Supplementary Material
for
A New Nonparametric Estimate of the Risk-Neutral Density with Applications to Variance Swaps

A. PROOF OF PROPOSITION 2.1

We rewrite the call and put option prices in **Eqs 3, 4** in terms of $a_1, a_2, \ldots, a_q, a_{q+1}$ as follows

$$\begin{aligned} e^{R_{tT}}\hat{P}_i &= \int_{-\infty}^{\log K_i} (K_i - e^y) f_\Delta(y) dy \\ &= \sum_{l=1}^{q+1} \int_{\log K_{l-1}}^{\log K_l} (K_i - e^y) a_l dy \cdot \mathbb{1}(K_i \geq K_l) \\ &= \sum_{l=1}^{q+1} a_l [(K_i \log \frac{K_l}{K_{l-1}}) - (K_l - K_{l-1})] \cdot \mathbb{1}(K_i \geq K_l),\ i \in \mathcal{P} \end{aligned} \tag{S1}$$

$$\begin{aligned} e^{R_{tT}}\hat{C}_i &= \int_{\log K_i}^{\infty} (e^y - K_i) f_\Delta(y) dy \\ &= \sum_{l=1}^{q+1} \int_{\log K_{l-1}}^{\log K_l} (e^y - K_i) a_l dy \cdot \mathbb{1}(K_i \leq K_{l-1}) \\ &= \sum_{l=1}^{q+1} a_l [(K_l - K_{l-1}) - K_i \log \frac{K_l}{K_{l-1}}] \cdot \mathbb{1}(K_i < K_l),\ i \in \mathcal{C} \end{aligned} \tag{S2}$$

Let $X_{i,l}^{(p)} = [K_i \log(K_l/K_{l-1}) - (K_l - K_{l-1})] \cdot \mathbb{1}(K_i \geq K_l)$, $l = 1, 2, \ldots, q+1$ be an entry of the design matrix for put options; and $X_{i,l}^{(c)} = [(K_l - K_{l-1}) - K_i \log(K_l/K_{l-1})] \cdot \mathbb{1}(K_i < K_l)$, $l = 1, 2, \ldots, q+1$ for call options. From **Eq. 2**, a_{q+1} can be represented by $a_1, a_2, \ldots, a_q$, as

$$a_{q+1} = \left(1 - \sum_{l=1}^{q} a_l \log \frac{K_l}{K_{l-1}}\right) (\log c_K)^{-1} \tag{S3}$$

Plugging **Eq.** S3 into **Eqs** S1, S2, we obtain

$$\begin{aligned} e^{R_{tT}}\hat{P}_i &= \sum_{l=1}^{q+1} a_l X_{i,l}^{(p)} \\ &= a_1 X_{i,1}^{(p)} + a_2 X_{i,2}^{(p)} + \cdots + a_q X_{i,q}^{(p)} \\ &\quad + \left(1 - a_1 \log \frac{K_1}{K_0} - \cdots - a_q \log \frac{K_q}{K_{q-1}}\right) (\log c_K)^{-1} X_{i,q+1}^{(p)} \\ &= a_1 [X_{i,1}^{(p)} - (\log \frac{K_1}{K_0})(\log c_K)^{-1} X_{i,q+1}^{(p)}] \quad + \cdots \\ &\quad + a_q [X_{i,q}^{(p)} - (\log \frac{K_q}{K_{q-1}})(\log c_K)^{-1} X_{i,q+1}^{(p)}] + \frac{1}{\log c_K} X_{i,q+1}^{(p)} \\ &\triangleq a_1 X_{i,1}^{(P)} + a_2 X_{i,2}^{(P)} + \cdots + a_q X_{i,q}^{(P)} + X_{i,q+1}^{(P)},\ i \in \mathcal{P} \end{aligned} \tag{S4}$$

where $X_{i,l}^{(P)} = X_{i,l}^{(p)} - (\log K_l/K_{l-1})(\log c_K)^{-1} X_{i,q+1}^{(p)}$, $l = 1, 2, \ldots, q$ and $X_{i,q+1}^{(P)} = X_{i,q+1}^{(p)} / \log c_K$. Similarly for call options,

