)$, $nh_t h_x^5 = O(1)$, $h_t (nh_x)^{1-2r} \rightarrow 0$ with $r = \min(\rho, 1)$, and $(nh_t)^{1-\epsilon} h_x^{\frac{\lambda(2+\delta)+2+2\delta}{\lambda(2+\delta)+2}} = O(1)$ for some $\epsilon > 0$.

Theorem 1. Let (C1)-(C10) hold. Then for $(t, x) \in (h_t, 1 - h_t) \times S$

$$\sqrt{nh_t h_x} (\hat{m}_h(t, x; 0) - m(t, x) - B(t, x)) \xrightarrow{d} N(0, V(t, x)), \quad (13)$$

where the asymptotic bias is given by

$$B(t, x) = h_t^2 \mu_2(K) \left(\frac{\partial}{\partial t} m(t, x) \frac{\partial}{\partial t} \log(f(t, x) \nu(t)) + \frac{1}{2} \frac{\partial^2}{\partial t^2} m(t, x) \right) + h_x^2 \mu_2(K) \left(\frac{\partial}{\partial x} m(t, x) \frac{\partial}{\partial x} \log f(t, x) + \frac{1}{2} \frac{\partial^2}{\partial x^2} m(t, x) \right) + o(h_t^2 + h_x^2), \quad (14)$$

and the asymptotic variance by

$$V(t, x) = \frac{R^2(K) \sigma^2(t, x)}{nh_t h_x \nu(t) f(t, x)} + o\left(\frac{1}{nh_t h_x}\right). \quad (15)$$

This result shows that the asymptotic distribution of our estimator is the same as that when the observations would be independent of each other, since the covariance function of neither the covariates or the point process shows up in the asymptotic variance. We also see that compared to deterministic sampling times, random sampling times lead to the additional term $\nu(t)$ in the asymptotic bias and variance expressions. For example, a higher mean intensity leads to more observations and hence to a lower variance. Note that this factor depends on the distribution of the covariates whenever these contribute to the conditional intensity, as illustrated in the following example.

Example. Suppose that the conditional intensity is specified by the marked Hawkes process

$$\frac{\lambda_n(t)}{n} = \phi\pi(t) + \int_{-\infty}^t h(X_{\{u\},n}) g(n(t-u)) dN_n(u),$$

where $h(\cdot)$ is a nonnegative, Lipschitz continuous function that captures the impact of the covariate, and $g(\cdot)$ is the infectivity function as before, now with its argument scaled by n . Then it can be shown (see Appendix) that

$$\frac{\bar{\lambda}_n(t)}{n} \rightarrow \frac{\phi\pi(t)}{1 - \int_S h(x) f(t, x) dx \int_0^\infty g(\tau) d\tau} =: \nu(t), \quad (16)$$

provided $\int_0^\infty g(\tau) d\tau < \infty$, $\int_0^\infty \tau g(\tau) d\tau < \infty$, and the stationary condition

$$\int_S h(x) f(t, x) dx \int_0^\infty g(\tau) d\tau < 1. \quad (17)$$

Hence the limiting local intensity can be expressed in terms of the density $f(t, x)$ of the stationary approximation process. The likelihood functions of this and similar models are available (see Daley and Vere-Jones (2003, §7.3)) and can be used to estimate the parameters in $\nu(t)$. \square

3 Application: Estimating the State-Price Density from European options

The state-price density (SPD) is a central object in two main branches of asset pricing theory. In no arbitrage-based models, the SPD is the density of the Equivalent Martingale Measure (EMM), typically denoted \mathbb{Q} , under which all discounted price processes become martingales. The existence of the EMM follows from the absence of arbitrage, while its uniqueness follows from market completeness. In this context estimates of the density of \mathbb{Q} implicit in option prices can be used for the pricing and hedging of new or illiquid assets, or to filter out the pricing error in traded option prices (Chen and Xu, 2014). It can also be used to assess parametric models for the dynamics of the underlying asset under the \mathbb{Q} -measure. A comparison between the theoretical and the estimated SPD then reveals if the model is able to capture empirical features such as skewness or fat tails.

In consumption-based equilibrium models, the SPD is the product of the density of the objective measure \mathbb{P} , and a stochastic discount factor or pricing kernel. The latter is proportional to the marginal rate of substitution over states of a representative agent. Hence an estimator of the SPD combined with an estimator of the objective density allows computing so-called empirical pricing kernels (EPK) (Rosenberg and Engle, 2000). These can be used to infer the risk aversion of investors over different levels of wealth at a certain maturity. The finding that the implied Arrow-Pratt risk aversion coefficient does not monotonically decrease with wealth has been referred to as the ‘pricing kernel puzzle’ (Jackwerth, 2000).

The call pricing function can be estimated via a nonparametric regression of observed call prices on their trade characteristics stock price, strike price, and time-to-maturity (Aït-Sahalia and Lo, 1998). The SPD is then obtained from the second derivative of call pricing function with respect to the strike price, using the result by Breeden and Litzenberger (1978). The availability of intraday option prices greatly enhances the applicability of data-intensive nonparametric methods for estimating the SPD. Proposed methods include kernel smoothers (Aït-Sahalia and Lo, 1998; Aït-Sahalia and Duarte, 2003), smoothing splines (Yatchew and Härdle, 2006), and various methods based on the shape constraints required by the absence of arbitrage.

An important difference in the literature is the horizon over which option prices are pooled together. Aït-Sahalia and Lo (1998) assume that call prices over a period of about one year are generated by the same call pricing function. Their estimate of the SPD can therefore be interpreted as the average SPD over the sample period, rather than a point estimate for a given point in time. Alternatively, Aït-Sahalia and Duarte (2003) show how a cross-section of option prices at a given point in time can suffice to get feasible estimates. They use shape constraints such as monotonicity and convexity to reduce the variance of their estimator. However, since the

number of traded strikes is limited by the exchange, their estimator is vulnerable to small sample issues.

Yet whereas the SPD may not be constant in the long run, it likely does exhibit some persistency in the short run. The availability of high frequency options data allows exploiting this persistency by pooling together data from time points close to each other, say from the same day or the same hour. The SPD is less likely to vary heavily during small intradaily time windows than for longer sampling periods. Indeed, by now there are a number of studies that use intradaily options data to estimate the call pricing function, e.g. Härdle and Hlavka (2009) and Chen and Xu (2014). However, the call pricing function may still vary throughout the day due to time-of-the-day effects, or indeed changes in the SPD due to new information. Empirical evidence of such intraday patterns comes from autocorrelation in the errors of a time-homogeneous model (Härdle and Hlavka, 2009).

This motivates the application of a time-varying regression model for the call pricing function. This makes it possible to estimate the SPD at any time point of interest, and to study the dynamics of the SPD over a given sampling period. Rather than having to set an arbitrary pooling window, such a model is naturally capable to balance the trade-off between increasing the sample size to reduce variance while controlling the bias due to time-variation in the SPD.

The remainder of this section is organized as follows. First we provide some preliminaries on the object of interest. Then we demonstrate how its estimation can be done using the time-varying regression model we introduced in Section 2. Finally we discuss the empirical results when estimating our model with options data from the E-mini S&P 500.

3.1 Preliminaries

The First Fundamental Theorem of Asset Pricing states that if a market has no arbitrage opportunities, then there exists a measure $\mathbb{Q} \sim \mathbb{P}$ such that the price process of any traded asset discounted by the price of a riskfree asset is a \mathbb{Q} -martingale. See for example Karatzas and Shreve (1998) for details. Intuitively, this measure takes account of risk aversion by adjusting the probabilities in such a way that assets are priced by the expectation of their future payouts. If all investors would be riskneutral, then the \mathbb{P} and \mathbb{Q} -measure coincide, which is why \mathbb{Q} is also referred to as the ‘riskneutral’ measure. The Second Fundamental Theorem of Asset Pricing states that if the market is complete, that is if all contingent claims can be perfectly replicated, then this measure is unique. Combining these theorems implies that when a contingent claim with payout ξ_T is introduced in an arbitrage-free and complete market, its discounted price satisfies the martingale property

$$\frac{\xi_t}{B_t} = E^{\mathbb{Q}} \left(\frac{\xi_T}{B_T} \middle| \mathcal{F}_t \right), \quad (18)$$

Hence its unique no-arbitrage price is given by

$$\xi_t = E^{\mathbb{Q}} \left(\frac{B_t}{B_T} \xi_T \middle| \mathcal{F}_t \right) = e^{-r(T-t)} E^{\mathbb{Q}} (\xi_T | \mathcal{F}_t), \quad (19)$$

where r is the riskfree rate of the bond B .

A European call option is a contingent claim which gives the owner the right but not the obligation to buy the stock at maturity time T for a given strike price K . Hence its payoff at maturity equals $C_T = (S_T - K)^+$, and its unique no-arbitrage price is given by

$$C_t = e^{-r(T-t)} E^{\mathbb{Q}} ((S_T - K)^+ | \mathcal{F}_t). \quad (20)$$

Similarly, a European put option gives the owner the right to sell the stock at maturity for a given strike price K , and thus has payoff $P_T = (K - S_T)^+$, and no-arbitrage price

$$P_t = e^{-r(T-t)} E^{\mathbb{Q}} ((K - S_T)^+ | \mathcal{F}_t). \quad (21)$$

Note that $(K - S_T)^+ = K - S_T - (S_T - K)^+$, so the payoff of a put option can be replicated by a linear combination of the bond, the stock, and the call option. As a result, put-call parity states the following relation between the no-arbitrage prices of put and call options:

$$P_t = e^{-r(T-t)} K - S_t + C_t \quad (22)$$

Provided the measure \mathbb{Q} admits a density $f_t^{\mathbb{Q}}$ with respect to the Lebesgue measure for information set \mathcal{F}_t , (20) can be expressed as

$$C_t = e^{-r(T-t)} \int_K^{\infty} (S_T - K) f_t^{\mathbb{Q}}(S_T) dS_T. \quad (23)$$

By differentiating with respect to K , we derive

$$\frac{\partial C_t}{\partial K} = -e^{-r(T-t)} \int_K^{\infty} f_t^{\mathbb{Q}}(S_T) dS_T. \quad (24)$$

By differentiating once more, as in Breeden and Litzenberger (1978), we derive

$$\frac{\partial^2 C_t}{\partial K^2} = e^{-r(T-t)} f_{t,S_T}^{\mathbb{Q}}(K). \quad (25)$$

This relation tells that the SPD is proportional to the curvature of the call pricing function. It forms the theoretical basis for estimation of $f_{t,S_T}^{\mathbb{Q}}$.

