Online Supplemental Appendix to Optimal Candlestick-Based Spot Volatility Estimation: New Tricks and Feasible Inference Procedures

Tim Bollerslev* Jia Li† Qiyuan Li‡ Yifan Li§

November 20, 2025

Abstract

This supplemental appendix is divided into three sections. Section SA presents detailed results on the error bounds of the proposed computational methods. Section SB reports additional empirical findings, including results for different choices of the number of candlesticks. Section SC provides the details for implementing the proposed spot volatility estimators.

*Department of Economics, Duke University, Durham, NC 27708, and NBER; e-mail: boller@duke.edu.
†School of Economics, Singapore Management University; e-mail: jiali@smu.edu.sg.
‡Corresponding author. Faculty of Business and Economics, University of Hong Kong; e-mail: qiyuanli@hku.hk.
§Accounting & Finance Division, University of Manchester; e-mail: yifan.li@manchester.ac.uk.

SA Error Bound of the Proposed Computational Methods

Consider the density function of $W_t \in dr$ with $l \leq W_s \leq h$ for all $0 \leq s \leq t$,

$$F_t(r,h,l) = 4\sum_{k=-\infty}^{\infty}\big(\phi_t(r+2kR) - \phi_t(r-2l+2kR)\big), \tag{SA.1}$$

where $h \leq 0$ and $l \leq 0$ represent Brownian supremum and infimum, $l < r < h$ is the terminal value of the Brownian motion, and $R \equiv h - l$ is the Brownian range, all up to time t. $\phi_t(x) \equiv (2\pi t)^{-1/2}e^{-x^2/2t}$ is the Gaussian kernel. (SA.1) has the following alternative form (see Cox (2017))

$$F_t(r,h,l) = \frac{2}{R}\sum_{k=1}^{\infty} e^{-\frac{k^2\pi^2}{2R^2}t}\sin\Big(\frac{n\pi(r-l)}{R}\Big)\sin\Big(-\frac{n\pi l}{R}\Big). \tag{SA.2}$$

The trivariate density function is then

$$f_t(r,h,l) = -\frac{\partial^2 F_t}{\partial h \partial l} = 4\sum_{k=-\infty}^{\infty}\big(k^2\phi_t''(r+2kR) - k(k+1)\phi_t''(r-2l+2kR)\big\}, \tag{SA.3}$$

where $\phi_t''(x) \equiv (x^2/t^2 - 1)\phi_t(x)$. We are interested in properties of the following function

$$h(\lambda) = f_1(\lambda r, \lambda h, \lambda l) \equiv \lambda^{-3} f_{\lambda^{-2}}(r,h,l),$$

for some (r,h,l) such that $R > 0$, which happens almost surely.

In particular, we would like to compute the following improper integral

$$I^p \equiv \int_0^{\infty} \lambda^p h(\lambda) d\lambda. \tag{SA.4}$$

As $h(\lambda)$ is an infinite sum over an infinite interval, we need to truncate the infinite sum in h and also the integral limit. We prove that the effect of such truncation is asymptotically negligible. To this end, we first introduce the following truncated version of h for some integer M,

$$h_M(\lambda) = 4\sum_{k=-M}^{M}\big(k^2\phi_1''(r+2kR) - k(k+1)\phi_1''(r-2l+2kR)\big).$$

We then approximate I^p by integrating $h_M(\lambda)$ over some finite interval $[\epsilon, \Lambda]$,

$$I^p_{M,\epsilon,\Lambda} \equiv \int_{\epsilon}^{\Lambda} \lambda^p h_M(\lambda) d\lambda. \tag{SA.5}$$

We show in the following theorem that the approximation error vanishes for sufficiently large M and Λ, and sufficiently small ϵ.

Theorem S1. *For any $R > 0$ and $p \in \mathbb{R}$, there exist $K > 0$ depending only on (r, h, l) such that, for $(M \vee \Lambda) \to \infty$ and $\epsilon \to 0$,*

$$\left|I^p - I^p_{M,\epsilon,\Lambda}\right| \leq E^{(1)}_\epsilon + E^{(2)}_{M,\epsilon,\Lambda} + E^{(3)}_\Lambda,$$

where, with $(x)_+ \equiv x \vee 0$ and $u \equiv \Lambda/\epsilon$,

$$\begin{aligned} E^{(1)}_\epsilon &\leq K\epsilon^{(p-4)_+} e^{-\frac{\pi^2}{4R^2\epsilon^2}}, \\ E^{(2)}_{M,\epsilon,\Lambda} &\leq KM\epsilon^{p-1} u^{(p+1)_+} e^{-\frac{R^2\epsilon^2 M^2}{4}} \log u, \\ E^{(3)}_\Lambda &\leq K\Lambda^{(p+1)_+} e^{-\frac{R^2\Lambda^2}{2}}. \end{aligned}$$

Proof. The following decomposition is natural

$$\left|I^p - I^p_{M,\epsilon,\Lambda}\right| \leq \underbrace{\int_0^\epsilon \lambda^p h(\lambda) d\lambda}_{E^{(1)}_\epsilon} + \underbrace{\int_\epsilon^\Lambda \lambda^p |h_M(\lambda) - h(\lambda)| d\lambda}_{E^{(2)}_{M,\epsilon,\Lambda}} + \underbrace{\int_\Lambda^\infty \lambda^p h(\lambda) d\lambda}_{E^{(3)}_\Lambda}.$$

The behavior of $E^{(1)}_\epsilon$ and $E^{(3)}_\Lambda$ are related to the tail properties of h near 0 and ∞, which we shall establish in the following lemmas.

