Finance and Economics Discussion Series

Federal Reserve Board, Washington, D.C.
ISSN 1936-2854 (Print)
ISSN 2767-3898 (Online)

Local Estimation for Option Pricing: Improving Forecasts with Market State Information

Hyung Joo Kim, Dong Hwan Oh

2025-076

Please cite this paper as:
Kim, Hyung Joo, and Dong Hwan Oh (2025). "Local Estimation for Option Pricing: Improving Forecasts with Market State Information," Finance and Economics Discussion Series 2025-076. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2025.076.

Local Estimation for Option Pricing: Improving Forecasts with Market State Information*

Hyung Joo Kim† Dong Hwan Oh‡

July 30, 2025

Abstract

We propose a novel estimation framework for option pricing models that incorporates local, state-dependent information to improve out-of-sample forecasting performance. Rather than modifying the underlying option pricing model, such as the Heston-Nandi GARCH or the Heston stochastic volatility framework, we introduce a local M-estimation approach that conditions on key state variables including VIX, realized volatility, and time. Our method reweights historical observations based on their relevance to current market conditions, using kernel functions with bandwidths selected via a validation procedure. This adaptive estimation improves the model's responsiveness to evolving dynamics while maintaining tractability. Empirically, we show that local estimators substantially outperform traditional non-local approaches in forecasting near-term option implied volatilities. The improvements are particularly pronounced in low-volatility environments and across the cross-section of options. The local estimators also outperform the non-local estimators in explaining future option returns. Our findings suggest that local information, when properly incorporated into the estimation process, can enhance the accuracy and robustness of option pricing models.

*The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors.

†Federal Reserve Board; hyungjoo.kim@frb.gov

‡Federal Reserve Board; donghwan.oh@frb.gov

1 Introduction

Accurately pricing options is essential for trading, hedging, risk management, and regulatory purposes. While the ability to forecast future option prices or their distributions is often more relevant for these applications, much of the literature on option valuation has focused on fitting historical or in-sample option prices more closely by developing increasingly complex models.[1] Since the seminal work of Black and Scholes (1973), numerous models have been proposed to address the empirical limitations of the Black-Scholes (BS) framework—most notably its inability to capture volatility smiles and skews due to the assumption of constant volatility. Models such as the stochastic volatility (SV) model of Heston (1993), the GARCH-based model of Heston and Nandi (2000), and their extended models have made significant strides in matching observed option prices, underlying return dynamics, and other option moments while offering tractable pricing formulas.[2] Nonetheless, relatively little attention has been paid to their forecasting performance.

Furthermore, the role of the estimation method itself has received limited attention. Most standard approaches use fixed-parameter estimation techniques, which implicitly treat all historical observations as equally informative. This uniform weighting can be problematic in evolving or regime-shifting markets, where stale observations may dilute the influence of relevant information. Recent work by Dendramis, Kapetanios, and Marcellino (2020) and Oh and Patton (2024) shows that reweighting observations based on current market conditions can improve forecasting accuracy in macroeconomic and financial applications. These advances motivate our core research question: can option pricing models be improved by modifying the estimation method, without changing the model?

In this paper, we develop a novel local M-estimation framework for option pricing models that conditions on market state variables using kernel-based weighting. Rather than altering the model structure, we adapt parameter estimates to reflect the current market environment. Our framework

[1] For example, regulatory applications and central counterparties (CCPs) often require models that can forecast potential exposures under future market scenarios. Forecasted option values are used in margin requirement calibration, stress testing, and systemic risk assessments.

[2] An extensive set of studies include, but not limited to, Bakshi, Cao, and Chen (1997), Bates (2000, 2019), Duffie, Pan, and Singleton (2000), Pan (2002), Eraker (2004), Christoffersen, Jacobs, Ornthanalai, and Wang (2008), Christoffersen, Heston, and Jacobs (2009, 2013), Christoffersen, Jacobs, and Mimouni (2010), Andersen, Fusari, and Todorov (2015), Babaoğlu, Christoffersen, Heston, and Jacobs (2018), Bardgett, Gourier, and Leippold (2019), Du and Luo (2019), Aït-Sahalia, Karaman, and Mancini (2020), Aït-Sahalia, Li, and Li (2021), Gruber, Tebaldi, and Trojani (2021), Heston, Jacobs, and Kim (2024).

allows for observation-specific weighting based on a range of informative variables, including the VIX, realized volatility, risk-neutral skewness and kurtosis, VIX volatility index (VVIX), and the variance risk premium (VRP), or weighting that places greater emphasis on more recent data points. Bandwidths or tuning parameters governing the weighting scheme are selected using a split-sample validation procedure to avoid look-ahead bias.

We apply this estimation technique to both GARCH and SV option pricing models and evaluate its forecasting performance using out-of-sample (OOS) option implied volatilities from 2015 to 2023. Our empirical analysis demonstrates that the local estimation approach significantly outperforms traditional non-local methods in both time-series and cross-sectional dimensions.

Our empirical findings demonstrate that local estimation substantially improves the accuracy of forecasting option prices or implied volatilities. Across multiple forecast periods from 2015 to 2023, local estimators consistently outperform their non-local counterparts in both time-series and cross-sectional environments. In the GARCH model, conditioning on VIX leads to the greatest improvements in implied volatility forecasting, while in the SV model, time-based local estimations perform best. These enhancements are especially pronounced in low-volatility periods, such as 2017, when market conditions evolve gradually. Although the outperformance of local estimation is less noticeable during high-volatility periods, it performs at least at par or slightly better relative to non-local estimation.

In addition to reduced time-series forecast errors, local estimation leads to overall improved and more balanced pricing performance across different moneyness levels and maturities. We confirm that this improvement stems from the better alignment of model-implied risk-neutral distributions with nonparametric benchmarks. Moreover, local estimators yield expected option returns that match ex-post realized option returns better, indicating improvements in the joint modeling of physical and risk-neutral dynamics. These results not only support the theoretical value of state-dependent estimation but also demonstrate its practical utility for traders, margin models, and short-horizon risk forecasts.

This paper contributes to the literature in three key directions:

First, our paper contributes to the growing literature on forecasting option moments. One strand of this literature focuses on forecasting option-implied volatility surfaces (e.g., Almeida, Fan, Freire,

and Tang, 2023; Chen, Grith, and Lai, 2025; Dufays, Jacobs, and Rombouts, 2025). While these studies generate the entire volatility surface and then forecast it using various volatility forecasting models or their extensions, our approach emphasizes incorporating local state information into the estimation of dynamic option pricing models to better forecast implied volatilities. Instead of directly forecasting the volatility surface, we use a better-estimated option pricing model to generate the distribution of future underlying prices and variances, and then compute expected implied volatilities from those distributions. This approach allows us to forecast implied volatility, internally consistent with a pricing model. Another strand of the literature focuses on forecasting the cross-section of option returns (e.g., Zhan, Han, Cao, and Tong, 2022; Bali, Beckmeyer, Moerke, and Weigert, 2023; Cao, Vasquez, Xiao, and Zhan, 2023). Though our goal is not to find forecasting models for option returns, we compute model-implied expected option returns and show that our local estimation model forecasts option returns better than a benchmark (non-local) model.

Second, we improve option pricing performance by incorporating current information through a novel estimation method. Rather than extending or modifying existing models, we employ a local estimation method that naturally incorporates recent market information, which proves beneficial for pricing options or forecasting implied volatilities.[3] This approach allows us to capture the most up-to-date market dynamics without increasing model complexity.[4] By focusing on estimation rather than model structure, we maintain simplicity while enhancing adaptability to changing market conditions. This method is particularly effective in volatile or rapidly evolving markets where traditional models might struggle to adapt quickly. Furthermore, our approach can potentially reduce estimation time, parameter uncertainty, and model risk, which are common challenges in more complex option pricing models.

Third, we bridge two strands of research—dynamic volatility modeling and local estimation methods in econometrics—and apply them to a forward-looking option pricing framework.[5] Our

[3]Relatedly, Kim (2025) and Schreindorfer and Sichert (2025) exploit state information to better estimate the conditional pricing kernel implied by index returns and options, and Oh and Patton (2024) show that applying a local estimation method improves volatility forecasting.

[4]As an example of capturing state-dependent market dynamics in option pricing with a modest increase in model complexity, Hansen and Tong (2025) improve the GARCH option pricing model fit by allowing a time-varying risk preference parameter.

[5]This perspective is particularly relevant for practitioners involved in margin setting and short-horizon trading. For example, clearinghouses often rely on forecasts of near-term option prices conditional on current market conditions to determine initial and variation margins. Similarly, trading desks require real-time pricing that reflects current volatility and tail risk exposure, which static parameter estimates may fail to capture.

approach aligns with how options are used in practice, such as in margin setting or short-horizon forecasting, where the objective is not to fit past prices but to forecast near-term values conditional on current information. By incorporating forward-looking local information into the estimation step, we improve the responsiveness of otherwise fixed-parameter models and enhance their forecasting performance without adding structural complexity.

The remainder of the paper is organized as follows. Section 2 describes the benchmark option pricing models and the local M-estimation framework, along with implementation details including kernel choices, bandwidth selection, and validation procedures. Section 3 introduces the data, and Section 4 reports the empirical results. Section 5 concludes.

2 Model and Estimation

This section outlines the option pricing models and the local estimation framework. We specify the physical and risk-neutral dynamics for the Heston-Nandi GARCH and Heston SV models, followed by a local M-estimation approach that reweights observations based on state variables. This method preserves the model structure while adapting parameters to current market conditions. We also describe kernel weighting, bandwidth selection, and the validation procedure used for out-of-sample evaluation.

2.1 Stock price dynamics and transition likelihoods

We focus on two widely used option pricing models: the Heston and Nandi (2000) GARCH option pricing model and the Heston (1993) SV option pricing model, as well as their underlying dynamics. The former is based on the discrete-time dynamic of the stock price, where its variance follows a GARCH process derived from historical returns and past volatilities. The latter uses the continuous-time dynamic of the stock price in conjunction with its stochastic volatility, which follows a square-root process. Both approaches are essential for pricing options and analyzing volatility dynamics because they incorporate the time-varying nature of volatility into the valuation process. They have also received ample attention in the literature, not only for their time-varying volatility but also for their quasi-closed-form option pricing formulas and their sufficient performance in terms of average option fits.

For estimation, we specify the physical dynamics and maximize the return transition likelihoods. It is worth noting that we do not use options data in model parameter estimations; instead, we use them for model validation and out-of-sample analysis. We discuss more details on how we convert the optimal physical dynamics to risk-neutral dynamics and how we price options in Section 2.2.

2.1.1 The GARCH dynamics

We employ the physical stock price and variance dynamics of Heston and Nandi (2000) and Christoffersen, Heston, and Jacobs (2013) as our first benchmark model:

$$\begin{aligned} \ln S(t+1) &= \ln S(t) + \left(r + \left(\mu - \frac{1}{2}\right) v(t+1)\right) + \sqrt{v(t+1)} z(t+1), \\ v(t+1) &= \omega + \beta v(t) + \alpha \left(z(t) - \gamma \sqrt{v(t)}\right)^2, \end{aligned} \tag{1}$$

where each time interval is one day, r is the risk-free rate, and $z(t)$ follows a standard normal distribution. In this model, μ governs the equity risk premium. In the variance dynamic, α identifies the reactivity of variance to stock price, γ provides the leverage effect, and β partially captures the persistence of variance.

Since the next-period daily variance $v(t+1)$ is determined once the daily return $\ln S(t) - \ln S(t-1)$ or return shock $z(t)$ is realized, it enables us to formulate the return transition likelihood recursively. At each point in time $t+1$, the log likelihood is given as

$$\ln \mathcal{L}(t+1) = -\frac{1}{2} \ln \left(2\pi v(t+1)\right) - \frac{1}{2} \left(\frac{\ln S(t+1) - \ln S(t) - \left(r + \left(\mu - \frac{1}{2}\right) v(t+1)\right)}{\sqrt{v(t+1)}} \right)^2, \tag{2}$$

where the variance is given as

$$v(t+1) = \omega + \beta v(t) + \alpha \left(\frac{\ln S(t) - \ln S(t-1) - \left(r + \left(\mu - \frac{1}{2}\right) v(t)\right)}{\sqrt{v(t)}} - \gamma \sqrt{v(t)} \right)^2. \tag{3}$$

2.1.2 The SV dynamics

As our second model, we adopt the continuous-time stochastic volatility model of Heston (1993). Its physical dynamics are given as

$$\begin{aligned} d\ln S(t) &= \left(r + \left(\mu - \frac{1}{2}\right) v(t)\right) dt + \sqrt{v(t)} dz_1(t), \\ dv(t) &= \kappa(\theta - v(t))dt + \sigma\sqrt{v(t)} dz_2(t), \end{aligned} \tag{4}$$

where z_1 and z_2 follow Brownian motions with correlation coefficient ρ. The square-root process or the CIR process used for the variance dynamic fundamentally captures a mean-reverting property of variance, where κ and θ identify the mean-reverting speed and the long-run mean, respectively.

For model estimation, the presence of latent stochastic variance makes implementation relatively more complex than that of the GARCH dynamics, as it requires filtering the latent variable.[6] To reduce the computational burden, we instead adopt the method of Heston, Jacobs, and Kim (2023), which exploits the linear relationship between the squared VIX and the instantaneous stochastic variance implied by this SV model.[7] Specifically, we assume

$$v(t) = \eta_0 + \eta_1 \mathrm{VIX}^2(t), \tag{5}$$

where η_0 and η_1 are free parameters to be estimated.

To characterize the transition likelihood function, we apply the Euler discretization to equation (4), resulting in:

$$\begin{aligned} \ln R(t+\Delta) &= \left[r + \left(\mu - \frac{1}{2}\right) v(t)\right] \Delta + \epsilon_1(t+\Delta), \\ v(t+\Delta) - v(t) &= \kappa(\theta - v(t))\Delta + \epsilon_2(t+\Delta), \end{aligned} \tag{6}$$

where $R(t+\Delta) = S(t+\Delta)/S(t)$ denotes the gross return and $\Delta = 1/252$.[8] The error terms

[6] To filter stochastic variance, the literature employs various approaches, including the Kalman filter, particle filter, and Markov Chain Monte Carlo (MCMC) techniques. For examples, see Ghysels and Jasiak (1994) and Ruiz (1994) for applications of the Kalman filter, Christoffersen, Jacobs, and Mimouni (2010) and Dufays, Jacobs, Liu, and Rombouts (2023) for the particle filter, and Andersen, Benzoni, and Lund (2002), Eraker, Johannes, and Polson (2003), Jones (2003), Eraker (2004), and Bates (2006) for MCMC methods.

[7] The exact linear relationship involves a specific functional form with a measurement error. However, to avoid the additional filtering problem required to model this error, they further simplify the setup by relaxing the functional form and omitting the measurement error.

[8] Note that $\ln R(t+\Delta)$ is the daily log return between t and $t+\Delta$ while $v(t)$ is the annualized variance at time t.