The fact that $f^{\mathbb{Q}}$ is a density also puts some restrictions on the call pricing function. In particular, from (24) and the nonnegativity of the density and its integrability to one, it must

hold that

$$-e^{-r(T-t)} \leq \frac{\partial C_t}{\partial K} \leq 0, \quad (26)$$

while (25) implies that

$$\frac{\partial^2 C_t}{\partial K^2} \geq 0. \quad (27)$$

Hence the call pricing function must be decreasing and convex. Any local violation implies negative state-prices, and hence arbitrage opportunities. No-arbitrage bounds can also be obtained for the level of the call pricing function, see Ait-Sahalia and Duarte (2003) for details.

3.2 Estimation method

From 23 it can be seen that the no arbitrage price of a call option is determined by the SPD $f_t^{\mathbb{Q}}(S_T)$, and by the observable variables strike price K , and time-to-maturity $T - t$. The Breeden and Litzenberger (1978) result 25 gives the relation (up to integration constants) between the SPD and the price of a call option C as a function of its strike price K . Hence an estimator of the SPD follows directly from an estimator of the call pricing function.

However, both the call pricing function and the SPD are random, as they are a conditional expectation (resp. density) given information in the filtration \mathcal{F}_t . Therefore it is convenient to make a Markov assumption, which states that all relevant information from the past is summarized by the present level of some state variables. These can either be observed, such as the current stock price S_t , or unobserved, such as the spot volatility (Song and Xiu, 2012). As we wish to avoid specifying which state variables are relevant, we propose to use the current level of the stock price as the only random state variable, but in addition include time itself in the regressor set. This means that the randomness in the call prices can only be generated by the underlying stock price and other observable trade characteristics, yet we are able to capture smooth changes in the functional form of the dependency of the call price on these variables. This leads to the nonparametric regression model

$$C_t = C(t, \tau, S_t, K_t, r_{t,T}) + \epsilon_t, \quad (28)$$

where ϵ_t is an error term that satisfies $E(\epsilon_t | \mathcal{F}_t) = 0$.³

3.2.1 Dimension reduction

Since nonparametric regression is vulnerable to the curse of dimensionality, we make some dimension reductions that are common in the literature. Firstly, we will be dealing with options

³Even more generally, the option price may depend on the dividend yield $\delta_{t,T}$. For options that are written on the futures prices, these are incorporated and hence do not affect option prices separately. For options written on stock prices, they may do so. Still it can be reasonably assumed that the dividend yield enters the option price only via the futures price $F_{t,T} = S_t e^{(r_{t,T} - \delta_{t,T})\tau}$, see Ait-Sahalia and Lo (1998) for details.

written on the price of a futures contract F_T , rather than price of a stock S_T .⁴ This has the advantage that the interest rate $r_{t,T}$ only affects the discount factor and not the SPD of the futures price, so it can simply be removed by discounting the observed call price.⁵ Specifically, if we work with the inversely discounted prices $C_t^d = e^{r_{t,T}(T-t)}C_t$, then the interest rate $r_{t,T}$ drops out of the regressor set:

$$C_t^d = C(t, \tau, F_t, K_t) + \epsilon_t. \quad (29)$$

Second, smoothing in the time-to-maturity dimension τ requires combining information on European option contracts with different maturities. Since the gaps between the maturities that are traded may take several weeks or even months, a long sampling period is needed to get repeated observations of the same value of τ . When working with high frequency data from only one or a few days, this gap cannot be closed and hence it is not possible to cover the points in between the maturity times. Therefore, high frequency studies typically produce separate estimators for each maturity T , rather than smoothing between them (e.g. Härdle and Hlavka, 2009). Therefore we restrict attention to options with the same maturity date T , so that we do not smooth explicitly in time-to-maturity dimension τ , and are left with the three-dimensional regression function $C(t, F_t, K_t)$. As a caution, note that when only one maturity time T is in the sample, then t and τ are not separately identified. In this case only the combined effect of time variation due to being closer to maturity and new information arising is identified. However, for short intradaily windows the time-to-maturity effect is likely to be small, and might otherwise be controlled for by an appropriate rescaling of the resulting densities.

Finally, one more dimension reduction can be done by a normalization of the strike prices K_t by the futures price level F_t at the time of trade. The rationale for this is provided by Merton (1973, Thm. 9), who shows that if the distribution of the *return* of the underlying asset S is independent from the *level* of S , then the call pricing function is homogeneous of order one in F and K , i.e.

$$C_t(F, K) = FC_t(1, K/F). \quad (30)$$

The condition that the return distribution of the underlying asset is independent from its level can be straightforwardly checked within a parametric model for the dynamics of the underlying asset. The class of models that satisfies this assumption includes the widely used geometric Brownian motion and many stochastic volatility models (Joshi, 2001). Yet in general, this condition may not be easy to test as it requires measuring changes in the return distribution. Now define the

⁴If the futures contract has the same maturity as the option, then the prices of a European option written on the stock and on the futures contract are the same (Aït-Sahalia and Lo, 1998). If the futures contract expires after the maturity of the option, this need no longer hold. Yet this only matters insofar the resulting estimate is now the SPD of the futures price, which only differs from that of the stock price by a known discount factor.

⁵Alternatively, for short, intradaily time windows the interest rate is unlikely to change heavily and can therefore safely be ignored.

‘moneyness’ ratio $M = K/F$, and the normalized option price $\tilde{C}_t = C_t^d/F_t$. Then, if we assume this condition for homogeneity holds, we can write the regression function in (29) as

$$\tilde{C}_t = m(t, M_t) + \epsilon_t, \quad (31)$$

where we assume $E(\epsilon_t|M_t) = 0$ and $\text{Var}(\epsilon_t|M_t) = \sigma^2(t, M_t)$. This means that we have reduced the regression model to a function of only two variables, time t and moneyness M_t .⁶

Moreover, homogeneity of F and K implies that the second derivative of the call pricing function becomes

$$\frac{\partial^2 C_t^d}{\partial K^2} = \frac{\partial^2 F_t m(t, K/F_t)}{\partial K^2} = \frac{1}{F_t} \frac{\partial^2 m(t, M)}{\partial M^2}. \quad (32)$$

Remembering the Breeden and Litzenberger (1978) result (25), and using the substitution

$$f_{F_T/F_t}^{\mathbb{Q}}(K/F_t) = f_{F_T}^{\mathbb{Q}}(K)F_t, \quad (33)$$

we derive that

$$f_{F_T/F_t}^{\mathbb{Q}}(x) = \left. \frac{\partial^2 \tilde{m}(t, M)}{\partial M^2} \right|_{M=x}. \quad (34)$$

This provides a direct relation between the SPD of the gross return F_T/F_t and our regression function $m(\cdot)$.

3.2.2 Time-varying kernel regression

We can now employ the theoretical results in Section 2 on time-varying kernel regression to estimate the bivariate regression model (31). For this purpose assume that we have a sample of option prices C_{t_i} observed at normalized times $0 \leq t_1 < \dots < t_{N_n(1)} \leq 1$ with strike prices K_i and futures price level F_i . As before, we assume the trading times have been generated by a point process model, whose mean intensity is controlled by an unobserved index n . Hence the observed sample size is $N_n(1)$, where we have normalized $N_n(0) = 0$. We then propose to estimate the reduced call pricing function at time t and moneyness level x by

$$\hat{m}(t, x) = \frac{\sum_{i=1}^N K_{h_t}(t - t_i) K_{h_M}(x - M_{t_i}) \tilde{C}_{t_i}}{\sum_{i=1}^N K_{h_t}(t - t_i) K_{h_M}(x - M_{t_i})}, \quad (35)$$

where $K_h(\cdot) = 1/hK(\cdot/h)$ with $K(\cdot)$ a kernel and h_t and h_M bandwidths for the time and moneyness dimension, respectively. This estimator takes a weighted average over call prices, where the weights depend on the distance in terms of time and moneyness.

This estimator can be generalized to local polynomial smoothing of order p , where $\hat{m}(t, x; p)$

⁶The homogeneity assumption also has been used recently by Chen and Xu (2014), who use intradaily data to estimate the call pricing function nonparametrically as a function of moneyness and time-to-maturity only.

is the constant coefficient of a fitted p -th order polynomial around (t, x) , where the weight of observation i is determined by $K_{h_t}(t - t_i)K_{h_M}(x - M_{t_i})$. Although we have not formally derived the theory for this generalization, the asymptotic normality result in Theorem 1 can be proven to go through under very similar conditions, except that this requires some more smoothness on the regression function m , and the bias and variance order terms now depend on higher order partial derivatives. In particular, we also use the local linear estimator with $p = 1$, which removes a first-order bias term and is less vulnerable to boundary effects (Fan and Gijbels, 1996). In our setting it is given by the minimizer of β_0 of the sum of squared residuals of the fitted polynomial around (t, x)

$$\sum_{i=1}^N \left(\tilde{C}_{t_i} - \beta_0 - \beta_1(t - t_i) - \beta_2(x - M_{t_i}) \right)^2 K_{h_t}(t - t_i)K_{h_M}(x - M_{t_i}) \quad (36)$$

Another advantage of higher order polynomials in our setting is that they give a direct method to estimate the SPD from higher coefficients in the polynomial. For example, using the local linear method, the estimator for the moneyness coefficient $\hat{\beta}_2$ is a direct estimator of the first partial derivative of $m(t, x)$ w.r.t. moneyness x . Hence using (34), we propose the estimator of the SPD given by

$$\hat{f}_{F_T/F_t}^{\mathbb{Q}}(x) = \frac{\partial}{\partial x} \hat{\beta}_2(x). \quad (37)$$

In practice this estimator can be implemented by estimating $\beta_2(x)$ on a grid of values of x and taking numerical differences.