$$\begin{aligned} e^{R_{tT}}\hat{C}_i &= \sum_{l=1}^{q+1} a_l X_{i,l}^{(c)} \\ &= a_1 X_{i,1}^{(c)} + a_2 X_{i,2}^{(c)} + \cdots + a_q X_{i,q}^{(c)} \\ &\quad + \left(1 - a_1 \log \frac{K_1}{K_0} - \cdots - a_q \log \frac{K_q}{K_{q-1}}\right) (\log c_K)^{-1} X_{i,q+1}^{(c)} \\ &= a_1 [X_{i,1}^{(c)} - (\log \frac{K_1}{K_0})(\log c_K)^{-1} X_{i,q+1}^{(c)}] \quad + \cdots \\ &\quad + a_q [X_{i,q}^{(c)} - (\log \frac{K_q}{K_{q-1}})(\log c_K)^{-1} X_{i,q+1}^{(c)}] + \frac{1}{\log c_K} X_{i,q+1}^{(c)} \\ &\triangleq a_1 X_{i,1}^{(C)} + a_2 X_{i,2}^{(C)} + \cdots + a_q X_{i,q}^{(C)} + X_{i,q+1}^{(C)},\ i \in \mathcal{C} \end{aligned} \tag{S5}$$

where $X_{i,l}^{(C)} = X_{i,l}^{(c)} - (\log K_l/K_{l-1})(\log c_K)^{-1} X_{i,q+1}^{(c)}$, $l = 1, \ldots, q$ and $X_{i,q+1}^{(C)} = X_{i,q+1}^{(c)} / \log c_K$. $\square$

B. PROOF OF THEOREM 3.1

Given $\epsilon > 0$, let $\delta_1 = \sqrt{\epsilon} e^{R_{tT}} / [3(1 + c_K + e)] > 0$. There exists $-\infty < A < 0 < B < \infty$, such that,

$$\int_{-\infty}^{A} f_{\mathbb{Q}}(x)dx < \delta_1,\ \int_{-\infty}^{A} e^x f_{\mathbb{Q}}(x)dx < \delta_1,\ \int_{B}^{\infty} f_{\mathbb{Q}}(x)dx < \delta_1,\ \int_{B}^{\infty} e^x f_{\mathbb{Q}}(x)dx < \delta_1$$

Let $\delta_2 = \sqrt{\epsilon} e^{R_{tT} - B - 1} / [3(B - A + 2)] > 0$. Since $f_{\mathbb{Q}}$ is continuous, there exists a $\delta > 0$, such that, for any $x_1, x_2 \in [A - 1, B + 1]$,

$$|f_{\mathbb{Q}}(x_1) - f_{\mathbb{Q}}(x_2)| < \delta_2$$

as long as $|x_1 - x_2| < \delta$.

For small enough $K_1, |\Delta|$ and large enough q, K_q, there exist integers u, v, such that, $1 < u < u+1 < v < v+1 < q$, $\log K_u \leq A < \log K_{u+1}$, $\log K_v < B \leq \log K_{v+1}$, $|\Delta| < \delta$.

We construct a f_Δ by defining

$$\begin{aligned} a_1 &= (\log c_K)^{-1} \int_{-\infty}^{\log K_1} f_{\mathbb{Q}}(x)dx \geq 0 \\ a_i &= [\log(K_i/K_{i-1})]^{-1} \int_{\log K_{i-1}}^{\log K_i} f_{\mathbb{Q}}(x)dx \geq 0, \quad i = 2, \ldots, q \\ a_{q+1} &= (\log c_K)^{-1} \int_{\log K_q}^{\infty} f_{\mathbb{Q}}(x)dx \geq 0 \end{aligned}$$

It can be verified that $\int_{-\infty}^{\infty} f_\Delta(x)dx = \sum_{i=1}^{q+1} a_i \log(K_i/K_{i-1}) = 1$. Let

$$\Delta_f = \max_{u \leq i \leq v} \left(\max_{\log K_i \leq x \leq \log K_{i+1}} f_{\mathbb{Q}}(x) - \min_{\log K_i \leq x \leq \log K_{i+1}} f_{\mathbb{Q}}(x) \right)$$

Then $|\Delta| < \delta$ implies $\Delta_f \leq \delta_2$. It can be verified that

$$\begin{aligned} |\hat{C}_i - \tilde{C}_i| &< \begin{cases} \sqrt{\epsilon}/3, & \text{for } i = v+1, \ldots, q \\ 2\sqrt{\epsilon}/3, & \text{for } i = u, \ldots, v \\ \sqrt{\epsilon}, & \text{for } i = 1, \ldots, u-1 \end{cases} \\ |\hat{P}_i - \tilde{P}_i| &< \begin{cases} \sqrt{\epsilon}/3, & \text{for } i = 1, \ldots, u \\ 2\sqrt{\epsilon}/3, & \text{for } i = u+1, \ldots, v+1 \\ \sqrt{\epsilon}, & \text{for } i = v+2, \ldots, q \end{cases} \end{aligned}$$