$\epsilon(t+\Delta) = (\epsilon_1(t+\Delta),\ \epsilon_2(t+\Delta))'$ follow a bivariate normal distribution with mean and covariance matrix given as

$$\mathbf{0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \mathbf{\Sigma}(t) = \begin{pmatrix} v(t) & \sigma\rho v(t) \\ \sigma\rho v(t) & \sigma^2 v(t) \end{pmatrix} \Delta.$$

Then, the joint log likelihood at $t+\Delta$ is expressed as:

$$\begin{aligned} \ln \mathcal{L}(t+\Delta) &= \ln f\left(\log R(t+\Delta), \mathrm{VIX}^2(t+\Delta) | \mathrm{VIX}^2(t)\right) \\ &= \ln\left[f(\log R(t+\Delta), v(t+\Delta)|v(t)) \times J(t+\Delta)\right] \\ &= -\ln(2\pi) - \frac{1}{2}\ln|\mathbf{\Sigma}(t)| - \frac{1}{2}\epsilon(t+\Delta)'\mathbf{\Sigma}^{-1}(t)\epsilon(t+\Delta) + \ln\eta_1, \end{aligned} \tag{7}$$

where $f(\ln R(t+\Delta), v(t+\Delta)|v(t))$ represents the conditional density of the discretized $\log R(t+\Delta)$ and $v(t+\Delta)$, $J(t+\Delta)$ is the Jacobian term relating $\mathrm{VIX}^2(t+\Delta)$ to $v(t+\Delta)$, given by η_1 from equation (5). Time t is measured in days.

2.2 Risk-neutral dynamics and option pricing

We now turn to the discussion of how we price options under the two dynamics (GARCH and SV), respectively. It is noteworthy to mention that our physical dynamics in equations (1) and (4) alone are insufficient for option pricing. Either the pricing kernel or risk-neutral dynamics are necessary to compute model-implied option prices. In this section, we describe our choice of the pricing kernel and the internally consistent risk-neutral dynamics, followed by the option valuation procedure.

To compute option prices, we first specify the pricing kernel and then derive the corresponding risk-neutral stock price dynamics. Following Heston and Nandi (2000), we assume that the representative agent is risk-averse to negative aggregate market returns. Specifically, we adopt the power utility function, which is parsimonious and widely used in the literature. The pricing kernel is then given by

$$M(t) = M(0)e^{-\delta t}\left(\frac{S(t)}{S(0)}\right)^{-\xi}, \tag{8}$$

where ξ represents the risk aversion parameter, and δ represents the time-preference parameter.

Inferred from the given physical dynamics and the pricing kernel in equation (8), we obtain the

In contrast, following the convention, $v(t)$ in equation (1) represents the daily variance in the GARCH model.

following risk-neutral dynamics.[9] For the GARCH model,

$$\begin{aligned} \ln S(t+1) &= \ln S(t) + \left(r - \frac{1}{2}v(t+1)\right) + \sqrt{v(t+1)}z^*(t+1), \\ v(t+1) &= \omega + \beta v(t) + \alpha\left(z^*(t) - \gamma^*\sqrt{v(t)}\right)^2, \end{aligned} \tag{9}$$

where

$$\begin{aligned} z^*(t) &= z(t) + \xi\sqrt{v(t)}, \\ \gamma^* &= \gamma + \xi, \\ \xi &= \mu. \end{aligned}$$

For the SV model,

$$\begin{aligned} d\ln S(t) &= \left(r - \frac{1}{2}v(t)\right)dt + \sqrt{v(t)}dz_1^*(t), \\ dv(t) &= \kappa^*(\theta^* - v(t))dt + \sigma\sqrt{v(t)}dz_2^*(t), \end{aligned} \tag{10}$$

where

$$\begin{aligned} dz_1^*(t) &= dz_1(t) + \xi\sqrt{v(t)}dt, \\ dz_2^*(t) &= dz_2(t) + \xi\rho\sigma\sqrt{v(t)}dt, \\ \kappa^* &= \kappa + \xi\rho\sigma, \\ \theta^* &= \kappa\theta/\kappa^*, \\ \xi &= \mu. \end{aligned}$$

Using the risk-neutral dynamics in equations (9) and (10), we follow the fast Fourier transformation of Carr and Madan (1999) for option valuation. The price of a call option with its strike

[9] Christoffersen, Heston, and Jacobs (2013) and Heston, Jacobs, and Kim (2023) provide more detailed derivation of these relationships for the GARCH and the SV models, respectively. They also discuss a more generalized pricing kernel that preserves the same functional form of the risk-neutral dynamics.

price K and maturity τ is given by a quasi-closed-form solution involving a numerical integration:

$$C(S(t), v(t), t) = \frac{e^{-ak}}{\pi} \int_0^\infty Re\left[e^{-iuk}\psi(u)\right] du, \tag{11}$$

where i is the imaginary unit and k is the natural log of K. The function $\psi(u)$ is the Fourier transform of a modified call price, which is the call price multiplied by e^{ak} for $a > 0$.[10] We calculate $\psi(u)$ as follows:

$$\psi(u) = \frac{e^{-r\tau} f_\tau^{CH}(u - i(a+1)|S(t), v(t))}{(a + iu)(a + 1 + iu)},$$

where $f_\tau^{CH}(\phi|S(t), v(t)) = E_t^*\left[e^{i\phi \log S(t+\tau)}\right]$ represents the risk-neutral conditional characteristic function of $\log S(t+\tau)$. We derive closed-form expressions for $f_\tau^{CH}(\phi|S(t), v(t))$ following Heston and Nandi (2000) for the GARCH model and Heston (1993) for the SV model. The price of a put option with the same strike and maturity can be determined using put-call parity. Note that the pricing formula in equation (11) does not account for future dividends. In line with the existing literature, we adjust the current index level by discounting it with the dividend yield; that is, we use $S(t)e^{-q\tau}$, where q is the continuously compounded dividend yield at time t. Lastly, we convert each model-implied option price into the Black–Scholes implied volatility to facilitate consistent comparisons across the cross section and over time.

2.3 Local M-estimation

Most stock price dynamics are typically estimated using M-estimation, a flexible framework for parameter estimation that minimizes or maximizes a specified objective function:

$$\hat{\theta}_T = \arg\min_\theta \frac{1}{T} \sum_{t=1}^{T} L\left(Y_t, g_{t-1}(\theta)\right), \tag{12}$$

where $\theta \in \Theta \subseteq \mathbb{R}^p$. In this framework, L represents the loss function, Y denotes the target variable, and $g(\theta)$ defines a parametric model of the target functional. M-estimation encompasses a wide range of methods, including maximum likelihood estimation (MLE) and least squares estimation. Under mild regularity conditions, M-estimators exhibit desirable properties such as consistency and

[10] We use $a = 4$, following Heston, Jacobs, and Kim (2023).

$\sqrt{T}$-asymptotic normality; see Newey and McFadden (1994).

However, the standard M-estimation framework assumes implicitly that all observations contribute equally to the objective function, which limits its ability to incorporate time-specific information that could improve parameter estimation. In the current pricing environment, certain historical data points may hold greater informational value for parameter estimation. As demonstrated by Oh and Patton (2024), maintaining the baseline model while optimally weighting observations can enhance forecasting performance. Following a similar approach, we preserve the standard option pricing model but assign greater weights to past data points that are more relevant to the specific day of pricing.

Building on the approaches of Dendramis, Kapetanios, and Marcellino (2020) and Oh and Patton (2024), we define the estimator as:

$$\tilde{\theta}_{h,T}(s) = \arg\min_{\theta} \frac{1}{T} \sum_{t=1}^{T} L\left(Y_t, g_{t-1}(\theta)\right) \mathcal{K}\left(s - S_{t-1}; h_T\right), \text{ for } s \in Int(\mathcal{S}), \tag{13}$$

where $\mathcal{K}$ represents the kernel function, h_T is a bandwidth parameter that decreases with sample size, s is a predetermined value of the state variable, typically set to its current value at the time the forecast is generated, and $\text{Int}(\mathcal{S})$ denotes the interior of the support of the state variable $\mathcal{S}$. The weights are determined by a kernel function based on the distance between the current and past levels of state variables. Past values that deviate significantly from the current level are assigned smaller weights, while those closer to the current level receive larger weights. This approach allows the underlying dynamics to adapt effectively to the current environment, potentially improving option price forecasts.

This local approach offers significant advantages over traditional estimation methods, particularly in the context of option pricing. There are two main reasons for its effectiveness:

1. Resolution of model misspecification: Standard option pricing models often suffer from misspecification, as they may not fully capture the complex dynamics of volatilities of underlying assets. The local estimation approach helps mitigate this issue by allowing the model parameters to vary with the state variable, thus adapting to prevailing market conditions where the option pricing is conducted. This flexibility enables the model to adjust to different market regimes and conditions, effectively reducing the impact of model misspecification. By giving more weight to observations

that are similar to the current market state, the method implicitly allows for nonlinearities and time-varying relationships that may not be captured by a fixed parametric model.

2. Informative state variables: The effectiveness of this approach relies on the choice of informative state variables that contain valuable information about potential model misspecification. These state variables serve as proxies for changing market conditions or regimes that affect option prices. For instance, variables such as option implied volatility, skewness and kurtosis, or macroeconomic indicators such as realized variance of S&P 500 returns can provide crucial information about the current market environment. By conditioning the estimation on these state variables, the local approach incorporates this additional information into the pricing model, leading to more accurate and adaptive option price estimates. The state variables essentially guide the model to focus on the most relevant historical data for the current pricing scenario, improving its predictive power.

By addressing model misspecification and leveraging informative state variables, this local estimation approach can potentially lead to more accurate option pricing, especially in dynamic and evolving market conditions where traditional fixed-parameter models may struggle.[11] The efficacy of this local estimation approach hinges critically on the judicious selection of state variables and the determination of appropriate bandwidth parameters. In the following section, we present a detailed exposition of our methodology for selecting these crucial elements.

In our empirical analyses, we use the negative log likelihood ($-\ln \mathcal{L}$) in equation (2) or (7) as a loss function L, in conjunction with the realized stock return as a target variable Y and either the GARCH dynamics or the SV dynamics for a parametric model $g(\theta)$. The set of parameters θ is $\{\mu,\ \omega,\ \beta,\ \alpha,\ \gamma\}$ for the GARCH dynamics and $\{\mu,\ \kappa,\ \theta,\ \sigma,\ \rho,\ \eta_0,\ \eta_1\}$ for the SV dynamics. Furthermore, we impose a couple of restrictions. To relevantly capture the unconditional level of equity risk premium, we fix μ at the level of the time-series average equity premium between January 1996 and August 2023. To preserve variance positivity, we impose $\omega \geq 0$ and $\beta \geq 0$ for the GARCH model and the Feller (1951) condition ($2\kappa\theta > \sigma^2$) for the SV model.

[11] For a comprehensive theoretical analysis of the improvements afforded by this local estimation approach, we refer readers to Section 2 of Oh and Patton (2024), which provides detailed discussions on the asymptotic properties of the estimator.

2.4 Implementing Local Estimation and Model Evaluation: Methodological Details

To implement the local estimation through equation (13), we need to specify the weighting function and select appropriate state variables. We begin by detailing the kernel functions for different types of state variables, then describe our process for bandwidth determination and model evaluation, and finally discuss the specific state variables used in our study.

For stochastic state variables, we adopt a Gaussian kernel:

$$\mathcal{K}_G(x;h) = \exp\left\{-\frac{x^2}{2h^2}\right\}, \text{ for } x \in \mathbb{R} \text{ and } h > 0 \tag{14}$$

We examine the bandwidth values, h, ranging from $0.05\sigma_S$ to $5\sigma_S$, where σ_S represents the standard deviation of the state variable. Smaller bandwidths make the model parameters more 'localized,' but they reduce precision due to a smaller effective sample size. Conversely, as h increases, the local estimator converges to the non-local methods equivalent to the original estimation methods proposed by Heston and Nandi (2000) and Heston (1993). While h is a time-insensitive parameter, we implement a dynamic approach for σ_S to enhance precision. Specifically, we compute σ_S at each point in time as the rolling standard deviation of the state variable over the preceding three-year window.

When employing time as the state variable, we use a one-sided exponential kernel defined by the bandwidth parameter λ and a window length of m:

$$\mathcal{K}_E(j;\lambda) = \frac{\lambda^j(1-\lambda)}{(1-\lambda^m)} 1\{j < m\}, \text{ for } j \in \{0,1,2,...\} \text{ and } 0 < \lambda < 1 \tag{15}$$

The values of λ explored range from 0.98 to 0.9999. Lower λ values result in reduced emphasis on older observations during estimation, leading to more temporally localized model parameters but increased estimation variability. As λ approaches 1, the weighting function flattens, and the local estimator converges to the non-local benchmark estimators described in Heston and Nandi (2000) and Heston (1993).

The process for determining bandwidth parameters and evaluating option pricing models follows

a structured approach. We divide our dataset into two distinct subsets: an estimation sample and an out-of-sample (OOS) set. The estimation sample is used to determine the optimal bandwidth parameter for each state variable. Subsequently, the selected bandwidth parameters are employed to evaluate option pricing errors during the OOS period, ensuring that bandwidth parameters are free from look-ahead bias.[12]

More specifically, our methodology follows a two-step process: first, identifying the optimal bandwidth, and second, evaluating the option pricing model in the OOS period.

Step 1: Optimal Bandwidth Selection

To determine the optimal bandwidth parameter for each state variable, we further subdivide the estimation sample into two parts: a 'training sample' (the first half) used to estimate model parameters across various bandwidths, and a 'validation sample' (the second half) used to identify the optimal bandwidth parameter. For each fixed bandwidth value[13], we estimate the model using the training sample of underlying stock returns data through equation (13). We then calculate option pricing errors in the validation sample of option data using these estimates and pricing formulas. The bandwidth value yielding the lowest option pricing errors in the validation sample is selected as optimal.

Step 2: OOS Model Evaluation

Given the optimal bandwidth parameter for each state variable, we re-estimate the model using the entire estimation sample of underlying return data (i.e., combined training and validation samples) via equation (13). Given these parameter estimates with pricing formulas, we then calculate option pricing errors using option price data from the OOS period. To assess the performance of various models in our application, we compare their OOS average losses. It is important to note that these OOS losses are unweighted, which ensures that the local estimator does not have any inherent advantage in the evaluation process. This approach provides a fair and unbiased comparison between local and non-local estimation methods.

[12]Look-ahead bias occurs when a model incorporates information that would not have been available at the time of prediction. This could lead to unrealistically accurate estimates by inadvertently using future data in parameter estimation.

[13]We test bandwidth values ranging from $0.05\sigma_s$ to $5\sigma_s$ with 0.05 increments for stochastic state variables, or from 0.98 to 0.9999 with 0.0005 increments for the time-based state variable.

For both validation and OOS evaluation, we first need to forecast option prices or implied volatilities. In this paper, we particularly rely on implied volatility. As a forecast measure for the i-day ahead implied volatility, we calculate the expected implied volatility by using parameter estimates at time t. To this end, we simulate the underlying price and variance for the next three days using their physical dynamics in equation (1) or (4), calculate option prices and implied volatilities over the joint distribution of the underlying price and variance, and then take their expectations for each time $t+i$.

To evaluate pricing errors, we focus on forecasting performance for the next three-day option implied volatilities.[14] Specifically, we evaluate option pricing performance by defining the the relative implied volatility error (IVR). Let IVR_t denote the mean squared relative errors of the next three-day option implied volatilities:

$$IVR_t = \frac{1}{3}\sum_{i=1}^{3}\frac{1}{N_t}\sum_{j=1}^{N_t}\left(\frac{IV_{t+i,j}^{data} - E_t\left[IV_{t+i,j}^{model}\right]}{IV_{t+i,j}^{data}}\right)^2, \tag{16}$$

where $IV_{t+i,j}^{data}$ is the Black-Scholes implied volatility for option j at time $t+i$ in data, and $E_t\left[IV_{t+i,j}^{model}\right]$ is the time-t expected Black-Scholes implied volatility obtained from the simulation discussed above. N_t represents the number of options included in the evaluation at time t. We then define IVR as the time series average of IVR_t over an evaluation period, that is,

$$IVR = \frac{1}{T}\sum_{t=1}^{T} IVR_t. \tag{17}$$

We assess that a model is superior to another model when the former IVR is statistically smaller than the latter's.