Provided the conditions of Theorem 1 are valid, our estimator is asymptotically normal with bias and variance as given in the theorem. Therefore we now comment on why the local stationarity and the mixing conditions are indeed sensible for the moneyness covariate. Recall first that the moneyness covariate is the ratio of two components, the traded strike price and the futures prices at the time of trade. The range of strike prices that is listed on the exchange is typically set on a daily basis in reaction to movements in the underlying.⁷ Hence moneyness will be recentered around one, which guarantees stationarity from day to day. However, it may still be nonstationary within a certain day. To see this note for a given strike price K' that is listed on a certain day, we have $\Delta \log M_{t_i} = \Delta \log \frac{F_{t_i}}{K'} = \log \frac{F_{t_i}}{F_{t_{i-1}}}$. So if intraday log futures returns are stationary then $\log M_t$ has a unit root and hence M_t is nonstationary. Yet, also within a day, we might expect traded strike prices to vary along with the underlying. When this happens in such a way that the relative amount of trading at different levels of moneyness is (roughly) constant during the day, M_t will be (locally) stationary. Regarding the mixing condition, the effective recentering of the moneyness level around one similarly rules out any long-term memory

⁷Our dataset is from the E-mini S&P 500, which lists strike prices at regular intervals within $\pm 50\%$ of the closing price of the futures on the previous day. See Group (2014) for details.

in the futures price level. Hence as long as the strike prices being traded do not possess some long-memory property, the mixing condition for the moneyness covariate seems sensible. This would for example be satisfied if the strike prices follow a finite-order Markov chain, regardless of the trading intensity. The next section describes some empirical properties of M_t to illustrate how these conditions apply to our setting.

3.3 Empirical results

3.3.1 Data Description

Our dataset consists of the transaction prices and bid and ask quotes of European options on a futures contract on the E-mini S&P 500 equity index, and the transaction prices and bid and ask quotes of the futures contract on which these options are written. The bid and ask quotes are ‘best’ quotes, that is, only ask (bid) quotes entering the order book that are lower (higher) than the previously best quote are recorded. They are traded on the electronic GLOBEX platform and obtained from the CME Group. The sample period is November 2013. The options are End-of-Month options, which means they deliver the futures contract at 3:00 PM Chicago Time on the last trading day of the month, cash-settled with the strike price. Note the futures contract is the nearest-expiring quarterly E-mini S&P 500 futures after maturity of the option. Hence the futures contract expires beyond the maturity time of the option. For example, for the November End-of-Month options the relevant futures contract is the December futures contract. However, since there exists a simple no-arbitrage relation between the futures prices and the equity index, see footnote 3, we can restrict attention to the SPD of the futures price at maturity of the option. This should be the same as the SPD of the equity index up to a known discount factor.

The options and futures on the E-mini S&P 500 are traded around the clock, except a daily maintenance interval from 4:15-5:00 PM. The time stamps are recorded in seconds. Since several transactions often take place within the same second, the data are ordered in terms of their trade sequence number. This numbering is reset at 5:00 PM every day, so that observations after 5:00 PM on a particular day are classified to the next trading day.

The dataset not only contains all transactions of option and futures contracts on each trading day, but also all updates on the highest bid price and the lowest ask price. Table 1 gives the number of transactions, bid updates, and ask updates, for futures and options on November 1, 2013.

Table 1: Number of observations for E-mini S&P 500 futures and option prices in November 2013.

	Futures	Calls	Puts
Transactions	6,922,393	11,329	15,713
Bid Prices	20,883,086	10,947,785	14,223,270
Ask Prices	20,883,071	11,472,708	15,395,095

As seen from the table, the number of updated bid and ask quotes of the futures and option prices is of the same order, with about twice as many observations for futures. However, whilst the number of futures transactions is in the order of millions, there are only a few ten thousands of option transactions. This indicates the huge potential of adding bid and ask prices to the sample. Other observations are that there are roughly as many bid updates as ask updates, and that put option are slightly more actively traded than call options.

Fig. 1 shows a scatter plot of moneyness levels of option transactions during November 2013 versus their trading times. These are the realizations of the covariate used to construct our estimator. The unevenly spaced nature of the transaction times is clearly observed, in particular from the difference between day and night time (five vertical lines per week), and the weekends when the markets are closed. Note also that closer to maturity, the traded moneyness levels are more centered around one.

Figure 2 shows time series plots of traded strike prices (upper panel), futures prices (mid panel), and their moneyness ratio (lower panel). The futures prices are synchronized with the option transactions by taking the average of the last recorded best bid and ask quotes. The upper panel shows an intradaily pattern in trading activity, which is high between the working hours 9:00AM-6:00PM and low during the night hours. In particular, during night hours almost only strike prices that are close to at-the-money are quoted, whereas during working hours almost the full range of listed strike prices is quoted upon. The scale of the mid panel shows that there were no large shocks in the futures price on this day, so that the time series of the moneyness ratio in the lower panel closely resembles the time series of the strike price. This suggests that intraday futures price movements are unlikely to induce severe nonstationarity in moneyness during intradaily time windows.

Figure 3 shows recorded call and put prices as a function of their moneyness level prices, for both transaction prices and best bid and ask quotes. Under homogeneity, the option pricing function should not be affected by scaling both the option price and the strike price by the futures price. The data clearly reflect that, by the absence of arbitrage, call prices should be convexly decreasing with strike price and put prices should be convexly increasing. By put-call parity (22), the put prices may be translated into call prices so that one convexly increasing function is displayed. The upper panel shows the best bid and ask quotes, from which it is clear that most activity is at strikes that are ‘at-the-money’, i.e. around with moneyness levels around one. For some intervals of there are remarkably high levels of the best ask prices. This can be explained

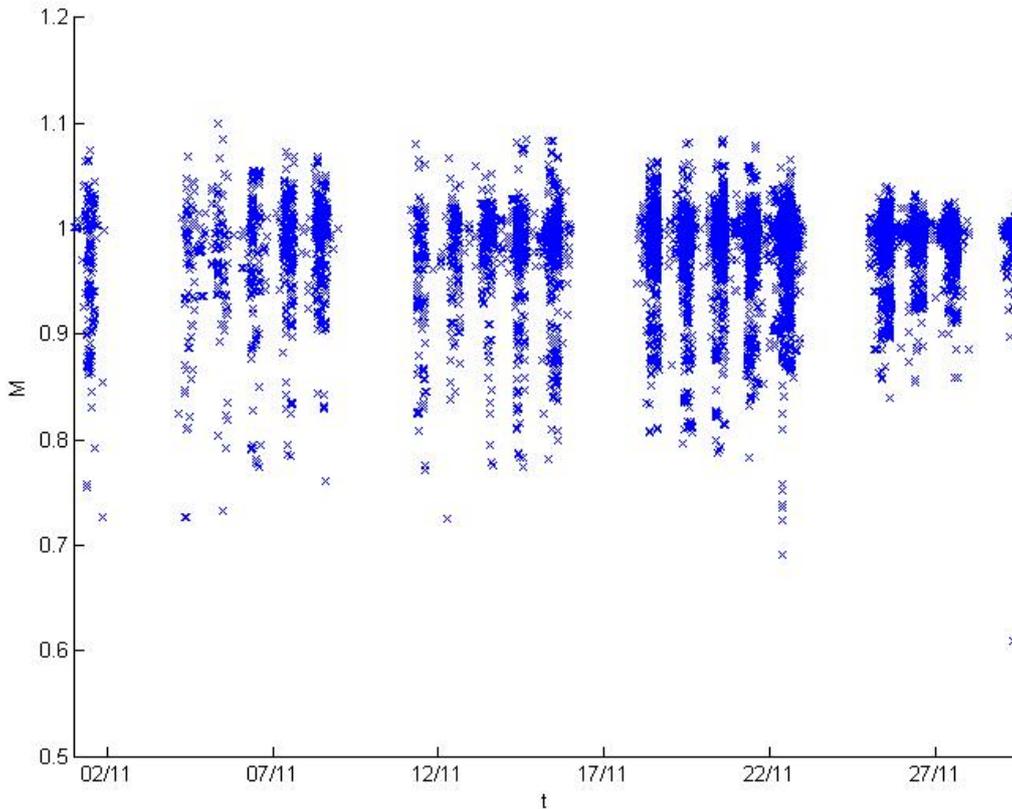


Figure 1: Time series plot of moneyness levels of options transactions during November 2013, for options expiring November 29, 2013.

by a lack of liquidity during parts of the day. It warns us that bid and ask prices, or even their averages, can be poor proxies of the actual market price. The lower panel shows transaction prices, from which it is even more clear that trading activity is centred around at-the-money strike prices. The transaction prices, however, do not display similar outliers as the bid and ask quotes do.