Lemma 1. *For any $R > 0$, it holds for all small enough $\lambda > 0$ that*

$$|h(\lambda)| \leq K\lambda^{-7} e^{-c\lambda^{-2}},$$

where $c \equiv \pi^2/2R^2$.

Proof. Write $t = \lambda^{-2} \geq 1$. We have $h(\lambda) = \lambda^{-3} f_t(r, h, l)$. Hence, it suffices to prove that, for all large enough t, we have

$$|f_t(r, h, l)| \leq Kt^2 e^{-ct}.$$

Notice that $f_t(r, h, l)$ can also be obtained by differentiating the form in (SA.2) w.r.t. h and l. The partial differentiations only produce polynomial factors in k and t which are dominated by the corresponding exponential term, but do not alter the order of the exponential term. As a result, the $k = 1$ term dominates the infinite sum for large enough t, such that

$$|f_t(r, h, l)| \leq K \frac{\partial}{\partial h \partial l} \Big\{ \frac{2}{R} e^{-ct} \sin\Big(\frac{\pi(r-l)}{R}\Big) \sin\Big(-\frac{\pi l}{R}\Big) \Big\}.$$

A routine calculate of the derivative shows that the leading term takes the following form for sufficiently large t,

$$|f_t(r, h, l)| \leq K \frac{2\pi^4}{R^7} \sin\Big(\frac{\pi(r-l)}{R}\Big) \sin\Big(-\frac{\pi l}{R}\Big) t^2 e^{-ct} \leq Kt^2 e^{-ct}.$$

Substituting t by λ^{-2} yields the desired result. □

From Lemma 1, we conclude that

$$E_\epsilon^{(1)} \le K \int_0^\epsilon \lambda^{p-7} e^{-c\lambda^{-2}} d\lambda \asymp \int_{c\epsilon^{-2}}^\infty u^{-\beta} e^{-u} du \le K \int_{c\epsilon^{-2}}^\infty u^q e^{-u} du,$$

where the second estimate follows from a change of variable $u = c/\lambda^2$ and $\beta = p/2 - 2$, and in the last estimate we set $q = 0 \vee (-\beta) \ge 0$. Now note that $\sup_{v\ge 0} v^q e^{-v/2} = (2q)^q e^{-q} < \infty$ for $q \ge 0$. Therefore, we have

$$E_\epsilon^{(1)} \le K \int_{c\epsilon^{-2}}^\infty e^{-u/2} du \asymp e^{-c/2\epsilon^2}, \tag{SA.6}$$

which is the desired asymptotic bound for $E_\epsilon^{(1)}$.

For h near infinity, we have

Lemma 2. *For any $R > 0$, it holds for all large enough λ that*

$$|h(\lambda)| \le K\lambda^2 e^{-R^2\lambda^2/2}.$$

Proof. By comparing the exponential orders of the summands in (SA.3), it is not hard to see that the $k = \pm 1$ terms with the smallest exponents dominate the exponential order. These exponents can be further bounded below by R since $|r| \le R$, hence we have

$$|h(\lambda)| \le K\phi_1''(R\lambda) \le K\lambda^2 e^{-R^2\lambda^2/2},$$

where the last term follows from the classic estimate $|\phi_1''(\lambda x)| \le K\lambda^2 e^{-\lambda^2 x^2/2}$ for all $\lambda \ge 1$, and we are done. □

From Lemma 2 and integration by parts, we deduce that

$$E_M^{(3)} \le K \int_{\Lambda_M}^\infty \lambda^{2+p} e^{-\lambda^2 R^2/2} d\lambda \le K\Lambda^{(p+1)\vee(-1)} e^{-R^2\Lambda^2/2} \le K\Lambda^{(p+1)_+} e^{-R^2\Lambda^2/2}, \tag{SA.7}$$

as desired.

