Implied Stochastic Volatility Models*

Yacine Aït-Sahalia†
Department of Economics
Princeton University and NBER

Chenxu Li‡
Guanghua School of Management
Peking University

Chen Xu Li§
Bendheim Center for Finance
Princeton University

This Version: February 18, 2019

Abstract

This paper proposes to build "implied stochastic volatility models" designed to fit option-implied volatility data, and implements a method to construct such models. The method is based on explicitly linking shape characteristics of the implied volatility surface to the specification of the stochastic volatility model. We propose and implement parametric and nonparametric versions of implied stochastic volatility models.

Keywords: implied volatility surface, stochastic volatility, jumps, (generalized) method of moments, kernel estimation, closed-form expansion.

JEL classification: G12; C51; C52.

1 Introduction

No-arbitrage pricing arguments for options most often start with an assumed dynamic model that serves as the data generating process for the option's underlying asset price. Most often again,

*We benefited from the comments of participants at the 2017 Stanford-Tsinghua-PKU Conference in Quantitative Finance, the 2017 Fifth Asian Quantitative Finance Conference, the 2017 BCF-QUT-SJTU-SMU Conference on Financial Econometrics, the Second PKU-NUS Annual International Conference on Quantitative Finance and Economics, the 2017 Asian Meeting of the Econometric Society, the Third Annual Volatility Institute Conference at NYU Shanghai, the 2018 Review of Economic Studies 30th Anniversary Conference and the 2018 FERM Conference. The research of Chenxu Li was supported by the Guanghua School of Management, the Center for Statistical Science, and the Key Laboratory of Mathematical Economics and Quantitative Finance (Ministry of Education) at Peking University, as well as the National Natural Science Foundation of China (Grant 71671003). Chen Xu Li is grateful for a graduate scholarship and funding support from the Graduate School of Peking University as well as support from the Bendheim Center for Finance at Princeton University.

†Address: JRR Building, Princeton, NJ 08544, USA. E-mail address: yacine@princeton.edu.

‡Address: Guanghua School of Management, Peking University, Beijing, 100871, P. R. China. E-mail address: cxli@gsm.pku.edu.cn.

§Address: JRR Building, Princeton, NJ 08544, USA. E-mail address: chenxul@princeton.edu.

that model is of the stochastic volatility type, see, e.g., Hull and White (1987), Heston (1993), Bates (1996), Duffie et al. (2000), and Pan (2002). Unfortunately, the relationship between the market data, namely option prices or equivalently, implied volatilities, and the model is not fully explicit. Implied volatilities can only be computed numerically or approximated, even under the affine stochastic volatility models, see, e.g., Duffie et al. (2000) and the references therein, for which option prices admit analytical Fourier transforms.

As the variety of affine or non-affine specifications suggest, there is no accepted consensus on the model specifications in the literature. There is however agreement that a stochastic volatility model should produce option prices (or equivalently, implied volatilities) with the features that are observed in the empirical data. A prevalent approach relies on fitting pre-specified models with particular dynamics to data by estimation or calibration, with goodness-of-fit determined by likelihood or mean-squared pricing errors. Alternatively, models can be calibrated to fit a set (a continuum is often required) of options or other derivative prices exactly. Prominent examples of the latter approach are the local volatility model of Dupire (1994) and the results of Andersen and Andreasen (2000), Carr et al. (2004), Carr and Cousot (2011), and Carr and Cousot (2012) including local Lévy jumps.

In the same spirit, we ask in this paper whether it is possible to use the information contained in implied volatility data to conduct inference about an underlying stochastic volatility (rather than a local volatility) model. At each point in time, implied volatility data take the form of a surface representing the implied volatility of the option as a function of its moneyness and time to maturity. We will show that it is possible to use a small number of observable and practically useful "shape characteristics" of the implied volatility surface, including but not limited to the slope of the implied volatility smile, to fully characterize the underlying stochastic volatility model.

For this purpose, we will rely on an expansion of the implied volatility surface in terms of time-to-maturity and log-moneyness. Various types of expansions for implied volatilities or option prices, obtained using different methods, are available in the literature. They include: small volatility-of-volatility expansions, near a non-stochastic volatility, also known as small ε or small noise expansions, see Kunitomo and Takahashi (2001) and Takahashi and Yamada (2012); expansion based on slow-varying volatility, see Sircar and Papanicolaou (1999) and Lee (2001); expansion based on fast-varying and slow-varying analysis, see Fouque et al. (2016); short maturity expansions, see Medvedev and Scaillet (2007) (for an expansion with respect to the square of time-to-maturity with expansion term sorted in terms of moneyness scaled by volatility), Durrleman (2010) (with a correction due to Pagliarani and Pascucci (2017)), and Lorig et al. (2017); expansion using PDE methods, see Berestycki et al. (2004); singular perturbation expansion, see Hagan and Woodward (1999); expansion around an auxiliary model, see Kristensen and Mele (2011); expansion using transition density expansion, see Gatheral et al. (2012) and Xiu (2014); expansion of the characteristic function, see

Jacquier and Lorig (2015). Some of these methods apply generally, while others apply only to specific models, such as the Heston model, as in Forde et al. (2012) (short maturity), Forde and Jacquier (2011) (long maturity), or exponential Lévy models as in Andersen and Lipton (2013). The asymptotic behavior of implied volatilities as time-to-maturity approaches zero is important: for the continuous case, see Ledoit et al. (2002) and Berestycki et al. (2002), and with jumps, see Carr and Wu (2003) and Durrleman (2008). Finally, a number of asymptotic results concerning long-dated, short-dated, far out of the money strike, and jointly-varying strike-expiration regimes are available, see Lee (2004), Gao and Lee (2014), and Tehranchi (2009).

The expansion we employ for our purposes is different from existing ones; it takes the form of a bivariate series in time-to-maturity and log-moneyness, applies to general stochastic volatility models and produces closed-form expressions for arbitrary stochastic volatility models with or without jumps. Given the extensive literature on expansions, however, the novelty in this paper is not its expansion (although it is new) but rather the use of such an expansion as a means to conduct inference on the underlying stochastic volatility model. The existing literature on implied volatility expansions has been primarily concerned with the derivation of the expansion and its properties, but rarely with using the expansion for the purposes of estimating or testing the model that underlies the expansion.

Said differently, the main use of expansions in the literature has been in the following direction: assuming a given stochastic volatility model, what can be said about the implied volatilities that this model generates? In this paper, we take the reverse direction: taking the observed implied volatility surface as given market data, what can be said about the stochastic volatility model that generated the data? We answer this question by constructing "implied stochastic volatility models", which are stochastic volatility models whose characteristics have been completely estimated to reproduce salient characteristics of the implied volatility data. Our approach consists in casting a small number of observable shape characteristics of the implied volatility surface (its level, slope and convexity along the moneyness dimension, as well as its slope along the term-structure dimension) as a set of restrictions on the specification of the stochastic volatility model. If one is interested in estimating a parametric stochastic volatility model, we show how to set up these restrictions as moment conditions in GMM. If one is not willing to parametrize the model, we show how the functions characterizing the stochastic volatility model can be recovered nonparametrically from the shape characteristics of the implied volatility surface.

Applying the proposed method to S&P 500 index options, we construct an implied stochastic volatility model with the following empirical features: a strong leverage effect between the innovations in returns and volatility, mean reversion in volatility, monotonicity and state dependency in volatility of volatility, while matching the features of the implied volatility surfaces: level, smile and convexity in the log-moneyness direction, and slope in the term structure direction.

The paper is organized as follows. Section 2 sets up the problem we are studying, the notation,

and describes the set of relationships between the stochastic volatility model and the implied volatility surface. We then use these relationships to propose two methods to construct an implied stochastic volatility model, first parametric in Section 3 and then nonparametric in Section 4. We implement these methods in Monte Carlo simulations in Section 5, showing that both the parametric and nonparametric estimation methods are accurate, and then on real data in Section 6. Section 7 extends the analysis to allow for jumps in the returns dynamics of the stochastic volatility model and discusses the empirical challenges that this poses. Section 8 concludes, while mathematical details are contained in the Appendix.

2 Stochastic volatility models and implied volatility surfaces

Consider a generic continuous bivariate stochastic volatility (SV thereafter) model. Under an assumed risk-neutral measure, the price of the underlying asset S_t and its volatility v_t jointly follow a diffusion process

$$\frac{dS_t}{S_t} = (r-d)dt + v_t dW_{1t}, \tag{1a}$$

$$dv_t = \mu(v_t)dt + \gamma(v_t)dW_{1t} + \eta(v_t)dW_{2t}. \tag{1b}$$

We will add jumps in returns to the model in Section 7 below. Here, r and d are the risk-free rate and the dividend yield of the underlying asset, both assumed constant for simplicity, and observable; W_{1t} and W_{2t} are two independent standard Brownian motions; μ, γ, and η are scalar functions. The generic specification (1a)–(1b) nests all existing continuous bivariate SV models. For models conventionally expressed in terms of instantaneous variance rather than volatility (e.g., the model of Heston (1993)), it is straightforward to obtain the equivalent form of (1a)–(1b) by Itô's lemma. Our objective is to fully identify the model, that is, v_t at each discrete instant at which data sampling occurs, and the unknown functions $\mu(\cdot)$, $\gamma(\cdot)$ and $\eta(\cdot)$. This is a natural extension to stochastic volatility models of the question answered in Dupire (1994) for local volatility models, which relied on a method which cannot be used in the stochastic volatility context.[1]

We are also interested in the leverage effect coefficient function as the correlation function between asset returns and innovations in spot volatility, defined as in Aït-Sahalia et al. (2013) by

$$\rho(v_t) = \frac{\gamma(v_t)}{\sqrt{\gamma(v_t)^2 + \eta(v_t)^2}}. \tag{2}$$

This coefficient function is identified once the other components of the model are. In general, $\rho(v_t)$ is empirically found to be negative, and is in general stochastic since the dependence in v_t need

[1]Local volatility models are of the form $dS_t/S_t = (r-d)dt + \sigma(S_t)dW_t$. The approach of Dupire (1994), based on inverting the pricing equation for the function $\sigma(\cdot)$, cannot be extended from the local to the stochastic volatility situation: when employed in a stochastic volatility setting, it can only characterize $\mathbb{E}[v_T|S_T, S_0]$ rather than the full dynamics (1b).

not cancel out between the numerator and denominator in (2): see, e.g., the models of Jones (2003) and Chernov et al. (2003), among others. For $\rho(v_t)$ to be independent of v_t, i.e., $\rho(v) \equiv \rho$ for some constant ρ, it must be that $\eta(v) = \rho\gamma(v)/\sqrt{1-\rho^2}$, i.e., the two functions $\eta(v)$ and $\gamma(v)$ are uniformly proportional to each other. This is the case in the model of Heston (1993), for instance.

The arbitrage-free price of an European-style put option with maturity T, i.e., time-to-maturity $\tau = T - t$, and exercise strike K is (in terms of log-moneyness $k = \log(K/S_t)$)

$$P(\tau, k, S_t, v_t) = e^{-r\tau}\mathbb{E}_t[\max(S_t e^k - S_T, 0)],$$

where $\mathbb{E}_t$ denotes the risk-neutral conditional expectation given the information up to time t. In practice, the market price of an option is typically quoted through its Black-Scholes implied volatility (IV thereafter) Σ, i.e., the value of the volatility parameter which, when plugged into the Black-Scholes formula $P_{\mathrm{BS}}(\tau, k, S_t, \sigma)$, leads to a theoretical value equal to the observed market price of the option[2]:

$$P_{\mathrm{BS}}(\tau, k, S_t, \Sigma) = P(\tau, k, S_t, v_t).$$

Viewed simply as mapping actual option prices into a different unit, using implied volatilities does not require that the assumptions of the Black-Scholes model be satisfied, and has a few advantages: implied volatilities are independent of the scale of the underlying asset value or strike price, deviations from a flat IV surface denote deviations from the Black-Scholes model (or equivalently deviations from the Normality of log-returns), and such deviations can be monotonically interpreted (the higher the IV above the flat level, the more expensive the option, and similarly below), so differences in IV allow for relative value comparisons between options.

2.1 From stochastic volatility to implied volatility

The IV depends on S_t only through k, that is, $\Sigma = \Sigma(\tau, k, v_t)$. This is because the option price can be written in a form proportional to the time-t price S_t,

$$P(\tau, k, S_t, v_t) = S_t \bar{P}(\tau, k, v_t) \text{ with } \bar{P}(\tau, k, v_t) = e^{-r\tau}\mathbb{E}_t\left[\max\left(e^k - \frac{S_T}{S_t}, 0\right)\right]. \tag{3}$$

For a given log-moneyness k, the function $\bar{P}(\tau, k, v_t)$ is independent of the initial underlying asset price S_t since the dynamics (1a) of the underlying asset price imply that the ratio S_T/S_t is independent of S_t, so is the expectation function (3) for defining $\bar{P}(\tau, k, v_t)$. Writing $P_{\mathrm{BS}}(\tau, k, S_t, \sigma) = S_t \bar{P}_{\mathrm{BS}}(\tau, k, \sigma)$, the IV Σ is determined by

$$\bar{P}_{\mathrm{BS}}(\tau, k, \Sigma) = \bar{P}(\tau, k, v_t), \tag{4}$$

and therefore

$$\Sigma(\tau, k, v_t) = \bar{P}_{\mathrm{BS}}^{-1}(\tau, k, \bar{P}(\tau, k, v_t)). \tag{5}$$

[2]Model implied volatilities calculated from put and call options are identical by put-call parity.

The mapping $(\tau, k) \longmapsto \Sigma(\tau, k, v_t)$ at a given t is the (model) IV surface at that time. We will consider several shape characteristics of the IV surface such as its slope and convexity along the log-moneyness and the term-structure dimensions, defined by the partial derivative $\partial^{i+j}\Sigma/\partial\tau^i\partial k^j$ for integers $i, j \geq 0$. In particular, we will focus on the at-the-money ($k = 0$) and short maturity ($\tau \to 0$) shape characteristics

$$\Sigma_{i,j}(v_t) = \lim_{\tau\to 0} \frac{\partial^{i+j}}{\partial\tau^i\partial k^j}\Sigma(\tau, 0, v_t). \tag{6}$$

To illustrate, we show in Figure 1 the S&P 500 IV surface on a given day along with the two slopes $\Sigma_{0,1}(v_t)$ (log-moneyness slope, or IV smile) and $\Sigma_{1,0}(v_t)$ (term-structure slope) as red and blue dashed lines, respectively.

The idea in this paper is to treat the shape characteristics of the IV surface (6) as observable from market data, and to use them to determine the SV model (1a)–(1b) that is compatible with them. The tool we call upon for that purpose is that of IV asymptotic expansions, which express the shape characteristics $\Sigma_{i,j}(\cdot)$ in terms of v_t and the functions $\mu(\cdot)$, $\gamma(\cdot)$ and $\eta(\cdot)$. The function Σ admits an expansion of the form

$$\Sigma^{(J,\mathbf{L}(J))}(\tau, k, v_t) = \sum_{j=0}^{J}\sum_{i=0}^{L_j} \sigma^{(i,j)}(v_t)\tau^i k^j, \tag{7}$$

up to some integer expansion orders J and $\mathbf{L}(J) = (L_0, L_1, \cdots, L_J)$ with $L_j \geq 0$, and therefore

$$\sigma^{(i,j)}(v_t) = \frac{1}{i!j!}\Sigma_{i,j}(v_t). \tag{8}$$

For a given SV model, the coefficients $\sigma^{(i,j)}(\cdot)$ can be derived in closed form to arbitrary order one after the other. We provide the precise mathematical details in the Appendix. In a nutshell, we first note that the $(0,0)$th order term must be given by the instantaneous volatility

$$\sigma^{(0,0)}(v) = v, \tag{9}$$

which is a well-known fact (see, e.g., Ledoit et al. (2002) and Durrleman (2008).) The purpose of an IV expansion is to compute the higher order coefficients in (7) for an arbitrary SV model. We describe in Appendix A our method for achieving this goal; the advantage in our view compared to many of the existing alternative approaches described in the Introduction is that this method yields fully explicit coefficients, and does so for arbitrary SV models.

We illustrate the approach by focusing on the at-the-money level $\sigma^{(0,0)}$, slope $\sigma^{(0,1)}$, and convexity $\sigma^{(0,2)}$ (up to a constant equal to 2) along the log-moneyness dimension, as well as the slope $\sigma^{(1,0)}$ along the term-structure dimension, all for short time-to-maturity. These four basic shape characteristics construct a skeleton of the IV surface, and thus conversely they can be extracted from an IV surface. Set $(J, \mathbf{L}(J)) = (2, (1, 0, 0))$ in (7), that is,

$$\Sigma^{(2,(1,0,0))}(\tau, k, v_t) = \sigma^{(0,0)}(v_t) + \sigma^{(1,0)}(v_t)\tau + \sigma^{(0,1)}(v_t)k + \sigma^{(0,2)}(v_t)k^2. \tag{10}$$

We show in Appendix A that

$$\sigma^{(0,1)}(v_t) = \frac{1}{2v_t}\gamma(v_t),\ \sigma^{(0,2)}(v_t) = \frac{1}{12v_t^3}[2v_t\gamma(v_t)\gamma'(v_t) + 2\eta(v_t)^2 - 3\gamma(v_t)^2], \tag{11}$$

$$\sigma^{(1,0)}(v_t) = \frac{1}{24v_t}[2\gamma(v_t)(6(d-r) - 2v_t\gamma'(v_t) + 3v_t^2) + 12v_t\mu(v_t) + 3\gamma(v_t)^2 + 2\eta(v_t)^2]. \tag{12}$$

These expressions provide the expansion of the IV surface that corresponds to a given specification of the SV model. The main idea in this paper is to use conversely the IV surface expansion to estimate the unknown coefficients functions of the SV model. In other words, treating the coefficients $\sigma^{(0,0)}(\cdot)$, $\sigma^{(0,1)}(\cdot)$, $\sigma^{(0,2)}(\cdot)$, and $\sigma^{(1,0)}(\cdot)$ (and higher order coefficients if necessary) as observable from options data, how can we use the data and these formulae to estimate the unknown functions $\mu(\cdot)$, $\gamma(\cdot)$, and $\eta(\cdot)$?