3.3.2 Bandwidth choice

To implement the kernel estimators we have to select the bandwidths h_t and h_x . A possible method would be to use plug-in bandwidths, which minimize the asymptotic Mean Square Error (AMSE) which can be computed from Theorem 1. However, the AMSE depends on unknown quantities, such as the density $f(t, x)$ and the mean intensity $\nu(t)$, which would have to be estimated first. An alternative is to use cross-validation, which is a completely data-driven method to choose the bandwidth. The bandwidths are then chosen to minimize objective functions of

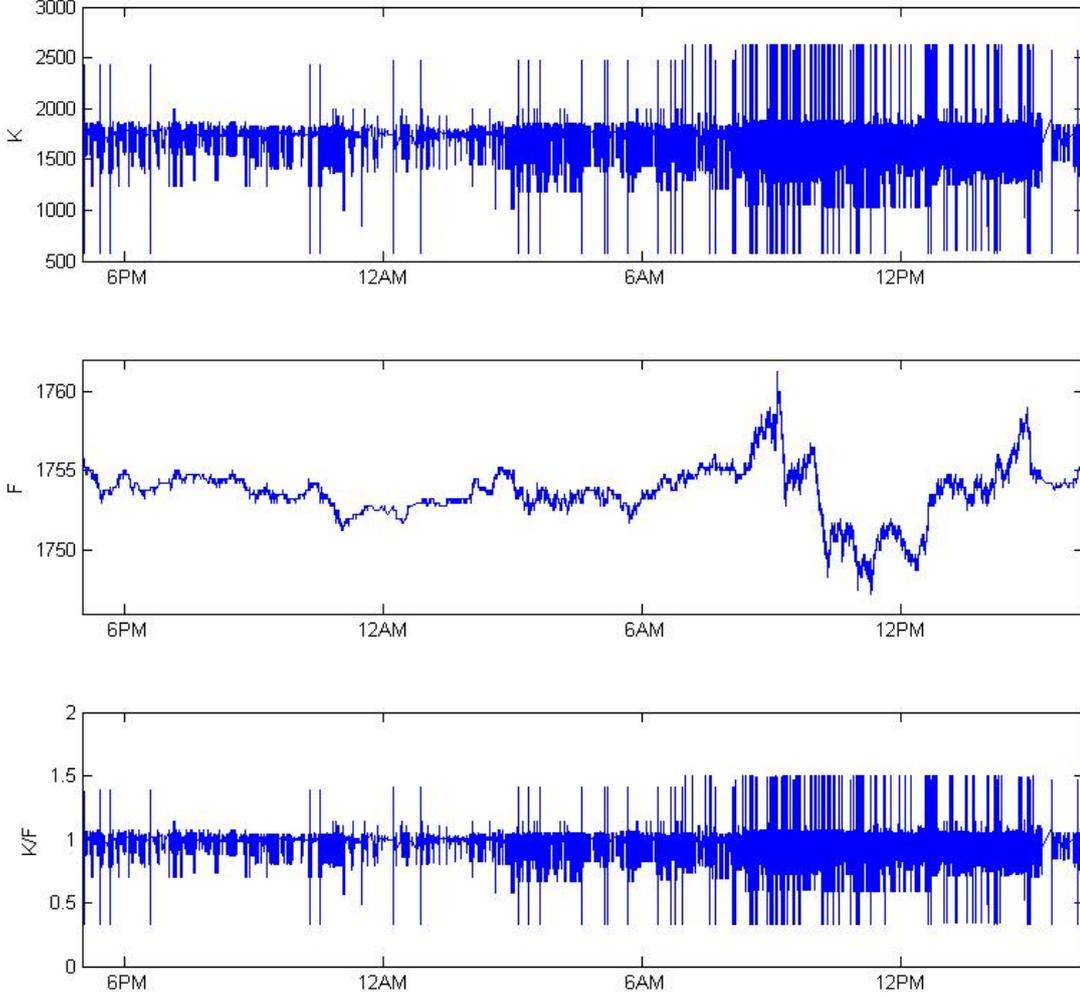


Figure 2: Time series plots of traded strike prices (upper panel), synchronized futures prices (mid panel), and their moneyness ratio (lower panel), for options traded on November 1, 2013.

the form

$$CV(h_t, h_x) = \sum_{i=1}^N (\tilde{C}_{t_i} - \hat{m}_{-i}(t_i, M_{t_i}; h_t, h_x))^2, \quad (38)$$

where \hat{m}_{-i} is the leave-one-out-estimator which uses all observations except i .

An important feature of transaction data is that the amount of data points can differ heavily from period to period. In our dataset this holds especially true as E-mini S&P 500 options are traded overnight, with the trading activity during the night much lower than during the day. With a globally cross-validated bandwidth h_t , this would lead to large differences in the effective sample size for the local weighing at different time points of interest. To balance this, and more

generally to take into account that the AMSE-optimal bandwidth varies over time t (as well as with the state x), we choose our bandwidths via *local* cross-validation. In particular, we divide the total time span covered by our data in blocks, and for each block choose the cross-validated bandwidth using (38) applied to the data in that block. Note that both data in and outside the block is used to calculate the leave-one-out estimators. For large samples, the choice of the block size is of minor importance compared to the effect of the bandwidth on the final estimate (Hall and Schucany (1989)). There is also an important computational advantage of using local cross-validation, as the ordering of the data in time means that only a subset of the data has to be used for each block.

3.3.3 Estimation results

We now apply the locally constant estimator (35) as well as the local linear estimator (37) to our dataset. Fig. 4 show the resulting estimates using all quotes on one particular trading day, November 7, 2013. This day is chosen as there was a large downturn of the futures prices during that day. The locally cross-validated bandwidths lead to effective sampling windows varying between five minutes and half an hour for bid and ask quotes during trading time. The upper panel shows the estimated call pricing function $\hat{m}(t, x)$ based on the locally constant estimator (35). These produce smooth surfaces which represent closely the payoff function $(F_T - K)^+$ of the European call options, but show a clear convex pattern which represents the time value of the options. The lower panels show the estimated state-price densities, the quantities that are of direct economic relevance as they indicate how valuable a unit payoff is to investors in different states of the economy (as proxied for by the S&P 500). From the lower panel in Fig. 4 a clear negative skewness and a fat left tail of the SPDs can be observed. They are obtained by taking numerical differences of (37) on a grid of values for moneyness. Whereas the call pricing and is a smooth surface, the surface plots for the SPD are much more wiggly. An explanation for this is that the differentiation aggravates differences, so that, *ceteris paribus*, a larger sample size is needed to estimate derivatives with the same amount of smoothness as the original function. Since the bandwidths have been chosen to optimize the fit of the call pricing function itself, rather than its derivatives, this has likely led to undersmoothed estimates of the SPD. A plug-in bandwidth selector might be employed to overcome this.

Regarding time variation, the fact that the cross validated bandwidth do not exceed more than a hour for quotes data, is evidence itself of the true regression function not being constant throughout the whole month. Formal testing procedures can point out whether the resulting changes in the SPD are statistically significant, as well as economically, which might be tested using improved pricing and hedging performance.

4 Conclusion

In this paper we proposed a method to estimate the state-price density from high frequency options data. Emphasis has been on the time-varying nature of the SPD, and the choice of a time-smoothing bandwidth when trading activity is varying from day to day and within days. For this purpose asymptotic theory of a time-varying locally constant kernel regressor under random sampling times is derived.

In the application, we see that locally constant and locally linear estimators are not yet able to produce smooth estimates of the SPD, which is the second derivative of the call pricing function. It may therefore be of interest to estimate this second derivative directly using a local cubic polynomial, as discussed by Fan and Gijbels (1996). The bandwidth can then be chosen to optimize the precision of the SPD. Furthermore practical plug-in bandwidth estimators may be implemented, for example using Hawkes-based models for the conditional intensity function. Another important topics left undiscussed is formal testing of this estimator, for example, to reject a time-invariant regression function or to test against parametric specifications.

Appendix

Proof of (6). Define for $(t, s) \in \mathbb{R}^2$

$$\mu_n^{(c)}(t, s) = \delta(t - s)\bar{\lambda}_n(t) + \mu_n(t, s), \quad (39)$$

with $\delta(\cdot)$ the Dirac delta function. Then we can write

$$\begin{aligned} \text{Var} \left(\frac{1}{n} \int_{t-h_t}^{t+h_t} K_{h_t}(t-u) N_n(du) \right) &= \frac{1}{n^2} \int_{t-h_t}^{t+h_t} \int_{t-h_t}^{t+h_t} K_{h_t}(t-u) K_{h_t}(t-v) \mu_n^{(c)}(u, v) dudv \\ &= \frac{1}{n} \int_{t-h_t}^{t+h_t} K_{h_t}^2(t-v) \frac{\bar{\lambda}_n(v)}{n} dv \\ &\quad + \int_{t-h_t}^{t+h_t} \int_{t-h_t}^{t+h_t} K_{h_t}(t-u) K_{h_t}(t-v) \frac{\mu_n(u, v)}{n^2} dudv \\ &= \frac{\nu(t)}{nh_t} R(K) + o\left(\frac{1}{nh_t}\right) \\ &\quad + \int_{-1}^1 \int_{-nh_t}^{nh_t} K(z) K(z+y/n) \frac{\mu_n(t+h_t z, t+h_t z+y/n)}{n^2} dy dz \\ &= \frac{R(K)}{nh_t} \left(\nu(t) + \int_{-\infty}^{\infty} \gamma_t \tau d\tau \right) + o\left(\frac{1}{nh_t}\right), \end{aligned}$$

using the dominated convergence theorem in the last step. \square

Proof of Theorem 1. Firstly, we decompose

$$\hat{m}_h(t, x; 0) - m(t, x) = \frac{\hat{g}^V(t, x) + \hat{g}^B(t, x)}{\hat{f}(t, x)}, \quad (40)$$

with

$$\hat{f}(t, x) = \frac{1}{n} \sum_{i=1}^{N_n(1)} K_{h_t}(t - t_{i,n}) K_{h_x}(x - X_{i,n}) \quad (41)$$

$$\hat{g}^V(t, x) = \frac{1}{n} \sum_{i=1}^{N_n(1)} K_{h_t}(t - t_{i,n}) K_{h_x}(x - X_{i,n}) \epsilon_{i,n} \quad (42)$$

$$\hat{g}^B(t, x) = \frac{1}{n} \sum_{i=1}^{N_n(1)} K_{h_t}(t - t_{i,n}) K_{h_x}(x - X_{i,n}) (m(t_{i,n}, X_{i,n}) - m(t, x)) \quad (43)$$

The proof now consists of the following three parts:

- (i) $\hat{f}(t, x) \xrightarrow{P} \nu(t) f(t, x)$
- (ii) $\hat{g}^B(t, x) \xrightarrow{P} \nu(t) f(t, x) B(t, x)$
- (iii) $\sqrt{nh_t h_x} \hat{g}^V(t, x) \xrightarrow{d} \nu(t) f(t, x) \times N(0, V(t, x))$.