Finally, to bound $E_M^{(2)}$, we need an almost uniform bound on the approximation error of $h(\lambda)$ by $h_M(\lambda)$. We have the following result:

Lemma 3. *It holds for all large enough M that*

$$\sup_{\lambda\ge\epsilon} |h(\lambda) - h_M(\lambda)| \le KM^2 e^{-R^2/4\cdot M}.$$

Proof. Define the reminder

$$R_M(\lambda) = 4 \sum_{k>|M|} \big(k^2\phi_1''(r + 2kR) - k(k+1)\phi_1''(r - 2l + 2kR)\big). \tag{SA.8}$$

We shall also need the following bound for all $y \in \mathbb{R}$ and some $C > 0$ independent of y and M:

$$\sup_{\lambda \geq \epsilon} \phi_1''(\lambda y) \leq C e^{-y^2\epsilon^2/4}. \tag{SA.9}$$

To see this, note that

$$|\phi_1''(\lambda y)| \leq C_0(1+\lambda^2 y^2)e^{-\lambda^2 y^2/2} \leq 2C_0 e^{\lambda^2 y^2/4} e^{-\lambda^2 y^2/2} = C e^{-y^2\lambda^2/4}, \tag{SA.10}$$

with the bound $(1+z^2) \leq 2e^{z^2/2}$. Now use the monotonicity of the last estimate in λ to conclude that the supremum of the bound can be achieved at $\lambda = \epsilon$.

Notice that for all fixed (r, h, l), there exists some finite $K > 0$ such that $|r + 2kR| \geq R|k|$ and $|r - 2l + 2kR| \geq R|k|$ for all $|k| \geq K$. Now take some $M > K$. For all $\lambda \geq \epsilon$, we have by (SA.9)

$$|R_M(\lambda)| \leq C \sum_{|k|>M} k^2 e^{-k^2\epsilon^2 R^2/4}, \tag{SA.11}$$

where we have used the relation $k^2 + |k(k+1)| \leq 3k^2$. Notice that for all $a > 0$, $f(x) = x^2 e^{-ax^2}$ is eventually decreasing in x for $x \geq (2a)^{-1/2}$. Therefore, for all positive k in the above sum, with $a = \epsilon^2 R^2/4$ we have

$$\begin{aligned} \sum_{k>M} k^2 e^{-ak^2} &\leq \int_M^\infty x^2 e^{-ax^2} dx = \frac{M}{2a} e^{-aM^2} + \frac{\sqrt{\pi}}{4a^{3/2}} \mathrm{Erfc}(\sqrt{a}M) \\ &\leq \Big(\frac{M}{2a} + \frac{1}{4a^2 M}\Big) e^{-aM^2} \leq K \frac{M}{\epsilon^2} e^{-R^2\epsilon^2 M^2/4}. \end{aligned}$$

Repeating the same argument for all negative k gives the desired result. □

Given Lemma 3, we have

$$E^{(2)}_{M,\epsilon,\Lambda} \leq K \frac{M}{\epsilon^2} e^{-R^2\epsilon^2 M^2/4} \int_\epsilon^\Lambda \lambda^p d\lambda. \tag{SA.12}$$

For the last integral, we have with $u = \Lambda/\epsilon$,

$$\int_\epsilon^\Lambda \lambda^p d\lambda = \epsilon^{p+1} \frac{u^{p+1} - 1}{p+1} \cdot \mathbf{1}_{\{p \neq -1\}} + \log u \cdot \mathbf{1}_{\{p=-1\}} \leq K \epsilon^{p+1} u^{(p+1)_+} \log u,$$

where the last inequality holds with all $u \geq 1$. This combined with (SA.12) yields the desired bound for $E^{(2)}_{M,\epsilon,\Lambda}$, and the proof is complete. *Q.E.D.*

SB Additional Empirical Results

SB.1 Different Choices of the Number of Candlesticks k

We do the empirical analysis for $k \in \{3, 10, 15\}$. Unsurprisingly, the qualitative results remains the same, see Figure S1-S6. The new optimal high-frequency candlestick-based estimators generally point to a sharp immediate spike along with significantly higher volatility for up to thirty minutes after the news releases. By comparison, the crude volatility estimates obtained by summing the high-frequency squared returns over short time intervals that are sometimes used in the literature provide a much "noisier" picture, making it difficult to tell how the market volatility changed in response to the news announcements.

SC Implementation Details

SC.1 Implementation Code

The MATLAB and Python code, together with a minimal walkthrough example, is available at https://lqyjasonlee.github.io/files/candle_spotvol.zip.

SC.2 Recommended Data Filters

In addition to the filter that remove extremal "doji" patterns mentioned in Footnote 5 of the main text, we recommend an auxiliary data filter that removes certain candlesticks to improve numerical stability.[1]

Specifically, compute the median range of adjacent candlesticks, e.g., the previous 30 minute-by-minute observations M_w, and discard candlesticks with $w_i/M_w \notin [0.3, 3.3]$. Under this filtering scheme, the excluded candles are virtually impossible (with probability $< 0.001\%$) to appear if volatility is approximately constant within blocks. The AMRE estimator exploits the fact that, over short blocks near time t, the volatilities of all candlesticks are close to σ_t, so that the Brownian approximation holds (for a technical argument, see Assumption 1 of Bollerslev et al. (2024)). Clearly, this assumption breaks down if the window contains an outlier candlestick representing an abrupt spike or drop in volatility that quickly dissipates. Such observations provide no useful information about σ_t and should therefore be filtered out.