In other words, there exist $a_1, \ldots, a_{q+1}$, such that, $(\hat{C}_i - \tilde{C}_i)^2 < \epsilon$, $(\hat{P}_i - \tilde{P}_i)^2 < \epsilon$, for $i = 1, \ldots, q$. It implies the $(a_1, \ldots, a_{q+1})$ that minimizes $L(a_1, \ldots, a_{q+1})$ also satisfies

$$\frac{1}{2q} \left[\sum_{i=1}^{q} (\hat{C}_i - \tilde{C}_i)^2 + \sum_{i=1}^{q} (\hat{P}_i - \tilde{P}_i)^2 \right] < \epsilon$$

which leads to the conclusion. $\square$

C. PROOF OF PROPOSITION 4.1

Since $\mathbb{E}_t^{\mathbb{Q}}[\sum_{i=1}^T R_i^2] = \sum_{i=1}^t R_i^2 + \sum_{i=t+1}^T \mathbb{E}_t^{\mathbb{Q}}[R_i^2]$, the key part

$$
\begin{aligned}
&\sum_{i=t+1}^{T} \mathbb{E}_t^{\mathbb{Q}}[R_i^2] = \sum_{i=t+1}^{T} \mathbb{E}_t^{\mathbb{Q}}[\log \frac{S_i}{S_{i-1}}]^2 \\
&= \sum_{i=t+1}^{T} [\mathbb{E}_t^{\mathbb{Q}}(\log S_i)^2 + \mathbb{E}_t^{\mathbb{Q}}(\log S_{i-1})^2 - 2\mathbb{E}_t^{\mathbb{Q}}(\log S_i)(\log S_{i-1})] \\
&= \sum_{i=t+1}^{T} \mathbb{E}_t^{\mathbb{Q}}(\log S_i)^2 + \sum_{i=t+1}^{T} \mathbb{E}_t^{\mathbb{Q}}(\log S_{i-1})^2 - 2\sum_{i=t+1}^{T} \mathbb{E}_t^{\mathbb{Q}}[\log S_{i-1} + \log(\frac{S_i}{S_{i-1}})][\log S_{i-1}] \\
&= \sum_{i=t+1}^{T} \mathbb{E}_t^{\mathbb{Q}}(\log S_i)^2 + \sum_{i=t+1}^{T} \mathbb{E}_t^{\mathbb{Q}}(\log S_{i-1})^2 - 2\sum_{i=t+1}^{T} \mathbb{E}_t^{\mathbb{Q}}(\log S_{i-1})^2 \\
&\quad - 2\sum_{i=t+1}^{T} \mathbb{E}_t^{\mathbb{Q}}[\log S_{i-1}][\log(\frac{S_i}{S_{i-1}})] \\
&= \mathbb{E}_t^{\mathbb{Q}}[\log S_T]^2 - [\log S_t]^2 - 2\sum_{i=t+1}^{T} \mathbb{E}_t^{\mathbb{Q}}[\log S_{i-1}][\log(\frac{S_i}{S_{i-1}})] \\
&= \mathbb{E}_t^{\mathbb{Q}}[\log S_T]^2 - [\log S_t]^2 - 2\sum_{i=t+1}^{T} \mathbb{E}_t^{\mathbb{Q}}[\log S_{i-1}]\mathbb{E}_t^{\mathbb{Q}}[\log(\frac{S_i}{S_{i-1}})] \\
&= \mathbb{E}_t^{\mathbb{Q}}[\log S_T]^2 - [\log S_t]^2 - 2\sum_{i=t+1}^{T} [\mathbb{E}_t^{\mathbb{Q}} \log S_{i-1}\mathbb{E}_t^{\mathbb{Q}} \log S_i - (\mathbb{E}_t^{\mathbb{Q}} \log S_{i-1})^2]
\end{aligned}
$$

Then **Eq. 12** can be obtained by plugging $\mathbb{E}_t^{\mathbb{Q}}[\sum_{i=1}^T R_i^2]$ into **Eq. 11**. □

D. LINEAR INTERPOLATION FOR 1ST AND 2ND MOMENTS IN SECTION 4.1

Mean imputation Suppose the trading day is t and the expiration day is T. We denote all possible expiration dates of traded contracts by $t+n_1, t+n_2, \ldots$. Suppose the time point to be imputed is $t+n_0$. Given all the information available at day t, $\log S_t$ can be regarded as its expectation at day t, $\mathbb{E}_t^{\mathbb{Q}} \log S_t$. Therefore, we consider cases separately according to whether or not $t+n_0$ is in the interval $[t, t+n_1]$ and then apply linear interpolation to obtain the mean of $\log S_{t+n_0}$. More specifically, there are two cases:

Case 1: $n_0 \in [0, n_1]$ and $\mathbb{E}_t^{\mathbb{Q}}(\log S_{t+n_1})$ has been calculated.