2.2 From implied volatility to stochastic volatility

It is possible in fact to fully characterize the SV model from observations on the level, log-moneyness slope and convexity, and term-structure slope of the IV surface. In other words, we view (11)–(12) as a system of equations to be solved for $\gamma(\cdot)$, $\eta(\cdot)$, and $\mu(\cdot)$, given the IV surface characteristics. These equations lead to a useful estimation method because it turns out that they can be inverted in closed form, so no further approximation, numerical solution of a differential equation or other numerical inversion is required. First, observe that (11)–(12) imply

$$\gamma(v_t) = 2\sigma^{(0,0)}(v_t)\sigma^{(0,1)}(v_t), \tag{13a}$$

and

$$\eta(v_t) = \left[2\left(6\sigma^{(0,0)}(v_t)^3\sigma^{(0,2)}(v_t) - 2\sigma^{(0,0)}(v_t)\gamma(v_t)\gamma'(v_t) + 3\gamma(v_t)^2\right)\right]^{-1/2}, \tag{13b}$$

$$\mu(v_t) = 2\sigma^{(1,0)}(v_t) + \frac{\gamma(v_t)}{6}(2\gamma'(v_t) - 3\sigma^{(0,0)}(v_t)) - \frac{\gamma(v_t)}{\sigma^{(0,0)}(v_t)}\left(d - r + \frac{1}{4}\gamma(v_t)\right) - \frac{\eta(v_t)^2}{6\sigma^{(0,0)}(v_t)}. \tag{13c}$$

Second, plug in (13a) into (13b), and then plug in both expressions into (13c) to obtain:

Theorem 1. *The coefficient functions $\gamma(\cdot)$, $\eta(\cdot)$, and $\mu(\cdot)$ of the SV model (1a)–(1b) can be recovered in closed form as functions of the coefficients of the IV expansion (10) as follows:*

$$\gamma(v_t) = 2\sigma^{(0,0)}(v_t)\sigma^{(0,1)}(v_t), \tag{14a}$$

and

$$\eta(v_t) = 2\sigma^{(0,0)}(v_t)\left[3\sigma^{(0,0)}(v_t)\sigma^{(0,2)}(v_t) + 2\sigma^{(0,1)}(v_t)^2 - 4\sigma^{(0,0)}(v_t)\sigma^{(0,1)}(v_t)\sigma^{(0,1)\prime}(v_t)\right]^{-1/2}, \tag{14b}$$

$$\mu(v_t) = \sigma^{(0,0)}(v_t)^2\left[\sigma^{(0,1)}(v_t)(2\sigma^{(0,1)\prime}(v_t) - 1) - \frac{1}{2}\sigma^{(0,2)}(v_t)\right] - 2(d-r)\sigma^{(0,1)}(v_t) + 2\sigma^{(1,0)}(v_t). \tag{14c}$$

where $\sigma^{(0,1)\prime}(v_t)$ represents the first order derivative of $\sigma^{(0,1)}(v_t)$ with respect to v_t.

We note a few interesting implications of this result. First, (14a) shows that for a given IV level $\sigma^{(0,0)}(v_t)$, the slope $\sigma^{(0,1)}(v_t)$ plays an important role in determining the volatility function $\gamma(v_t)$ attached to the common Brownian shocks W_{1t} of the asset price S_t and its volatility v_t. For a fixed level $\sigma^{(0,0)}(v_t)$, a steeper slope $\sigma^{(0,1)}(v_t)$ results in a higher absolute value of the volatility function $\gamma(v_t)$. Second, from (14b), a steeper slope $\sigma^{(0,1)}(v_t)$ has an effect on the volatility function $\eta(v_t)$ attached to the idiosyncratic Brownian shock W_{2t} in the volatility dynamics which can be of either sign. Besides the level $\sigma^{(0,0)}(v_t)$ and slope $\sigma^{(0,1)}(v_t)$, the convexity $\sigma^{(0,2)}(v_t)$ also matters for the volatility function $\eta(v_t)$. The total spot volatility of volatility is $\sqrt{\gamma(v_t)^2+\eta(v_t)^2}$, so for a fixed level $\sigma^{(0,0)}(v_t)$ and slope $\sigma^{(0,1)}(v_t)$, a greater convexity $\sigma^{(0,2)}(v_t)$ results in a larger volatility of volatility. Third, from (2) and (14a), we see that the sign of the leverage effect coefficient $\rho(v_t)$ is determined by the sign of the slope $\sigma^{(0,1)}(v_t)$: as is typically the case in the data, a downward-sloping IV smile, $\sigma^{(0,1)}(v_t)<0$, translates directly into $\rho(v_t)<0$. Further, $\rho(v_t)$ is monotonically decreasing in $\eta(v_t)$, so it follows from (14b) and (2) that, for a fixed level $\sigma^{(0,0)}(v_t)$ and slope $\sigma^{(0,1)}(v_t)$, a greater convexity $\sigma^{(0,2)}(v_t)$ leads to a larger volatility of volatility, and consequently, a weaker leverage effect $\rho(v_t)$. Finally, (14c) shows that for fixed levels of $\sigma^{(0,0)}(v_t)$, $\sigma^{(0,1)}(v_t)$, and $\sigma^{(0,2)}(v_t)$, an increase of the term-structure slope $\sigma^{(1,0)}(v_t)$ on the IV surface results in an increase in the drift $\mu(v_t)$, i.e., a faster expected change of the instantaneous volatility v_t.

3 Constructing a parametric implied stochastic volatility model

We now turn to using the above connection between the specification of the SV model and the resulting IV expansion in order to estimate the coefficient functions of a parametric SV model, doing so in such a way that the estimated model generates option prices that match the observed features of the IV surface.

We assume for now that the SV model (1a)–(1b) is a parametric one, so that $\mu(\cdot)=\mu(\cdot;\theta)$, $\gamma(\cdot)=\gamma(\cdot;\theta)$, and $\eta(\cdot)=\eta(\cdot;\theta)$, where θ denotes the vector of unknown parameters to be estimated in a compact space $\Theta\subset\mathbb{R}^K$, and θ_0 denotes their true values. We further assume that the parametric functions are known, and twice continuously differentiable in θ.

To estimate θ, we propose to use the closed-form IV expansion coefficients to form moment conditions. Assume that a total of n IV surfaces are observed with equidistant time interval Δ, without loss of generality. On day l, we observe n_l implied volatilities $\Sigma^{\text{data}}(\tau_l^{(m)},k_l^{(m)})$ along with time-to-maturity $\tau_l^{(m)}$ and log-moneyness $k_l^{(m)}$ for $m=1,2,\ldots,n_l$. We assume that the data are stationary and strong mixing with rate greater than two.

The moment functions we propose to use are

$$g^{(i,j)}(v_{l\Delta};\theta)=[\sigma^{(i,j)}]_l^{\text{data}}-[\sigma^{(i,j)}(v_{l\Delta};\theta)]^{\text{model}}, \tag{15}$$

where $[\sigma^{(i,j)}]_l^{\text{data}}$ (resp. $[\sigma^{(i,j)}(v_{l\Delta};\theta)]^{\text{model}}$) denote the data coefficients (resp. the closed-form for-

mulae given in (11)–(12) and additional higher orders if necessary) of the expansion terms $\sigma^{(i,j)}(v_{l\Delta})$ of (7).

We gather the different moment conditions $g^{(i.j)}$ into a vector

$$g(v_{l\Delta};\theta) = (g^{(i,j)}(v_{l\Delta};\theta))_{(i,j)\in I}$$

of moment conditions, for some integer index set I consisting of nonnegative integer pairs (i,j) such that $i+j \geq 1$: for example, $I = \{(1,0),(0,1),(0,2)\}$. The choice of moment conditions is flexible, depending on the shape characteristics one decides to fit, and the number of parameters to be estimated, and may include higher order terms. We assume that

$$\mathbb{E}[g(v_{l\Delta};\theta_0)] = 0$$

and $\mathbb{E}[g(v_{l\Delta};\theta)] \neq 0$ for $\theta \neq \theta_0$ holds. We also assume that θ_0 is in the interior of Θ. As the moments are given by coefficients of an expansion, a bias term of small order is left, an effect similar to that in Aït-Sahalia and Mykland (2003). We treat this term as negligible on the basis of fitting each IV surface near its at-the-money and short maturity point.

To extract $[\sigma^{(i,j)}]_l^{\text{data}}$ from the observed options data, recall the form of the expansion (7), which can be interpreted as a polynomial regression of IV on time-to-maturity τ and log-moneyness k. So, on any day l, we regress

$$\Sigma^{\text{data}}(\tau_l^{(m)}, k_l^{(m)}) = \sum_{j=0}^{J}\sum_{i=0}^{L_j} \beta_l^{(i,j)} (\tau_l^{(m)})^i (k_l^{(m)})^j + \epsilon_l^{(m)}, \text{ for } m = 1,2,\ldots,n_l, \tag{16}$$

where $\epsilon_l^{(m)}$ represent i.i.d. exogenous observation errors with zero means.[3] The coefficient $\sigma^{(i,j)}(v_{l\Delta})$ of the expansion (7) is then estimated by the regression coefficient $\beta_l^{(i,j)}$ in (16):

$$[\sigma^{(i,j)}]_l^{\text{data}} = \hat{\beta}_l^{(i,j)}, \text{ for } i,j \geq 0; \tag{17}$$

in particular, $v_{l\Delta} = [\sigma^{(0,0)}]_l^{\text{data}} = \hat{\beta}_l^{(0,0)}$. While the objects of interest (6) are derivatives of the IV surface Σ evaluated at $(\tau,k) = (0,0)$, the regression (16) includes observations with (τ,k) away from $(0,0)$ in order to estimate these derivatives.

To estimate the parameters θ by GMM (see Hansen (1982)) we construct the sample analog of $\mathbb{E}[g(v_{l\Delta};\theta)]$ as follows:

$$g_n(\theta) \equiv \frac{1}{n}\sum_{l=1}^{n} g(v_{l\Delta};\theta).$$

The estimator $\hat{\theta}$ is defined as the solution of the quadratic minimization problem

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, g_n(\theta)^{\intercal} W_n g_n(\theta), \tag{18}$$

[3]This is a generalization of the linear regression in Dumas et al. (1998) of implied volatilities on τ and $K = S_t e^k$.

where W_n is a positive definite weight matrix. If the number of moment conditions is equal to that of parameters to estimate, i.e., the model is exactly identified, the estimator $\hat{\theta}$ is the solution of the (system of) equations

$$g_n(\hat{\theta}) = 0,$$

and the choice of W_n does not matter. Otherwise, i.e., if the number of moment conditions is greater than that of parameters to estimate, the model is over-identified and the optimal choice of the weight matrix W_n follows from a standard two-step estimation.

The asymptotic behavior of $\hat{\theta}$ is given by

$$\hat{\theta} \xrightarrow{P} \theta_0 \text{ and } \sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}\left(0, V^{-1}(\theta_0)\right), \text{ as } n \to \infty, \tag{19}$$

where

$$V(\theta) = G(\theta)^\intercal \Omega^{-1}(\theta) G(\theta), \text{ with } G(\theta) = \mathbb{E}\left[\frac{\partial g\left(v_{l\Delta}; \theta\right)}{\partial \theta}\right], \ \Omega(\theta) = \Omega_0(\theta) + \sum_{j=1}^{n-1} \left(\Omega_j(\theta) + \Omega_j(\theta)^\intercal\right),$$

and

$$\Omega_j(\theta) = \mathbb{E}[g(v_{l\Delta}; \theta) g(v_{(l+j)\Delta}; \theta)^\intercal], \text{ for } j = 0, 1, 2, \ldots, n-1.$$

A consistent estimator of the matrix $\hat{V}(\theta)$ is given by

$$\hat{V}(\theta) = \hat{G}(\theta)^\intercal \hat{\Omega}^{-1}(\theta) \hat{G}(\theta), \text{ with } \hat{G}(\theta) = \frac{1}{n} \sum_{l=1}^{n} \frac{\partial g\left(v_{l\Delta}, \theta\right)}{\partial \theta}. \tag{20}$$

In the exactly identified case, the matrix $\hat{\Omega}(\theta)$ is the Newey-West estimator with ℓ lags:

$$\hat{\Omega}(\theta) = \hat{\Omega}_0(\theta) + \sum_{j=1}^{\ell} \left(\frac{\ell + 1 - j}{\ell + 1}\right) \left(\hat{\Omega}_j(\theta) + \hat{\Omega}_j(\theta)^\intercal\right), \tag{21}$$

where

$$\hat{\Omega}_0(\theta) = \frac{1}{n} \sum_{l=1}^{n} g(v_{l\Delta}; \theta) g(v_{l\Delta}; \theta)^\intercal \text{ and } \hat{\Omega}_j(\theta) = \frac{1}{n} \sum_{l=j+1}^{n} g(v_{l\Delta}; \theta) g(v_{(l-j)\Delta}; \theta)^\intercal, \text{ for } j = 1, 2, \ldots \ell.$$

In principle, the number of lags ℓ grows with n at the rate $\ell = \mathcal{O}(n^{1/3})$. In the over-identified case, the optimal choice of W_n ought to be a consistent estimator of $\hat{\Omega}^{-1}(\theta_0)$. For this, the estimator $\hat{\theta}$ is obtained by the following two steps: First, set the initial weight matrix W_n in (18) as the identity matrix and arrive at a consistent estimator $\tilde{\theta}$. Second, compute $\hat{\Omega}(\tilde{\theta})$ according to (21), so that its inverse $\hat{\Omega}^{-1}(\tilde{\theta})$ is a consistent estimator of $\Omega^{-1}(\theta_0)$. Then set the weight matrix W_n in (18) as $\hat{\Omega}^{-1}(\tilde{\theta})$ and update the estimator to $\hat{\theta}$.

We provide below in Section 5.1 an example showing how to construct a Heston implied stochastic volatility model, and the results of Monte Carlo simulations where the model is either exactly

identified or over-identified. We find that for each parameter, the bias of the estimator is less than the corresponding finite-sample standard deviation and that the estimator $\sqrt{\hat{V}^{-1}(\hat{\theta})/n}$ of the asymptotic standard deviations, calculated according to (20), provides a reliable way of approximating standard errors for the parameters.

4 Constructing a nonparametric implied stochastic volatility model

We now turn to the case where no parametric form is assumed for the coefficient functions $\mu(\cdot)$, $\gamma(\cdot)$, and $\eta(\cdot)$ of the SV model, and show how the coefficients of the IV expansion (10) can be employed to recover them.

Theorem 1 can now be employed to construct the following explicit nonparametric estimation method for SV models. As in the parametric case of Section 3, the data for the four expansion terms $\sigma^{(0,0)}$, $\sigma^{(0,1)}$, $\sigma^{(0,2)}$, and $\sigma^{(1,0)}$ are regarded as input and obtained by a polynomial regression (16) of IV on time-to-maturity and log-moneyness. As in (15), we denote by $v_{l\Delta} = [\sigma^{(0,0)}]_l^{\text{data}}$, $[\sigma^{(0,1)}]_l^{\text{data}}$, $[\sigma^{(0,2)}]_l^{\text{data}}$, and $[\sigma^{(1,0)}]_l^{\text{data}}$ these data at time $l\Delta$.

To estimate $\gamma(\cdot)$ nonparametrically, we rely on (14a). Let

$$[\gamma]_l^{\text{data}} = 2[\sigma^{(0,0)}]_l^{\text{data}}[\sigma^{(0,1)}]_l^{\text{data}}, \tag{22}$$

and consider the nonparametric regression

$$[\gamma]_l^{\text{data}} = \gamma(v_{l\Delta}) + \epsilon_l, \tag{23}$$

where $v_{l\Delta}$ is the explanatory variable, and ϵ_l represents the exogenous observation error. The function $\gamma(\cdot)$ can be estimated based on (23) using a local polynomial kernel regression (see, e.g., Fan and Gijbels (1996).)