[1] We thank Vaishnav Garg and Roberto Renò for raising this point.

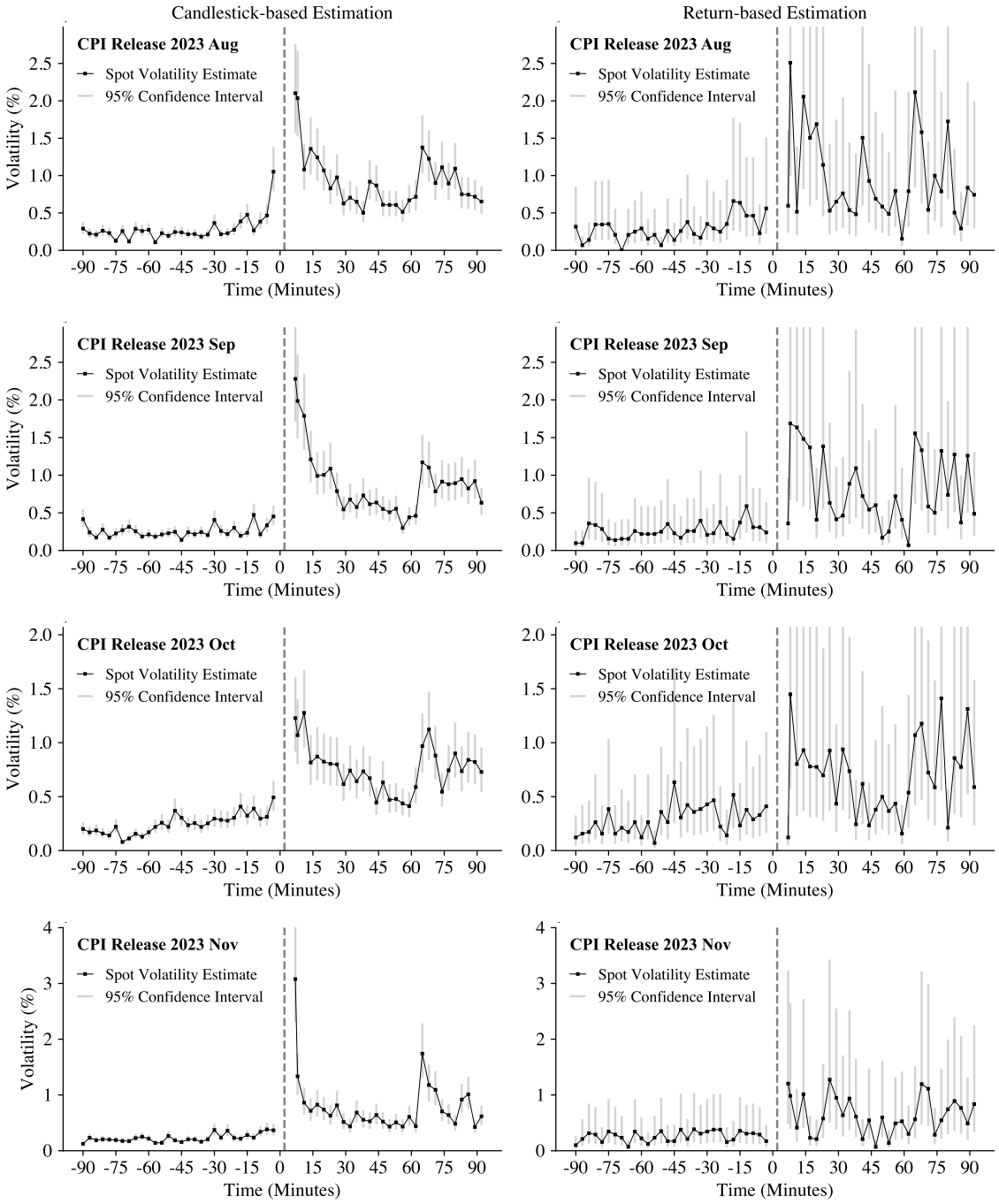

Figure S1: Spot volatility estimation for E-mini S&P 500 futures around CPI releases. Each spot volatility estimator is based on $k = 3$ consecutive candlesticks sampled at a one-minute frequency.