Having established our methodology for local estimation and bandwidth determination, we now introduce the set of state variables considered to completely implement the local estimation. We consider six stochastic state variables that are well-established in the literature as informative and useful for pricing options or explaining option returns. These include two measures of volatility:

[14] We evaluate the performance every week. Excluding the evaluation date and a holiday if exists, we forecast the next three-day option implied volatilities.

the VIX and the 5-minute realized volatility (RV5) of the S&P 500 index. Furthermore, we use the VVIX, which captures volatility-of-volatility by measuring the expected volatility of the 30-day forward price of the VIX.[15] As tail measures, we incorporate model-free risk-neutral skewness and kurtosis based on Bakshi, Kapadia, and Madan (2003). We also include the variance risk premium (VRP), a measure of risk compensation that investors and policymakers use to assess sentiment regarding uncertainty. In addition to these six variables associated with a kernel function equation (14), we consider time as a state variable associated with a kernel function in equation (15). This results in a total of seven possible state variables.[16]

3 Data and Empirical Setup

In both the validation and out-of-sample (OOS) evaluation, we use daily S&P 500 call and put options with maturities between 14 and 183 days. The total sample period for the options data is from January 2011 to August 2023. We obtain the data from OptionMetrics and apply the following initial filters: discarding (i) options with implied volatility below 5% or above 150%, (ii) options with volume or open interest of fewer than ten contracts, (iii) options with mid price below \$0.50 or a bid price below \$0.375 to avoid low-valued options, and (iv) options with data errors, where the bid price exceeds the offer price or where a negative price is implied by put-call parity.

Under our empirical setup, where we evaluate pricing performances for subsequent days of a week, we select one day per week to determine a weekly set of options. Following existing studies (see, e.g., Heston and Nandi, 2000; Christoffersen, Heston, and Jacobs, 2013), we choose Wednesday because it is the day of the week least likely to be a holiday. It is also less likely than other days to be affected by day-of-the-week effects. Every Wednesday, we apply additional filters by retaining only out-of-the-money (OTM) call and put options and excluding those with moneyness below 0.75 or above 1.25.

To enhance computational efficiency, instead of using all available data points, we select representative options that adequately span the implied volatility surface each Wednesday. We follow

[15] This forward price represents the price of a hypothetical VIX futures contract expiring in 30 days.

[16] The use of these state variables are supported by extant papers. For instance, Bali, Beckmeyer, Moerke, and Weigert (2023), Cao, Vasquez, Xiao, and Zhan (2023), and Horenstein, Vasquez, and Xiao (2025) highlight the importance of volatility measures, including implied volatility, realized volatility, volatility of volatility, and variance risk premium; Kelly and Jiang (2014), tail risk. Oh and Patton (2024) also bolster the significance of volatility and time state variables in forecasting volatility.

the methodology of Dufays, Jacobs, Liu, and Rombouts (2023) to construct a balanced dataset. Specifically, we create eight buckets for moneyness (S/K) and five buckets for maturity.[17] This results in 40 buckets, and we keep the most actively traded option in each bucket, although not all buckets are populated every week. Comparing the filtered set of Wednesday options with those available throughout the following week, we retain only the common options for each week from Wednesday. The final filtered dataset comprises 64,970 option contracts. Panels A and B of Table 1 provide descriptive statistics of the filtered option dataset.

[Table 1 AROUND HERE]

We obtain S&P 500 index returns from CRSP. The sample period of the return data is from January 1996 to August 2023. We use a longer return series than options data to better capture the time series dynamics of the underlying index's variance. While extracting the underlying variance begins in 1996, our likelihood-based estimation sample starts in 2007, as all local estimation state variables are available from that year onward. We obtain the VIX and VVIX from Bloomberg. We calculate RV5 as the sum of open-to-close squared five-minute returns. The VRP is calculated as the difference between the forward-looking market variance implied by option prices (VIX^2) and the annualized one-month forecast of realized variance over time (Bollerslev, Tauchen, and Zhou, 2009; Carr and Wu, 2009). We also compute model-free risk-neutral skewness (SKEW) and kurtosis (KURT) following Bakshi, Kapadia, and Madan (2003). Panel C of Table 1 reports descriptive statistics of these six state variables.

The time series for the risk-free rate is proxied by the one-month Treasury Bill rates obtained from CRSP. Following the standard implementation in the literature, options are valued using a maturity-specific risk-free rate obtained from OptionMetrics.

[Table 2 AROUND HERE]

We conduct model estimation and evaluation across multiple periods using both expanding and rolling window approaches, as summarized in Table 2. In the expanding window approach, to determine an optimal bandwidth parameter, we initially estimate our models using data from

[17] In the moneyness dimension, we define buckets with moneyness levels spaced every 0.03 between 0.91 and 1.09, and assign two additional buckets for options with moneyness below 0.91 or above 1.09. In the maturity dimension, we define buckets with maturities of 30 days or less, 31–60 days, 61–90 days, 91–121 days, and more than 122 days.

January 2007 to December 2010 (training sample) and validate them using the subsequent period from January 2011 to December 2014 (validation sample). After identifying the optimal bandwidth parameter from this estimation sample, we conduct out-of-sample (OOS) evaluation for the period of January 2015 to December 2016. Our validation sample and OOS evaluation procedure employs an expanding window approach. We estimate the model at time t and evaluate it using calculated option prices from $t+1$ to $t+3$. We then expand the window to include $t+1$ data to estimate the model at $t+1$ and repeat the evaluation using $t+2$ to $t+4$ data. This method ensures our model is continuously updated with the most recent data. As illustrated in Table 2, we progressively expand our sample periods. For each subsequent period (P2, P3, and P4), we extend both the training and validation samples by one year and evaluate the next two years out-of-sample. This progressive expansion enables us to maintain the most appropriate bandwidth parameter that reflects recent market conditions. By regularly updating our estimation window, we can potentially achieve better results for both local and non-local estimations, as the model adapts to evolving market dynamics. It is important to note that for each subsequent period (P2, P3, and P4), we use a different bandwidth parameter, optimally chosen in the respective estimation sample. This approach ensures that our model remains responsive to changing market conditions across different time frames.

In the rolling window approach, we fix the window size at 1,512 observations (approximately six years) for validation sample evaluation, and 2,016 observations (approximately eight years) for OOS evaluation. As the window advances one week at a time, we re-estimate the model using the most recent 1,512 or 2,016 data points, depending on the evaluation stage, and assess its performance over the subsequent three days. This procedure mirrors the evaluation horizon used in the expanding window setup but preserves a constant sample size throughout. By relying on a fixed-length, moving estimation sample, the rolling window approach offers a complementary perspective on model performance under evolving market conditions.

4 Empirical Results

We report the out-of-sample (OOS) performance of the GARCH and SV models, specifically comparing several local estimation models with the non-local estimation approach. We then compare the best-performing local specification with the non-local model across various dimensions and

examine their associated economic implications. Finally, we assess the robustness of the local estimation model's outperformance by examining alternative OOS performance metrics.

4.1 Out-of-sample option price forecasting performance

Panel A of Table 3 provides a detailed evaluation of the pricing performance of Heston-Nandi GARCH option pricing models during the OOS period, spanning from January 2015 to August 2023, inferred from the expanding window estimation. These models are estimated using two distinct approaches: the non-local quasi-maximum likelihood (QML) method by Heston and Nandi (2000), denoted by ⧫, and the local QML method described in Section 2.3. The table organizes the rows based on the average OOS loss, IVR, as defined in equation (17). Each local QML estimator employs a specific set of state variables and bandwidth parameters, which are outlined in the second and third columns, respectively. As described in Section 2.4, these bandwidth parameters are optimized based on the performance measured by the same loss function IVR in the validation sample.

[Table 3 AROUND HERE]

To further assess the relative performance of the models, the table reports Diebold-Mariano (DM) t-statistics based on Newey-West standard errors with five lags, following Diebold and Mariano (1995), which compare each model's average OOS loss against that of the benchmark non-local QML method (denoted by ⧫). Negative t-statistics indicate that a given model achieves a lower average loss compared to the benchmark. The final column provides an additional layer of evaluation by indicating, with a checkmark, whether a given model is included in the 95% model confidence set (MCS)—a procedure proposed by Hansen, Lunde, and Nason (2011).[18]

The benchmark method, which relies on non-local QML, performs the seventh among the eight estimation methods considered. Four local methods demonstrate significantly lower OOS losses compared to the benchmark, as evidenced by the DM test results, with their DM t-statistics falling below -2. The local method that achieves the best performance in the validation sample uses VIX as a state variable and also exhibits the lowest average loss during the OOS period. A direct

[18]To construct the MCS, we implement the stationary bootstrap, adopting an average block duration of ten observations.

comparison between the benchmark method and the best-performing local method selected from the validation sample yields a DM t-statistic of -5.85, providing strong evidence that the local approach outperforms the non-local benchmark. Among all estimation methods evaluated, the local methods utilizing VIX and time as state variables are included in the model confidence set.

Panel B of Table 3 evaluates the pricing performance of SV option pricing models on the same dataset, inferred from the expanding window estimation. The benchmark method, which relies on non-local QML, ranks the seventh among the eight estimation methods. All local methods except the one with VVIX show significantly lower OOS losses compared to the benchmark, as indicated by the DM test results, with all DM t-statistics below -2. The local method that performs best in the validation sample incorporates time as a state variable and also exhibits the lowest average loss during the OOS period. Across the entire set of estimation methods, the local methods utilizing time, RV5, and VIX as state variables are included in the model confidence set.

The two option pricing model classes, the GARCH and SV models, do not necessarily have to have the same state variable for the best-performing local estimation specification because one model can capture a certain characteristic better than the other, and vice versa. As a result, we find that the optimal state variables for the GARCH and SV option pricing models are VIX and time, respectively. However, within the estimation specifications of the GARCH option pricing model, the time state variable is also included in the model confidence set, implying that the importance of the local time information is not statistically different from that of VIX. Similarly, within the SV model specifications, VIX is also included in the model confidence set.

Additionally, Table 4 reports the OOS option pricing performance based on the rolling window estimation. Panel A presents results for the Heston-Nandi GARCH models, and Panel B for the Heston SV models. Interestingly, the MCS of each model class includes the same state variables as those identified using the expanding window approach. Although the most optimal state variable for the GARCH model under the rolling window is time, whereas VIX is the most optimal under the expanding window, both time and VIX are included in the MCS under both approaches. For the SV model, the state variables, time, RV5, and VIX, consistently constitute the MCS, with their rankings remaining unchanged. Similar to the expanding window approach, the non-local benchmark is ranked very low, namely seventh or eighth, with significantly higher OOS average losses compared to those from local estimation methods. Regardless of whether the expanding or

rolling window is used, the overall OOS pricing performances are highly consistent across the two approaches.

[Table 4 AROUND HERE]

4.2 Model performance comparison and economic implications

We further investigate when, where, and why a local estimation outperforms the non-local estimation. Specifically, we compare the time series and cross-sectional pricing errors of different specifications to see when and where performance improvement is pronounced. We also discuss how parameter estimates vary over time. Furthermore, simulated variance and return distributions help us better understand the differences between the local and non-local estimation performances. Lastly, since we estimate physical dynamics of the stock price and variance but evaluate the model using option prices, we are able to exploit the link between the physical distributions and the risk-neutral distributions for an external validity test. Specifically, we employ expected option returns that measure the ratio between the physical expectation and the risk-neutral expectation of future payoffs. For brevity, in the following analyses, we focus on comparing the results between the non-local benchmark and the best-performing local estimation in either the GARCH or SV model.

4.2.1 Time series and cross sectional pricing errors

Figure 1 presents the time series of out-of-sample (OOS) pricing errors. Panels A and C depict the errors under the GARCH models, based on the expanding window and rolling window approaches, respectively. Similarly, Panels B and D show the errors under the SV models, using the expanding window and rolling window approaches. As discussed in Tables 3 and 4, the local QML estimation with VIX as a state variable demonstrates the best performance for the expanding-window GARCH model in terms of the average loss (IVR) out of sample. For the other three models, the local estimation with time as a state variable performs best. The local estimation in all four panels generally outperforms the non-local benchmark during the pre-2019 period, while the difference is less pronounced during the post-2019 period, on average. Nevertheless, we still observe a statistically significant difference between the local and non-local approaches for the rolling-window GARCH model during the post-2019 period. Across all panels, the biggest difference occurs around

2017, when the market was relatively calm.

[Figure 1 AROUND HERE]

Comparing the pricing errors between Panels A and C or between Panels B and D reveals that the time series dynamics of the non-local pricing errors are more dependent on the estimation data window than those of the local pricing errors. For instance, the non-local pricing errors during 2017 and 2018 decrease substantially when the estimation scheme changes from the expanding window to the rolling window, whereas the local pricing errors during the same period change little. This observation highlights a methodological advantage of local estimation. Unlike non-local estimation, which assigns equal weights to every data point, local estimation places more weight on states similar to current conditions, making it less dependent on the estimation sample size or period.

In addition to comparing the time-series errors, we also examine the cross-sectional option pricing errors. Panels A and B of Table 5 present the OOS pricing errors from the expanding-window GARCH model across moneyness and maturity buckets. Panel A shows the results from the local estimation using VIX, while Panel B displays the results from the non-local estimation. For deep OTM put options ($S/K > 1.06$), the non-local estimation error is slightly smaller, but for all other buckets, the local estimation error is significantly smaller. Furthermore, the non-local estimation exhibits larger variation in pricing errors across the moneyness dimension, whereas the local estimation shows much less variation overall. In other words, non-local estimation tends to focus more on reducing errors for OTM puts, while local estimation results in a more balanced allocation of importance across the cross-section. Ultimately, the use of local information incurs a slight performance cost for deep OTM puts but maintains small errors for those options while reducing overall error and improving performance.

[Table 5 AROUND HERE]

Panels C and D of Table 5 present the cross-sectional option pricing errors for the expanding-window SV model. As shown in Table 3, the performance gap between the optimal local estimation and the non-local estimation is smaller in the SV model than in the GARCH model. The cross-sectional pricing errors follow the same pattern: the estimation using time-local information

statistically outperforms the non-local estimation overall, but the magnitude of improvement is smaller than in the GARCH model. Interestingly, unlike the GARCH model, the variation in errors under non-local estimation is more pronounced along the maturity dimension than the moneyness dimension for the SV model. When local estimation is used, the variation in errors across different maturities is significantly reduced, leading to more balanced results. Taken together, the SV option pricing model is already more improved compared to the GARCH model, and thus the marginal improvement from using local information is relatively smaller. Nevertheless, we still observe statistical improvement in the SV model as well.

Table 6 repeats the cross-sectional pricing error analysis for the rolling-window GARCH and SV models. The rolling window approach produces highly consistent cross-sectional pricing error patterns with those from the expanding window approach, with one exception: the local GARCH model outperforms the non-local GARCH model for deep OTM put options. We again observe that local estimation leads to more balanced pricing errors across option maturity and moneyness compared to non-local estimation. Given the overall similarities between the results from the expanding window and rolling window approaches, we focus on further analyses based on the expanding window approach.