To estimate the coefficient functions $\eta(\cdot)$ and $\mu(\cdot)$, we implement the closed-form relations (13b)–(13c).[4] Note that these equations require to estimate both the function γ and its derivative γ'. One advantage of local polynomial kernel regression is that it provides in one pass not only an estimator of the regression function but also of its derivative(s). Consider specifically locally linear kernel regression. For two arbitrary points v and w, suppose that $\gamma(w)$ can be approximated by its first order Taylor expansion around $w = v$, i.e., $\gamma(w) \approx \gamma(v) + \gamma'(v)(w - v)$. Then, for any arbitrary value v of the independent variable, $[\gamma]_l^{\text{data}}$ is regarded as being approximately generated from the local linear regression as follows:

$$[\gamma]_l^{\text{data}} \approx \alpha_0 + \alpha_1(v_{l\Delta} - v) + \epsilon_l,$$

where the localization argument makes the intercept α_0 and slope α_1 coincide with γ and its first

[4] It is mathematically equivalent to implement the closed-form formulae (14b)–(14c) in Theorem 1.

order derivative γ' evaluated at v, respectively, i.e.,

$$\hat{\gamma}(v) = \hat{\alpha}_0 \text{ and } \hat{\gamma}'(v) = \hat{\alpha}_1.^5$$

The estimators $\hat{\alpha}_0$ and $\hat{\alpha}_1$ are obtained from the following weighted least squares minimization problem

$$(\hat{\alpha}_0, \hat{\alpha}_1) = \underset{\alpha_0,\alpha_1}{\operatorname{argmin}} \sum_{l=1}^{n} ([\gamma]_l^{\text{data}} - \alpha_0 - \alpha_1(v_{l\Delta} - v))^2 \mathcal{K}\left(\frac{v_{l\Delta} - v}{h}\right), \tag{24}$$

where $\mathcal{K}$ denotes a kernel function and h the bandwidth. In practice, we use the Epanechnikov kernel

$$\mathcal{K}(z) = \frac{3}{4}(1 - z^2)1_{\{|z|<1\}},$$

and a bandwidth h selected either by the standard rule of thumb or by standard cross-validation, which minimizes the sum of leave-one-out squared errors. The sum of leave-one-out squared errors, e.g., for the volatility function γ, is given by $\sum_{l=1}^{n}([\gamma]_l^{\text{data}} - \hat{\alpha}_{0,-l})^2$, where $\hat{\alpha}_{0,-l}$ is the local linear estimator $\hat{\alpha}_0$, at $v = v_{l\Delta}$, obtained from the weighted least squares problem (24) but without using the lth observation $(v_{l\Delta}, [\gamma]_l^{\text{data}})$.[6]

Next, in light of (13b), we define

$$[\eta]_l^{\text{data}} = \left[2\left(6([\sigma^{(0,0)}]_l^{\text{data}})^3[\sigma^{(0,2)}]_l^{\text{data}} - 2[\sigma^{(0,0)}]_l^{\text{data}}\hat{\gamma}(v_{l\Delta})\hat{\gamma}'(v_{l\Delta}) + 3\hat{\gamma}(v_{l\Delta})^2\right)\right]^{-1/2},$$

given $[\sigma^{(0,0)}]_l^{\text{data}}$ and $[\sigma^{(0,2)}]_l^{\text{data}}$, i.e., those of the expansion terms $\sigma^{(0,0)}$ and $\sigma^{(0,2)}$, as well as the estimators of γ and γ' obtained previously. In practice, on the right hand side of the above equation, the quantity inside the bracket $[\cdot]^{-1/2}$ may take a negative value, owing to sampling noise in the data $[\sigma^{(0,0)}]_l^{\text{data}}$ and $[\sigma^{(0,2)}]_l^{\text{data}}$. To solve this problem, we work instead with $[\eta^2]_l^{\text{data}}$ defined as

$$[\eta^2]_l^{\text{data}} = \left[2\left(6([\sigma^{(0,0)}]_l^{\text{data}})^3[\sigma^{(0,2)}]_l^{\text{data}} - 2[\sigma^{(0,0)}]_l^{\text{data}}\hat{\gamma}(v_{l\Delta})\hat{\gamma}'(v_{l\Delta}) + 3\hat{\gamma}(v_{l\Delta})^2\right)\right]^{-1}. \tag{25}$$

We then estimate the coefficient functions $\eta^2(\cdot)$ at each value v by a kernel regression that localizes the data $[\eta^2]_l^{\text{data}}$ at each point $v = v_{l\Delta}$, as we did in (24) for $\gamma(\cdot)$. In our experience, the estimator $\hat{\eta}^2(\cdot)$ is always nonnegative thanks to the kernel smoothing (even though a small number of data points $[\eta^2]_l^{\text{data}}$ may be negative.) We then define $\hat{\eta}(\cdot) \equiv \left[\hat{\eta}^2(\cdot)\right]^{1/2}$.

[5] Note that $\hat{\gamma}'(v)$ is an estimator of $\gamma'(v)$ but is not the derivative of $\hat{\gamma}(v)$.

[6] For a choice of kernel function $\mathcal{K}$ with bandwidth h, the solution of the weighted least squares problem (24) is explicitly given by

$$\hat{\alpha}_0 = \left(\sum_{i,j=1}^{n} s_{ij}(v)(v_{i\Delta} - v)\right)^{-1}\left(\sum_{i,j=1}^{n} s_{ij}(v)(v_{i\Delta} - v)y_{j\Delta}\right) \text{ and } \hat{\alpha}_1 = -\left(\sum_{i,j=1}^{n} s_{ij}(v)(v_{i\Delta} - v)\right)^{-1}\left(\sum_{i,j=1}^{n} s_{ij}(v)y_{j\Delta}\right),$$

where

$$s_{ij}(v) = \mathcal{K}\left(\frac{v_{i\Delta} - v}{h}\right)\mathcal{K}\left(\frac{v_{j\Delta} - v}{h}\right)(v_{i\Delta} - v_{j\Delta}).$$

Finally, in light of (13c), we define

$$[\mu]_l^{\text{data}} = 2[\sigma^{(1,0)}]_l^{\text{data}} + \frac{\hat{\gamma}(v_{l\Delta})}{6}(2\hat{\gamma}'(v_{l\Delta}) - 3[\sigma^{(0,0)}]_l^{\text{data}}) - \frac{\hat{\eta}(v_{l\Delta})^2}{6[\sigma^{(0,0)}]_l^{\text{data}}} - \frac{\hat{\gamma}(v_{l\Delta})}{[\sigma^{(0,0)}]_l^{\text{data}}}\left(d - r + \frac{1}{4}\hat{\gamma}(v_{l\Delta})\right). \tag{26}$$

given the estimators of γ, γ', and η^2 obtained previously and estimate the coefficient function $\mu(\cdot)$ at each value v using on the data (26) the same kernel localization procedure (24) as employed for $\hat{\gamma}(\cdot)$ and $\hat{\eta}^2(\cdot)$.

5 Monte Carlo simulation results

In this Section, we conduct Monte Carlo simulations to determine whether the coefficient functions of the SV model can be accurately recovered, either parametrically or nonparametrically, using the methods we proposed in Sections 3 and 4.

5.1 An implied Heston model

Consider first the parametric case, which we illustrate with the SV model of Heston (1993). Under the assumed risk-neutral measure, the underlying asset price S_t and its spot variance $V_t = v_t^2$ follow

$$\frac{dS_t}{S_t} = (r-d)dt + \sqrt{V_t}dW_{1t}, \tag{27a}$$

$$dV_t = \kappa(\alpha - V_t)dt + \xi\sqrt{V_t}[\rho dW_{1t} + \sqrt{1-\rho^2}dW_{2t}], \tag{27b}$$

where W_{1t} and W_{2t} are independent standard Brownian motions. Here, the parameter vector is $\theta = (\kappa, \alpha, \xi, \rho)$ and we assume that Feller's condition holds: $2\kappa\alpha > \xi^2$. The leverage effect parameter is $\rho \in [-1, 1]$.

To estimate the four parameters in $\theta = (\kappa, \alpha, \xi, \rho)$, we successively employ the four moment conditions in $g = (g^{(1,0)}, g^{(0,1)}, g^{(0,2)}, g^{(1,1)})^\intercal$ to exactly identify the parameters or employ the five moment conditions in $g = (g^{(1,0)}, g^{(0,1)}, g^{(0,2)}, g^{(1,1)}, g^{(2,0)})^\intercal$ to over-identify the parameters. We impose $\alpha > 0$, $\kappa > 0$, $\xi > 0$ and Feller's condition as constraints during the GMM minimization (18).

Itô's lemma applied to $v_t = \sqrt{V_t}$ yields

$$\mu(v) = \frac{\kappa(\alpha - v^2)}{2v} - \frac{\xi^2}{8v},\ \gamma(v) = \frac{\xi\rho}{2},\ \eta(v) = \frac{\xi\sqrt{1-\rho^2}}{2}. \tag{28}$$

Then, applying the results of Section 2.1 and the general method for deriving higher orders in Appendix A, we can calculate the expansion terms $\sigma^{(0,1)}(v)$, $\sigma^{(0,2)}(v)$, $\sigma^{(1,0)}(v)$, $\sigma^{(1,1)}(v)$, and $\sigma^{(2,0)}(v)$:

$$\sigma^{(0,0)}(v) = v,\ \sigma^{(0,1)}(v) = \frac{\rho\xi}{4v},\ \sigma^{(0,2)}(v) = -\frac{1}{48v^3}\left(5\rho^2 - 2\right)\xi^2, \tag{29}$$

and

$$\begin{aligned}
\sigma^{(1,0)}(v) &= \frac{1}{96v}\left(\xi\left(24\rho(d-r)+\xi\left(\rho^2-4\right)\right)+v^2(12\xi\rho-24\kappa)+24\kappa\alpha\right),\\
\sigma^{(1,1)}(v) &= -\frac{\xi}{384v^3}\left(16\left(2-5\rho^2\right)(r-d)\xi+\rho\left(40\kappa\alpha+3\left(3\rho^2-4\right)\xi^2+v^2(4\rho\xi-8\kappa)\right)\right),\\
\sigma^{(2,0)}(v) &= \frac{1}{30720v^3}\left[\xi^2\left(-640(r^2+d^2)\left(5\rho^2-2\right)+80d\left(3\rho\left(4-3\rho^2\right)\xi+16\left(5\rho^2-2\right)r\right)\right.\right.\\
&\quad +\left(59\rho^4-88\rho^2-16\right)\xi^2+240\rho\left(3\rho^2-4\right)r\xi\Big)+320v^4\left(5\kappa^2-5\kappa\rho\xi+\left(2\rho^2-1\right)\xi^2\right)\\
&\quad -80\kappa\alpha\xi\left(40d\rho+\left(5\rho^2-8\right)\xi-40\rho r\right)-40v^2(2\kappa-\rho\xi)\left(\xi\left(-8d\rho+3\rho^2\xi+8\rho r\right.\right.\\
&\quad \left.\left.\left.-4\xi\right)+8\kappa\alpha\right)-960\kappa^2\alpha^2\right].
\end{aligned}$$

We now generate a time series of (S_t, V_t) with $n = 1,000$ consecutive samples at the daily frequency, i.e., with time increment $\Delta = 1/252$, by subsampling higher frequency data simulated using the Euler scheme. The parameter values are $r = 0.03$, $d = 0$, $\kappa = 3$, $\alpha = 0.04$, $\xi = 0.2$, and $\rho = -0.7$. Each day, we calculate option prices with time-to-maturity τ equal to 5, 10, 15, 20, 25, and 30 days and for each time-to-maturity τ, include 20 log-moneyness values k within $\pm v_t\sqrt{\tau}$, where τ is annualized and v_t is the spot volatility. The principles for judiciously choosing such a region of (τ, k) for simulation will be intensively discussed in the next paragraph. Due to the affine nature of the model of Heston (1993), these option prices can be calculated by Fourier transform inversion and compute the corresponding IV values. To mimic a realistic market scenario, we add observation errors to these implied volatilities, sampled from a Normal distribution with mean zero and constant standard deviation equals to 15 bps and further assumed to be uncorrelated across time-to-maturity and log-moneyness, as well as over time. Then, for each IV surface, we follow the regression procedure described around (16) to extract the estimated coefficients $\hat{\beta}_l^{(i,j)}$ of the bivariate regression (16).

In practice, one needs to choose the orders $J, L_0, L_1, \ldots, L_J$ in the bivariate polynomial regression in (16) and the region in (τ, k) of the IV surface data to compute the regression. On the one hand, we need at a minimum to include enough orders in the regression to estimate the coefficients of interest for the estimation method; recall that we need the terms $\sigma^{(0,0)}$, $\sigma^{(0,1)}$, $\sigma^{(0,2)}$, $\sigma^{(1,0)}$, and $\sigma^{(1,1)}$ for constructing an exactly identified Heston model, and need to include an additional term $\sigma^{(2,0)}$ for constructing an over-identified one. But we can consistently estimate all these lower order coefficients from a higher order regression, discarding the estimates of the higher order coefficients. On the other hand, the orders cannot be chosen as too high and the region in (τ, k) cannot be chosen as too narrow to avoid over-fitting the regression. Specifically, we set the order to be $(J, \mathbf{L}(J)) = (2, (2, 2, 1))$, so:

$$\begin{aligned}
\Sigma^{\text{data}}(\tau_l^{(m)}, k_l^{(m)}) &= \beta_l^{(0,0)} + \beta_l^{(1,0)}\tau_l^{(m)} + \beta_l^{(2,0)}(\tau_l^{(m)})^2 + \beta_l^{(0,1)}k_l^{(m)} + \beta_l^{(1,1)}\tau_l^{(m)}k_l^{(m)}\\
&\quad + \beta_l^{(2,1)}(\tau_l^{(m)})^2k_l^{(m)} + \beta_l^{(0,2)}(k_l^{(m)})^2 + \beta_l^{(1,2)}\tau_l^{(m)}(k_l^{(m)})^2 + \epsilon_l^{(m)}, \qquad (30)
\end{aligned}$$

for $m = 1, 2, \ldots, n_l$. The estimated coefficients from this regression estimate the IV surface characteristics that we need (recall (17)).

We then implement the method proposed in Section 3 to estimate the model parametrically. We consider two cases. The first one is exactly identified using $g = (g^{(1,0)}, g^{(0,1)}, g^{(0,2)}, g^{(1,1)})^\intercal$, while the second adds one more moment condition, $g^{(2,0)}$, to over-identify the parameters. Table 1 summarizes the results. We find that, for each parameter, the absolute bias is relatively small and is less than the corresponding finite-sample standard deviation. In the exactly identified (resp. over-identified) case, we compare for each parameter the finite-sample standard deviation exhibited in the fourth (resp. sixth) column of Table 1 with the consistent estimator of its asymptotic counterpart, based on $\hat{V}(\hat{\theta})$ given in (20). Figure 2 (resp. 3) compares the finite-sample standard deviation for each parameter with the distribution of sample-based asymptotic counterparts in the exactly identified (resp. over-identified) case. Consider the upper left panel of Figure 2 as an example. The histogram characterizes the distribution of sample-based asymptotic standard deviation $\sqrt{\hat{V}_{11}^{-1}(\hat{\theta})/n}$ for parameter κ, where $\hat{V}_{11}^{-1}$ represents the $(1,1)$th entry of matrix $\hat{V}^{-1}$. The red star marks the corresponding finite-sample standard deviation shown in the fourth cell from the first row of Table 1. As shown from Figures 2 and 3, for each parameter, the finite-sample standard deviation falls within the range of its sampled-based asymptotic counterparts in both cases. As the sample size further increases, the finite-sample standard deviation and its sampled-based asymptotic counterpart tend to converge to each other, and shows that the sampled-based approximation $\sqrt{\hat{V}^{-1}(\hat{\theta})/n}$ of the asymptotic standard deviations is a reasonable estimator of the standard errors.

5.2 Nonparametric implied stochastic volatility model

Next, we apply the nonparametric method of Section 4 to the simulated data that was generated under the Heston model. In Figure 4, the upper left, upper right, middle left, and middle right panels exhibit the results for nonparametrically estimating the functions μ, $-\gamma$, η^2, and η of model (1a)–(1b), respectively. Consider the upper left panel for the function μ. We perform local polynomial regression at equidistantly distributed values of v in the interval $[0.1, 0.3]$. For each $v \in [0.1, 0.3]$, we mark the true value of $\mu(v)$ by a blue dot, according to its equation given in (28), and plot the mean of estimators of $\mu(v)$ on a black solid curve. Then, we generate each point on the upper (resp. lower) dashed curve from vertically upward (resp. downward) shifting the corresponding one on the mean curve by a distance equal to twice of the corresponding finite-sample standard deviation. As seen from the figure, the shape of estimated nonparametric function resembles that of the true one on average. Besides, the two dashed curves sandwich the true curve. This indicates that, at each point of interest, the nonparametric estimator is sufficiently close to the true value that the estimation bias is less than twice of the corresponding standard deviation.

We then combine the estimators $\hat{\gamma}(\cdot)$ and $\hat{\eta}^2(\cdot)$ to estimate the leverage effect $\rho(v_t)$ under the

nonparametric implied stochastic volatility model (1a)–(1b) by

$$\hat{\rho}(v_t) = \frac{\hat{\gamma}(v_t)}{\sqrt{\hat{\gamma}(v_t)^2 + \hat{\eta}(v_t)^2}}. \tag{31}$$

As in the other four panels of Figure 4, we exhibit the estimation results for $\rho(v)$ in the lowest panel. We find that the shape of the estimated function $\rho(v)$ is approximately constant at the level of parameter ρ, as it ought to be under the model of Heston (1993).