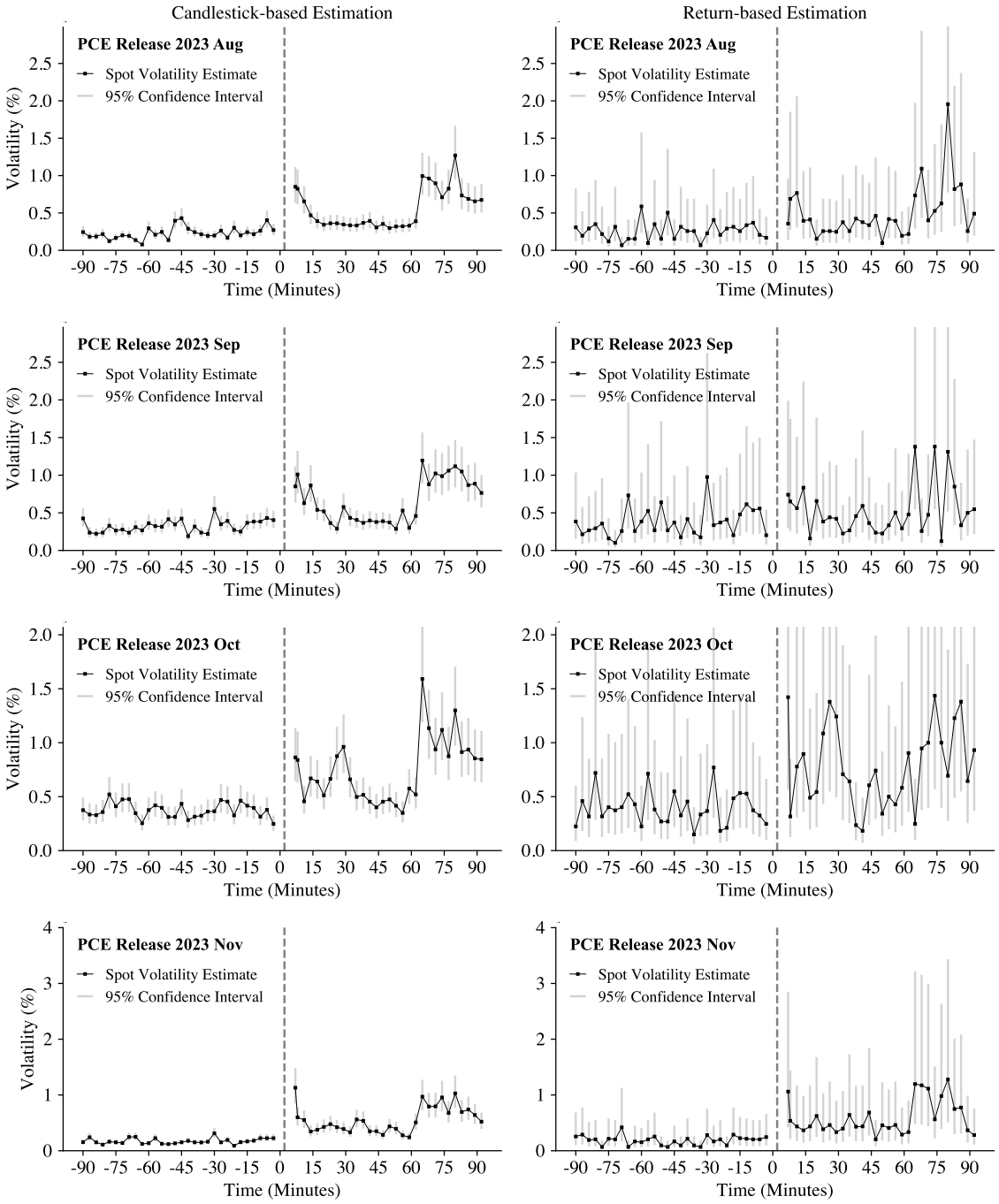

Figure S2: Spot volatility estimation for E-mini S&P 500 futures around PCE releases. Each spot volatility estimator is based on $k = 3$ consecutive candlesticks sampled at a one-minute frequency.