[Table 6 AROUND HERE]

4.2.2 Time series of parameter estimates

Panels A through D of Figure 2 depict the time series of the parameter estimates of ω, β, α, and γ, in order under the GARCH model. To get a better sense of economic intuition, we also report the variance persistence measured by $\beta + \alpha\gamma^2$ and the long-run variance measured by $(\omega + \alpha)/(1 - \beta - \alpha\gamma^2)$ in Panels E and F, respectively. The parameter estimates from the local estimation are highly time-varying, and they sometimes far deviate from the non-local estimates.[19] For example, around 2017, which is a relatively calm period, the local estimate of the reactivity parameter α is approximately half of its non-local estimate, suggesting that variance is locally less reactive to the return shock during such a period. On the other hand, since the COVID-19 shock in 2020, we observe many data points where α is more than double of its non-local estimate, indicating that

[19] Note that the non-local parameter estimates are also slightly time-varying because we estimate them every week using the expanding window scheme.

variance is locally more reactive to the return shock.

[Figure 2 AROUND HERE]

Furthermore, variance persistence implied by the non-local estimation is set at between 0.95 and 0.98 over the course of the OOS period. In contrast, variance persistence implied by the local estimation ranges from 0.8 to 0.98, especially with low persistence during a calm period in 2017. This result indicates that variance tends to revert to its mean level relatively quickly during a period with low conditional variance, while it reverts to its mean level relatively slowly during a period with high conditional variance. Interestingly, the persistence implied by the local estimation rarely exceeds that implied by the non-local estimation.

Figure 3 depicts the time series of parameter estimates under the SV model. As in the GARCH model, we also observe that the local parameter estimates are time-varying and highly volatile. First, κ represents the variance mean reversion speed, and $1 - \kappa/252$ corresponds to the variance's daily persistence. In Panel A, we see that the local estimate of κ is almost always higher than the non-local estimate, which implies that the variance persistence implied by the local estimation is lower than that implied by the non-local estimation—consistent with the results from the GARCH model estimation. One difference is that while the persistence was lowest in 2017 under the GARCH model, it is lowest during 2018–2019 under the SV model. Looking at θ, which corresponds to the long-run variance, the local estimate is consistently lower than the non-local estimate before the 2020 COVID-19 shock, but shows a sharp spike at the onset of COVID-19 and thereafter stays above or at a similar level to the non-local estimate. Compared to the GARCH model estimation results, one interesting point is that the locally estimated long-run variance in the GARCH model follows market volatility patterns like the VIX, while this pattern is relatively less pronounced in the SV model.

[Figure 3 AROUND HERE]

Most of the SV model's local parameter estimates exhibit two jumps—one during the 2018 Volmageddon[20] and another during the 2020 COVID-19 shock. Looking at the correlation parameter

[20]It refers to the sudden spike in volatility on February 5, 2018, which led to large losses in inverse volatility products.

ρ in Panel D, we see that the negative correlation between return and variance shocks changes significantly at these two events, with the magnitude of the correlation decreasing.[21] Although not as large as in the SV model results, in fact, Panel D of Figure 2 shows that the leverage parameter γ in the GARCH model also exhibits a negative jump during the same periods, which likewise suggests a reduction in the size of the negative correlation.

4.2.3 Simulated return and variance distributions

We have confirmed that both the GARCH and SV models improve their OOS pricing performances in both time series and cross section when local information is incorporated in estimation. Then, why does model performance improve when local information is used? To investigate this, we simulate returns and variances using each model's estimates and examine their forward risk-neutral distributions. We select two representative days from our sample period—one with low conditional volatility and one with high. The year 2017 is a low-volatility period, and during this time, the difference between local and non-local estimation errors appears consistently, making it a reasonable period for this analysis. Additionally, the peak of the COVID-19 crisis is undoubtedly a high-volatility period. As a result, we select July 12, 2017, and March 18, 2020, as the respective low- and high-volatility days for analysis.

For each model and each of the selected dates, we take the parameter estimates and run simulations to obtain the τ-horizon forward risk-neutral distribution.[22] To assess how well these model-implied distributions match the actual distribution, we compare them against the nonparametric risk-neutral distribution of Breeden and Litzenberger (1978). Since we evaluate models out-of-sample by checking whether the model estimated today can price options accurately over the next three days, we compare the model-implied distributions with the data-driven nonparametric risk-neutral return distribution corresponding to the day after each of the two representative dates.

Figures 4 and 5 show the one-month forward risk-neutral distributions from the GARCH and SV models, respectively.[23] Panels A and B describe the return and variance distributions for the low

[21] This observation is in line with Pyun (2019).

[22] We run 50,000 simulation draws at a daily frequency.

[23] The overall implications remain consistent for other option maturities or horizons; we report the three-month distributions in Figures A.1 and A.2 in Appendix.

conditional volatility case, while Panels C and D correspond to the high conditional volatility case. In all panels, we show both the optimal local and non-local estimation results, and for the return distributions, we also display the nonparametric distribution for comparison. Since our analysis focuses only on OTM options, the portion of the distribution with negative returns (Return < 1) corresponds to the state prices used to compute OTM put option prices, while the positive return region (Return > 1) corresponds to that used for OTM call options.

[Figure 4 AROUND HERE]

Looking at Panels A and C of Figure 4, we can see that in the positive return region, the return distribution derived from the local estimation using VIX is much closer to the nonparametric distribution than that from the non-local estimation. This is consistent with the results from Table 5, where the difference in OTM call pricing errors between local and non-local estimations is substantial. Furthermore, in Panel C, the return distribution inferred from the local estimation is also closer to the nonparametric distribution in the negative return region, while in Panel A, the results are more mixed or similar. This partially aligns with the earlier observation that local estimation incurs some performance cost for deep OTM puts in the cross-sectional error analysis.

[Figure 5 AROUND HERE]

Moreover, Figure 5 shows that in the SV model, the results are similar under low conditional volatility, but under high volatility, the differences between the distributions inferred from the local and non-local estimations become much smaller. Indeed, Figure 3 shows that during the COVID-19 period, the difference in errors is quite minimal. This suggests that the SV model itself is already potentially more beneficial for capturing return and volatility dynamics during high-volatility periods.

Taking a step further, why do the return distributions differ between local and non-local estimation? We can find a clue by looking at the variance distributions in Panels B and D. In the low conditional volatility case, the variance distribution from the local estimation is narrower and more concentrated around lower variances. As a result, the return distribution from the local estimation also becomes narrower and closer to the nonparametric distribution than that from the non-local estimation. In contrast, in the high conditional volatility case, the variance distribution from the

local estimation is much wider and more concentrated around higher variances. This makes the return distribution from the local estimation also wider and closer to the nonparametric distribution compared to the non-local estimation. The reason why the variance distributions change in this way is that as seen in Figure 2, the parameter estimates move appropriately in response to local information. For example, during low-volatility periods, both the long-run variance and persistence estimated by the local method are lower than those from the non-local estimation, resulting in patterns like Panel B. During high-volatility periods, the opposite pattern appears.

4.2.4 Expected option returns

So far, we have explored various ways in which local estimation outperforms non-local estimation, focusing on how well each approach explains out-of-sample (OOS) option prices. A noteworthy aspect of our estimation framework is that the model parameters are estimated solely using return data to capture the physical dynamics, while the implied risk-neutral dynamics are derived under a power utility assumption to compute option prices. In the end, our validation and OOS performance assessments have concentrated on how well the risk-neutral dynamics align with actual outcomes or how accurately they explain future option prices. However, we have not yet examined the implications for the physical dynamics used in the estimation process.

Therefore, we now aim to investigate whether both the physical and risk-neutral dynamics jointly contribute to performance improvement. To do this, we evaluate how well the local and non-local estimation methods explain option returns. Specifically, we compare the holding-to-expiration realized option returns with the model-implied expected option returns. Since we focus exclusively on index options, which are European-style, all payoffs are determined at maturity without early exercise. The option payoff is driven by the comparison between the index spot price at maturity and the strike price, and thus is determined by the physical dynamics. In contrast, the current option price is determined under risk-neutral dynamics. Consequently, the option return reflects information from both the physical and risk-neutral dynamics as of the current date.

We compute the realized option return as follows:

$$Ret_{t,T}^{call} = \frac{(S_T - K)^+}{E_t^* \left[e^{-r(T-t)} (S_T - K)^+ \right]}, \quad \text{and} \quad Ret_{t,T}^{put} = \frac{(K - S_T)^+}{E_t^* \left[e^{-r(T-t)} (K - S_T)^+ \right]}, \tag{18}$$

and the expected option return is calculated as:

$$ERet_{t,T}^{call} = \frac{E_t\left[(S_T - K)^+\right]}{E_t^*\left[e^{-r(T-t)}(S_T - K)^+\right]}, \quad \text{and} \quad ERet_{t,T}^{put} = \frac{E_t\left[(K - S_T)^+\right]}{E_t^*\left[e^{-r(T-t)}(K - S_T)^+\right]}. \tag{19}$$

In the model-implied expected option return, the numerator is the expectation under the physical measure. This can be derived in closed form using the same characteristic function employed for option pricing but using the physical dynamics instead of the risk-neutral dynamics.

To ensure a more consistent comparison – given that characteristics such as maturity and moneyness vary daily in the market data – we utilize the Volatility Surface data from OptionMetrics. This dataset provides interpolated market option prices on a uniform grid of Black-Scholes Deltas and maturities. We conduct our analysis across 1-month and 3-month horizons. The sample period matches our OOS window, and we consistently use option data from Wednesdays only to compute both realized and expected option returns. We then compare their time-series averages. Figure 6 presents the results.

[Figure 6 AROUND HERE]

We find that the expected option returns computed from the local estimation consistently align more closely with the realized option returns than those from the non-local estimation. This pattern holds across all panels, regardless of the underlying option pricing model (GARCH or SV) or maturity. When comparing local- and non-local-implied option returns, we observe that in-the-money options (roughly $|\text{Delta}| > 50$), whether calls (positive Delta) or puts (negative Delta), show relatively similar results between the two methods. However, as we move toward out-of-the-money options ($|\text{Delta}| < 50$), the divergence between the two expected returns grows in both calls and puts, with the local-implied return aligning more closely with the realized return. This pattern becomes more pronounced as the maturity increases.

This finding connects benefits of the optimal physical dynamics obtained through the local estimation approach with the improved OOS option pricing performance. Taken together, they demonstrate that the local estimation provides superior economic implications in terms of option returns compared to the non-local estimation.

4.3 Robustness to choice of performance measures

In this section, we investigate whether local estimation continues to robustly outperform non-local estimation even when alternative performance evaluation measures are used.

4.3.1 Performance measure: IVR given future stock prices and variances

First, we assess the robustness of our main results by calculating future option prices and the corresponding implied volatilities under the assumption that future stock prices and variances are known and given. In the main analysis, we computed the expected value of implied volatility without access to any future information at the time of evaluation. However, forecasting option prices inherently involves two components: one related to predicting the option pricing model or its parameters, and the other related to the future evolution of the underlying prices and variances. Since we forecast implied volatilities directly without explicitly forecasting the underlying variables better, it is difficult to clearly disentangle the sources of uncertainty associated with each component. To address this concern, we conduct an additional analysis in which future option values are computed based on ex-post realized future stock prices and variances, and examine whether the results remain robust.

Table 7 presents the results. Overall, the findings are highly consistent with our main out-of-sample performance results reported in Table 3. For the GARCH and SV models, local estimation using the VIX and time as the state variable, respectively, continues to deliver the best performance in both the validation and out-of-sample periods. One notable difference is that, while the MCS set under the SV model includes time, RV5, and VIX in Table 3, only time and RV5 are included in Table 7. Nevertheless, VIX still ranks third and yields highly negative DM statistics. Looking at average losses, we find that the GARCH models show a reduction of approximately 0.001 to 0.005 across different specifications compared to the main results. For the SV models, the reduction ranges from about 0.002 to 0.005. While these improvements reflect the benefit of knowing future underlying prices, the relative performance of different estimation methods remains largely unchanged. This suggests that the ranking of estimation methods is robust to this assumption—that is, whether or not future stock prices and variances are known, the local estimation method continues to outperform the non-local approach.

[Table 7 AROUND HERE]

4.3.2 Performance measure: variance of delta-hedged portfolio return

Second, we test the robustness of our approach using an alternative performance measure in place of the relative implied volatility error, which we used in the main analysis. As an alternative measure, we consider the variance of returns from a delta-hedged portfolio.[24] Delta hedging is commonly used in option trading to reduce or neutralize directional risk. However, as emphasized by Hull and White (2017), reducing the variance of the portfolio's return may be more important than simply achieving zero Delta, as this accounts for risk from both price changes and expected changes in volatility. Furthermore, while in the Black and Scholes (1973) model the delta-hedged portfolio has zero Delta, this is no longer the case under models such as the Heston-Nandi GARCH or the Heston SV model. Therefore, rather than simply analyzing the returns of the hedged portfolio, we use the variance of those returns as our performance measure to test the effectiveness of the estimation approach.

Table 8 presents the results. Looking at Panel A, which shows the performance of the Heston-Nandi GARCH model, we find that using time as the local information variable yields the best performance in both the validation and OOS samples under this new performance measure. However, the results from using VIX as the local variable are also very close to those of the time specification. VIX is included in the MCS set and significantly outperforms the non-local estimation in a statistically meaningful way. Compared to the main results in Panel A of Table 3, although the rankings have shifted slightly, it is strongly consistent that VIX and time remain the two best-performing state variables despite the change in performance measure.

Panel B shows the results for the Heston SV model using the new performance measure. Here, time performs best in the validation sample, whereas in the OOS sample, VIX leads, followed by RV5 and time, in that order. Nevertheless, the differences in OOS average loss are small, all three state variables are included in the MCS set, and they statistically outperform the non-local estimation. Comparing this with Panel B of Table 3, where IVR is used as the performance measure,

[24] The delta-hedged call return calculation follows Almeida, Freire, and Hizmeri (2025):

$$\frac{\max(S(t+\tau)-K,0)-Call(t,\tau,K)-\Delta(S(t+\tau)-S(t))-(Call(t,\tau,K)-\Delta S(t))r\tau}{S(t)}.$$

we again find that the three state variables consist of the MCS set.

In sum, we evaluate 1- to 3-day-ahead option forecast performance using two different option pricing models and two different performance measures. Across all cases, a local estimation consistently and statistically outperforms the non-local estimation, demonstrating the effectiveness of the local approach for option forecasting. It is also encouraging to see that, despite using different performance metrics, the best-performing state variables or those included in the MCS set remain virtually unchanged.

[Table 8 AROUND HERE]

5 Conclusion

This paper develops and empirically evaluates a local M-estimation framework that enhances out-of-sample option pricing accuracy by incorporating state-dependent information into standard pricing models. Rather than modifying the model structure, the framework reweights historical observations based on their relevance to current market conditions using kernel-based techniques. We implement this approach for both GARCH and stochastic volatility (SV) models, conditioning on a range of variables including VIX, realized volatility (RV5), variance risk premium (VRP), VVIX, skewness, kurtosis, and calendar time.

Our empirical analysis, based on data from 2015 to 2023, shows that the local estimation method consistently outperforms traditional non-local approaches in forecasting near-term option prices. These improvements are statistically significant in both time-series and cross-sectional dimensions. For the GARCH model, VIX- or time-based local estimation achieves the lowest implied volatility forecast errors and is included in the model confidence set. For the SV model, time-based local estimation performs best, with RV5 and VIX also yielding strong results. The gains are especially pronounced in calm periods around 2017, where local parameter estimates differ substantially from their non-local counterparts. In addition, local estimation reduces pricing error dispersion across moneyness and maturities, resulting in more balanced cross-sectional performance.