We propose in what follows a bootstrap estimator of standard error. It is based on multiple bootstrap replications out of one simulation trial, for mimicking an empirical estimation scenario. In each bootstrap replication, we reproduce an IV surface for each day. The reproduced surface contains the same number of implied volatilities as that of the original surface, and the implied volatilities on the reproduced surface are sampled as i.i.d. replications of the volatilities on the original surface. Based on the bootstrap "data", we apply the same estimation method proposed in Section 4 to obtain the bootstrap estimators of functions $\mu(\cdot)$, $-\gamma(\cdot)$, $\eta^2(\cdot)$, $\eta(\cdot)$, and $\rho(\cdot)$. The bootstrap standard error of each function is accordingly calculated as the standard deviations of its multiple bootstrap estimators.

To validate this method, which we will employ below in real data, we randomly select one simulation trial and calculate the bootstrap standard error of each coefficient function out of 500 corresponding bootstrap estimators. Figure 5 summarizes the estimation result of this trial. In each panel of Figure 5, the blue dotted (resp. black solid) curve represents the true function (resp. nonparametric estimator.) Each point on the upper (resp. lower) dashed curve is plotted from vertically upward (resp. downward) shifting the corresponding one on the black solid curve by a distance equal to twice of the corresponding bootstrap standard error. Figure 5 suggests that our nonparametric estimators are all accurate, as they are close to the corresponding true functions. More importantly, the bootstrap method appears to be valid from a comparison of each panel in Figure 5 with the corresponding one in Figure 4. Compare the upper right panels of Figures 5 and 4 as an example. For any v, the lengths of intervals bounded by the two dashed curves in these two panels are close to each other. Thus, the bootstrap standard errors seem to provide a reliable way for calculating standard deviations in the coming empirical analysis.

6 Empirical results

We now employ S&P 500 options data covering the period from January 2, 2013 to December 29, 2017, obtained from OptionMetrics. Guided by the simulations evidence discussed above, we select options with time-to-maturity between 15 and 60 calendar days, thereby excluding both extremely short-maturity ones which are subject to significant trading effects and biases, and long-maturity ones for which the IV expansion becomes less accurate. Table 2 reports the basic descriptive statistics of

the sample of 269,622 observations. Table 2 divides the data into three (calendar) days-to-expiration categories and six log-moneyness categories. For each category, we report the total number, mean, and standard deviation of implied volatilities therein.

As in the simulations, we implement each day the regression (16) of implied volatilities with time-to-maturities between 15 and 60 days, and log-moneyness within $\pm v_t\sqrt{\tau}$. Here, τ is the annualized time-to-maturity and v_t is the instantaneous volatility, which is estimated by the observed IV with both the time-to-maturity τ and the log-moneyness k closest to 0 on that day. We run the regression only if at least four different time-to-maturities between 15 and 60 days are available; otherwise, we do not include that day in the sample. We end up with $n = 1,002$ IV surfaces at the daily frequency $\Delta = 1/252$. Moreover, for choosing the order of polynomial regression (16), a reasonable compromise is to set $(J, \mathbf{L}(J)) = (2, (2, 2, 0))$, i.e.,

$$\begin{aligned}\Sigma^{\text{data}}(\tau_l^{(m)}, k_l^{(m)}) = \beta_l^{(0,0)} + \beta_l^{(1,0)}\tau_l^{(m)} + \beta_l^{(2,0)}(\tau_l^{(m)})^2 + \beta_l^{(0,1)}k_l^{(m)} \\ + \beta_l^{(1,1)}\tau_l^{(m)}k_l^{(m)} + \beta_l^{(2,1)}(\tau_l^{(m)})^2k_l^{(m)} + \beta_l^{(0,2)}(k_l^{(m)})^2 + \epsilon_l^{(m)}. \qquad (32)\end{aligned}$$

Comparing with the bivariate regression (30) employed in the Monte Carlo experiments, we extend the time-to-maturity τ of the employed IV data to 60 days owing the deficiency of data with τ less than 30 days in practice, and remove a high order regression coefficient $\beta_l^{(1,2)}$ to reduce the standard errors of the estimators of other low order coefficients without loss of accuracy.

Figure 6 plots a histogram of the R^2 values achieved by the parametric regressions (32) across the full sample of IV surfaces. We find that for over 95% of the sample the R^2 are greater than 0.96, and essentially none are lower than 0.90, suggesting that (32) is quite successful at fitting the data. Incidentally, practitioners often use polynomial regression to fit the short-maturity near at-the-money region of the IV surfaces in their own internal models[7], so it is not surprising that the market data we collect end up reflecting this feature. As an example, Figure 7 plots the IV data and the corresponding fitted surface produced by regression (32) on a randomly selected day in our sample (January 3, 2017).

6.1 Parametric implied stochastic volatility model

We now implement the method of Section 3 to estimate a parametric implied stochastic volatility model of the Heston (1993) type. Table 3 reports the GMM results for both of the exactly identified and over-identified cases. First, the estimators of ρ are around -0.6 in both cases, consistent with what can be heuristically inferred directly from the $[\sigma^{(0,0)}]^{\text{data}}$ and $[\sigma^{(0,2)}]^{\text{data}}$, depicted by the corresponding histograms in Figure 8. The mean and standard deviation of the multiplicative data $([\sigma^{(0,0)}]^{\text{data}})^3[\sigma^{(0,2)}]^{\text{data}}$ are 1.55×10^{-4} and 7.19×10^{-3}, respectively. Thus, there is no evidence for

[7] See, e.g., Gatheral (2006).

the mean of $([\sigma^{(0,0)}]^{\text{data}})^3[\sigma^{(0,2)}]^{\text{data}}$ to be significantly different from zero. On the other hand, it follows from the closed-form formulae for $\sigma^{(0,0)}(v)$ and $\sigma^{(0,2)}(v)$ given in (29) that

$$\sigma^{(0,0)}(v)^3\sigma^{(0,2)}(v) = -\frac{1}{48}\left(5\rho^2 - 2\right)\xi^2.$$

Heuristically, moment matching by equating the estimated zero mean requires $-\left(5\rho^2 - 2\right)\xi^2/48 = 0$. This would approximately estimate ρ as -0.63, independently of the values of v and ξ.

Second, the estimator of ξ is around 1 (resp. 0.8) in the exactly identified (resp. over-identified) case. We find that, for both of these two cases, the estimators of ξ are somewhat greater than those in the literature, which are usually less than 0.55 (see, e.g., Eraker et al. (2003), Aït-Sahalia and Kimmel (2007), and Christoffersen et al. (2010) among others.) As pointed out in, e.g., Eraker et al. (2003), the Heston model tends to underestimate the slope of the IV smile with small estimators of ξ. However, our implied stochastic volatility model forces the model to fit this slope by construction. Recall that the closed-form formula for the slope $\sigma^{(0,1)}$, given in (29), is $\sigma^{(0,1)}(v) = \rho\xi/(4v)$. Thus, for fitting the usually steep slope, the corresponding moment condition requires (given ρ) ξ to be larger than other methods, and this is what our GMM estimation procedure produces. Furthermore, based on the data $[\sigma^{(0,0)}]^{\text{data}}$ and $[\sigma^{(0,1)}]^{\text{data}}$ shown in Figure 8, the mean of the multiplicative data $[\sigma^{(0,0)}]^{\text{data}}[\sigma^{(0,2)}]^{\text{data}}$ is -0.17. On the other hand, it follows from (29) that

$$\sigma^{(0,0)}(v)\sigma^{(0,1)}(v) = \frac{\rho\xi}{4}.$$

Similar to the aforementioned determination of ρ via the heuristic moment matching, we plug the estimated mean -0.17 of the multiplicative data and the estimator -0.619 of ρ as shown Table 3 into the above formula to solve the parameter ξ as 1.1, which basically agrees with our GMM estimator.

Third, in both the exactly identified and over-identified cases, the estimators for κ are larger than 10, which are larger values than those estimated in the literature. This is necessary given the large values of the volatility of variance ξ, to keep the volatility process v_t mean-reverting sufficiently fast and consequently diminish the likelihood of having extreme volatilities. Fourth, again in both cases, the estimators of the long term variance value α are around 0.02, which is consistent with the low values recorded by the S&P 500 volatility during the sample period.

6.2 Nonparametric implied stochastic volatility model

Using the same data, and the same expansion terms $\sigma^{(0,0)}$, $\sigma^{(0,1)}$, $\sigma^{(0,2)}$, and $\sigma^{(1,0)}$ estimated from (32), we now follow the method proposed in Section 4 to construct a nonparametric implied stochastic volatility model.

The results are summarized in Figure 9. The upper left, upper right, and middle left panels show the estimators $\hat{\mu}(\cdot)$, $-\hat{\gamma}(\cdot)$, and $\hat{\eta}^2(\cdot)$, respectively. The different elements in each panel are as follows. Consider the upper left panel. The dots represent realizations of $[\mu]^{\text{data}}$, which we recall are

calculated according to (26) while the nonparametric estimator of the function μ is shown as the solid curve. The confidence intervals on the curve are pointwise and represent two standard errors. The standard errors are calculated by a standard bootstrap procedure based on 500 bootstrap replications as proposed and validated in Section 5.2. Given the nonparametric estimator of η^2 obtained based on the real sample (resp. bootstrap sample), we calculate the corresponding estimator of η by taking a square root. The result for this calculation are presented in the middle right panel. Finally, given the estimators of γ and η^2 based on the real sample (resp. bootstrap sample), we calculate the corresponding estimator of the leverage effect function ρ, i.e., $\hat{\rho}(v_t) = \hat{\gamma}(v_t)/\sqrt{\hat{\gamma}(v_t)^2 + \hat{\eta}(v_t)^2}$. The results are shown in the lowest panel. Likewise, the standard error of the estimators $\hat{\eta}$ (resp. $\hat{\rho}$) is calculated by the standard deviation of 500 bootstrap estimators of η (resp. ρ.)

We find that $\hat{\mu}(\cdot)$ is positive (resp. negative) when its argument is relatively small (resp. large), consistent with mean reversion in v_t. The upper right panel indicates that the coefficient function $\hat{\gamma}(\cdot)$ is always negative (the upper right panel shows $-\hat{\gamma}(\cdot)$) and approximately linear. As shown in the middle right panel, $\hat{\eta}(\cdot)$ is always positive and concave, as opposed to being approximately linear as γ is. The leverage effect estimator $\hat{\rho}(\cdot)$ is consistently negative, non constant, and more negative when v_t increases. The negativeness of $\rho(\cdot)$ is a direct consequence of that of γ given (2). This non constant shape of $\hat{\rho}(\cdot)$ versus v_t implies that the leverage effect $\rho(v_t)$ is indeed stochastic, unlike the assumption in the Heston model.

Finally, we verify that the goodness-of-fit of the expansion terms $\sigma^{(i,j)}$ involved in (32). In each panel of Figure 10, we plot the data of expansion term $\sigma^{(i,j)}$ as well as its fitted values $\hat{\sigma}^{(i,j)}$. Here, the data are inferred from bivariate regression (32), while the fitted values $\hat{\sigma}^{(i,j)}$ are obtained by plugging in $\hat{\mu}$, $\hat{\gamma}$ and $\hat{\eta}$, as well as $\hat{\gamma}'$ (which we recall is estimated at the same time as $\hat{\gamma}$ by locally linear kernel regression) in the corresponding formula for $\sigma^{(i,j)}$ given in (11)–(12). The fitted expansion terms $\hat{\sigma}^{(0,1)}$, $\hat{\sigma}^{(0,2)}$, and $\hat{\sigma}^{(1,0)}$ match the data well, which is expected since they are inputs in the construction. Surprisingly, however, we find that the fitted expansion terms $\hat{\sigma}^{(1,1)}$ and $\hat{\sigma}^{(2,0)}$ also match the data well, as shown in the middle right and lowest panels of Figure 10, even though the expansion terms $\sigma^{(1,1)}$ and $\sigma^{(2,0)}$ (corresponding to the mixed slope $\Sigma_{1,1}$ and term-structure convexity $\Sigma_{2,0}$ of the IV surface up some constants according to (8)) are not employed in the nonparametric construction of the implied model, and require higher order derivatives of the coefficient functions. This indicates that the nonparametric implied stochastic volatility model is flexible enough to reproduce all the second order shape characteristics of IV surface, or equivalently that all the shape characteristics of the IV surface up to the second order are consistent with the nonparametric implied stochastic volatility model. All the aforementioned six shape characteristics, that our implied model fits, are more than enough for characterizing an IV surface in the short-maturity and near at-the-money region.

7 Extension: Adding jumps to the model

We now generalize our approach to include jumps in returns to the model (1a)–(1b):

$$\frac{dS_t}{S_{t-}} = (r - d - \lambda(v_t)\bar{\mu})dt + v_t dW_{1t} + (\exp(J_t) - 1)dN_t, \tag{33a}$$

$$dv_t = \mu(v_t)dt + \gamma(v_t)dW_{1t} + \eta(v_t)dW_{2t}. \tag{33b}$$

N_t is a doubly stochastic Poisson process (or Cox process) with stochastic intensity $\lambda(v_t)$. J_t represents the size of log-price jump, which is assumed to be independent of the asset price S_t. When a jump occurs at time t, the log-price $\log S_t$ changes according to $\log S_t - \log S_{t-} = J_t$, i.e., $S_t - S_{t-} = (\exp(J_t) - 1)S_{t-}$, where S_{t-} denotes the pre-jump price of the asset. The constant $\bar{\mu}$ is

$$\bar{\mu} = \mathbb{E}[\exp(J_t)] - 1,$$

where $\mathbb{E}$ denotes risk-neutral expectation. Based on this choice of $\bar{\mu}$, the drift term $-\lambda(v_t)\bar{\mu}dt$ compensates the jump component $(\exp(J_t)-1)dN_t$ in the sense that the process $\int_0^t(\exp(J_s)-1)dN_s - \int_0^t \lambda(v_s)\bar{\mu}ds$ becomes a martingale under the risk-neutral measure.

A typical example (as in Merton (1976)) is one where the jump size J_t is normally distributed with mean μ_J and variance σ_J^2, in which case

$$\bar{\mu} = \exp\left(\mu_J + \frac{\sigma_J^2}{2}\right) - 1. \tag{34}$$

For future reference, we also define

$$\mu_+ = \frac{\mu_J}{\sigma_J} + \sigma_J \text{ and } \mu_- = \frac{\mu_J}{\sigma_J}, \tag{35}$$

and let $\mathcal{N}$ denote the standard Normal cumulative distribution function.

Adding jumps to the volatility dynamics, or infinite activity jumps to either returns or volatility dynamics, has the potential to improve the fit and realism of the model even further but would substantially alter the approach we employ to derive the IV expansion. So for now we consider only the case of jumps in returns and leave these further extensions to future work.

7.1 The effect of jumps on the implied volatility expansion

Following the same analysis as in Section 2, it is straightforward to see that the IV Σ remains as in the continuous model a trivariate function of τ, k, and v_t in the form given by (5). A generalization of the expansion (7) of the IV surface $\Sigma(\tau, k, v_t)$ to the case of jumps will now incorporate the square root of time-to-maturity $\sqrt{\tau}$, as well as possibly its negative powers

$$\Sigma^{(J,\mathbf{L}(J))}(\tau, k, v_t) = \sum_{j=0}^{J} \sum_{i=\min(0,1-j)}^{L_j} \varphi^{(i,j)}(v_t)\tau^{\frac{i}{2}}k^j, \tag{36}$$

where J and $\mathbf{L}(J) = (L_0, L_1, \cdots, L_J)$ with $L_j \geq \min(0, 1-j)$ are integers. Expansion (36) includes negative powers of $\sqrt{\tau}$ if the lowest power $\min(0, 1-J)$ of $\sqrt{\tau}$ in the double summation is less than or equal to -1, i.e., if $J \geq 2$. With the presence of jumps in return, the away-from-the-money IV will possibly explode to infinity as the time-to-maturity shrinks to zero: this limiting behavior was noted by Carr and Wu (2003), who used this divergence to construct a test for the presence of jumps in the data, and by Andersen and Lipton (2013).