It will be useful to represent the sums in (40) as integrals with respect to the counting process $N_n(t)$. Throughout we use the convention $dN_n(t) = N_n(t - dt, t)$. We will be most detailed for the density estimation part (i), the main arguments for the other parts being similar. For each of the parts we use the following asymptotic expansion. Note that similar to (??) and (6), it can be shown that for general r

$$\frac{1}{n} \int_{t-h_t}^{t+h_t} K_{h_t}(t-u)(t-u)^r N_n(du) = \begin{cases} h_t^r \mu_r(K) \nu(t) + o_P(h_t^r) & \text{if } r \text{ even} \\ h_t^{r+1} \mu_{r+1}(K) \nu'_n(t) + o_P(h_t^{r+1}) & \text{if } r \text{ odd.} \end{cases}$$

Hence using a Taylor expansion it follows that for any continuously differentiable function $g : [0, 1] \rightarrow \mathbb{R}$, $u \mapsto g(u)$,

$$\frac{1}{n} \int_{t-h_t}^{t+h_t} K_{h_t}(t-u) g(u) N_n(du) = g(t) \nu(t) + O_P\left(\frac{1}{\sqrt{nh_t}}\right) + o_P(h_t). \quad (44)$$

- (i) We can write the denominator term (41) as

$$\hat{f}(t, x) = \frac{1}{n} \int_{t-h_t}^{t+h_t} K_{h_t}(t-u) K_{h_x}(x - X_{\{u\},n}) N_n(du).$$

Its expectation is

$$\begin{aligned} E(\hat{f}(t, x)) &= \frac{1}{n} \int_{t-h_t}^{t+h_t} K_{h_t}(t-u) E(K_{h_x}(x - X_{\{u\},n}) N_n(du)) \\ &= \frac{1}{n} \int_{t-h_t}^{t+h_t} K_{h_t}(t-u) E(K_{h_x}(x - X_{\{u\},n}) | dN_n(u) = 1) \bar{\lambda}_n(u) du \end{aligned}$$

Now we approximate each $X_{\{u\},n}$ by its stationary counterpart $X_{\{u\},n}^*(u)$ at the time of observation $t_{i,n} = u$. For brevity from now on we write $X_{\{u\},n}^*(u) = X_{\{u\},n}^*$, that is, unless mentioned otherwise we assume the process is approximated around its own observation time. Following Vogt (2012, Thm. 4.2), let \bar{K} be a Lipschitz continuous function on $[-q, q]$, $q > 1$ with $\bar{K}(x) = 1$ for $x \in [-1, 1]$, and decompose

$$\begin{aligned} E(K_{h_x}(x - X_{\{u\},n}) | N_n(du) = 1) &= E\left(\bar{K}_{h_x}(x - X_{\{u\},n}) \{K_{h_x}(x - X_{\{u\},n}) - K_{h_x}(x - X_{\{u\},n}^*)\} | N_n(du) = 1\right) \\ &\quad + E\left(\{\bar{K}_{h_x}(x - X_{\{u\},n}) - \bar{K}_{h_x}(x - X_{\{u\},n}^*)\} K_{h_x}(x - X_{\{u\},n}^*) | N_n(du) = 1\right) \\ &\quad + E\left(K_{h_x}(x - X_{\{u\},n}^*) | N_n(du) = 1\right) \end{aligned}$$

For the first term, note that

$$\begin{aligned} \left| K\left(\frac{x - X_{\{u\},n}}{h_x}\right) - K\left(\frac{x - X_{\{u\},n}^*}{h_x}\right) \right| &\leq C \left| K\left(\frac{x - X_{\{u\},n}}{h_x}\right) - K\left(\frac{x - X_{\{u\},n}^*}{h_x}\right) \right|^r \\ &\leq CL \left| \frac{X_{\{u\},n} - X_{\{u\},n}^*}{h_x} \right|^r \\ &\leq \left| \frac{U_{\{u\},n}(u)}{nh_x} \right|^r, \end{aligned}$$

by boundedness and Lipschitz continuity of K , and local stationarity of $\{X_{\{u\},n}\}$. Here C denotes some constant which may take different values in different places. Hence from the existence of conditional moments of $U_{\{u\},n}$ it follows that the first term is $O_P\left(\frac{1}{n^r h_x^r}\right)$. Similarly it follows that the second term is also $O_P\left(\frac{1}{n^r h_x^r}\right)$, so with a standard approximation for the third term it follows that

$$E(K_{h_x}(x - X_{\{u\},n}) | N_n(du) = 1) = f(u, x) + o_P(h_x) + O_P\left(\frac{1}{n^r h_x^r}\right). \quad (45)$$

Using (44) with $g(u) = f(u, x)$ it follows that

$$E(\hat{f}_h(t, x)) = \bar{\lambda}_n(t) f(t, x) + o(h_t + h_x) + O\left(\frac{1}{n^r h_x^r}\right).$$

Its variance is given by

$$\begin{aligned}
\text{Var}(\hat{f}_h(t, x)) &= \frac{1}{(nh_t h_x)^2} \text{Var} \left(\int_{t-h_t}^{t+h_t} K \left(\frac{t-u}{h_t} \right) K \left(\frac{x-X_{\{u\},n}}{h_x} \right) dN_n(u) \right) \\
&= \frac{1}{(nh_t h_x)^2} \int_{t-h_t}^{t+h_t} K^2 \left(\frac{t-u}{h_t} \right) \text{Var} \left(K \left(\frac{x-X_{\{u\},n}}{h_x} \right) dN_n(u) \right) \\
&\quad + \frac{1}{(nh_t h_x)^2} \iint_{\substack{t-h_t \leq u, v \leq t+h_t, \\ u \neq v}} K \left(\frac{t-u}{h_t} \right) K \left(\frac{t-v}{h_t} \right) \\
&\quad \times \text{Cov} \left(K \left(\frac{x-X_{\{u\},n}}{h_x} \right) dN_n(u), K \left(\frac{x-X_{\{v\},n}}{h_x} \right) dN_n(v) \right) \\
&\equiv (V) + (CV).
\end{aligned}$$

The variance increments can be computed using the following conditioning argument, with the shorthand $K_{u,n}(x) = K((x - X_{\{u\},n})/h_x)$,

$$\begin{aligned}
\text{Var}(K_{u,n}(x)dN_n(u)) &= \text{Var}(E(K_{u,n}(x)|dN_n(u))dN_n(u)) \\
&\quad + E(\text{Var}(K_{u,n}(x)|dN_n(u))(dN_n(u))^2) \\
&= \text{Var}((h_x f(u, x) + o_p(h_x))dN_n(u)) + E(h_x R(K)f(u, x)dN_n(u)(1 + o_p(1))) \\
&= \bar{\lambda}_n(u)h_x R(K)f(u, x)du + o(nh_x du),
\end{aligned}$$

so that

$$\begin{aligned}
(V) &= \frac{R(K)}{nh_t h_x} \int_{t-h_t}^{t+h_t} K^2 \left(\frac{t-u}{h_t} \right) \frac{\bar{\lambda}_n(u)}{n} R(K)f(u, x)du + o\left(\frac{1}{nh_t h_x}\right) \\
&= \frac{\mu_2^2(K)f(t, x)\nu(t)}{nh_t h_x} + o\left(\frac{1}{nh_t h_x}\right).
\end{aligned}$$

For the covariance increments we use the law of total covariance to get, for $u < v$,

$$\begin{aligned}
\text{Cov}\{K_{u,n}(x)dN_n(u), K_{v,n}(x)dN_n(v)\} &= E\{\text{Cov}(K_{u,n}(x), K_{v,n}(x)|N_n(du), N_n(dv))N_n(du)N_n(dv)\} \\
&\quad + \text{Cov}\{E(K_{u,n}(x)|N_n(du), N_n(dv))N_n(du), \\
&\quad \quad \quad E(K_{v,n}(x)|N_n(du), N_n(dv))N_n(dv)\} \\
&= (\star) + (\star\star).
\end{aligned}$$

Here using $\mathbb{P}(N_n(du) = N_n(dv) = 1) = \bar{\lambda}_n(u)\bar{h}_n(u, v)dudv$ for $u < v$,

$$(\star) = \text{Cov}\{K_{u,n}(x), K_{v,n}(x)|N_n(du) = N_n(dv) = 1\} \bar{\lambda}_n(u)\bar{h}_n(u, v)dudv,$$

while the nonnegativity of the jumps implies that we can bound

$$(\star\star) \leq Ch_x^2 \text{Cov} \{N_n(du), N_n(dv)\} = Ch_x^2 \mu_n(u, v).$$

Now note the equality of events $\{N_n(du) = N_n(dv) = 1\} = \{t_{N_n(u),n} = u, t_{N_n(v),n} = v\}$. By definition, we have $t_{N_n(u),n} \leq u$ for every u . Yet note that when $n \rightarrow \infty$, the grid becomes arbitrarily dense and hence $t_{N_n(u),n} \xrightarrow{P} u$. This can be used to show, by conditioning on $t_{N_n(u),n}$ and $t_{N_n(v),n}$, that for every $u, v \in (0, 1)$ and $x, y \in S$

$$\left| f_{X_{\{u\},n}, X_{\{v\},n} | N_n(du)=N_n(dv)=1}(x, y) - f_{X_{\{u\},n}, X_{\{v\},n}}(x, y) \right| \rightarrow 0.$$

This implies that the distribution of the $X_{\{u\},n}$ and $X_{\{v\},n}$, which are the last observed values at time u and v , resp., is asymptotically independent from the fact that there are jumps at u and v . In particular, in the limit their covariance is not affected by these jumps.