$$
\begin{aligned}
\mathbb{E}_t^{\mathbb{Q}}(\log S_{t+n_0}) &= \mathbb{E}_t^{\mathbb{Q}}(\log S_{t+n_1}) - \frac{(n_1-n_0)[\mathbb{E}_t^{\mathbb{Q}}(\log S_{t+n_1}) - \log S_t]}{n_1} \\
&= \frac{n_0\mathbb{E}_t^{\mathbb{Q}}(\log S_{t+n_1}) + (n_1-n_0)\log(S_t)}{n_1}
\end{aligned}
$$

Case 2: $n_0 \in [n_i, n_{i+1}]$ for some $i = 1, 2, \ldots$. The expectations $\mathbb{E}_t^{\mathbb{Q}}(\log S_{t+n_i})$ and $\mathbb{E}_t^{\mathbb{Q}}(\log S_{t+n_{i+1}})$ have already been calculated.

$$\begin{aligned}\mathbb{E}_t^{\mathbb{Q}}(\log S_{t+n_0}) &= \frac{(n_0 - n_i)[\mathbb{E}_t^{\mathbb{Q}}(\log S_{t+n_{i+1}}) - \mathbb{E}_t^{\mathbb{Q}}(\log S_{t+n_i})]}{n_{i+1} - n_i} + \mathbb{E}_t^{\mathbb{Q}}(\log S_{t+n_i}) \\ &= \frac{(n_0 - n_i)\mathbb{E}_t^{\mathbb{Q}}(\log S_{t+n_{i+1}}) + (n_{i+1} - n_0)\mathbb{E}_t^{\mathbb{Q}}(\log S_{t+n_i})}{n_{i+1} - n_i}\end{aligned}$$

Variance Imputation In order to calculate the variance $\mathbb{V}_t^{\mathbb{Q}}(\log S_{t+n_0})$ at day t, we use a similar interpolation based on the available variances of log returns at day t with expiration T. Based on the scatterplot (not shown here) of all available variances that we have from the existing contracts, the trend of variances has a curved pattern against the number of days to expiration. More specifically, it is roughly a quadratic curve. Before we implement a linear interpolation, we first perform a square-root transformation of variances.

Case 1: $n_0 \in [0, n_1]$. $\mathbb{V}_t^{\mathbb{Q}}(\log S_{t+n_1})$ has been calculated. Then

$$\sqrt{\mathbb{V}_t^{\mathbb{Q}}(\log S_{t+n_0})} = \frac{n_0\sqrt{\mathbb{V}_t^{\mathbb{Q}}(\log S_{t+n_1})}}{n_1}$$

Case 2: $n_0 \in [n_i, n_{i+1}]$ for some $i = 1, 2, \ldots$. The values $\mathbb{V}_t^{\mathbb{Q}}(\log S_{t+n_i})$ and $\mathbb{V}_t^{\mathbb{Q}}(\log S_{t+n_{i+1}})$ have been calculated. Then

$$\begin{aligned}&\sqrt{\mathbb{V}_t^{\mathbb{Q}}(\log S_{t+n_0})} \\ &= \sqrt{\mathbb{V}_t^{\mathbb{Q}}(\log S_{t+n_0})} - \sqrt{\mathbb{V}_t^{\mathbb{Q}}(\log S_{t+n_i})} + \sqrt{\mathbb{V}_t^{\mathbb{Q}}(\log S_{t+n_i})} \\ &= \frac{(n_0 - n_i)\left[\sqrt{\mathbb{V}_t^{\mathbb{Q}}(\log S_{t+n_{i+1}})} - \sqrt{\mathbb{V}_t^{\mathbb{Q}}(\log S_{t+n_i})}\right]}{n_{i+1} - n_i} + \sqrt{\mathbb{V}_t^{\mathbb{Q}}(\log S_{t+n_i})} \\ &= \frac{(n_0 - n_i)\sqrt{\mathbb{V}_t^{\mathbb{Q}}(\log S_{t+n_{i+1}})} + (n_{i+1} - n_0)\sqrt{\mathbb{V}_t^{\mathbb{Q}}(\log S_{t+n_i})}}{n_{i+1} - n_i}.\end{aligned}$$

Then the second moment is

$$\mathbb{E}_t^{\mathbb{Q}}(\log S_{t+n_0})^2 = [\mathbb{E}_t^{\mathbb{Q}}(\log S_{t+n_0})]^2 + \mathbb{V}_t^{\mathbb{Q}}(\log S_{t+n_0})$$

A fair price of variance swap $VS_{t,T}$ can be obtained by the pricing formula **Eq. 11**.