We show in Appendix B that with Normally distributed jumps, the $(3, (2, 1, 0, -2))$th order expansion of (36) is given by

$$\begin{aligned}\Sigma^{(3,(2,1,0,-2))}(\tau, k, v_t) = \varphi^{(0,0)}(v_t) + \varphi^{(1,0)}(v_t)\tau^{\frac{1}{2}} + \varphi^{(2,0)}(v_t)\tau + \varphi^{(0,1)}(v_t)k + \varphi^{(1,1)}(v_t)\tau^{\frac{1}{2}}k \\ + \varphi^{(-1,2)}(v_t)\tau^{-\frac{1}{2}}k^2 + \varphi^{(0,2)}(v_t)k^2 + \varphi^{(-2,3)}(v_t)\tau^{-1}k^3,\end{aligned}$$

where the $(0,0)$th order term $\varphi^{(0,0)}(v_t)$ coincides with the spot volatility v_t, and the closed-form formulae of all other terms are given by

$$\varphi^{(-1,2)}(v_t) = \frac{\lambda(v_t)\sqrt{\pi}}{2\sqrt{2}v_t^2}(-\bar{\mu} + 2(\bar{\mu}+1)\mathcal{N}(\mu_+) - 2\mathcal{N}(\mu_-)),\ \varphi^{(-2,3)}(v_t) = \frac{\lambda(v_t)\bar{\mu}}{3v_t^2}, \tag{37a}$$

$$\varphi^{(0,1)}(v_t) = \frac{1}{2v_t}[2\lambda(v_t)\bar{\mu} + \gamma(v_t)],\ \varphi^{(1,0)}(v_t) = 2v_t^2\varphi^{(-1,2)}(v_t), \tag{37b}$$

and

$$\begin{aligned}\varphi^{(1,1)}(v_t) = \frac{\sqrt{\pi}\lambda(v_t)}{2\sqrt{2}v_t^2}[2(r-d)\bar{\mu} + 2(\bar{\mu}+1)\mathcal{N}(\mu_+)(2(d-r) - v_t^2) + \bar{\mu}v_t^2 + 2v_t^2 \\ - 2\mathcal{N}(\mu_-)(2(d-r) + v_t^2)],\end{aligned} \tag{37c}$$

$$\begin{aligned}\varphi^{(0,2)}(v_t) = \frac{1}{12v_t^3}[-3\gamma(v_t)^2 + 2v_t\gamma(v_t)\gamma'(v_t) - 3\pi\lambda(v_t)^2(\bar{\mu} - 2(\bar{\mu}+1)\mathcal{N}(\mu_+) + 2\mathcal{N}(\mu_-))^2 \\ + 2\eta(v_t)^2 + 6\lambda(v_t)(\bar{\mu}(2(d-r) - \gamma(v_t)) - 2(\bar{\mu}+2)v_t^2)],\end{aligned} \tag{37d}$$

as well as

$$\begin{aligned}\varphi^{(2,0)}(v_t) = \frac{1}{24v_t}[6v_t^2\gamma(v_t) + 2\eta(v_t)^2 + 12\lambda(v_t)(\bar{\mu}(2(d-r) + \gamma(v_t)) - (\bar{\mu}+2)v_t^2) + 3\gamma(v_t)^2 \\ + 12(d-r)\gamma(v_t) + 2v_t(6\mu(v_t) - 2\gamma(v_t)\left(3\bar{\mu}\lambda'(v_t) + \gamma'(v_t)\right)) + 12\bar{\mu}^2\lambda(v_t)^2].\end{aligned} \tag{37e}$$

Note that if we set the jump intensity function $\lambda(v)$ to zero, the expansion (36) reduces to the expansion (7) under the continuous SV model: under the model (1a)–(1b), the expansion term $\varphi^{(i,j)}(v)$ is identically zero for any negative or odd integer i and the expansion term $\varphi^{(i,j)}(v)$ coincides with $\sigma^{(i/2,j)}(v)$ for any nonnegative even integer i.

7.2 Example: The Merton jump-diffusion model

In the special case of the jump-diffusion model of Merton (1976):

$$\frac{dS_t}{S_{t-}} = (r - d - \lambda\bar{\mu})dt + v_0 dW_t + (\exp(J_t) - 1)dN_t,$$

where λ represents a constant jump intensity and v_0 a constant volatility. We obtain expansion (36) under this model simply by letting the SV components be zero and let the jump intensity function be the constant λ, i.e.,

$$v_t = v_0 \text{ and } \lambda(v_t) = \lambda. \tag{38}$$

The expansion terms $\varphi^{(0,1)}(v_t)$, $\varphi^{(-1,2)}(v_t)$, and $\varphi^{(1,1)}(v_t)$ reduce to

$$\varphi^{(0,1)}(v_0) = \frac{\lambda\bar{\mu}}{v_0},\ \varphi^{(-1,2)}(v_0) = \frac{\lambda\sqrt{\pi}}{2\sqrt{2}v_0^2}(-\bar{\mu} + 2(\bar{\mu}+1)\mathcal{N}(\mu_+) - 2\mathcal{N}(\mu_-)),$$

and

$$\begin{aligned}\varphi^{(1,1)}(v_0) = \frac{\sqrt{\pi}\lambda}{2\sqrt{2}v_0^2}[&2(r-d)\bar{\mu} + 2(\bar{\mu}+1)\mathcal{N}(\mu_+)(2(d-r) - v_0^2) + \bar{\mu}v_0^2 + 2v_0^2 \\ &- 2\mathcal{N}(\mu_-)(2(d-r) + v_0^2)],\end{aligned}$$

from the general formulae provided in (37b), (37a), and (37c), respectively.

7.3 From implied volatility to stochastic volatility and jumps

Returning to the general model, the terms $\varphi^{(i,j)}$ correspond to at-the-money IV shape characteristics or combinations thereof, as the time-to-maturity shrinks to zero, up to time scalings. For instance, the expansion terms $\varphi^{(0,0)}(v_t)$, $\varphi^{(-2,3)}(v_t)$, $\varphi^{(0,1)}(v_t)$, $\varphi^{(1,1)}(v_t)$, and $\varphi^{(-1,2)}(v_t)$ satisfy

$$\varphi^{(0,0)}(v_t) = \lim_{\tau\to 0}\Sigma(\tau,0,v_t),\ \varphi^{(-2,3)}(v_t) = \lim_{\tau\to 0}\frac{\tau}{6}\frac{\partial^3\Sigma}{\partial k^3}(\tau,0,v_t),\ \varphi^{(0,1)}(v_t) = \lim_{\tau\to 0}\frac{\partial\Sigma}{\partial k}(\tau,0,v_t), \tag{39a}$$

$$\varphi^{(1,1)}(v_t) = \lim_{\tau\to 0}2\sqrt{\tau}\frac{\partial^2\Sigma}{\partial\tau\partial k}(\tau,0,v_t),\ \varphi^{(-1,2)}(v_t) = \lim_{\tau\to 0}\frac{\sqrt{\tau}}{2}\frac{\partial^2\Sigma}{\partial k^2}(\tau,0,v_t), \tag{39b}$$

while the expansion terms $\varphi^{(0,2)}(v_t)$ and $\varphi^{(2,0)}(v_t)$ satisfy

$$\varphi^{(0,2)}(v_t) = \lim_{\tau\to 0}\left(\frac{1}{2}\frac{\partial^2\Sigma}{\partial k^2}(\tau,0,v_t) + \tau\frac{\partial^3\Sigma}{\partial k^2\partial\tau}(\tau,0,v_t)\right), \tag{39c}$$

and

$$\varphi^{(2,0)}(v_t) = \lim_{\tau\to 0}\left(\frac{\partial\Sigma}{\partial\tau}(\tau,0,v_t) + 2\tau\frac{\partial^2\Sigma}{\partial\tau^2}(\tau,0,v_t)\right). \tag{39d}$$

Formulae (39a)–(39d) hinge on the univariate expansions of at-the-money shape characteristics $\partial^{i+j}\Sigma(\tau,0,v_t)/\partial\tau^i\partial k^j$ with respect to $\sqrt{\tau}$, while these expansions can be obtained simply by differentiating both sides of bivariate expansion (36) i times with respect to τ and j times with respect to k, and then set k to zero. We provide the details for these calculations at the end of Appendix B.

The following result then generalizes Theorem 1 to the case where jumps are present. It establishes that the coefficient functions $\mu(\cdot)$, $\gamma(\cdot)$, $\eta(\cdot)$, and $\lambda(\cdot)$, as well as the jump size parameters μ_J and σ_J, can be recovered explicitly in terms of the shape characteristics (6) of the IV surface, or equivalently in terms of the relevant coefficients $\varphi^{(i,j)}$:

Theorem 2. *The jump size parameters μ_J and σ_J of the model (33a)–(33b) can be recovered by the following coupled algebraic equations*

$$\frac{\bar{\mu}+2-2(\bar{\mu}+1)\mathcal{N}(\mu_+)-2\mathcal{N}(\mu_-)}{-\bar{\mu}+2(\bar{\mu}+1)\mathcal{N}(\mu_+)-2\mathcal{N}(\mu_-)}=\frac{1}{\varphi^{(0,0)}(v_t)^2}\left[\frac{\varphi^{(1,1)}(v_t)}{\varphi^{(-1,2)}(v_t)}+2(r-d)\right] \tag{40a}$$

and

$$\frac{\sqrt{2}\bar{\mu}}{3\sqrt{\pi}[-\bar{\mu}+2(\bar{\mu}+1)\mathcal{N}(\mu_+)-2\mathcal{N}(\mu_-)]}=\frac{\varphi^{(-2,3)}(v_t)}{\varphi^{(-1,2)}(v_t)}. \tag{40b}$$

The coefficient functions $\lambda(\cdot)$, $\gamma(\cdot)$, $\eta(\cdot)$, and $\mu(\cdot)$ can be recovered in closed form as

$$\lambda(v_t)=\frac{\sqrt{2}\varphi^{(0,0)}(v_t)^2\varphi^{(-1,2)}(v_t)}{\sqrt{\pi}[-\bar{\mu}+2(\bar{\mu}+1)\mathcal{N}(\mu_+)-2\mathcal{N}(\mu_-)]}, \tag{40c}$$

$$\gamma(v_t)=2\varphi^{(0,0)}(v_t)\varphi^{(0,1)}(v_t)-2\lambda(v_t)\bar{\mu}, \tag{40d}$$

and

$$\begin{aligned}\eta(v_t)=\Bigg[&6\varphi^{(0,0)}(v_t)^3\varphi^{(0,2)}(v_t)-\varphi^{(0,0)}(v_t)\gamma(v_t)\gamma'(v_t)+\frac{3}{2}\pi\lambda(v_t)^2(\bar{\mu}-2(\bar{\mu}+1)\mathcal{N}(\mu_+)\\&+2\mathcal{N}(\mu_-))^2+\frac{3}{2}\gamma(v_t)^2+3\lambda(v_t)(2\bar{\mu}(r-d)+(\bar{\mu}+2)\varphi^{(0,0)}(v_t)^2+\bar{\mu}\gamma(v_t))\Bigg]^{\frac{1}{2}},\end{aligned} \tag{40e}$$

as well as

$$\begin{aligned}\mu(v_t)=&\frac{1}{12\varphi^{(0,0)}(v_t)}[24\varphi^{(0,0)}(v_t)\varphi^{(2,0)}(v_t)-6\varphi^{(0,0)}(v_t)^2\gamma(v_t)-2\eta(v_t)^2-12\lambda(v_t)\\&\times(\bar{\mu}(2(d-r)+\gamma(v_t))-(\bar{\mu}+2)\varphi^{(0,0)}(v_t)^2)-12(d-r)\gamma(v_t)-3\gamma(v_t)^2\\&-12\bar{\mu}^2\lambda(v_t)^2+4\varphi^{(0,0)}(v_t)\gamma(v_t)\left(3\bar{\mu}\lambda'(v_t)+\gamma'(v_t)\right)].\end{aligned} \tag{40f}$$

Equations (40a)–(40f) constitute a complete mapping from the expansion terms $\varphi^{(i,j)}(v_t)$ of the IV surface to the specification of the SV model (33a)–(33b).

Here is how the jump size parameters μ_J and σ_J are determined from the IV surface. By assembling the algebraic equation system (40a)–(40b) and the geometric interpretations of the involved expansion terms $\varphi^{(0,0)}$, $\varphi^{(-2,3)}$, $\varphi^{(1,1)}$, and $\varphi^{(-1,2)}$ provided in (39a)–(39b), we obtain the following mapping from the shape characteristics (on the right hand side) to the jump parameters μ_J and σ_J (on the left hand side):

$$\frac{\bar{\mu}+2-2(\bar{\mu}+1)\mathcal{N}(\mu_+)-2\mathcal{N}(\mu_-)}{-\bar{\mu}+2(\bar{\mu}+1)\mathcal{N}(\mu_+)-2\mathcal{N}(\mu_-)}=\frac{1}{\lim\limits_{\tau\to 0}\Sigma(\tau,0,v_t)^2}\left[\frac{\lim\limits_{\tau\to 0}2\sqrt{\tau}\frac{\partial^2\Sigma}{\partial\tau\partial k}(\tau,0,v_t)}{\lim\limits_{\tau\to 0}\frac{\sqrt{\tau}}{2}\frac{\partial^2\Sigma}{\partial k^2}(\tau,0,v_t)}-2(d-r)\right] \tag{41a}$$

and

$$\frac{\sqrt{2}\bar{\mu}}{3\sqrt{\pi}[-\bar{\mu}+2(\bar{\mu}+1)\mathcal{N}(\mu_+)-2\mathcal{N}(\mu_-)]}=\frac{\lim\limits_{\tau\to 0}\frac{\tau}{6}\frac{\partial^3\Sigma}{\partial k^3}(\tau,0,v_t)}{\lim\limits_{\tau\to 0}\frac{\sqrt{\tau}}{2}\frac{\partial^2\Sigma}{\partial k^2}(\tau,0,v_t)}, \tag{41b}$$

where we recall that $\bar{\mu}$, μ_+, and μ_- are deterministic functions of μ_J and σ_J defined in (34) and (35). According to these equations, one needs various at-the-money IV shape characteristics in both log-moneyness and time-to-maturity dimensions to pin down μ_J and σ_J, without requiring any prior identification of any of the coefficient functions $\lambda(\cdot)$, $\mu(\cdot)$, $\gamma(\cdot)$, or $\eta(\cdot)$. In particular, it deserves noting that the third order derivative $\partial^3\Sigma/\partial k^3$ plays a crucial role. This is somewhat unfortunate from an empirical perspective as it implies that the jump size parameters μ_J and σ_J depend on a higher order structure of the IV surface that will be difficult to estimate precisely in the absence of large amounts of high quality options data.

The stochastic intensity function $\lambda(v_t)$ is characterized by:

$$\lambda(v_t) = \frac{\sqrt{2}\lim_{\tau\to 0}\Sigma(\tau,0,v_t)^2 \cdot \lim_{\tau\to 0}\frac{\sqrt{\tau}}{2}\frac{\partial^2\Sigma}{\partial k^2}(\tau,0,v_t)}{\sqrt{\pi}[-\bar{\mu}+2(\bar{\mu}+1)\mathcal{N}(\mu_+)-2\mathcal{N}(\mu_-)]}. \tag{42}$$

So the short-maturity at-the-money IV convexity $\partial^2\Sigma(\tau,0,v_t)/\partial k^2$ is involved in determining the stochastic intensity function $\lambda(v_t)$ but not any third order characteristics, except of course that those were already needed to identify $\bar{\mu}$, μ_+, and μ_-, which enter (42). In the continuous case, the at-the-money convexity is finite as the time-to-maturity shrinks to zero, since equations (8) and (6) imply that $\lim_{\tau\to 0}\partial^2\Sigma(\tau,0,v_t)/\partial k^2 = 2\sigma^{(0,2)}(v_t)$. By contrast, under the discontinuous model, the convexity $\partial^2\Sigma(\tau,0,v_t)/\partial k^2$ explodes to infinity as the time-to-maturity shrinks to zero, since the last equation in (39b) directly implies that $\partial^2\Sigma(\tau,0,v_t)/\partial k^2 \sim 2\varphi^{(-1,2)}(v_t)/\sqrt{\tau}$ as $\tau\to 0$ with $\varphi^{(-1,2)}(v_t)$ finite. The formula (42) remains valid in the limiting case where the intensity $\lambda(v_t)$ tends to zero, i.e., the jumps degenerate. This is because the convexity behaves in that case according to $\lim_{\tau\to 0}\sqrt{\tau}\partial^2\Sigma(\tau,0,v_t)/\partial k^2 = 0$, which obviously results in the right hand side of (42) tending to zero.

The volatility function $\gamma(v_t)$ and $\eta(v_t)$ and the drift function $\mu(v_t)$ are all affected by the presence of jumps. Compared to the continuous case, the third order mixed partial derivative $\partial^3\Sigma/\partial k^2\partial\tau$ (resp. term-structure slope $\partial\Sigma/\partial\tau$ and term-structure convexity $\partial^2\Sigma/\partial\tau^2$) participate in determining the volatility function $\eta(v_t)$ (resp. the drift function $\mu(v_t)$.) By contrast in the continuous case, the term structure slope $\partial\Sigma/\partial\tau$ is the only IV characteristic along the term-structure dimension that matters.

Equations (41a), (41b), and (42) (equivalently, (40a)–(40c) in Theorem 2) apply to the jump-diffusion model of Merton (1976), by plugging in the specification assumptions (38). One is able to identify all the model components, i.e., the constant volatility v_0, intensity λ, as well as jump size parameters μ_J and σ_J. Combining the following equations

$$\frac{\lambda\bar{\mu}}{v_0} = \lim_{\tau\to 0}\frac{\partial\Sigma}{\partial k}(\tau,0,v_t),\ \frac{\lambda\sqrt{\pi}}{2\sqrt{2}v_0^2}(-\bar{\mu}+2(\bar{\mu}+1)\mathcal{N}(\mu_+)-2\mathcal{N}(\mu_-)) = \lim_{\tau\to 0}\frac{\sqrt{\tau}}{2}\frac{\partial^2\Sigma}{\partial k^2}(\tau,0,v_t),$$
$$\frac{\sqrt{\pi}\lambda}{2\sqrt{2}v_0^2}[2(r-d)\bar{\mu}+(\bar{\mu}+2)v_0^2+2(\bar{\mu}+1)\mathcal{N}(\mu_+)(2(d-r)-v_0^2)-2\mathcal{N}(\mu_-)(2(d-r)+v_0^2)]$$

$$= \lim_{\tau \to 0} 2\sqrt{\tau} \frac{\partial^2 \Sigma}{\partial \tau \partial k}(\tau, 0, v_t)$$

with the first equation in (39a), i.e., $v_0 = \lim_{\tau \to 0} \Sigma(\tau, 0, v_0)$, we can identify the parameters of the Merton model v_0, λ, μ_J, and σ_J, given observations on the following four observable short-maturity IV shape characteristics – at-the-money level Σ, slope $\partial\Sigma/\partial k$, convexity $\partial^2\Sigma/\partial k^2$, and the mixed slope $\partial^2\Sigma/\partial\tau\partial k$, all evaluated at $(\tau, 0, v_0)$. If employing equation (41b) instead, the much less easily observable third order shape characteristic $\partial^3\Sigma/\partial k^3$ would become necessary.