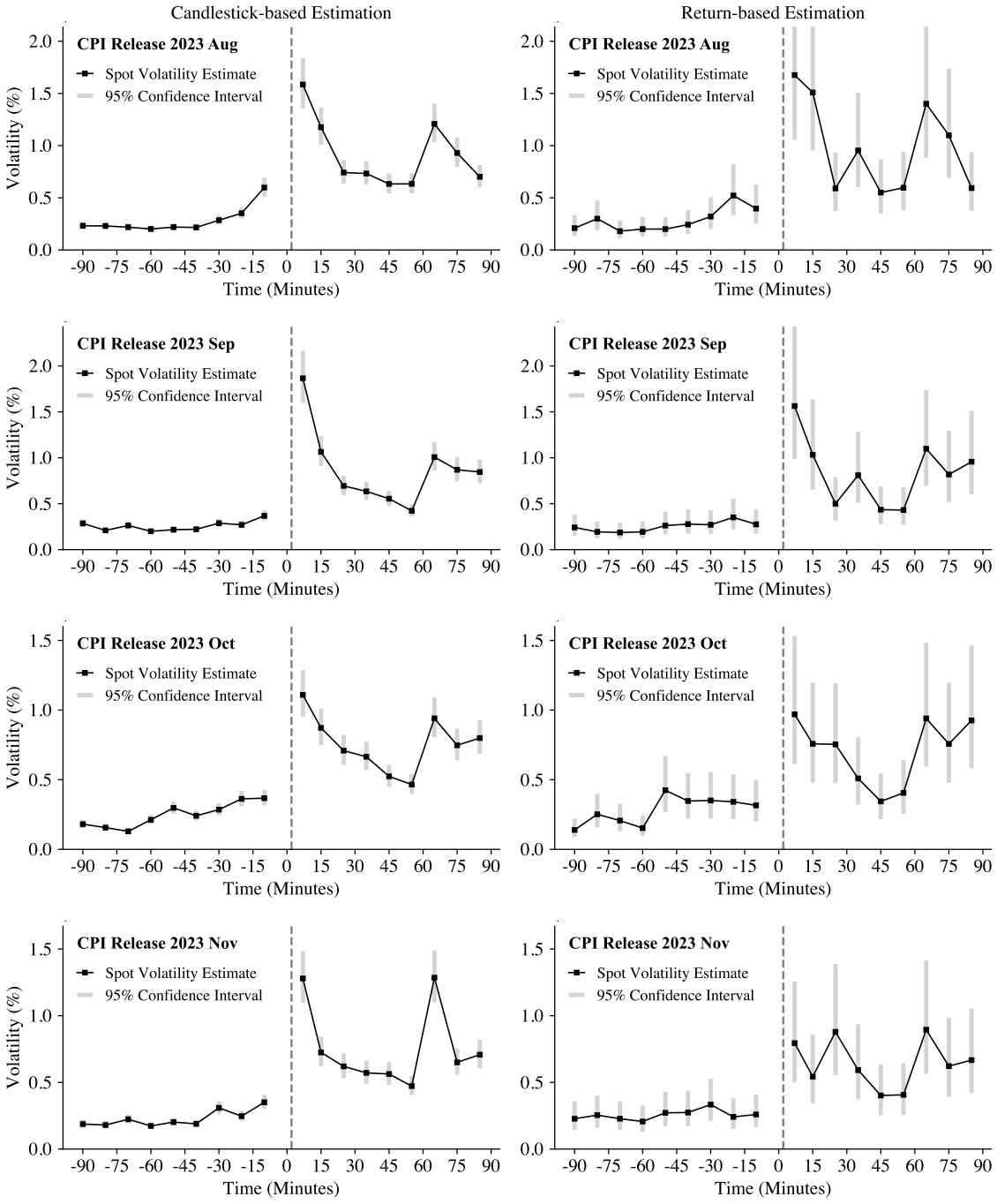

Figure S3: Spot volatility estimation for E-mini S&P 500 futures around CPI releases. Each spot volatility estimator is based on $k = 10$ consecutive candlesticks sampled at a one-minute frequency.