Furthermore, local estimation produces risk-neutral return distributions more closely aligned with nonparametric benchmarks, particularly in low-volatility environments. Model-implied expected option returns under local estimation more accurately match realized returns, reflecting

improved alignment between physical and risk-neutral dynamics. Lastly, robustness checks using alternative evaluation metrics—including forecasts given ex-post realized underlyings and delta-hedged return variances—corroborate the consistency and superiority of the local approach.

Overall, our findings underscore the practical value of adapting estimation methods to prevailing market conditions. Without increasing model complexity, the local estimation framework yields substantial gains in predictive accuracy, making it particularly well-suited for applications such as risk management, margin setting, and regulatory forecasting. Future research may extend this framework to incorporate jumps, alternative pricing kernels, or machine learning techniques for automated state variable selection and bandwidth tuning.

References

Aït-Sahalia, Yacine, Mustafa Karaman, and Loriano Mancini, 2020, The Term Structure of Equity and Variance Risk Premia, *Journal of Econometrics*.

Aït-Sahalia, Yacine, Chenxu Li, and Chen Xu Li, 2021, Closed-form implied volatility surfaces for stochastic volatility models with jumps, *Journal of Econometrics* 222, 364–392.

Almeida, Caio, Jianqing Fan, Gustavo Freire, and Francesca Tang, 2023, Can a machine correct option pricing models?, *Journal of Business & Economic Statistics* 41, 995–1009.

Almeida, Caio, Gustavo Freire, and Rodrigo Hizmeri, 2025, 0DTE asset pricing, Working paper, Princeton University.

Andersen, Torben G, Luca Benzoni, and Jesper Lund, 2002, An empirical investigation of continuous-time equity return models, *The Journal of Finance* 57, 1239–1284.

Andersen, Torben G., Nicola Fusari, and Viktor Todorov, 2015, Parametric Inference and Dynamic State Recovery from Option Panels, *Econometrica* 83, 1081–1145.

Babaoğlu, Kadir, Peter Christoffersen, Steven Heston, and Kris Jacobs, 2018, Option valuation with volatility components, fat tails, and nonmonotonic pricing kernels, *The Review of Asset Pricing Studies* 8, 183–231.

Bakshi, Gurdip, Charles Cao, and Zhiwu Chen, 1997, Empirical Performance of Alternative Option Pricing Models, *Journal of Finance* 52, 2003–2049.

Bakshi, Gurdip, Nikunj Kapadia, and Dilip Madan, 2003, Stock return characteristics, skew laws, and the differential pricing of individual equity options, *Review of Financial Studies* 16, 101–143.

Bali, Turan G, Heiner Beckmeyer, Mathis Moerke, and Florian Weigert, 2023, Option return predictability with machine learning and big data, *Review of Financial Studies* 36, 3548–3602.

Bardgett, Chris, Elise Gourier, and Markus Leippold, 2019, Inferring Volatility Dynamics and Risk Premia from the S&P 500 and VIX Markets, *Journal of Financial Economics* 131, 593–618.

Bates, David S., 2000, Post-'87 Crash Fears in the S&P 500 Futures Option Market, *Journal of Econometrics* 94, 181–238.

Bates, David S., 2006, Maximum Likelihood Estimation of Latent Affine Processes, *Review of Financial Studies* 19, 909–965.

Bates, David S., 2019, How Crashes Develop: Intradaily Volatility and Crash Evolution, *Journal of Finance* 74, 193–238.

Black, Fischer, and Myron Scholes, 1973, The Pricing of Options and Corporate Liabilities, *Journal of Political Economy* 81, 637–654.

Bollerslev, Tim, George Tauchen, and Hao Zhou, 2009, Expected Stock Returns and Variance Risk Premia, *Review of Financial Studies* 22, 4463–4492.

Breeden, Douglas T., and Robert H. Litzenberger, 1978, Prices of State-Contingent Claims Implicit in Option Prices, *Journal of Business* 51, 621–651.

Cao, Jie Jay, Aurelio Vasquez, Xiao Xiao, and Xintong Eunice Zhan, 2023, Why does volatility uncertainty predict equity option returns?, *Quarterly Journal of Finance* 13, 2350005.

Carr, Peter, and Dilip Madan, 1999, Option Valuation Using the Fast Fourier Transform, *Journal of Computational Finance* 2, 61–73.

Carr, Peter, and Liuren Wu, 2009, Variance risk premiums, *Review of Financial Studies* 22, 1311–1341.

Chen, Ying, Maria Grith, and Hannah LH Lai, 2025, Neural tangent kernel in implied volatility forecasting: a nonlinear functional autoregression approach, *Journal of Business & Economic Statistics* pp. 1–15.

Christoffersen, Peter, Steven Heston, and Kris Jacobs, 2009, The Shape and Term Structure of the Index Option Smirk: Why Multifactor Stochastic Volatility Models Work So Well, *Management Science* 55, 1914–1932.

Christoffersen, Peter, Steven Heston, and Kris Jacobs, 2013, Capturing Option Anomalies with a Variance-Dependent Pricing Kernel, *Review of Financial Studies* 26, 1963–2006.

Christoffersen, Peter, Kris Jacobs, and Karim Mimouni, 2010, Volatility Dynamics for the S&P500: Evidence from Realized Volatility, Daily Returns, and Option Prices, *Review of Financial Studies* 23, 3141–3189.

Christoffersen, Peter, Kris Jacobs, Chayawat Ornthanalai, and Yintian Wang, 2008, Option valuation with long-run and short-run volatility components, *Journal of Financial Economics* 90, 272–297.

Dendramis, Y., G. Kapetanios, and M. Marcellino, 2020, A Similarity-Based Approach for Macroeconomic Forecasting, *Journal of the Royal Statistical Society Series A: Statistics in Society* 183, 801–827.

Diebold, Francis X., and Roberto S. Mariano, 1995, Comparing Predictive Accuracy, *Journal of Business & Economic Statistics* 13, 253–263.

Du, Du, and Dan Luo, 2019, The pricing of jump propagation: Evidence from spot and options markets, *Management Science* 65, 2360–2387.

Dufays, Arnaud, Kris Jacobs, Yuguo Liu, and Jeroen Rombouts, 2023, Fast Filtering with Large Option Panels: Implications for Asset Pricing, *Journal of Financial and Quantitative Analysis* pp. 1–56.

Dufays, Arnaud, Kris Jacobs, and Jeroen Rombouts, 2025, A Framework for Real-Time Modeling and Forecasting of Large Unbalanced Option Implied Volatility Surfaces, Working paper, University of Houston.

Duffie, Darrell, Jun Pan, and Kenneth Singleton, 2000, Transform Analysis and Asset Pricing for Affine Jump-Diffusions, *Econometrica* 68, 1343–1376.

Eraker, Bjørn, 2004, Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices, *Journal of Finance* 59, 1367–1404.

Eraker, Bjørn, Michael Johannes, and Nicholas Polson, 2003, The Impact of Jumps in Volatility and Returns, *Journal of Finance* 58, 1269–1300.

Feller, William, 1951, Two Singular Diffusion Problems, *The Annals of Mathematics* 54, 173–182.

Ghysels, Eric, and Joanna Jasiak, 1994, Stochastic volatility and time deformation: an application of trading volume and leverage effects, *Cahier de recherche, Université de Montréal. Département de sciences économiques.*

Gruber, Peter H, Claudio Tebaldi, and Fabio Trojani, 2021, The price of the smile and variance risk premia, *Management Science* 67, 4056–4074.

Hansen, Peter R., Asger Lunde, and James M. Nason, 2011, The Model Confidence Set, *Econometrica* 79, 453–497.

Hansen, Peter Reinhard, and Chen Tong, 2025, Option Pricing with Time-Varying Volatility Risk Aversion, *Review of Financial Studies, Forthcoming.*

Heston, Steven, Kris Jacobs, and Hyung Joo Kim, 2023, The Pricing Kernel in Options, *Finance and Economics Discussion Series, Board of Governors of the Federal Reserve System* 2023-053.

Heston, Steven, Kris Jacobs, and Hyung Joo Kim, 2024, A new closed-form discrete-time option pricing model with stochastic volatility, Working paper, University of Maryland.

Heston, Steven L., 1993, A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options, *Review of Financial Studies* 6, 327–343.

Heston, Steven L., and Saikat Nandi, 2000, A Closed-Form GARCH Option Valuation Model, *Review of Financial Studies* 13, 585–625.

Horenstein, Alex R, Aurelio Vasquez, and Xiao Xiao, 2025, Common factors in equity option returns, Working paper, University of Cambridge.

Hull, John, and Alan White, 2017, Optimal delta hedging for options, *Journal of Banking & Finance* 82, 180–190.

Jones, Christopher S., 2003, The Dynamics of Stochastic Volatility: Evidence from Underlying and Options Markets, *Journal of Econometrics* 116, 181–224.

Kelly, Bryan, and Hao Jiang, 2014, Tail risk and asset prices, *The Review of Financial Studies* 27, 2841–2871.

Kim, Hyung Joo, 2025, Characterizing the conditional pricing kernel: A new approach, Working paper.

Newey, Whitney K., and Daniel McFadden, 1994, Chapter 36 Large sample estimation and hypothesis testing, in (Elsevier,).

Oh, Dong Hwan, and Andrew J. Patton, 2024, Better the devil you know: Improved forecasts from imperfect models, *Journal of Econometrics* 242, 105767.

Pan, Jun, 2002, The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study, *Journal of Financial Economics* 63, 3–50.

Pyun, Sungjune, 2019, Variance risk in aggregate stock returns and time-varying return predictability, *Journal of Financial Economics* 132, 150–174.

Ruiz, Esther, 1994, Quasi-maximum likelihood estimation of stochastic volatility models, *Journal of econometrics* 63, 289–306.

Schreindorfer, David, and Tobias Sichert, 2025, Conditional risk and the pricing kernel, *Journal of Financial Economics, Forthcoming*.

Zhan, Xintong, Bing Han, Jie Cao, and Qing Tong, 2022, Option return predictability, *The Review of Financial Studies* 35, 1394–1442.

Table 1: Summary statistics

Panel A: Option data by moneyness				
	Number of contracts	Average IV (%)	Average price	Average spread
$S/K \leq 0.91$	2,472	18.015	11.799	0.678
$0.91 < S/K \leq 0.94$	4,857	15.011	17.993	0.818
$0.94 < S/K \leq 0.97$	8,562	13.857	25.958	0.923
$0.97 < S/K \leq 1.00$	10,045	14.913	54.112	1.267
$1.00 < S/K \leq 1.03$	10,138	17.054	62.736	1.415
$1.03 < S/K \leq 1.06$	10,084	19.489	42.598	1.109
$1.06 < S/K \leq 1.09$	9,129	21.280	30.880	0.966
$S/K > 1.09$	9,683	25.867	21.640	0.836
All	64,970	18.471	37.546	1.058
Panel B: Option data by maturity				
	Number of contracts	Average IV (%)	Average price	Average spread
DTM ≤ 30	10,720	18.977	16.246	0.630
$30 <$ DTM ≤ 61	19,200	18.373	26.190	0.883
$61 <$ DTM ≤ 91	15,575	18.207	38.477	1.140
$91 <$ DTM ≤ 122	10,420	18.565	52.693	1.304
DTM > 122	9,055	18.422	67.814	1.510
Panel C: State variables				
	Mean	Standard deviation	Skewness	Kurtosis
VIX (%)	20.357	8.241	2.081	10.743
RV5	0.023	0.053	13.284	312.590
VRP	0.012	0.022	5.276	78.072
VVIX (%)	93.280	16.429	1.177	5.847
SKEW	-2.544	1.295	-1.411	4.817
KURT	21.091	20.733	2.272	8.361

Notes: Panels A and B present descriptive statistics for closing OTM option contracts for the January 2011 to August 2023 period. Panel C reports on the six state variables for the January 2007 to August 2023 period. VIX and VVIX are in percentage terms. Mean and standard deviation of the first four state variables (VIX, RV5, VRP, and VVIX) are annualized.
Source: Bloomberg Finance LP; OptionMetrics, Ivy DB US, via WRDS; and authors' calculations.

Table 2: Sample periods for estimation and out-of-sample evaluation

	Estimation sample				Out-of-sample (OOS)	
	Training sample		Validation sample			
	Start	End	Start	End	Start	End
Panel A: Expanding window						
P1	Jan 2007	Dec 2010	Jan 2011	Dec 2014	Jan 2015	Dec 2016
P2	Jan 2007	Dec 2011	Jan 2012	Dec 2016	Jan 2017	Dec 2018
P3	Jan 2007	Dec 2012	Jan 2013	Dec 2018	Jan 2019	Dec 2020
P4	Jan 2007	Dec 2013	Jan 2014	Dec 2020	Jan 2021	Aug 2023
Panel B: Rolling window						
P1	Jan 2007	Dec 2012	Jan 2013	Dec 2014	Jan 2015	Dec 2016
P2	Jan 2009	Dec 2014	Jan 2015	Dec 2016	Jan 2017	Dec 2018
P3	Jan 2011	Dec 2016	Jan 2017	Dec 2018	Jan 2019	Dec 2020
P4	Jan 2013	Dec 2018	Jan 2019	Dec 2020	Jan 2021	Aug 2023

Notes: This table outlines the sequential sample periods used for model estimation and out-of-sample (OOS) evaluation with an expanding window (Panel A) and a rolling window (Panel B). Each row in each panel corresponds to a forecasting period (P1 through P4), in which the model is first trained on historical data (training sample), then validated on a holdout window (validation sample) to select optimal bandwidths, and finally evaluated using an OOS test. The estimation sample includes both the training and validation periods. The validation and OOS periods involve weekly forecasts of three-day-ahead option prices and are strictly forward-looking, ensuring no look-ahead bias in model assessment.

Table 3: Out-of-sample option pricing performance: expanding window approach

Panel A: Heston-Nandi GARCH model								
Rank	State variable	Optimal bandwidth parameters				Out-of-sample performance		
		P1	P2	P3	P4	Avg. loss	DM stat	MCS
1*	VIX	1.51	1.31	1.16	1.16	0.0315	-5.85	✓
2	time	0.9945	0.9945	0.9945	0.9945	0.0315	-5.32	✓
3	VRP	0.50	0.50	0.50	0.65	0.0448	-3.70	
4	RV5	1.01	1.01	0.96	1.01	0.0473	-3.21	
5	BKMskew	0.65	0.55	0.60	0.60	0.0575	-1.00	
6	BKMkurt	0.70	0.45	0.50	0.70	0.0630	-0.09	
7	♦					0.0637		
8	VVIX	1.50	1.35	0.50	1.20	0.0659	0.38	
Panel B: Heston SV model								
Rank	State variable	Optimal bandwidth parameters				Out-of-sample performance		
		P1	P2	P3	P4	Avg. loss	DM stat	MCS
1*	time	0.9980	0.9975	0.9975	0.9975	0.0181	-4.82	✓
2	RV5	2.71	2.31	2.01	2.01	0.0192	-5.34	✓
3	VIX	4.51	3.36	2.86	2.16	0.0207	-5.22	✓
4	VRP	3.15	1.95	1.70	1.30	0.0226	-4.37	
5	BKMskew	1.45	1.30	0.95	0.85	0.0228	-4.77	
6	BKMkurt	1.85	1.60	0.80	0.85	0.0269	-2.78	
7	♦					0.0310		
8	VVIX	2.00	1.90	1.95	1.90	0.0315	0.36	

Notes: This table presents metrics evaluating pricing performance during the out-of-sample (OOS) period, January 2015 to August 2023, for Heston-Nandi GARCH option pricing models (Panel A) and Heston SV option pricing models (Panel B). These models are estimated using either the (non-local) quasi-maximum likelihood (QML) approach, indicated by ♦, or the local QML approach. In each panel, the rows are ordered by the average OOS loss, calculated as the relative implied volatility error in equation (17). The local method that demonstrates the best performance in the validation sample (the latter half of the estimation sample) is highlighted with an asterisk (*) in the first column. Local estimators utilize state variables listed in the second column and bandwidth parameters specified in the next four columns, which are optimized using the validation sample. All models are estimated using an expanding window. The penultimate column reports Diebold-Mariano t-statistics for each model compared to the benchmark non-local method (identified as ♦), where negative values suggest lower average loss. Finally, the last column marks models included in the 95% model confidence set with a checkmark.