7.4 Implied stochastic volatility models with jumps

The analysis in Section 7.3 provides a theoretical foundation for constructing parametric and nonparametric implied stochastic volatility models with jumps. In practice, to construct a parametric model, based on the closed-form formulae for the expansion terms $\varphi^{(i,j)}$, we can use the moment conditions

$$\mathbb{E}[g^{(i,j)}(v_{l\Delta}; \theta_0)] = 0, \text{ with } g^{(i,j)}(v_{l\Delta}; \theta) = [\varphi^{(i,j)}]_l^{\text{data}} - [\varphi^{(i,j)}(v_{l\Delta}; \theta)]^{\text{model}},$$

where $[\varphi^{(i,j)}]_l^{\text{data}}$ denotes the data of $\varphi^{(i,j)}(v_{l\Delta})$. Then apply the GMM estimation approach proposed in Section 3 to estimate the parameters θ.

To construct a nonparametric model, we can in principle estimate the jump size parameters μ_J and σ_J before estimating the coefficient functions $\lambda(\cdot)$, $\mu(\cdot)$, $\gamma(\cdot)$, and $\eta(\cdot)$ as discussed. Indeed, the estimators of μ_J and σ_J can be obtained by the two (exactly identified) conditions as the sample analogs of algebraic equations (40a) and (40b)

$$\frac{\bar{\mu} + 2 - 2(\bar{\mu}+1)\mathcal{N}(\mu_+) - 2\mathcal{N}(\mu_-)}{-\bar{\mu} + 2(\bar{\mu}+1)\mathcal{N}(\mu_+) - 2\mathcal{N}(\mu_-)} = \frac{1}{n}\sum_{l=1}^{n} \frac{1}{[\varphi^{(0,0)}]_l^{\text{data}}} \left(\frac{[\varphi^{(1,1)}]_l^{\text{data}}}{[\varphi^{(-1,2)}]_l^{\text{data}}} + 2(r-d) \right),$$

and

$$\frac{\sqrt{2}\bar{\mu}}{3\sqrt{\pi}[-\bar{\mu} + 2(\bar{\mu}+1)\mathcal{N}(\mu_+) - 2\mathcal{N}(\mu_-)]} = \frac{1}{n}\sum_{l=1}^{n} \frac{[\varphi^{(-2,3)}]_l^{\text{data}}}{[\varphi^{(-1,2)}]_l^{\text{data}}}.$$

Then, regarding the estimators of jump size parameters as inputs, equations (40c)–(40f) allow us to estimate coefficient functions $\lambda(\cdot)$, $\gamma(\cdot)$, $\eta(\cdot)$, and $\mu(\cdot)$ one after another iteratively, by following a similar approach proposed in Section 4 for constructing a nonparametric SV model without jumps.

7.5 Empirical challenges when jumps are present in the model

So, we have shown that it is possible in theory to imply a SV model with jumps from the shape characteristics of the IV surface. However, given the current liquidity of options markets and resulting availability of options data, one would encounter significant practical challenges when implementing

the above strategy. As we did in the continuous case, it is natural to interpret the expansion (36) as the following regression

$$\Sigma^{\text{data}}(\tau_l^{(m)}, k_l^{(m)}) = \sum_{j=0}^{J} \sum_{i=\min(0,1-j)}^{L_j} \beta_l^{(i,j)} (\tau_l^{(m)})^{\frac{i}{2}} k^j + \epsilon_l^{(m)}, \text{ for } m = 1, 2, \ldots, n_l, \tag{44}$$

and the estimator of the coefficient $\beta_l^{(i,j)}$ serves as the data $[\varphi^{(i,j)}]_l^{\text{data}}$.

Similar to the case for regression (16), the choice of the orders J, L_0, L_1, $\ldots$, L_J and the regions of IV surfaces data employed in regression (44) should strike a balance between, on the one hand, the accuracy of the expansion $\Sigma^{(J,\mathbf{L}(J))}$ and on the other hand, over-fitting the regression to the IV data. Most importantly, the presence of jumps necessitates the estimation of third order characteristics of the IV surface, which in our experience is effectively impossible to do accurately given the limitations of the data currently available. A substantially denser set of observations on option prices or implied volatilities would be necessary to accurately estimate third order derivatives without the error introduced by the strike and maturity surface interpolation implicit in (44). Furthermore, the divergence of the IV surface due to the presence of negative powers of τ also requires very short maturity options to be accurately observed (as in Carr and Wu (2003)'s test for the presence of jumps in options data); such data can be affected by trading patterns specific to options with, e.g., time-to-maturity τ less than one week, and log-moneyness k within $\pm 0.1 v\sqrt{\tau}$, where v represents the instantaneous volatility. This makes inferring the desired data $[\varphi^{(i,j)}]_l^{\text{data}}$ from the IV surface and the subsequent procedures for constructing implied stochastic volatility models substantially more difficult since we do not need to just identify the divergence as in Carr and Wu (2003) but also estimate higher order coefficients.

8 Conclusions and future directions

We proposed to construct implied stochastic volatility models to be consistent with observed shape characteristics of the implied volatility market data. In the construction of a parametric model, all parameters are estimated in one pass, regardless of how they get involved in expansion terms. In the construction of a nonparametric model, the coefficient functions are estimated one after another iteratively, based on the closed-form relationships we derived. At least in principle, implied stochastic volatility models in higher dimensions can be constructed using the same principle, although a bivariate nonparametric implied stochastic volatility model, as we considered and empirically illustrated, is flexible enough in terms of fitting the observable and practically useful shape characteristics of the implied volatility surface (the level, slope and convexity along the moneyness dimension, as well as the slope along the term-structure dimension.)

When jumps are introduced to the model, we showed that the same ideas continue to work in principle and that a full characterization of the stochastic volatility model can still be obtained in

closed form, at least for models with jumps only in the returns dynamics. However, higher order shape characteristics become necessary, whose estimation require substantially denser options observations in both time and moneyness than is currently available, even though options with shorter maturities, such as weekly, have recently become more liquid. Adding jumps to the volatility dynamics, or infinite activity jumps to either returns or volatility dynamics, would substantially alter the approach we employ to derive the implied volatility expansion as a tool, and require in practice more accurate and delicate shape characteristics for fully recovering the model components. We intend to pursue this line of inquiry in future research.

References

Aït-Sahalia, Y., 2002. Maximum-likelihood estimation of discretely-sampled diffusions: A closed-form approximation approach. Econometrica 70, 223–262.

Aït-Sahalia, Y., Fan, J., Li, Y., 2013. The leverage effect puzzle: Disentangling sources of bias at high frequency. Journal of Financial Economics 109, 224–249.

Aït-Sahalia, Y., Kimmel, R., 2007. Maximum likelihood estimation of stochastic volatility models. Journal of Financial Economics 83, 413–452.

Aït-Sahalia, Y., Mykland, P. A., 2003. The effects of random and discrete sampling when estimating continuous-time diffusions. Econometrica 71, 483–549.

Andersen, L., Andreasen, J., 2000. Jump-diffusion processes: Volatility smile fitting and numerical methods for option pricing. Review of Derivatives Research 4 (3), 231–262.

Andersen, L., Lipton, A., 2013. Asymptotics for exponential Lévy processes and their volatility smile: Survey and new results. International Journal of Theoretical and Applied Finance 16, 1–98.

Bates, D. S., 1996. Jumps and stochastic volatility: Exchange rate processes implicit in Deutsche Mark options. Review of Financial Studies 9, 69–107.

Berestycki, H., Busca, J., Florent, I., 2002. Asymptotics and calibration of local volatility models. Quantitative Finance 2, 61–69.

Berestycki, H., Busca, J., Florent, I., 2004. Computing the implied volatility in stochastic volatility models. Communications on Pure and Applied Mathematics 57, 1352–1373.

Brémaud, P., 1981. Point Processes and Queues: Martingale Dynamics. Springer-Verlag.

Carr, P., Cousot, L., 2011. A PDE approach to jump-diffusions. Quantitative Finance 11 (1), 33–52.

Carr, P., Cousot, L., 2012. Explicit constructions of martingales calibrated to given implied volatility smiles. SIAM Journal on Financial Mathematics 3 (1), 182–214.

Carr, P., Geman, H., Madan, D. B., Yor, M., 2004. From local volatility to local Lévy models. Quantitative Finance 4 (5), 581–588.

Carr, P., Wu, L., 2003. What type of process underlies options? A simple robust test. The Journal of Finance 58, 2581–2610.

Chernov, M., Gallant, A. R., Ghysels, E., Tauchen, G. T., 2003. Alternative models for stock price dynamics. Journal of Econometrics 116, 225–257.

Christoffersen, P., Jacobs, K., Mimouni, K., 2010. Volatility dynamics for the S&P500: Evidence from realized volatility, daily returns, and option prices. Review of Financial Studies 23 (8), 3141–3189.

Duffie, D., Pan, J., Singleton, K. J., 2000. Transform analysis and asset pricing for affine jump-diffusions. Econometrica 68, 1343–1376.

Dumas, B., Fleming, J., Whaley, R. E., 1998. Implied volatility functions: Empirical tests. The Journal of Finance 53, 2059–2106.

Dupire, B., 1994. Pricing with a smile. RISK 7, 18–20.

Durrleman, V., 2008. Convergence of at-the-money implied volatilities to the spot volatility. Journal of Applied Probability 45, 542–550.

Durrleman, V., 2010. From implied to spot volatilities. Finance and Stochastics 14 (2), 157–177.

Eraker, B., Johannes, M. S., Polson, N., 2003. The impact of jumps in equity index volatility and returns. The Journal of Finance 58, 1269–1300.

Fan, J., Gijbels, I., 1996. Local Polynomial Modelling and Its Applications. Chapman & Hall, London, U.K.

Forde, M., Jacquier, A., 2011. The large-maturity smile for the Heston model. Finance and Stochastics 17, 755–780.

Forde, M., Jacquier, A., Lee, R., 2012. The small-time smile and term structure of implied volatility under the Heston model. SIAM Journal on Financial Mathematics 3 (1), 690–708.

Fouque, J.-P., Lorig, M., Sircar, R., 2016. Second order multiscale stochastic volatility asymptotics: Stochastic terminal layer analysis and calibration. Finance and Stochastics 20, 543–588.

Gao, K., Lee, R., 2014. Asymptotics of implied volatility to arbitrary order. Finance and Stochastics 18 (2), 349–392.

Gatheral, J., 2006. The Volatility Surface: A Pactitioner's Guide. John Wiley and Sons, Hoboken, NJ.

Gatheral, J., Hsu, E. P., Laurence, P., Ouyang, C., Wang, T.-H., 2012. Asymptotics of implied volatility in local volatility models. Mathematical Finance 22 (4), 591–620.

Hagan, P. S., Woodward, D. E., 1999. Equivalent Black volatilities. Applied Mathematical Finance 6, 147–157.

Hansen, L. P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

Heston, S., 1993. A closed-form solution for options with stochastic volatility with applications to bonds and currency options. Review of Financial Studies 6, 327–343.

Hull, J., White, A., 1987. The pricing of options on assets with stochastic volatilities. The Journal of Finance 42, 281–300.

Jacquier, A., Lorig, M., 2015. From characteristic functions to implied volatility expansions. Advances in Applied Probability 47, 837–857.

Jones, C. S., 2003. The dynamics of stochastic volatility: Evidence from underlying and options markets. Journal of Econometrics 116, 181–224.

Karlin, S., Taylor, H. M., 1975. A First Course in Stochastic Processes, 2nd Edition. Academic Press.

Kristensen, D., Mele, A., 2011. Adding and subtracting Black-Scholes: A new approach to approximating derivative prices in continuous-time models. Journal of Financial Economics 102, 390–415.

Kunitomo, N., Takahashi, A., 2001. The asymptotic expansion approach to the valuation of interest rate contingent claims. Mathematical Finance 11 (1), 117–151.

Ledoit, O., Santa-Clara, P., Yan, S., 2002. Relative pricing of options with stochastic volatility. Tech. rep., University of California at Los Angeles.

Lee, R., 2001. Implied and local volatilities under stochastic volatility. International Journal of Theoretical and Applied Finance 4, 45–89.

Lee, R., 2004. The moment formula for implied volatility at extreme strikes. Mathematical Finance 14 (3), 469–480.

Li, C., 2014. Closed-form expansion, conditional expectation, and option valuation. Mathematics of Operations Research 39, 487–516.

Lorig, M., Pagliarani, S., Pascucci, A., 2017. Explicit implied volatilities for multifactor local-stochastic volatility models. Mathematical Finance 27, 927–960.

Medvedev, A., Scaillet, O., 2007. Approximation and calibration of short-term implied volatilities under jump-diffusion stochastic volatility. Review of Financial Studies 20 (2), 427–459.

Merton, R. C., 1976. Option pricing when underlying stock returns are discontinuous. Journal of Financial Economics 3, 125–144.

Pagliarani, S., Pascucci, A., 2017. The exact Taylor formula of the implied volatility. Finance and Stochastics 21, 661–718.

Pan, J., 2002. The jump-risk premia implicit in options: Evidence from an integrated time-series study. Journal of Financial Economics 63, 3–50.

Sircar, K. R., Papanicolaou, G. C., 1999. Stochastic volatility, smile & asymptotics. Applied Mathematical Finance 6, 107–145.

Takahashi, A., Yamada, T., 2012. An asymptotic expansion with push-down of Malliavin weights. SIAM Journal on Financial Mathematics 3 (1), 95–136.

Tehranchi, M. R., 2009. Asymptotics of implied volatility far from maturity. Journal of Applied Probability 46, 629–650.

Xiu, D., 2014. Hermite polynomial based expansion of European option prices. Journal of Econometrics 179, 158–177.

Appendix

Appendix A Implied volatility expansion for continuous models

In this appendix, we sketch on how to derive the IV expansion terms $\sigma^{(i,j)}$ in (7) in closed form for the continuous SV model (1a)–(1b). To simplify notations, we assume $S_t = s$ and $v_t = v$ at time t. The main idea hinges on expanding the both sides of identity (4) with respect to the square root of time-to-maturity $\epsilon = \sqrt{\tau}$ and log-moneyness k and then matching expansion terms of the same orders. Thus, as an indispensable preparation, we propose the following $(J, \mathbf{L}(J))$th order expansion of $\bar{P}(\tau, k, v_t)$ introduced in (3) and appearing on the right hand side of (4):

$$\bar{P}^{(J,\mathbf{L}(J))}(\epsilon^2, k, v) = \sum_{j=0}^{J} \sum_{i=1-j}^{L_j} p^{(i,j)}(v)\, \epsilon^i k^j, \text{ with } \epsilon = \sqrt{\tau}, \tag{A.1}$$

for any orders $J \geq 0$ and $L_j \geq 1-j$, $j = 0, 1, \ldots, J$. The coefficients $p^{(i,j)}$ can be calculated explicitly by following Li (2014), in which the option price $P(\epsilon^2, k, s, v)$ admits a pseudo univariate expansion with respect to ϵ with closed-form expansion terms depending on both ϵ and k. The bivariate expansion (A.1) follows from taking $s = 1$ in this univariate expansion and further expanding the coefficients with respect to k and ϵ.

Now, based on the bivariate expansion (A.1) of $\bar{P}(\tau, k, v_t)$, which appears on the right hand side of (4), in what follows, we accordingly expand the composite function $\bar{P}_{\text{BS}}(\tau, k, \Sigma(\tau, k, v))$ on the left hand side. By matching the expansion term on the both sides, we establish a set of iterations and solve the expansion terms $\sigma^{(i,j)}$ recursively.

We start from the following expansion of at-the-money IV $\Sigma(\epsilon^2, 0, v)$ with respect to ϵ :

$$\Sigma^{(L_0)}(\epsilon^2, 0, v) = \sum_{i=0}^{L_0} \sigma^{(i,0)}(v)\epsilon^{2i}, \tag{A.2}$$

which is obtained by setting $k = 0$ in the bivariate expansion (7). According to Durrleman (2008), $\Sigma(\epsilon^2, 0, v)$ converges to the instantaneous SV of the asset price v_t, as the time-to-maturity $\tau = \epsilon^2$ approaches to zero. Thus, $\sigma^{(0,0)}(v) = v$. By taking $\sigma^{(0,0)}(v)$ as the initial input, all other expansion terms can be solved recursively.