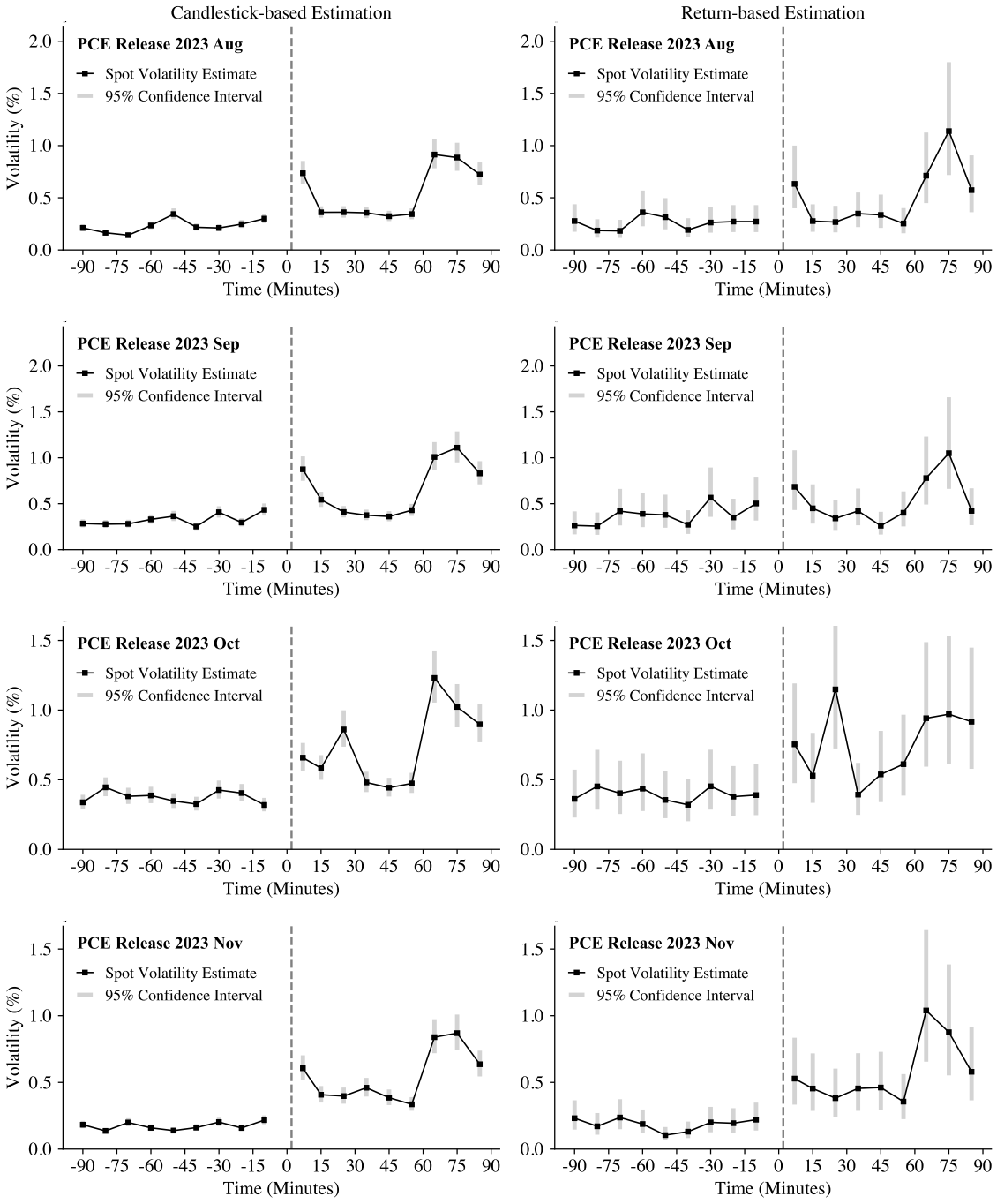

Figure S4: Spot volatility estimation for E-mini S&P 500 futures around PCE releases. Each spot volatility estimator is based on $k = 10$ consecutive candlesticks sampled at a one-minute frequency.

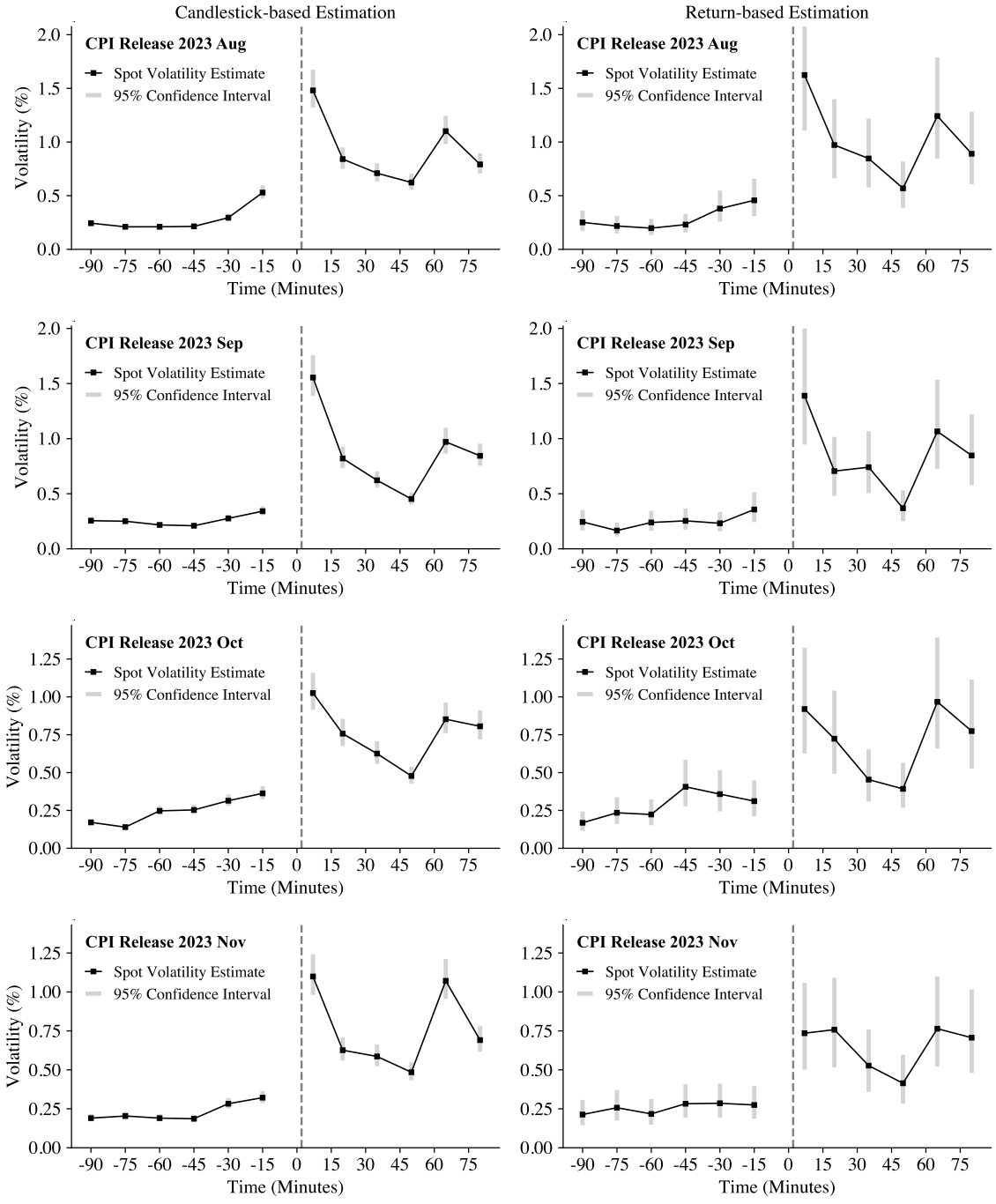

Figure S5: Spot volatility estimation for E-mini S&P 500 futures around CPI releases. Each spot volatility estimator is based on $k = 15$ consecutive candlesticks sampled at a one-minute frequency.