Table 4: Out-of-sample option pricing performance: rolling window approach

Panel A: Heston-Nandi GARCH model								
Rank	State variable	Optimal bandwidth parameters				Out-of-sample performance		
		P1	P2	P3	P4	Avg. loss	DM stat	MCS
1*	time	0.9940	0.9970	0.9965	0.9940	0.0329	-6.57	✓
2	VIX	1.36	1.56	1.61	1.26	0.0384	-6.50	✓
3	RV5	0.96	1.56	1.06	1.11	0.0541	-3.79	
4	BKMkurt	0.30	0.65	1.05	2.25	0.0550	-3.99	
5	BKMskew	0.15	0.30	0.90	2.25	0.0557	-3.62	
6	VRP	0.40	0.60	0.80	1.95	0.0595	-2.03	
7	⧫					0.0726		
8	VVIX	0.25	1.85	0.65	1.40	0.0797	0.83	
Panel B: Heston SV model								
Rank	State variable	Optimal bandwidth parameters				Out-of-sample performance		
		P1	P2	P3	P4	Avg. loss	DM stat	MCS
1*	time	0.9960	0.9985	0.9985	0.9970	0.0187	-4.81	✓
2	RV5	2.06	4.71	2.91	1.81	0.0194	-5.84	✓
3	VIX	2.66	4.96	4.96	1.41	0.0199	-5.48	✓
4	VRP	2.00	5.00	3.25	1.85	0.0217	-3.58	
5	BKMskew	1.35	5.00	1.85	2.20	0.0221	-4.24	
6	BKMkurt	1.50	5.00	1.65	1.90	0.0229	-3.94	
7	VVIX	1.45	3.65	3.00	2.60	0.0233	-1.47	
8	⧫					0.0243		

Notes: This table presents metrics evaluating pricing performance during the out-of-sample (OOS) period, January 2015 to August 2023, for Heston-Nandi GARCH option pricing models (Panel A) and Heston SV option pricing models (Panel B). These models are estimated using either the (non-local) quasi-maximum likelihood (QML) approach, indicated by ⧫, or the local QML approach. In each panel, the rows are ordered by the average OOS loss, calculated as the relative implied volatility error in equation (17). The local method that demonstrates the best performance in the validation sample (the latter half of the estimation sample) is highlighted with an asterisk (*) in the first column. Local estimators utilize state variables listed in the second column and bandwidth parameters specified in the next four columns, which are optimized using the validation sample. All models are estimated using a rolling window. The penultimate column reports Diebold-Mariano t-statistics for each model compared to the benchmark non-local method (identified as ⧫), where negative values suggest lower average loss. Finally, the last column marks models included in the 95% model confidence set with a checkmark.

Table 5: Out-of-sample option pricing error by maturity and moneyness: expanding window approach

Heston-Nandi GARCH model				
Panel A: Local - VIX				
	$S/K \leq 0.94$	$0.94 < S/K \leq 1.00$	$1.00 < S/K \leq 1.06$	$S/K > 1.06$
$\tau \leq 30$	0.0456	0.0558	0.0204	0.0429
$30 < \tau \leq 61$	0.0330	0.0435	0.0163	0.0337
$61 < \tau \leq 91$	0.0330	0.0322	0.0149	0.0342
$91 < \tau \leq 122$	0.0387	0.0237	0.0158	0.0368
$\tau > 122$	0.0361	0.0194	0.0160	0.0405
Panel B: Nonlocal				
	$S/K \leq 0.94$	$0.94 < S/K \leq 1.00$	$1.00 < S/K \leq 1.06$	$S/K > 1.06$
$\tau \leq 30$	0.0727	0.1206	0.0410	0.0314
$30 < \tau \leq 61$	0.0507	0.1192	0.0435	0.0243
$61 < \tau \leq 91$	0.0639	0.1217	0.0445	0.0227
$91 < \tau \leq 122$	0.0828	0.1077	0.0402	0.0240
$\tau > 122$	0.0896	0.0880	0.0395	0.0253
Heston SV model				
Panel C: Local - time				
	$S/K \leq 0.94$	$0.94 < S/K \leq 1.00$	$1.00 < S/K \leq 1.06$	$S/K > 1.06$
$\tau \leq 30$	0.0359	0.0310	0.0138	0.0359
$30 < \tau \leq 61$	0.0251	0.0209	0.0096	0.0212
$61 < \tau \leq 91$	0.0207	0.0184	0.0089	0.0169
$91 < \tau \leq 122$	0.0222	0.0167	0.0097	0.0179
$\tau > 122$	0.0240	0.0165	0.0110	0.0189
Panel D: Nonlocal				
	$S/K \leq 0.94$	$0.94 < S/K \leq 1.00$	$1.00 < S/K \leq 1.06$	$S/K > 1.06$
$\tau \leq 30$	0.0304	0.0445	0.0113	0.0252
$30 < \tau \leq 61$	0.0272	0.0470	0.0138	0.0120
$61 < \tau \leq 91$	0.0364	0.0615	0.0183	0.0085
$91 < \tau \leq 122$	0.0543	0.0661	0.0199	0.0099
$\tau > 122$	0.0693	0.0690	0.0264	0.0118

Notes: This table presents the cross-sectional out-of-sample (OOS) pricing errors from Heston-Nandi GARCH option pricing models (Panels A and B) and Heston SV option pricing models (Panels C and D) across moneyness and maturity buckets. Panels A and C show the results from the local estimation using the VIX and time as a state variable, respectively, while Panels B and D show the result from the non-local estimation. All models are estimated using an expanding window.

Table 6: Out-of-sample option pricing error by maturity and moneyness: rolling window approach

Heston-Nandi GARCH model				
Panel A: Local - time				
	$S/K \leq 0.94$	$0.94 < S/K \leq 1.00$	$1.00 < S/K \leq 1.06$	$S/K > 1.06$
$\tau \leq 30$	0.0912	0.0608	0.0262	0.0463
$30 < \tau \leq 61$	0.0492	0.0439	0.0210	0.0347
$61 < \tau \leq 91$	0.0347	0.0335	0.0192	0.0335
$91 < \tau \leq 122$	0.0412	0.0265	0.0193	0.0364
$\tau > 122$	0.0244	0.0172	0.0196	0.0355
Panel B: Nonlocal				
	$S/K \leq 0.94$	$0.94 < S/K \leq 1.00$	$1.00 < S/K \leq 1.06$	$S/K > 1.06$
$\tau \leq 30$	0.1238	0.1233	0.0503	0.0461
$30 < \tau \leq 61$	0.0726	0.1213	0.0530	0.0434
$61 < \tau \leq 91$	0.0784	0.1223	0.0529	0.0450
$91 < \tau \leq 122$	0.1009	0.1029	0.0474	0.0463
$\tau > 122$	0.0878	0.0791	0.0453	0.0509
Heston SV model				
Panel C: Local - time				
	$S/K \leq 0.94$	$0.94 < S/K \leq 1.00$	$1.00 < S/K \leq 1.06$	$S/K > 1.06$
$\tau \leq 30$	0.0347	0.0288	0.0139	0.0368
$30 < \tau \leq 61$	0.0240	0.0199	0.0103	0.0224
$61 < \tau \leq 91$	0.0185	0.0187	0.0106	0.0186
$91 < \tau \leq 122$	0.0207	0.0184	0.0120	0.0200
$\tau > 122$	0.0232	0.0183	0.0137	0.0212
Panel D: Nonlocal				
	$S/K \leq 0.94$	$0.94 < S/K \leq 1.00$	$1.00 < S/K \leq 1.06$	$S/K > 1.06$
$\tau \leq 30$	0.0305	0.0340	0.0123	0.0340
$30 < \tau \leq 61$	0.0252	0.0281	0.0102	0.0198
$61 < \tau \leq 91$	0.0325	0.0341	0.0129	0.0164
$91 < \tau \leq 122$	0.0455	0.0374	0.0157	0.0181
$\tau > 122$	0.0510	0.0405	0.0196	0.0207

Notes: This table presents the cross-sectional out-of-sample (OOS) pricing errors from Heston-Nandi GARCH option pricing models (Panels A and B) and Heston SV option pricing models (Panels C and D) across moneyness and maturity buckets. Panels A and C show the results from the local estimation using time as a state variable, while Panels B and D show the result from the non-local estimation. All models are estimated using a rolling window.

Table 7: Out-of-sample pricing performance given future stock prices and variances

Panel A: Heston-Nandi GARCH model								
Rank	State variable	Optimal bandwidth parameters				Out-of-sample performance		
		P1	P2	P3	P4	Avg. loss	DM stat	MCS
1*	VIX	1.51	1.36	1.21	1.16	0.0300	-5.78	✓
2	time	0.9950	0.9950	0.9955	0.9955	0.0313	-4.92	✓
3	VRP	0.60	0.50	0.50	0.65	0.0427	-3.27	
4	RV5	1.16	1.06	0.96	1.06	0.0445	-3.12	
5	BKMskew	1.05	0.55	0.65	0.65	0.0553	-0.62	
6	⧫					0.0586		
7	BKMkurt	0.70	0.45	0.70	0.70	0.0589	0.05	
8	VVIX	1.90	1.90	0.50	1.20	0.0616	0.51	
Panel B: Heston SV model								
Rank	State variable	Optimal bandwidth parameters				Out-of-sample performance		
		P1	P2	P3	P4	Avg. loss	DM stat	MCS
1*	time	0.9980	0.9980	0.9975	0.9980	0.0158	-5.07	✓
2	RV5	2.81	2.46	2.16	2.16	0.0171	-5.42	✓
3	VIX	4.61	4.16	3.36	2.56	0.0186	-4.99	
4	BKMskew	1.45	1.25	1.00	0.90	0.0190	-5.16	
5	VRP	3.00	1.95	1.65	1.25	0.0206	-3.61	
6	BKMkurt	1.80	1.55	0.80	0.85	0.0228	-2.82	
7	⧫					0.0265		
8	VVIX	2.00	1.90	2.00	1.95	0.0269	0.37	

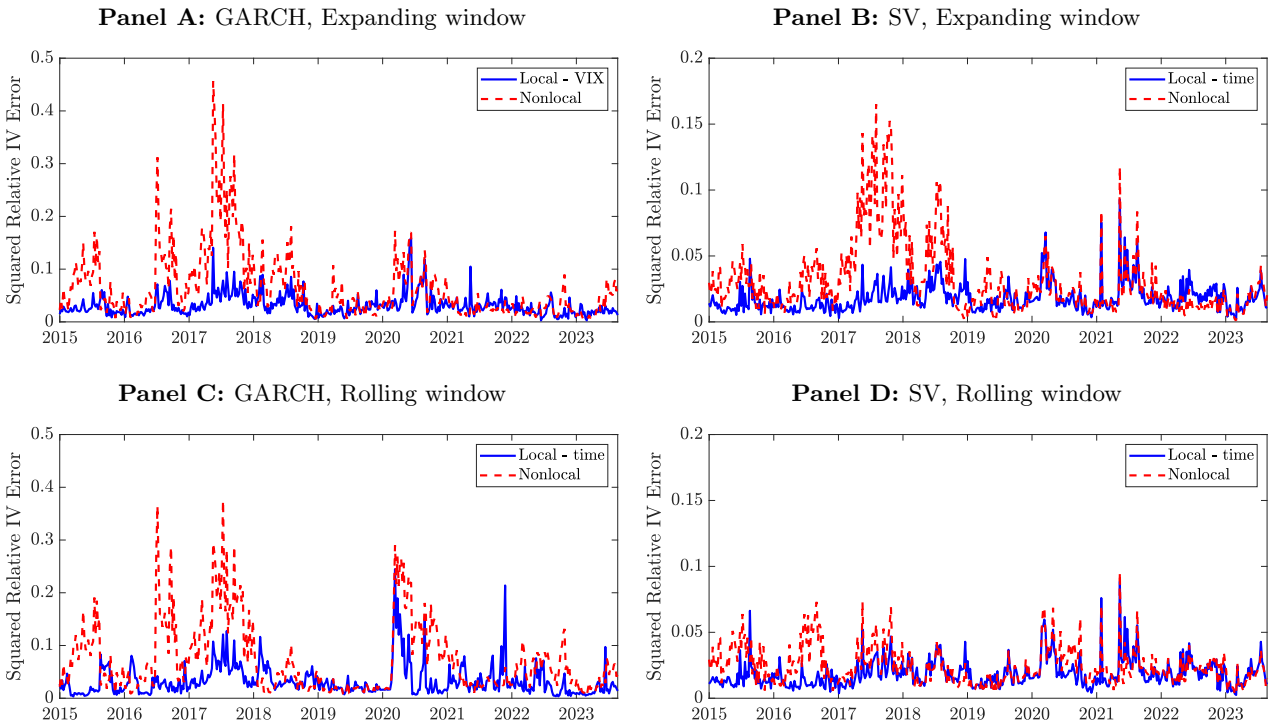
Notes: This table presents metrics evaluating pricing performance during the out-of-sample (OOS) period, January 2015 to August 2023, for Heston-Nandi GARCH option pricing models (Panel A) and Heston SV option pricing models (Panel B). These models are estimated using either the (non-local) quasi-maximum likelihood (QML) approach, indicated by ⧫, or the local QML approach. In each panel, the rows are ordered by the average OOS loss, calculated as the relative implied volatility error under the assumption that future stock prices and variances are known and given. The local method that demonstrates the best performance in the validation sample (the latter half of the estimation sample) is highlighted with an asterisk (*) in the first column. Local estimators utilize state variables listed in the second column and bandwidth parameters specified in the next four columns, which are optimized using the validation sample. All models are estimated using an expanding window. The penultimate column reports Diebold-Mariano t-statistics for each model compared to the benchmark non-local method (identified as ⧫), where negative values suggest lower average loss. Finally, the last column marks models included in the 95% model confidence set with a checkmark.