To compute the expansion terms $\sigma^{(i,0)}$, we apply the at-the-money condition $k = 0$ on the both sides of (4) to obtain

$$\bar{P}(\epsilon^2, 0, v) = \bar{P}_{\text{BS}}(\epsilon^2, 0, \Sigma(\epsilon^2, 0, v)). \tag{A.3}$$

Expanding the both sides of (A.3) with respect to ϵ and matching the coefficients, we can obtain a system of equations. The closed-form formulae of expansion terms $\sigma^{(i,0)}$ follows by solving the equations recursively. Indeed, for the left hand side of (A.3), the expansion of $\bar{P}(\epsilon^2, 0, v)$ with

respect to ϵ can be obtained from (A.1) by setting $k = 0$, i.e.,

$$\bar{P}^{(L_0)}(\epsilon^2, 0, v) = \sum_{l=0}^{L_0} p^{(l,0)}(v)\,\epsilon^l. \tag{A.4}$$

For the right hand side of (A.3), the expansion of $\bar{P}_{\mathrm{BS}}(\epsilon^2, 0, \Sigma(\epsilon^2, 0, v))$ with respect to ϵ follows by combining the expansion of the function $\bar{P}_{\mathrm{BS}}(\epsilon^2, 0, \sigma)$, which is obtained by expanding the explicit formula of $\bar{P}_{\mathrm{BS}}(\epsilon^2, 0, \sigma)$, and the expansion of at-the-money IV $\Sigma(\epsilon^2, 0, v)$, which is proposed in (A.2) with the undetermined expansion terms $\sigma^{(i,0)}$. Then, the closed-form expansion of $\bar{P}_{\mathrm{BS}}(\epsilon^2, 0, \Sigma(\epsilon^2, 0, v))$ is in the following form

$$\bar{P}^{(J)}_{\mathrm{BS}}(\epsilon^2, 0, \Sigma(\epsilon^2, 0, v)) = \sum_{l=1}^{J} \tilde{p}^{(l,0)}(v)\epsilon^l, \tag{A.5}$$

for any integer $J \geq 1$. In particular, for any odd integer $l \geq 3$, the expansion term $\tilde{p}^{(l,0)}$ by computation consists of IV expansion terms $\sigma^{(i,0)}$ for all $i \leq (l-1)/2$. Matching the coefficients of expansions (A.5) and (A.4) yields the following system of equations

$$p^{(l,0)}(v) = \tilde{p}^{(l,0)}(v), \text{ for any odd integer } l \geq 1.$$

For any integer $i \geq 1$, the closed-form formula of the expansion term $\sigma^{(i,0)}(v)$ follows from solving the above equation with $l = 2i + 1$.

Finally, to compute the expansion terms $\sigma^{(i,j)}$ for $j \geq 1$, we resort to the following identity

$$\frac{\partial^j}{\partial k^j}\bar{P}(\epsilon^2, 0, v) = f_j(\epsilon, v), \tag{A.6}$$

which is obtained from differentiating the identity (4) j times with respect to k and then applying the at-the-money condition $k = 0$. Here, the function f_j is defined by

$$f_j(\epsilon, v) = \sum_{0 \leq m_1 \leq m_2 \leq j} \binom{j}{m_2} \frac{\partial^{j-m_2+m_1}\bar{P}_{\mathrm{BS}}}{\partial k^{j-m_2}\partial \sigma^{m_1}}(\epsilon^2, 0, \Sigma(\epsilon^2, 0, v))G^{(m_1,m_2)}(\epsilon, v), \tag{A.7}$$

where the nonnegative integers m_1 and m_2 satisfy that $m_1 = 0$ if and only if $m_2 = 0$. The function $G^{(m_1,m_2)}$ is defined by $G^{(0,0)}(\epsilon, v) = 1$ and

$$G^{(m_1,m_2)}(\epsilon, v) = \sum_{\mathbf{l} \in \mathcal{S}_{m_1,m_2}} \frac{m_2!}{R(\mathbf{l})} \prod_{\ell=1}^{m_1} \frac{1}{i_\ell!} \frac{\partial^{i_\ell}\Sigma}{\partial k^{i_\ell}}(\epsilon^2, 0, v), \tag{A.8}$$

for $1 \leq m_1 \leq m_2$. Here, the integer index set $\mathcal{S}_{m_1,m_2}$ is given by

$$\mathcal{S}_{m_1,m_2} = \{(i_1, i_2, \cdots, i_{m_1}) : 1 \leq i_1 \leq i_2 \leq \cdots \leq i_{m_1},\ \textstyle\sum_{\ell=1}^{m_1} i_\ell = m_2\}, \tag{A.9}$$

and the function $R(\mathbf{l})$ is a constant defined by the product of factorials of the repeating times of distinct nonzero entries appearing for more than once in index $\mathbf{l}$. For example, in index $\mathbf{l} =$

$(1,1,2,2,2)$, distinct entries 1 and 2 appear twice and thrice, respectively. Then, the constant $R(\mathbf{l})$ is calculated as $2! \times 3! = 24$. Similar to the previous case of $j = 0$, by expanding the both sides of (A.6) with respect to ϵ and matching the coefficients, we can obtain a system of equations for solving the expansion terms $\sigma^{(i,j)}(v)$ recursively.

Indeed, the expansion of $\partial^j \bar{P}(\epsilon^2, 0, v)/\partial k^j$ on left hand side of (A.6) is

$$\frac{\partial^j \bar{P}^{(L_j)}}{\partial k^j}(\epsilon^2, 0, v) = \sum_{i=1-j}^{L_j} j! p^{(i,j)}(v)\, \epsilon^i, \tag{A.10}$$

which is obtained from differentiating expansion (A.1) j times with respect to k and then setting $k = 0$. According to the definition (A.7), the expansion of the function f_j on the right hand side of (A.6) hinges on the expansions of two types of ingredients

$$\frac{\partial^{j-m_2+m_1} \bar{P}_{\mathrm{BS}}}{\partial k^{j-m_2} \partial \sigma^{m_1}}(\epsilon^2, 0, \Sigma(\epsilon^2, 0, v)) \text{ and } G^{(m_1,m_2)}(\epsilon, v). \tag{A.11}$$

As to the first ingredient, its expansion can be obtained by combining the expansions of the Black-Scholes sensitivities $\partial^{j-m_2+m_1} \bar{P}_{\mathrm{BS}}(\epsilon^2, 0, \sigma)/\partial k^{j-m_2} \partial \sigma^{m_1}$, which is obtained based on the explicit formula of $\bar{P}_{\mathrm{BS}}$, and the expansion of at-the-money IV $\Sigma(\epsilon^2, 0, v)$, which is explicitly computed from the preceding iteration for $j = 0$. By combining these two types of expansions, we obtain the Jth order expansion of the first ingredient in (A.11) as

$$\frac{\partial^{j-m_2+m_1} \bar{P}^{(J)}_{\mathrm{BS}}}{\partial k^{j-m_2} \partial \sigma^{m_1}}(\epsilon^2, 0, \Sigma(\epsilon^2, 0, v)) = \sum_{l=1-j+m_2}^{J} H^{(j-m_2)}_{l,m_1} \epsilon^l, \tag{A.12}$$

for any integer order $J \geq 1 - j + m_2$, where the expansion terms $H^{(j-m_2)}_{l,m_1}$ consist of various Black-Scholes sensitivities and at-the-money IV expansion terms $\sigma^{(i,0)}$.

To obtain the expansion of the second ingredient $G^{(m_1,m_2)}(\epsilon, v)$ in (A.11), according to its definition (A.8), it suffices to combine the expansions of various at-the-money IV shape characteristics $\partial^{i_\ell} \Sigma(\epsilon^2, 0, v)/\partial k^{i_\ell}$, while the expansion of $\partial^{i_\ell} \Sigma(\epsilon^2, 0, v)/\partial k^{i_\ell}$ follows

$$\frac{\partial^{i_\ell} \Sigma^{(L_{i_\ell})}}{\partial k^{i_\ell}}(\epsilon^2, 0, v) = \sum_{l=0}^{L_{i_\ell}} i_\ell! \sigma^{(l,i_\ell)}(v) \epsilon^{2l},$$

by differentiating (7) i_ℓ times with respect to k and then setting $k = 0$. Then, the function $G^{(m_1,m_2)}$ admits a Jth order expansion in the form

$$G^{(m_1,m_2)}(\epsilon, v) = \sum_{l=0}^{J} G^{(m_1,m_2)}_l \epsilon^{2l}, \tag{A.13}$$

for any integer order $J \geq 0$. Here, the expansion term $G^{(m_1,m_2)}_l$ is defined according to

$$G^{(m_1,m_2)}_l = \sum_{\mathbf{l} \in \mathcal{S}_{m_1,m_2},\, \mathbf{v} \in \mathcal{T}_{\mathbf{l},l}} \frac{m_2!}{R(\mathbf{l})} \prod_{\ell=1}^{m_1} \sigma^{(v_\ell, i_\ell)}(v),$$

for any integers $m_2 \geq m_1 \geq 1$ and $l \geq 0$, with the integer index set $\mathcal{S}_{m_1,m_2}$ given in (A.9) and the function $R(\mathbf{l})$ provided right after (A.9). Moreover, for any index $\mathbf{l} \in \mathcal{S}_{m_1,m_2}$, the integer index set $\mathcal{T}_{\mathbf{l},l}$ is defined by

$$\mathcal{T}_{\mathbf{l},l} = \{\mathbf{v} = (v_1, v_2, \cdots, v_{m_1}) : v_1 + \cdots + v_{m_1} = l \text{ and } v_\ell \geq 0, \text{ for } \ell = 1, 2, \ldots, m_1\}.$$

Based on the expansions (A.12) and (A.13), it follows from the definition (A.7) that the function $f_j(\epsilon, v)$ admits the following Jth order expansion

$$f_j^{(J)}(\epsilon, v) = \sum_{l=1-j}^{J} \tilde{p}^{(l,j)}(v)\epsilon^l, \tag{A.14}$$

for any integer $J \geq 1 - j$, where the expansion term $\tilde{p}^{(l,j)}$ satisfies

$$\tilde{p}^{(l,j)}(v) = \sum_{0 \leq m_1 \leq m_2 \leq j} \binom{j}{m_2} \sum_{l_1+2l_2=l,\ l_1 \geq 1-j+m_2,\ l_2 \geq 0} H_{l_1,m_1}^{(j-m_2)} G_{l_2}^{(m_1,m_2)},$$

for any integer $l \geq 1-j$. In particular, for any odd integer $l \geq 1$, the expansion term $\tilde{p}^{(l,j)}(v)$ consists of IV expansion terms $\sigma^{(i,m)}$ for all $0 \leq m \leq j$ and $0 \leq i \leq (l-1)/2 + \lfloor j - m \rfloor /2$, where the notation $\lfloor a \rfloor$ represents the integer part of any arbitrary real number a. By matching the coefficients of expansions (A.14) and (A.10), we obtain the following system of equations

$$j! p^{(l,j)}(v) = \tilde{p}^{(l,j)}(v), \text{ for any integer } l \geq 1 - j.$$

For any integer $i \geq 1$, the closed-form formula of the expansion term $\sigma^{(i,j)}(v)$ follows from solving the above equation with $l = 2i + 1$.

Appendix B Implied volatility expansion for models with jumps

Similar to the derivation for the continuous case, the expansion terms $\varphi^{(i,j)}$ can be solved by iterations. These iterations can be obtained by expanding the both sides of identity (4) with respect to the square root of time-to-maturity $\epsilon = \sqrt{\tau}$ and log-moneyness k and then matching expansion terms of the same orders. Solving these matched equations leads to the desired iterations. Thus, by omitting the similar arguments, it suffices to the following indispensable ingredient for completing the derivation: Under the general SV model with jumps (33a)–(33b), we propose the following closed-form bivariate expansion of $\bar{P}(\tau, k, v_t)$ introduced in (3) and appearing on the right hand side of (4):

$$\bar{P}^{(J,\mathbf{L}(J))}(\epsilon^2, k, v) = \sum_{j=0}^{J} \sum_{i=1-j}^{L_j} \bar{p}^{(i,j)}(v)\, \epsilon^i k^j, \text{ with } \epsilon = \sqrt{\tau},$$

for any orders $J \geq 0$ and $L_j \geq 1 - j$, $j = 0, 1, 2, \ldots, J$. This expansion generalizes that for the continuous model (1a)–(1b) provided in (A.1) and can be developed from the following three steps.

Without loss of generality, by the time-homogeneity property of the model (33a)–(33b), the time span from t to T can be translated to that from 0 to $\tau = T - t$ for simplicity. We assume $S_0 = s$ and $v_0 = v$.

Step 1 – Representing $\bar{P}(\tau, k, v)$ under an auxiliary measure: We will rewrite the expectation representation of $\bar{P}(\tau, k, v)$ in (3) under an auxiliary probability measure, under which the expectation becomes easier to handle. We denote by $\mathbb{Q}$ the assumed risk-neutral measure and denote by $\mathcal{F}_t$ the filtration generated by the process $(S_t, v_t)^\top$. The new probability measure $\tilde{\mathbb{Q}}$ is induced by a Radon-Nikodým derivative Λ_t according to

$$\left.\frac{d\mathbb{Q}}{d\tilde{\mathbb{Q}}}\right|_{\mathcal{F}_t} = \Lambda_t, \text{ with } \Lambda_t \text{ defined as } \Lambda_t = \left(\prod_{i=1}^{N_t} \lambda(v_{\tau_i})\right) \exp\left\{t - \int_0^t \lambda(v_s) ds\right\}, \tag{B.1}$$

where τ_i denotes the arrival time of the ith jump, i.e., $\tau_i = \inf\{t \geq 0 : N_t = i\}$; in particular, $\Lambda_0 = 1$. According to Theorem T3 in Chapter VI of Brémaud (1981), N_t is a Poisson process with constant jump intensity 1 under the measure $\tilde{\mathbb{Q}}$. Changing the measure from $\mathbb{Q}$ to $\tilde{\mathbb{Q}}$ yields the following equivalent expectation representation of $\bar{P}(\tau, k, v)$:

$$\bar{P}(\tau, k, v) = e^{-r\tau} \tilde{\mathbb{E}}\left[\Lambda_\tau \max\left(e^k - \frac{S_\tau}{s}, 0\right)\right],$$

where $\tilde{\mathbb{E}}$ represents the expectation under the measure $\tilde{\mathbb{Q}}$. Then, by conditioning on the number of jumps between 0 and τ, we reformulate $\bar{P}(\tau, k, v_t)$ as the following summation form

$$\bar{P}(\tau, k, v) = \sum_{\ell=0}^{\infty} \tilde{\mathbb{Q}}(N_\tau = \ell) \bar{P}_\ell(\tau, k, v), \text{ with } \bar{P}_\ell(\tau, k, v) = \tilde{\mathbb{E}}^{(\ell)}\left[\Lambda_\tau \max\left(e^k - \frac{S_\tau}{s}, 0\right)\right], \tag{B.2}$$

where the multiplier $\tilde{\mathbb{Q}}(N_\tau = \ell)$, as the probability of $N_\tau = \ell$ under the measure $\tilde{\mathbb{Q}}$, can be explicitly calculated as $\tau^\ell e^{-\tau}/\ell!$, and the notation $\tilde{\mathbb{E}}^{(\ell)}[\cdot]$ serves as the abbreviation of the conditional expectation $\tilde{\mathbb{E}}[\cdot | N_\tau = \ell]$.

According to the relation (B.2), to expand $\bar{P}$, it suffices to multiply the expansion of $\tilde{\mathbb{Q}}(N_\tau = \ell) = \tau^\ell e^{-\tau}/\ell!$ with respect to τ, which is trivial, and the expansion of conditional expectation $\bar{P}_\ell$ with respect to $\epsilon = \sqrt{\tau}$ and k for any $\ell \geq 0$. To expand $\bar{P}_\ell$ for $\ell = 0$, in the beginning of *Step 2*, we propose a decomposition of Λ_τ and S_τ. Then, based on this decomposition, we apply the method proposed in Li (2014) and develop a pseudo expansion of $\bar{P}_0$ with respect to ϵ with coefficients depending on both ϵ and k. The desired bivariate expansion of $\bar{P}_0$ follows from further expanding those coefficients with respect to k and ϵ. To expand $\bar{P}_\ell$ for $\ell \geq 1$, based on the decomposition of Λ_τ and S_τ introduced in *Step 2*, we apply the operator-based expansion discussed in Aït-Sahalia (2002) to obtain the desired result in *Step 3*.

Step 2 – Expanding the conditional expectation $\bar{P}_\ell$ in (B.2) for $\ell = 0$: It follows from the dynamics (33a) that the underlying asset price S_u admits the following decomposition form

$$S_u = s S_u^c S_u^J, \tag{B.3}$$

where S_u^c and S_u^J are the continuous and jump components of S_u/s given by

$$S_u^c = \exp\left\{\int_0^u \left(r - d - \lambda(v_t)\bar{\mu} - \frac{1}{2}v_t^2\right)dt + \int_0^u v_t dW_{1t}\right\} \text{ and } S_u^J = \exp\left\{\sum_{i=1}^{N_u} J_{\tau_i}\right\}, \tag{B.4}$$

respectively. Likewise, the Radon-Nikodým derivative Λ_u by definition (B.1) is decomposed as

$$\Lambda_u = \Lambda_u^c \Lambda_u^J, \tag{B.5}$$

with the continuous component Λ_u^c and jump component Λ_u^J given by

$$\Lambda_u^c = \exp\left\{u - \int_0^u \lambda(v_t)dt\right\} \text{ and } \Lambda_u^J = \prod_{i=1}^{N_u} \lambda(v_{\tau_i}), \tag{B.6}$$

respectively. Apparently, the continuous components S_u^c and Λ_u^c satisfy

$$\frac{dS_u^c}{S_u^c} = (r - d - \lambda(v_u)\bar{\mu})du + v_u dW_u,\ S_0^c = 1, \tag{B.7}$$

and

$$d\Lambda_u^c = (1 - \lambda(v_u))\Lambda_u^c du,\ \Lambda_0^c = \Lambda_0 = 1, \tag{B.8}$$

respectively, with the volatility v_u governed by (33b).