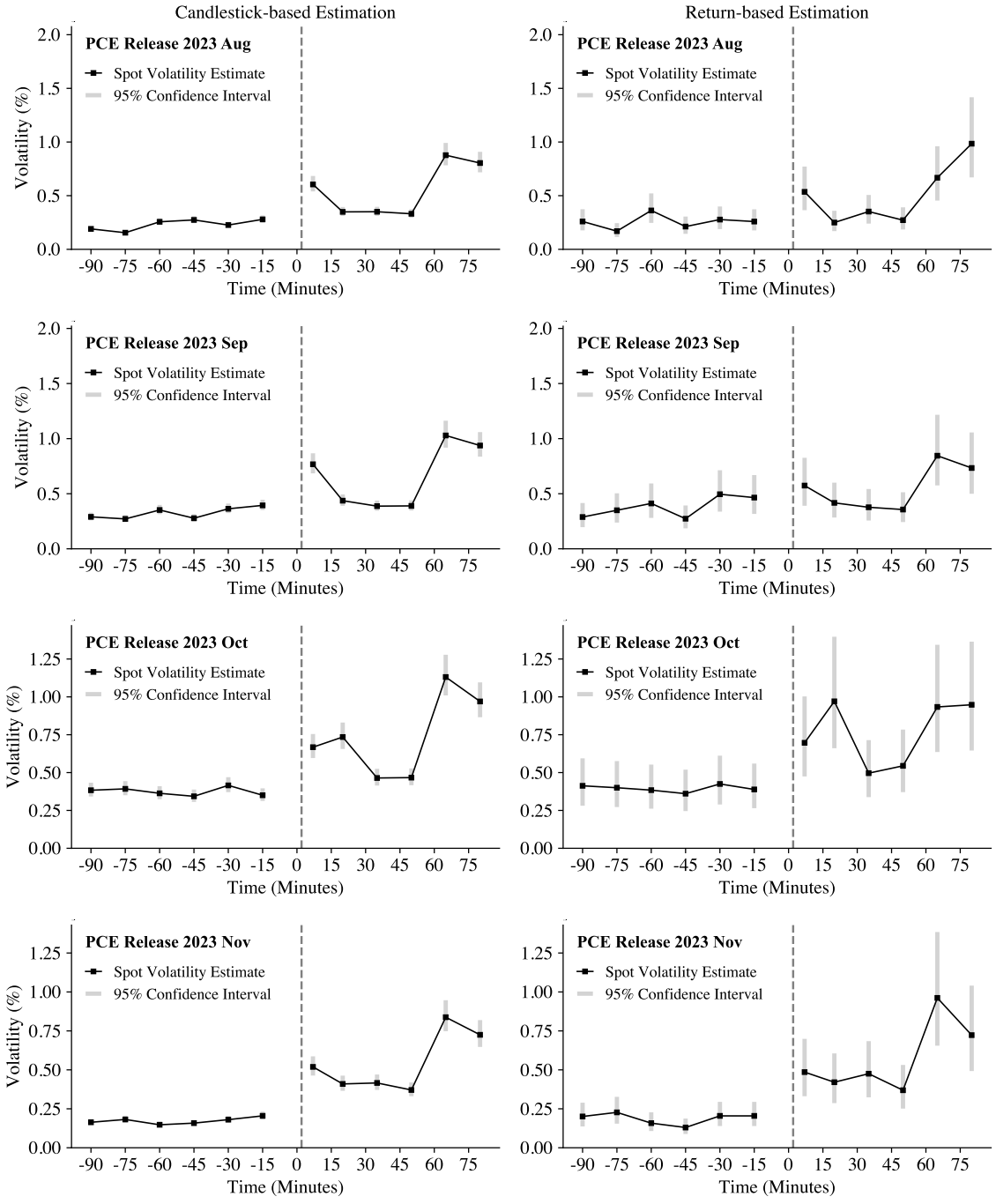

Figure S6: Spot volatility estimation for E-mini S&P 500 futures around PCE releases. Each spot volatility estimator is based on $k = 15$ consecutive candlesticks sampled at a one-minute frequency.

References

Bollerslev, T., J. Li, and Q. Li (2024): "Optimal Nonparametric Range-Based Volatility Estimation," *Journal of Econometrics*, 238, 105548.

Cox, D. R. (2017): *The theory of stochastic processes*, Routledge.