Now we can exploit the mixing conditions to bound the covariance using Davydov's lemma, which gives for $\delta > 0$

$$\begin{aligned} |\text{Cov}(K_{u,n}(x), K_{v,n}(x))| &\leq 8\alpha(n(v-u))^{\frac{\delta}{2+\delta}} E(K_{u,n}(x)^{2+\delta})^{\frac{1}{2+\delta}} E(K_{v,n}(x)^{2+\delta})^{\frac{1}{2+\delta}} \\ &\leq C\alpha(n(v-u))^{\frac{\delta}{2+\delta}} h_x^{\frac{2}{2+\delta}}. \end{aligned} \tag{46}$$

Hence

$$\begin{aligned}
nh_t h_x |(CV)| &\leq \frac{2}{nh_t h_x} \int_{t-h_t}^{t+h_t} \int_v^{t+h_t} K\left(\frac{t-u}{h_t}\right) K\left(\frac{t-v}{h_t}\right) \text{Cov}(K_{u,n}(x), K_{v,n}(x)) \bar{\lambda}_n(u) \bar{h}_n(u, v) du dv \\
&\quad + \frac{Ch_x}{nh_t} \int_{t-h_t}^{t+h_t} \int_v^{t+h_t} K\left(\frac{t-u}{h_t}\right) K\left(\frac{t-v}{h_t}\right) |\mu_n(u, v)| du dv \\
&= \frac{2nh_t}{h_x} \int_{-1}^1 \int_0^{1-z} K(z+y) K(z) \text{Cov}(K_{z,n}(x), K_{z+y,n}(x)) \\
&\quad \quad \quad \times \frac{\bar{\lambda}_n(t+h_t z) \bar{h}_n(t+h_t z, t+h_t(z+y))}{n} dy dz \\
&\quad + Cnh_t h_x \int_{-1}^1 \int_0^{1-z} K(z+y) K(z) \frac{|\mu_n(t+h_t z, t+h_t(z+y))|}{n^2} dy dz \\
&= \frac{2}{h_x} \int_{-1}^1 \int_0^{nh_t(1-z)} K\left(z + \frac{w}{nh_t}\right) K(z) \text{Cov}\left(K_{t+h_t z, n}(x), K_{t+h_t z + \frac{w}{n}, n}(x)\right) \\
&\quad \quad \quad \times \frac{\bar{\lambda}_n(t+h_t z) \bar{h}_n(t+h_t z, t+h_t z + \frac{w}{n})}{n} dw dz \\
&\quad + Ch_x \int_{-1}^1 \int_0^{nh_t(1-z)} K\left(z + \frac{w}{nh_t}\right) K(z) \frac{|\mu_n(t+h_t z, t+h_t z + \frac{w}{n})|}{n^2} dw dz \\
&\equiv (*) + (**),
\end{aligned}$$

We now bound the separate terms. We split up (*) using a sequence m_n such that $m_n \rightarrow \infty$ and $m_n h_x \rightarrow 0$, and use covariance inequality (46) to bound the covariance between terms far apart. Combined with boundedness of the kernel this yields

$$\begin{aligned}
(*) &\leq Cm_n h_x + Ch_x^{\frac{\delta}{2+\delta}} \int_{-1}^1 \int_{m_n}^{nh_t(1-z)} \alpha(w)^{1-\frac{2}{\delta}} \frac{\bar{\lambda}_n(t+h_t z) \bar{h}_n(t+h_t z, t+h_t z + \frac{w}{n})}{n} dw dz \\
&\leq Cm_n h_x + \frac{Ch_x^{\frac{\delta}{2+\delta}}}{m_n^\gamma} \int_{-1}^1 \int_{m_n}^{nh_t(1-z)} w^\gamma \alpha(w)^{1-\frac{2}{\delta}} \frac{\bar{\lambda}_n(t+h_t z) \bar{h}_n(t+h_t z, t+h_t z + \frac{w}{n})}{n} dw dz \\
&\rightarrow 0,
\end{aligned}$$

for example by setting $m_n^\gamma = h_x^{\frac{\delta}{2+\delta}}$, using the dominated convergence theorem and the condition on the mixing coefficients (12). For the second term we find

$$(**) \rightarrow C \int_{-1}^1 K^2(z) \int_0^\infty \gamma_t(w) dw dz < \infty,$$

again using the dominated convergence theorem. Hence it follows $(**) = O(h_x)$. Combining terms, we conclude that $(CV) = o\left(\frac{1}{nh_t h_x}\right)$, which means that the covariance terms are of smaller order than the variance terms. Hence we have $\text{Var}(\hat{f}(t, x)) \rightarrow 0$, and since the bias vanishes as well we have proven the first part.

(ii) The leading bias term of the estimator is given by

$$E(\hat{g}^B(t, x)) = E\left(\frac{1}{n} \int_{t-h_t}^{t+h_t} K_{h_t}(t-u) K_{h_x}(x - X_{\{u\},n})(m(u, X_{\{u\},n}) - m(t, x)) dN_n(u)\right).$$

In terms of the stationary approximation process we can write, similar to (45),

$$\begin{aligned} & E\left(K_{h_x}(x - X_{\{u\},n})(m(u, X_{\{u\},n}) - m(t, x)) dN_n(u)\right) \\ &= E\left(K_{h_x}(x - X_{\{u\},n}^*)(m(u, X_{\{u\},n}^*) - m(t, x)) | dN_n(u) = 1\right) \bar{\lambda}_n(u) du + O\left(\frac{1}{n^r h_x^r}\right) \end{aligned}$$

so that using a Taylor approximation of $m(u, X_{\{u\},n}^*)$ around $m(t, x)$ and (44) we find

$$\begin{aligned} E(\hat{g}^B(t, x)) &= h_t^2 \mu_2(K) \left(\frac{\partial m(t, x)}{\partial t} \frac{\partial f(t, x) \nu(t)}{\partial t} + \frac{1}{2} \frac{\partial^2 m(t, x)}{\partial t^2} f(t, x) \nu(t) \right) \\ &\quad + h_x^2 \mu_2(K) \left(\frac{\partial m(t, x)}{\partial x} \frac{\partial f(t, x) \nu(t)}{\partial x} + \frac{1}{2} \frac{\partial^2 m(t, x)}{\partial x^2} f(t, x) \nu(t) \right) + o(h_t^2 + h_x^2) + O\left(\frac{1}{n^r h_x^r}\right). \end{aligned}$$

Using the Lipschitz condition on $m(t, x)$ it can be shown that $\text{Var}(\hat{g}^B(t, x)) = o\left(\frac{1}{nh_t h_x}\right)$, which ensures that the variance of the bias term is asymptotically negligible.

(iii) For the variance term, we can write

$$\begin{aligned} nh_t h_x \text{Var}(\hat{g}^V(t, x)) &= \frac{1}{nh_t h_x} \text{Var}\left(\int_{t-h_t}^{t+h_t} K\left(\frac{t-u}{h_t}\right) K\left(\frac{x - X_{\{u\},n}}{h_x}\right) \epsilon_{\{u\},n} dN_n(u)\right) \\ &= \frac{1}{nh_t h_x} E\left(\int_{t-h_t}^{t+h_t} K^2\left(\frac{t-u}{h_t}\right) K^2\left(\frac{x - X_{\{u\},n}}{h_x}\right) \epsilon_{\{u\},n}^2 dN_n(u)\right) \\ &\quad + \frac{1}{nh_t h_x} E\left(\iint_{\substack{t-h_t \leq u, v \leq t+h_t, \\ u \neq v}} K\left(\frac{t-u}{h_t}\right) K\left(\frac{t-v}{h_t}\right) K\left(\frac{x - X_{\{u\},n}}{h_x}\right) \right. \\ &\quad \left. \times K\left(\frac{x - X_{\{v\},n}}{h_x}\right) \epsilon_{\{u\},n} \epsilon_{\{v\},n} dN_n(u) dN_n(v)\right) \\ &= \mu_2^2(K) f(t, x) \sigma^2(t, x) \nu(t) + o(1), \end{aligned}$$

using arguments similar as for computing $\text{Var}(\hat{f}_h(t, x))$, except in addition to bound the covariance term we use

$$E|K_{u,n}(x)\epsilon_{\{u\},n}|^{2+\delta} = E\left(E\left(\epsilon_{\{u\},n}^{2+\delta}(x)|X_{\{u\},n}\right)K_{u,n}^{2+\delta}(x)\right) < Ch_x.$$