Table 8: Out-of-sample delta hedging performance

Panel A: Heston-Nandi GARCH model								
Rank	State variable	Optimal bandwidth parameters				Out-of-sample performance		
		P1	P2	P3	P4	Avg. loss	DM stat	MCS
1*	time	0.9960	0.9935	0.9935	0.9905	8.22E-06	-3.99	✓
2	VIX	1.96	2.11	1.86	1.46	9.72E-06	-3.35	✓
3	RV5	1.56	1.61	1.41	1.41	1.08E-05	-2.07	
4	⧫					1.20E-05		
5	BKMkurt	0.70	0.55	0.70	1.30	1.29E-05	0.94	
6	BKMskew	1.05	0.50	0.30	1.35	1.33E-05	1.35	
7	VRP	0.55	0.50	0.50	1.15	1.38E-05	0.69	
8	VVIX	0.60	0.35	0.35	1.25	1.91E-05	0.95	
Panel B: Heston SV model								
Rank	State variable	Optimal bandwidth parameters				Out-of-sample performance		
		P1	P2	P3	P4	Avg. loss	DM stat	MCS
1	VIX	1.56	1.71	1.61	1.36	6.43E-06	-7.68	✓
2	RV5	2.06	1.91	1.76	1.56	6.50E-06	-8.23	✓
3*	time	0.997	0.9975	0.9975	0.9965	7.04E-06	-3.55	✓
4	VRP	4.75	1.95	1.75	1.35	7.67E-06	-3.59	
5	BKMskew	2.10	1.35	1.10	0.90	7.70E-06	-3.68	
6	BKMkurt	2.70	1.95	0.80	0.95	8.49E-06	-0.69	
7	⧫					8.67E-06		
8	VVIX	1.95	1.90	2.00	1.60	8.93E-06	0.95	

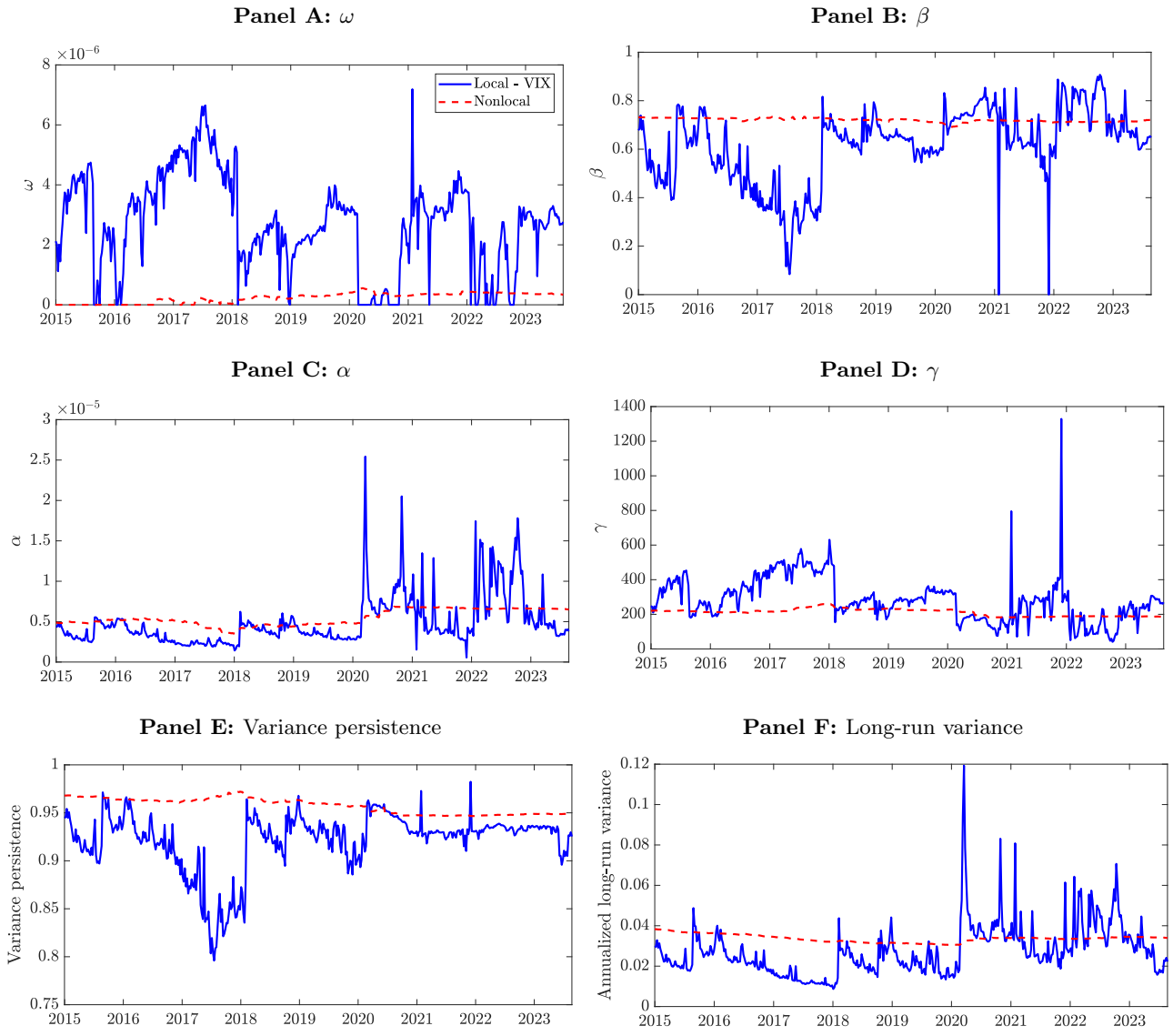
Notes: Notes: This table presents metrics evaluating delta hedging performance during the out-of-sample (OOS) period, January 2015 to August 2023, for Heston-Nandi GARCH option pricing models (Panel A) and Heston SV option pricing models (Panel B). These models are estimated using either the (non-local) quasi-maximum likelihood (QML) approach, indicated by ⧫, or the local QML approach. In each panel, the rows are ordered by the average OOS loss, calculated as the variance of returns from a delta-hedged portfolio. The local method that demonstrates the best performance in the validation sample (the latter half of the estimation sample) is highlighted with an asterisk (*) in the first column. Local estimators utilize state variables listed in the second column and bandwidth parameters specified in the next four columns, which are optimized using the validation sample. All models are estimated using an expanding window. The penultimate column reports Diebold-Mariano t-statistics for each model compared to the benchmark non-local method (identified as ⧫), where negative values suggest lower average loss. Finally, the last column marks models included in the 95% model confidence set with a checkmark.

Figure 1: Time-series out-of-sample option pricing error



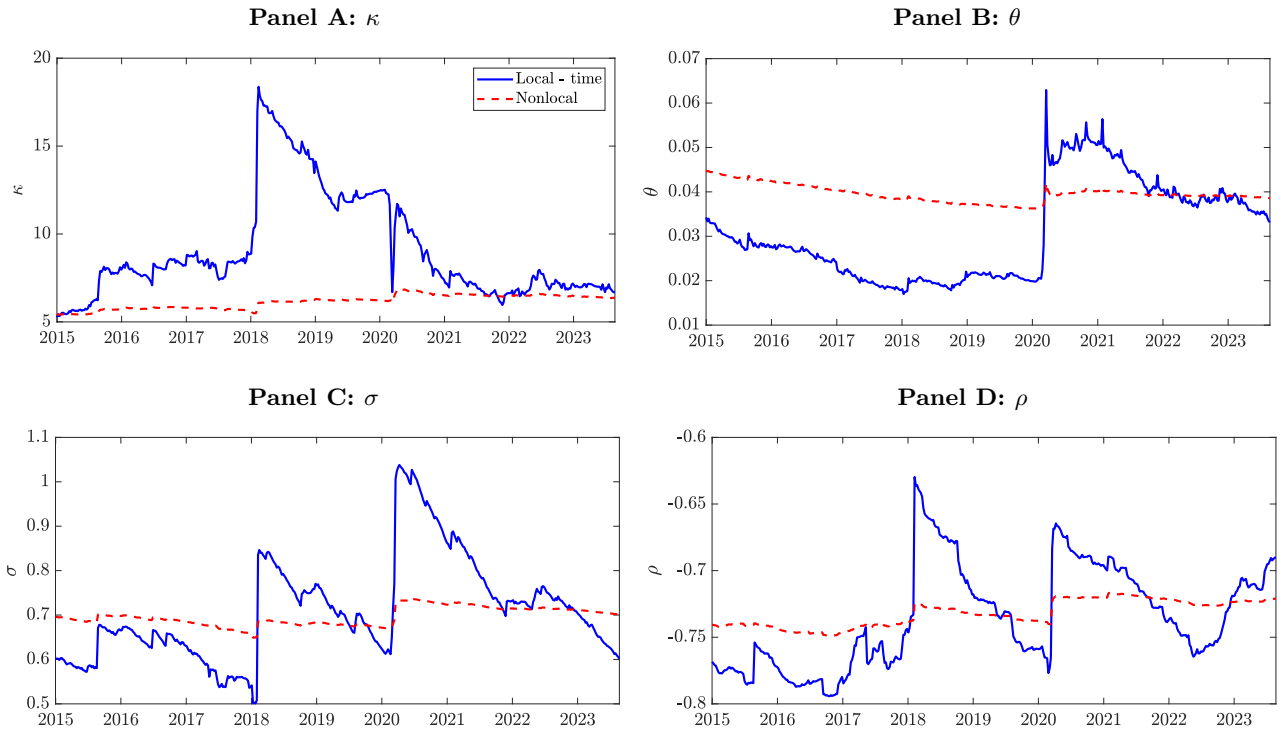
Notes: This figure shows the time-series out-of-sample (OOS) option pricing errors measured by the relative IV error in equation (16) for the Heston-Nandi GARCH model (Panels A and C) and for the Heston SV model (Panels B and D). Panels A and B present the results from the expanding window estimation, while Panels C and D present the results from the rolling window estimation. The solid blue lines represent the results from the local estimation using the VIX as the state variable, and the dashed red lines represent the results from the non-local estimation.

Figure 2: Time-series out-of-sample GARCH parameter estimates



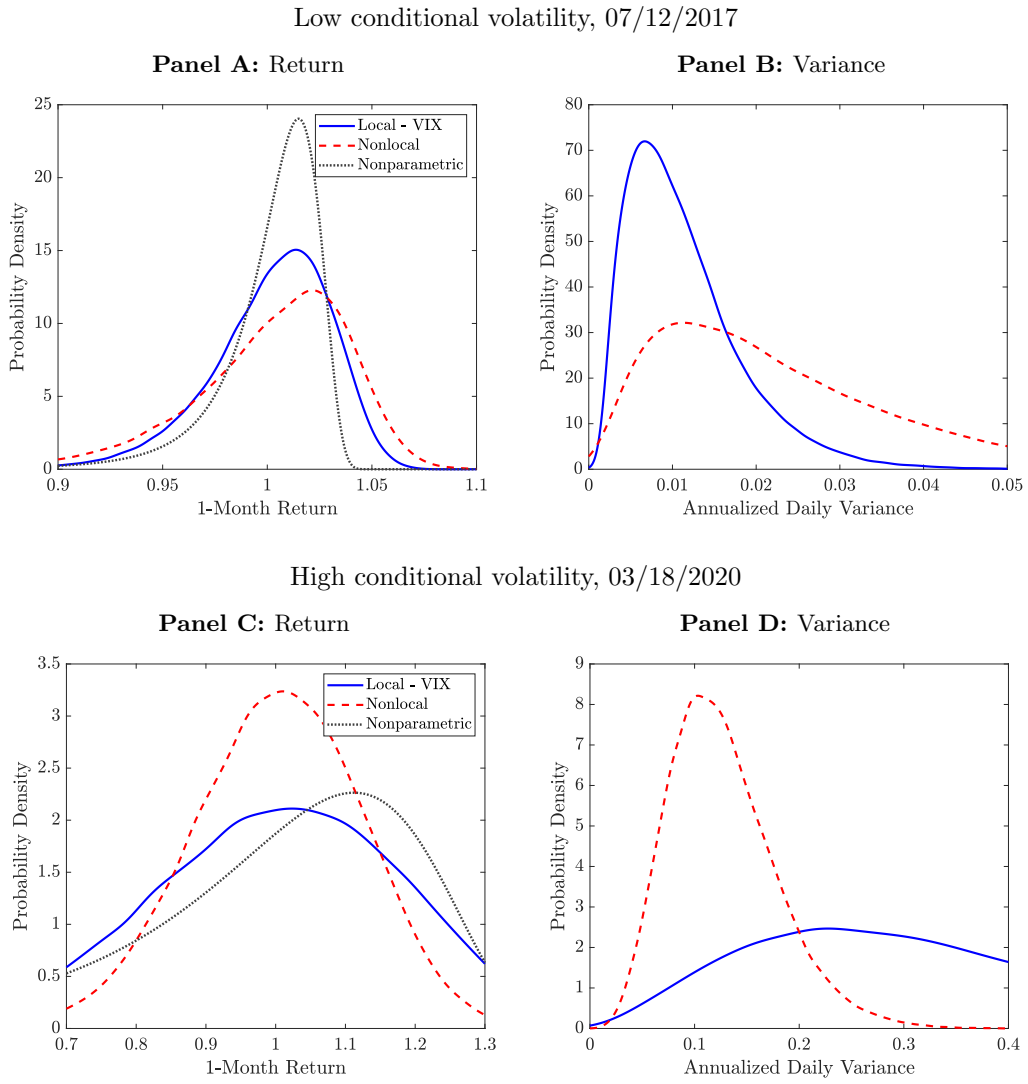
Notes: This figure shows the time-series out-of-sample (OOS) GARCH model parameter estimates. Panels A, B, C, and D plot the estimates of ω, β, α, and γ, respectively. Panels E and F plot the model-implied variance persistence and long-run variance, measured by $\beta + \alpha\gamma^2$ and $(\omega + \alpha)/(1 - \beta - \alpha\gamma^2)$, respectively. The solid blue lines represent the results from the local estimation using the VIX as the state variable, and the dashed red lines represent the results from the non-local estimation.

Figure 3: Time-series out-of-sample SV parameter estimates



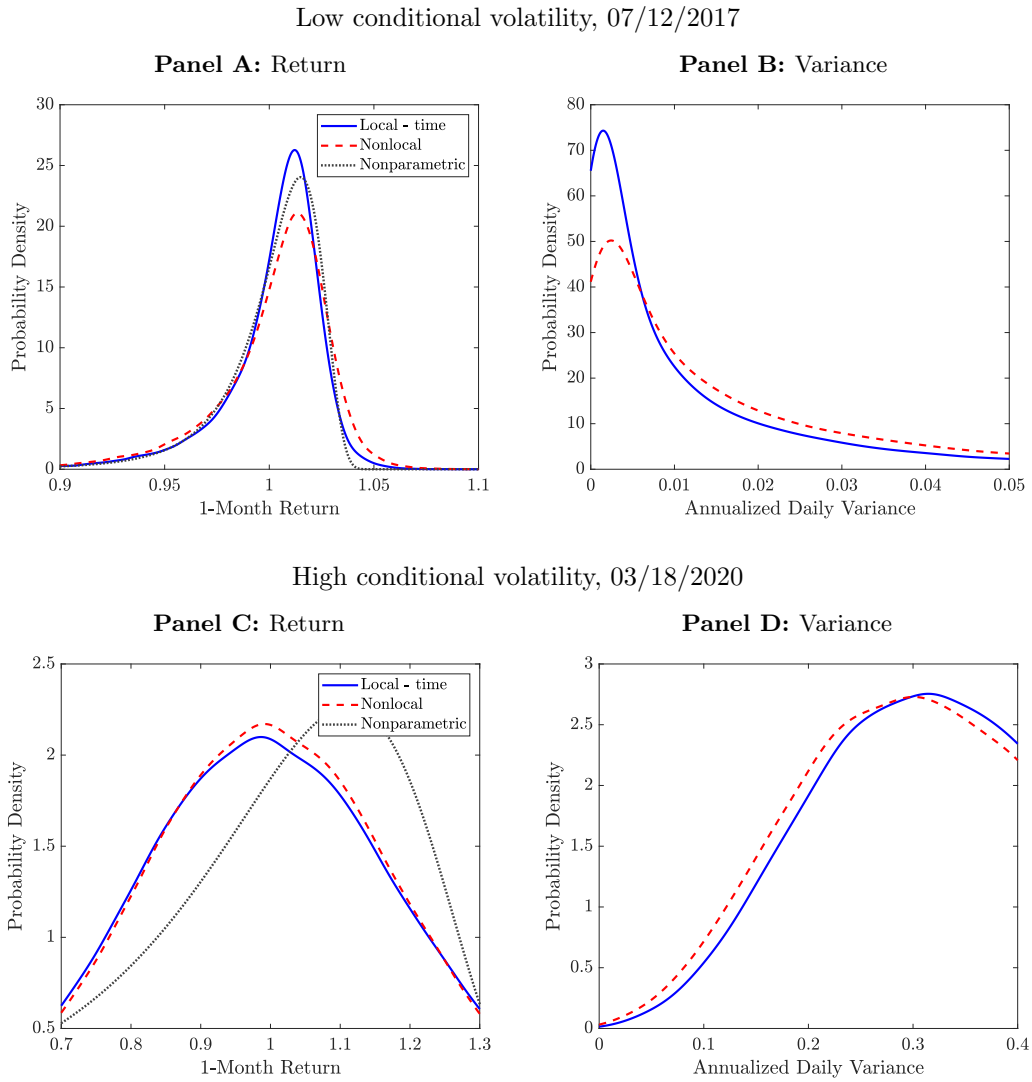
Notes: This figure shows the time-series out-of-sample (OOS) SV model parameter estimates. Panels A, B, C, and D plot the estimates of κ, θ, σ, and ρ, respectively. The solid blue lines represent the results from the local estimation using time as the state variable, and the dashed red lines represent the results from the non-local estimation.