The asymptotic normality follows from applying large-small block arguments to the integral form of $\hat{g}^V(t, x)$. In particular, write

$$\sqrt{nh_t h_x} \hat{g}^V(t, x) = \frac{1}{\sqrt{nh_t h_x}} \int_{-nh_t}^{nh_t} K\left(\frac{w}{nh_t}\right) K\left(\frac{x - X_{\{t+\frac{w}{n}\}}}{h_x}\right) \epsilon_{\{t+\frac{w}{n}\}} N_n(t + dw/n),$$

and divide the integral in k_n large blocks $\zeta_{j,n}$ of length τ_l , and k_n small blocks $\eta_{j,n}$ of length τ_s , which integrate, respectively, from $-nh_t + (j-1)(\tau_l + \tau_s)$ to $-nh_t + (j-1)(\tau_l + \tau_s) + \tau_l$, and from $-nh_t + (j-1)(\tau_l + \tau_s) + \tau_l$ to $-nh_t + j(\tau_l + \tau_s)$. Hence we have

$$\sqrt{nh_t h_x} \hat{g}^V(t, x) = \frac{1}{\sqrt{nh_t h_x}} \sum_{j=1}^{k_n} (\zeta_{j,n} + \eta_{j,n}),$$

where

$$k_n = \frac{2nh_t}{\tau_l + \tau_s}.$$

We set the block sizes such that $\tau_l \rightarrow \infty$, $\tau_s \rightarrow \infty$, $\frac{\tau_s}{\tau_l} \rightarrow 0$, $\frac{\tau_l}{\sqrt{nh_t h_x}} \rightarrow 0$, and $k_n \alpha(s_n) \rightarrow 0$. Based on the choice in (Fan and Yao, 2002, Thm. 2.22), it can be shown that

$$\tau_l = \frac{\sqrt{nh_t h_x}}{\log nh_t}, \quad \tau_s = \left(\sqrt{nh_t/h_x} \log nh_t\right)^{\frac{\delta}{(\lambda+1)(2+\delta)}}$$

satisfies these conditions under the bandwidth conditions in the theorem and the mixing condition. Note

$$\text{Var}\left(\frac{1}{\sqrt{nh_t h_x}} \sum_{j=1}^{k_n} \eta_{j,n}\right) < \frac{Ck_n \tau_s}{nh_t} \rightarrow 0,$$

so that the small blocks are asymptotically negligible. Since by the Volkonskii-Rozanov lemma

$$\left|E\left(\exp\left(it \frac{\sum_{j=1}^{k_n} \zeta_{j,n}}{\sqrt{nh_t h_x}}\right)\right) - \prod_{j=1}^{k_n} E\left(\exp\left(it \frac{\zeta_{j,n}}{\sqrt{nh_t h_x}}\right)\right)\right| \leq 16(k_n - 1)\alpha(\tau_s) \rightarrow 0,$$

the large blocks are asymptotically independent. It can be verified that

$$\frac{1}{nh_t h_x} \sum_{j=1}^{k_n} \text{Var}(\zeta_{j,n}) \rightarrow \mu_2^2(K) f(t, x) \sigma^2(t, x) \nu(t) > 0,$$

by decomposing

$$\text{Var} \left(\sqrt{nh_t h_x} \hat{g}^V(t, x) \right) = \frac{1}{nh_t h_x} \sum_{j=1}^{k_n} \text{Var}(\zeta_{j,n} + \eta_{j,n}) + \frac{1}{nh_t h_x} \sum_{j=1}^{k_n} \sum_{k \neq j}^{k_n} \text{Cov}(\zeta_{j,n} + \eta_{j,n}, \zeta_{k,n} + \eta_{k,n}),$$

and showing that all other terms on the right hand side vanish, in particular using that for adjacent blocks

$$\frac{1}{nh_t h_x} \text{Cov}(\zeta_{j,n}, \eta_{j,n}) < C \frac{\tau_s}{nh_t}.$$

The Lindeberg condition

$$\frac{1}{nh_t h_x} \sum_{j=1}^{k_n} E \left(\zeta_{j,n}^2 \mathbf{1}_{\{|\zeta_{j,n}| > \sqrt{nh_t h_x} \epsilon\}} \right) \rightarrow 0,$$

is satisfied for any $\epsilon > 0$, since $\tau_l = o(\sqrt{nh_t h_x})$ causes $\{|\zeta_{j,n}| > \sqrt{nh_t h_x} \epsilon\}$ to become an empty set for large n . The Lindeberg-Feller central limit theorem now gives the required result. \square

Proof of (16).

$$\begin{aligned} \bar{\lambda}_n(t) &= E \lambda_n(t) = n \phi \pi(t) + \int_{-\infty}^t E \left(h(X_{\{u\},n}) | N_n(du) = 1 \right) n g(n(t-u)) \bar{\lambda}_n(u) du, \\ &= n \phi \pi(t) + \int_0^\infty E \left\{ h \left(X_{\{t-\frac{w}{n}\},n} | N_n(t+dw/n) = 1 \right) \right\} g(w) \bar{\lambda}_n(t-w/n) dw + o(n), \end{aligned}$$

where we approximate with the stationary process around time t (all expectations are given $N_n(t+dw/n) = 1$)

$$\begin{aligned} E \left| h \left(X_{\{t-\frac{w}{n}\},n} \right) - h \left(X_{\{t-\frac{w}{n}\},n}^*(t) \right) \right| &\leq L E \left| X_{\{t-\frac{w}{n}\},n} - X_{\{t-\frac{w}{n}\},n}^*(t) \right| \\ &\leq C E \left| X_{\{t-\frac{w}{n}\},n} - X_{\{t-\frac{w}{n}\},n}^*(t) \right|^r \\ &\leq C \left(\frac{1+w}{n} \right)^\rho E \left| U_{(t-\frac{w}{n}),n}(t) \right|^r \\ &\leq C \left(\frac{1+w}{n} \right)^r, \end{aligned}$$

where the second inequality uses the compactness of S , and we recall $r = \min(\rho, 1)$. Hence

$$\nu(t) = \lim_{n \rightarrow \infty} \frac{\bar{\lambda}_n(t)}{n} = \phi \pi(t) + \nu(t) \int_S h(x) f(t, x) dx \int_0^\infty g(w) dw,$$

using the stationarity of $X_{\{t\},n}^*$ and the integrability conditions on $g(\cdot)$. \square

References

- Aït-Sahalia, Y. (2007). Estimating continuous-time models with discretely sampled data. *Econometric Society Monographs*, 43:261.
- Aït-Sahalia, Y. and Duarte, J. (2003). Nonparametric option pricing under shape restrictions. *Journal of Econometrics*, 116(1-2):9–47.
- Aït-Sahalia, Y. and Lo, A. W. (1998). Nonparametric estimation of state-price densities implicit in financial asset prices. *Journal of Finance*, 53(2):499–547.
- Aït-Sahalia, Y. and Lo, A. W. (2000). Nonparametric risk management and implied risk aversion. *Journal of Econometrics*, 94(1-2):9–51.
- Aït-Sahalia, Y. and Mykland, P. A. (2003). The effects of random and discrete sampling when estimating continuous-time diffusions. *Econometrica*, 71(2):483–549.
- Bakshi, G., Cao, C., and Chen, Z. (1997). Empirical performance of alternative option pricing models. *The Journal of Finance*, 52(5):2003–2049.
- Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, 76(6):1481–1536.
- Barndorff-Nielsen, O. E. and Shephard, N. (2004). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica*, 72(3):885–925.
- Beber, A. and Brandt, M. W. (2006). The effect of macroeconomic news on beliefs and preferences: Evidence from the options market. *Journal of Monetary Economics*, 53(8):1997–2039.
- Bosq, D. (1996). *Nonparametric statistics for stochastic processes*. Springer.
- Breedon, D. T. and Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *The Journal of Business*, 51(4):621–51.
- Broadie, M., Chernov, M., and Johannes, M. (2007). Model specification and risk premia: Evidence from futures options. *The Journal of Finance*, 62(3):1453–1490.
- Chen, S. X. and Xu, Z. (2014). On implied volatility for options: Some reasons to smile and more to correct. *Journal of Econometrics*, 179(1):1–15.
- Cranwell, R. and Weiss, N. (1978). A central limit theorem for mixing stationary point processes. *Stochastic Processes and their Applications*, 8(2):229–242.

- Daley, D. J. and Vere-Jones, D. (2003). *An introduction to the theory of point processes*, volume 1. Springer.
- Duffie, D. and Glynn, P. (2004). Estimation of continuous-time markov processes sampled at random time intervals. *Econometrica*, 72(6):1773–1808.
- Ellis, S. P. (1991). Density estimation for point processes. *Stochastic processes and their applications*, 39(2):345–358.
- Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*, volume 66. CRC Press.
- Fan, J. and Yao, Q. (2002). *Nonlinear time series*, volume 2. Springer.
- Group, C. (2014). E-mini s&p 500 eom options contract specs.
- Hall, P. and Schucany, W. R. (1989). A local cross-validation algorithm. *Statistics & Probability Letters*, 8(2):109–117.
- Härdle, W. and Hlavka, Z. (2009). Dynamics of state price densities. *Journal of Econometrics*, 150(1):1–15.
- Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90.
- Jackwerth, J. C. (2000). Recovering Risk Aversion from Option Prices and Realized Returns. *Review of Financial Studies*, 13(2).
- Joshi, M. (2001). Log-type models, homogeneity of option prices and convexity.
- Karatzas, I. and Shreve, S. E. (1998). *Methods of mathematical finance*, volume 39. Springer.
- Kristensen, D. (2010). Nonparametric filtering of the realized spot volatility: a kernel-based approach. *Econometric Theory*, 26(01):60–93.
- Li, Y., Mykland, P. A., Renault, E., Zhang, L., and Zheng, X. (2009). Realized volatility when sampling times are possibly endogenous.
- Merton, R. C. (1973). Theory of rational option pricing. *Bell Journal of Economics*, 4(1):141–183.
- Mykland, P. A., Zhang, L., et al. (2006). Anova for diffusions and ito processes. *The Annals of Statistics*, 34(4):1931–1963.
- Pawlas, Z. (2009). Empirical distributions in marked point processes. *Stochastic Processes and their Applications*, 119(12):4194–4209.

- Renault, E. and Werker, B. J. (2011). Causality effects in return volatility measures with random times. *Journal of Econometrics*, 160(1):272–279.
- Robinson, P. M. (1983). Nonparametric estimators for time series. *Journal of Time Series Analysis*, 4(3):185–207.
- Rosenberg, J. and Engle, R. F. (2000). Empirical pricing kernels. *Journal of Financial Economics*, (99-014).
- Song, Z. and Xiu, D. (2012). A tale of two option markets: State-price densities implied from s&p 500 and vix option prices. *Unpublished working paper, Federal Reserve Board and University of Chicago*.
- Vogt, M. (2012). Nonparametric regression for locally stationary time series. *The Annals of Statistics*, 40(5):2601–2633.
- Yatchew, A. and Härdle, W. (2006). Nonparametric state price density estimation using constrained least squares and the bootstrap. *Journal of Econometrics*, 133(2):579–599.
- Zhang, L. (2011). Estimating covariation: Epps effect, microstructure noise. *Journal of Econometrics*, 160(1):33–47.

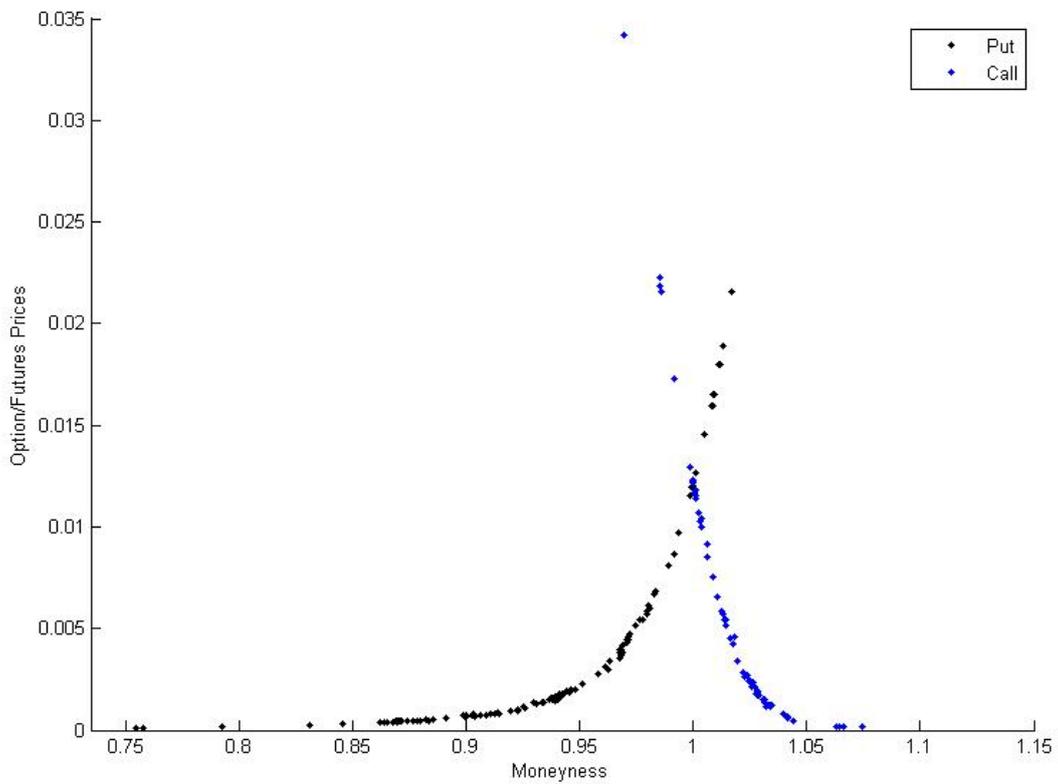
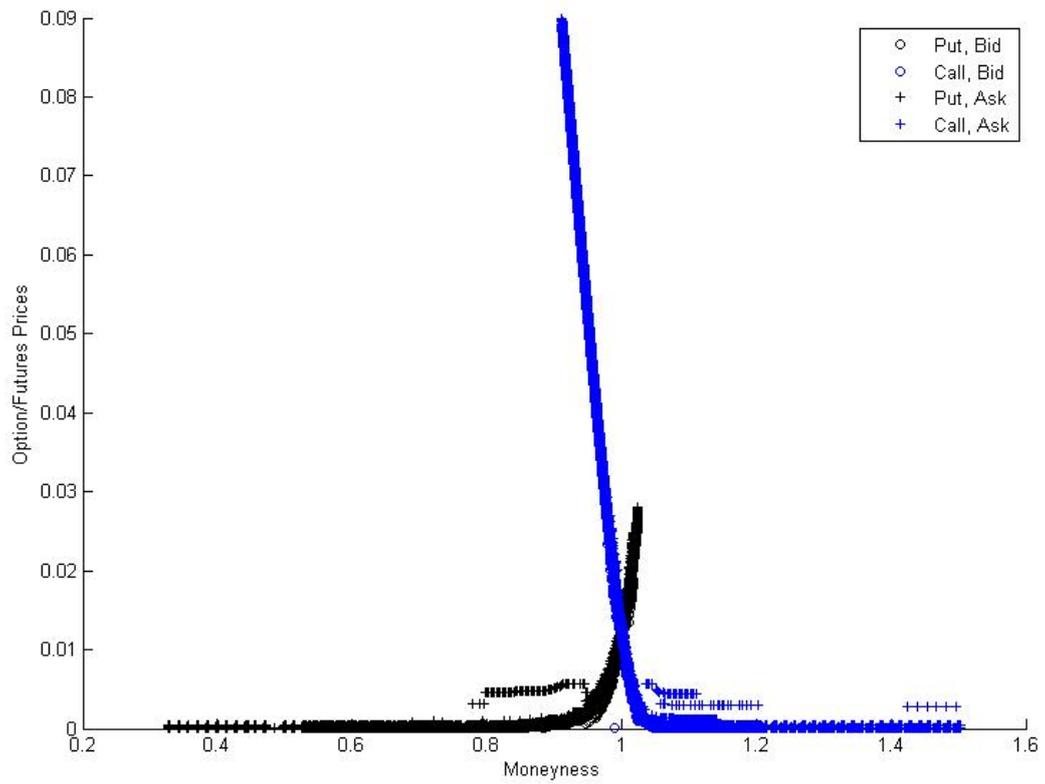


Figure 3: Call and put prices per unit of futures price as a function of moneyness, in terms of best bid and ask quotes (upper panel) and transaction prices (lower panel), for November 1, 2013.

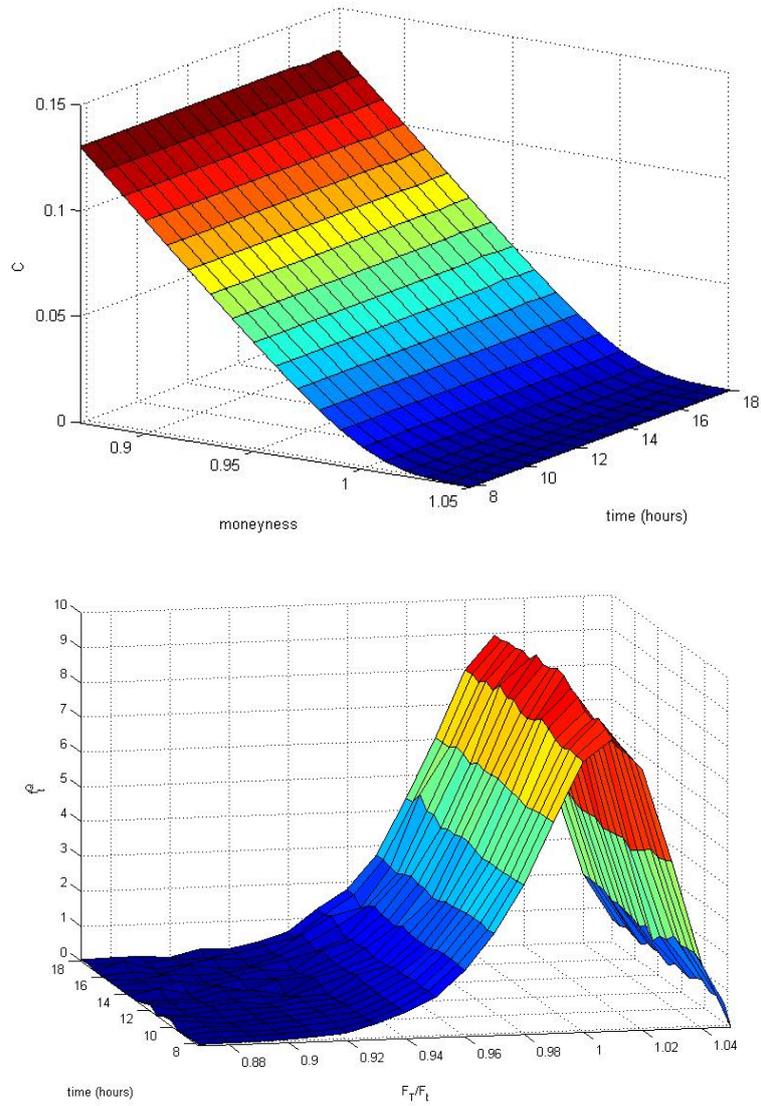


Figure 4: Time-varying kernel smoother of call price function (upper panel), and SPD (lower), using bid and ask quotes on November 7, 2013 of options expiring November 29, 2013. Bandwidths are chosen by local cross-validation with one-hour blocks between 9AM and 6PM.