Figure 4: GARCH-simulated one-month forward variance and return distributions



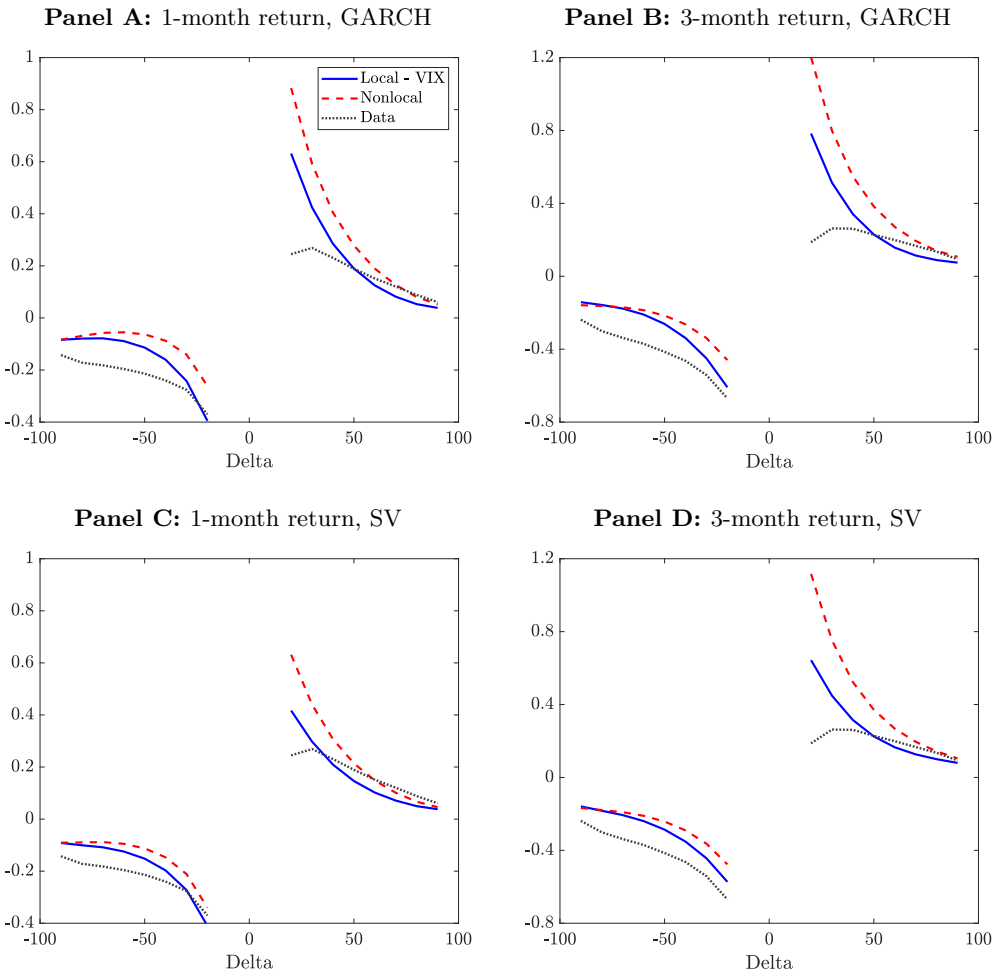
Notes: This figure shows the simulated risk-neutral distributions of the one-month forward return and annualized daily variance for Heston-Nandi GARCH option pricing models. Panels A and B display these distributions conditional on a low-volatility environment, using parameter estimates from July 12, 2017. Assuming the underlying stock price is observed the following day, the distributions of one-month forward returns and variances are simulated for that day. Panels C and D show the corresponding distributions under a high-volatility environment, using parameter estimates from March 18, 2020, with an identical simulation procedure to the low-volatility case. The solid blue lines represent the results from the local estimation using the VIX as the state variable, the dashed red lines represent the results from the non-local estimation, and the dotted black lines represent the nonparametric return distribution from data by Breeden and Litzenberger (1978).
Source: OptionMetrics, Ivy DB US, via WRDS; and authors' calculations.

Figure 5: SV-simulated one-month forward variance and return distributions



Notes: This figure shows the simulated risk-neutral distributions of the one-month forward return and annualized daily variance for Heston SV option pricing models. Panels A and B display these distributions conditional on a low-volatility environment, using parameter estimates from July 12, 2017. Assuming the underlying stock price is observed the following day, the distributions of one-month forward returns and variances are simulated for that day. Panels C and D show the corresponding distributions under a high-volatility environment, using parameter estimates from March 18, 2020, with an identical simulation procedure to the low-volatility case. The solid blue lines represent the results from the local estimation using time as the state variable, the dashed red lines represent the results from the non-local estimation, and the dotted black lines represent the nonparametric return distribution from data by Breeden and Litzenberger (1978).
Source: OptionMetrics, Ivy DB US, via WRDS; and authors' calculations.

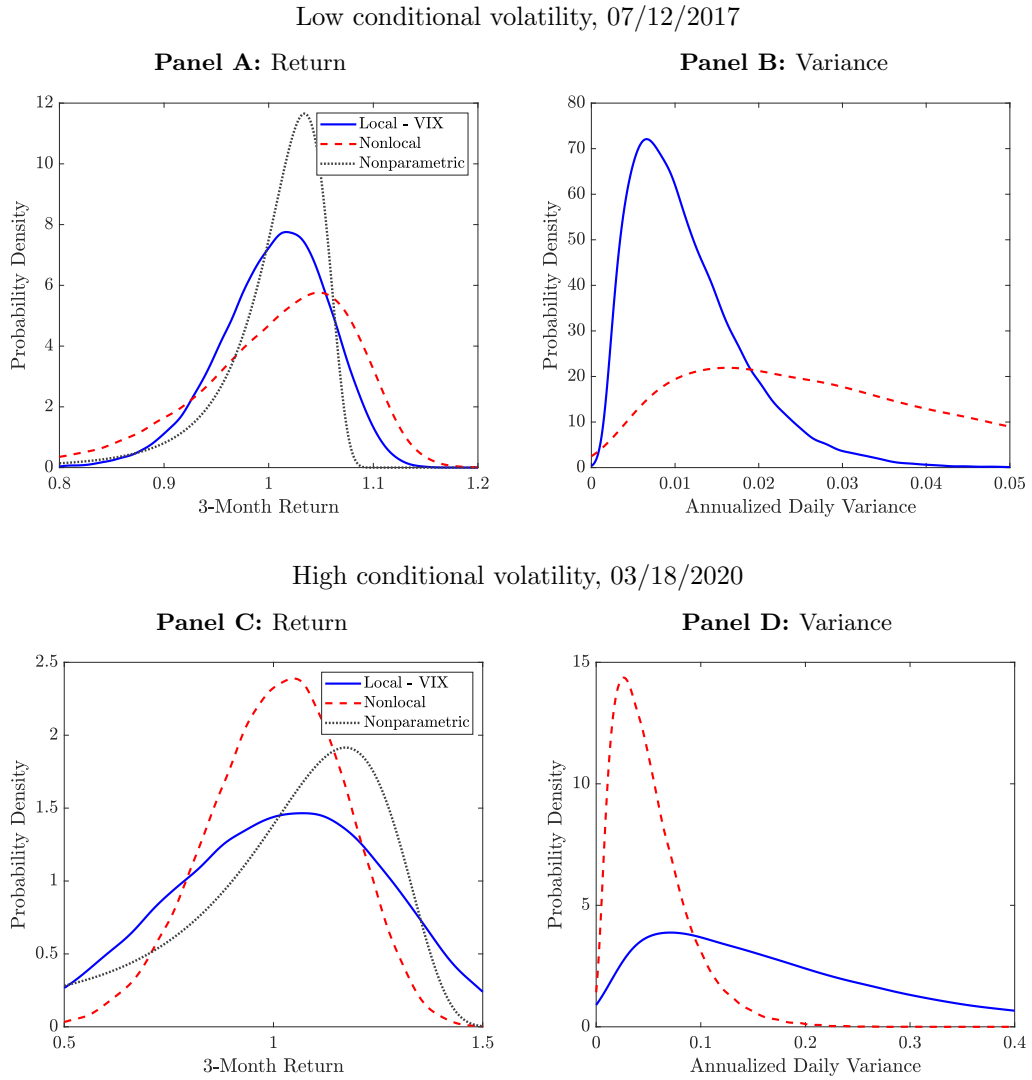
Figure 6: Expected option returns



Notes: This figure shows the comparison between the model-implied holding-to-expiration expected option returns with the realized option returns. Panels A and B present the one-month and three-month option returns under Heston-Nandi GARCH option pricing models across the Black-Scholes Delta dimension (x-axis). Panels C and D present the corresponding results for Heston SV option pricing models. The solid blue lines represent the result from the optimal local estimation, the dashed red lines represent the result from the non-local estimation, and the dotted black lines represent the data.
Source: OptionMetrics, Ivy DB US, via WRDS; Center for Research in Security Prices, CRSP 1925 US Indices Database and CRSP US Treasury Database - Daily/Monthly (updated monthly), via WRDS; and authors' calculations.

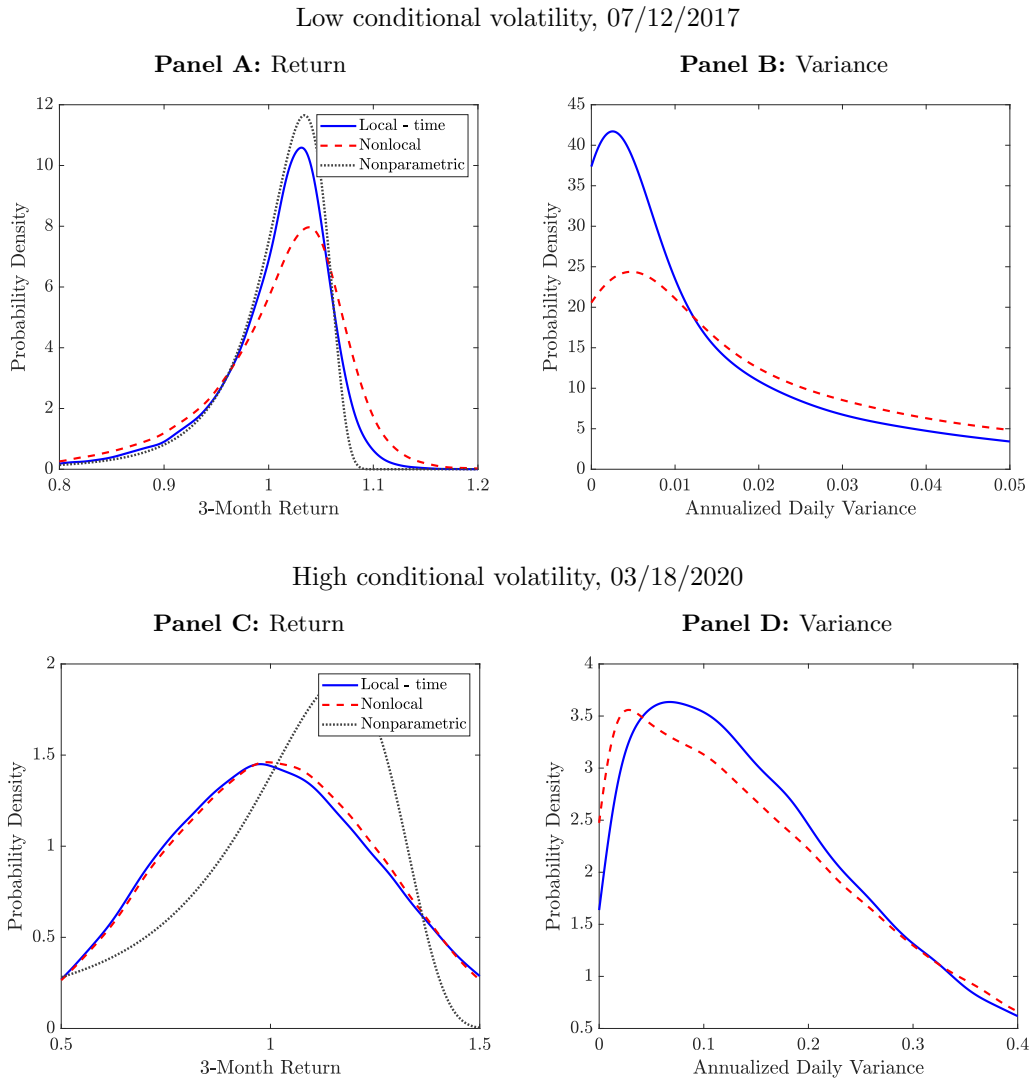
Appendix

Figure A.1: GARCH-simulated three-month forward variance and return distributions



Notes: This figure shows the simulated risk-neutral distributions of the three-month forward return and annualized daily variance for Heston-Nandi GARCH option pricing models. Panels A and B display these distributions conditional on a low-volatility environment, using parameter estimates from July 12, 2017. Assuming the underlying stock price is observed the following day, the distributions of three-month forward returns and variances are simulated for that day. Panels C and D show the corresponding distributions under a high-volatility environment, using parameter estimates from March 18, 2020, with an identical simulation procedure to the low-volatility case. The solid blue lines represent the results from the local estimation using the VIX as the state variable, the dashed red lines represent the results from the non-local estimation, and the dotted black lines represent the nonparametric return distribution from data by Breeden and Litzenberger (1978).
Source: OptionMetrics, Ivy DB US, via WRDS; and authors' calculations.

Figure A.2: SV-simulated three-month forward variance and return distributions



Notes: This figure shows the simulated risk-neutral distributions of the three-month forward return and annualized daily variance for Heston SV option pricing models. Panels A and B display these distributions conditional on a low-volatility environment, using parameter estimates from July 12, 2017. Assuming the underlying stock price is observed the following day, the distributions of three-month forward returns and variances are simulated for that day. Panels C and D show the corresponding distributions under a high-volatility environment, using parameter estimates from March 18, 2020, with an identical simulation procedure to the low-volatility case. The solid blue lines represent the results from the local estimation using time as the state variable, the dashed red lines represent the results from the non-local estimation, and the dotted black lines represent the nonparametric return distribution from data by Breeden and Litzenberger (1978).
Source: OptionMetrics, Ivy DB US, via WRDS; and authors' calculations.