In the case of $\ell = 0$, the jump components in the decompositions (B.3) and (B.5) are disabled, so that the conditional expectation $\bar{P}_0$ in (B.2) simplifies to

$$\bar{P}_0(\tau, k, v) = e^{-r\tau}\tilde{\mathbb{E}}\left[\Lambda_\tau^c \max(e^k - S_\tau^c, 0)\right],$$

since $S_\tau = sS_\tau^c$ and $\Lambda_\tau = \Lambda_\tau^c$, By regarding $\Lambda_\tau^c \max\left(e^k - S_\tau^c, 0\right)$ as the payoff function of a derivative security with the underlying asset $(S_\tau^c, \Lambda_\tau^c)$ evolving according to dynamics (B.7), (B.8), and (33b), we apply the method proposed in Li (2014) and arrive at the following Jth order univariate expansion of $\bar{P}_0(\tau, k, v)$:

$$\bar{P}_0^{(J)}(\epsilon^2, k, v) = e^{-r\epsilon^2}\epsilon v \sum_{l=0}^{J} \Phi_0^{(l)}\left(\frac{e^k - 1}{v\epsilon}\right)\epsilon^l,$$

where the coefficients $\Phi_0^{(l)}$ can be calculated in closed form. The desired bivariate expansion of $\bar{P}_0$ follows by further expanding the coefficients $\Phi_0^{(l)}$ with respect to k and ϵ.

Step 3 – Expanding the conditional expectation $\bar{P}_\ell$ in (B.2) for $\ell \geq 1$: Plugging in the decompositions (B.3) and (B.5) into (B.2) yields

$$\bar{P}_\ell(\tau, k, v) = \tilde{\mathbb{E}}^{(\ell)}\left[\Lambda_\tau^c \Lambda_\tau^J \max(e^k - S_u^c S_u^J, 0)\right].$$

Conditioning on Λ_τ^c, Λ_τ^J, and S_τ^c, we reformulate the above expectation as

$$\bar{P}_\ell(\tau, k, v) = \tilde{\mathbb{E}}^{(\ell)}[\Lambda_\tau^c \Lambda_\tau^J \tilde{\mathbb{E}}^{(\ell)}[\max(e^k - S_u^c S_u^J, 0)|\Lambda_\tau^c, \Lambda_\tau^J, S_\tau^c]]. \tag{B.9}$$

We note that the component S_τ^J inside the inner expectation is independent with all the conditioning arguments S_τ^c, Λ_τ^c and Λ_τ^J defined in (B.4), (B.6), and (B.6), respectively, simply because jump sizes J_{τ_i} are assumed to be independent with the asset price S_u, the volatility v_u, and the Poisson process N_u for any $u \in [0, \tau]$ under the measure $\tilde{\mathbb{Q}}$. Consequently, the inner expectation in (B.9) can be expressed as $\phi_\ell(k, S_\tau^c)$ for some function ϕ_ℓ determined by the following integral form

$$\phi_\ell(k, S_\tau^c) = \tilde{\mathbb{E}}^{(\ell)}[\max(e^k - S_u^c S_u^J, 0) | \Lambda_\tau^c, \Lambda_\tau^J, S_\tau^c] \tag{B.10}$$

$$= \int_{\mathcal{J}^\ell} \max(e^k - S_\tau^c e^{u_1+u_2+\cdots+u_\ell}) f(u_1) f(u_2) \cdots f(u_\ell) du_1 du_2 \cdots du_\ell, \tag{B.11}$$

where $\mathcal{J}$ and f represent the state space and the probability density function of the jump size J_t, respectively. The integral (B.11) can be explicitly calculated if, for example, the jump size J_t follows a normal distribution with mean μ_J and variance σ_J^2 as commonly employed since the breakthrough invention of the jump-diffusion model by Merton (1976). Under this case, the closed-form formula of the integral (B.11) is given by

$$\phi_\ell(k, S_\tau^c) = e^k \mathcal{N}\left(\frac{k - \log S_\tau^c - \ell\mu_J}{\sqrt{\ell}\sigma_J}\right) - S_\tau^c \exp\left(\ell\mu_J + \frac{\ell\sigma_J^2}{2}\right) \mathcal{N}\left(\frac{k - \log S_\tau^c - \ell\mu_J}{\sqrt{\ell}\sigma_J} - \sqrt{\ell}\sigma_J\right).$$

It follows from (B.9) and (B.10) that

$$\bar{P}_\ell(\tau, k, v) = \tilde{\mathbb{E}}^{(\ell)}[\Lambda_\tau^c \Lambda_\tau^J \phi_\ell(k, S_\tau^c)].$$

By conditioning on the components S_τ^c and Λ_τ^c, as well as the whole path of the volatility v_u for all $u \in [0, \tau]$, denoted by V for simplicity, the law of iterated expectation implies

$$\bar{P}_\ell(\tau, k, v) = \tilde{\mathbb{E}}[\Lambda_\tau^c \phi_\ell(k, S_\tau^c) \tilde{\mathbb{E}}^{(\ell)}[\Lambda_\tau^J | S_\tau^c, \Lambda_\tau^c, V]]. \tag{B.12}$$

Plugging in the explicit expression of the jump component Λ_τ^J given in (B.6), we write the inner expectation as

$$\tilde{\mathbb{E}}^{(\ell)}[\Lambda_\tau^J | S_\tau^c, \Lambda_\tau^c, V] \equiv \tilde{\mathbb{E}}\left[\prod_{i=1}^{\ell} \lambda(v_{\tau_i}) \middle| S_\tau^c, \Lambda_\tau^c, V, N_\tau = \ell\right]. \tag{B.13}$$

Given the conditions spelt in (B.13), the randomness of $\prod_{i=1}^{\ell} \lambda(v_{\tau_i})$ solely hinges on those of the jump arrival times τ_i. Since N_t follows a Poisson process with constant intensity 1 independent with S_τ^c, Λ_τ^c, and V under the measure $\tilde{\mathbb{Q}}$, the conditional joint distribution of $(\tau_1, \tau_2, \cdots, \tau_\ell)$ given S_τ^c, Λ_τ^c, V, and $N_\tau = \ell$ is equivalent to that of $(\tau_1, \tau_2, \cdots, \tau_\ell)$ given $N_\tau = \ell$, which distributes as the order statistics of ℓ independent observations sampled from the uniform distribution on $[0, \tau]$ (see, e.g., Theorem 2.3 in Chapter 4.2 of Karlin and Taylor (1975).) Then, direct computation leads to that

$$\tilde{\mathbb{E}}^{(\ell)}[\Lambda_\tau^J | S_\tau^c, \Lambda_\tau^c, V] = \left(\int_0^\tau \frac{1}{\tau} \lambda(v_u) du\right)^\ell = \left(1 - \frac{1}{\tau} \log \Lambda_\tau^c\right)^\ell, \tag{B.14}$$

where the second equality follows from the representation of Λ_τ^c in (B.6). Hence, by plugging (B.14) into (B.12), we simplify $\bar{P}_\ell(\tau, k, v)$ in (B.9) to

$$\bar{P}_\ell(\tau, k, v) = \tilde{\mathbb{E}}\left[\Lambda_\tau^c \phi_\ell(k, S_\tau^c)\left(1 - \frac{1}{\tau}\log \Lambda_\tau^c\right)^\ell\right].$$

Finally, based on the dynamics of S_u^c, v_u, and Λ_u^c given in (B.7), (33b), and (B.8), respectively, an application of the operator-based expansion discussed in Aït-Sahalia (2002) to the conditional expectation $\bar{P}_\ell(\tau, k, v)$ yields the Taylor expansion with respect $\tau = \epsilon^2$ in the form:

$$\bar{P}_\ell^{(J)}(\tau, k, v) = \sum_{l=0}^{J} \Phi_\ell^{(l)}(k)\tau^l,$$

for any integer order $J \geq 0$, where the expansion terms $\Phi_\ell^{(l)}$ are in closed form. The desired bivariate expansion of $\bar{P}_\ell$ follows from further expanding the coefficients $\Phi_\ell^{(l)}(k)$ with respect to k.

The last part of this Appendix shows the calculations to link the coefficients $\varphi^{(i,j)}$ to the IV surface shape characteristics in Section 7.3. It follows by setting $k = 0$ in the bivariate expansion (36) that

$$\Sigma^{(L_0)}(\tau, 0, v_t) = \sum_{i=0}^{L_0} \varphi^{(i,0)}(v_t)\tau^{\frac{i}{2}}. \tag{B.15}$$

Differentiating both sides of (36) with respect to k once, twice, or thrice, and then taking k to be zero, we obtain

$$\frac{\partial}{\partial k}\Sigma^{(L_1)}(\tau, 0, v_t) = \sum_{i=0}^{L_1} \varphi^{(i,1)}(v_t)\tau^{\frac{i}{2}}, \tag{B.16}$$

$$\frac{\partial^2}{\partial k^2}\Sigma^{(L_2)}(\tau, 0, v_t) = \sum_{i=-1}^{L_2} 2\varphi^{(i,2)}(v_t)\tau^{\frac{i}{2}}, \tag{B.17}$$

and

$$\frac{\partial^3}{\partial k^3}\Sigma^{(L_3)}(\tau, 0, v_t) = \sum_{i=-2}^{L_3} 6\varphi^{(i,3)}(v_t)\tau^{\frac{i}{2}}.$$

Equation (B.15) (resp. (B.16)) implies the first (resp. third) equation in (39a) as τ approaches to zero, i.e., $\varphi^{(0,0)}(v_t) = \lim_{\tau\to 0}\Sigma(\tau, 0, v_t)$ (resp. i.e., $\varphi^{(0,1)}(v_t) = \lim_{\tau\to 0}\partial\Sigma/\partial k(\tau, 0, v_t)$.) The rest of equations listed in (39a)–(39d) hinge on finding the univariate Taylor expansions with respect to $\sqrt{\tau}$ of the time-scaled shape characteristics or their combinations appearing on the right hand sides of these equations. Consider (39c). It follows from (B.17) that

$$\frac{1}{2}\frac{\partial^2}{\partial k^2}\Sigma^{(L_2)}(\tau, 0, v_t) = \sum_{i=-1}^{L_2} \varphi^{(i,2)}(v_t)\tau^{\frac{i}{2}} \text{ and } \tau\frac{\partial^3}{\partial k^2 \partial \tau}\Sigma^{(L_2)}(\tau, 0, v_t) = \sum_{i=-1}^{L_2} i\varphi^{(i,2)}(v_t)\tau^{\frac{i}{2}}.$$

Adding the above two equations yields

$$\frac{1}{2}\frac{\partial^2}{\partial k^2}\Sigma^{(L_2)}(\tau,0,v_t)+\tau\frac{\partial^3}{\partial k^2\partial\tau}\Sigma^{(L_2)}(\tau,0,v_t)=\sum_{i=0}^{L_2}(i+1)\varphi^{(i,2)}(v_t)\tau^{\frac{i}{2}},$$

which is a Taylor expansion with leading term $\varphi^{(0,2)}(v_t)$ and (39c) follows by letting τ approach zero.

Parameter	True	Exact identification		Over identification	
		Bias	Std. dev.	Bias	Std. dev.
κ	3.00	-0.031	0.554	-0.259	0.488
α	0.04	3.21×10^{-4}	0.0022	0.0012	0.0029
ξ	0.20	0.0021	0.0109	3.53×10^{-4}	0.0106
ρ	-0.70	0.0058	0.0374	0.0017	0.0374

Table 1: Monte Carlo results for parametric implied stochastic volatility model of type (27a)–(27b)

Note: In the fourth and sixth columns, the standard deviation of each parameter is calculated by the finite-sample standard deviation of estimators based on the 500 simulation trials.

	Number			Mean			Standard deviation		
Days-to-expiration	$[15, 30]$	$(30, 45]$	$(45, 60]$	$[15, 30]$	$(30, 45]$	$(45, 60]$	$[15, 30]$	$(30, 45]$	$(45, 60]$
Log-moneyness k									
$k < 5\%$	8,481	22,275	27,261	21.92	19.68	19.24	4.68	4.43	4.19
$-5\% \leq k \leq -2.5\%$	32,319	24,598	15,643	15.38	15.05	15.18	3.35	3.13	3.08
$-2.5\% \leq k < 0$	40,983	24,151	15,635	12.19	12.69	13.05	3.39	3.36	3.22
$0 \leq k < 2.5\%$	23,556	16,025	10,392	10.51	10.89	11.27	3.52	3.56	3.52
$2.5\% \leq k < 5\%$	2,417	3,015	2,205	14.10	12.58	11.92	3.50	3.64	3.83
$k \geq 5\%$	106	269	291	18.87	17.01	15.80	2.21	2.42	2.51
Total	107,862	90,333	71,427	13.59	14.75	15.60	4.66	4.82	4.79

Table 2: Descriptive statistics for the S&P500 index implied volatility data, 2013 – 2017

Note: The sample consists of daily implied volatilities of European options written on the S&P 500 index covering the period of January 2, 2013 – December 29, 2017. The columns "Mean" and "Standard deviation" are reported as percentages. The log-moneyness k is defined by $k = \log(K/S_t)$, where K is the exercise strike of the option and S_t the spot price of the S&P 500 index.

Parameter	Exact identification		Over identification	
	Estimator	Standard error	Estimator	Standard error
κ	15.2	1.95	13.5	1.64
α	0.023	0.0032	0.022	0.0030
ξ	0.98	0.065	0.77	0.052
ρ	-0.619	0.0021	-0.609	0.0038

Table 3: Parametric implied stochastic volatility model of type (27a)–(27b)

Note: In the third and fifth columns, the standard error of each parameter is calculated by the Newey-West (sample-based) estimator according to (19) and (20). For instance, the standard error of the parameter κ is given by $\sqrt{\hat{V}_{11}^{-1}(\hat{\theta})/n}$, where $\hat{V}_{11}^{-1}$ represents the $(1,1)$th entry of the matrix $\hat{V}^{-1}$.

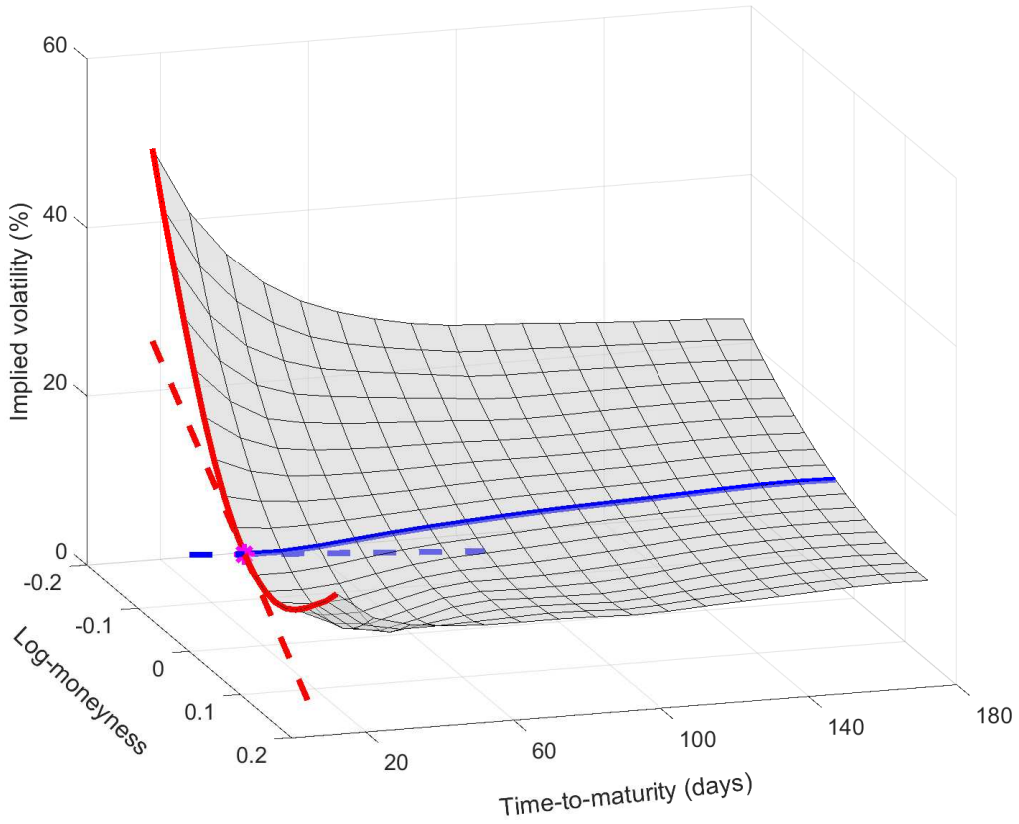

Figure 1: The implied volatility surface of S&P 500 index's options on January 3, 2017

Note: This plot represents the IV surface $(\tau, k) \longmapsto \Sigma(\tau, k, v_t)$ on January 3, 2017 for S&P 500 index options. The two slopes $\Sigma_{0,1}(v_t)$ (log-moneyness slope, or implied volatility smile) and $\Sigma_{1,0}(v_t)$ (term-structure slope) are approximated and represented as red and blue dashed lines, respectively, with each partial derivative $\partial^{i+j}\Sigma(\tau, 0, v_t)/\partial\tau^i\partial k^j$ evaluated at $\tau = 1$ month.

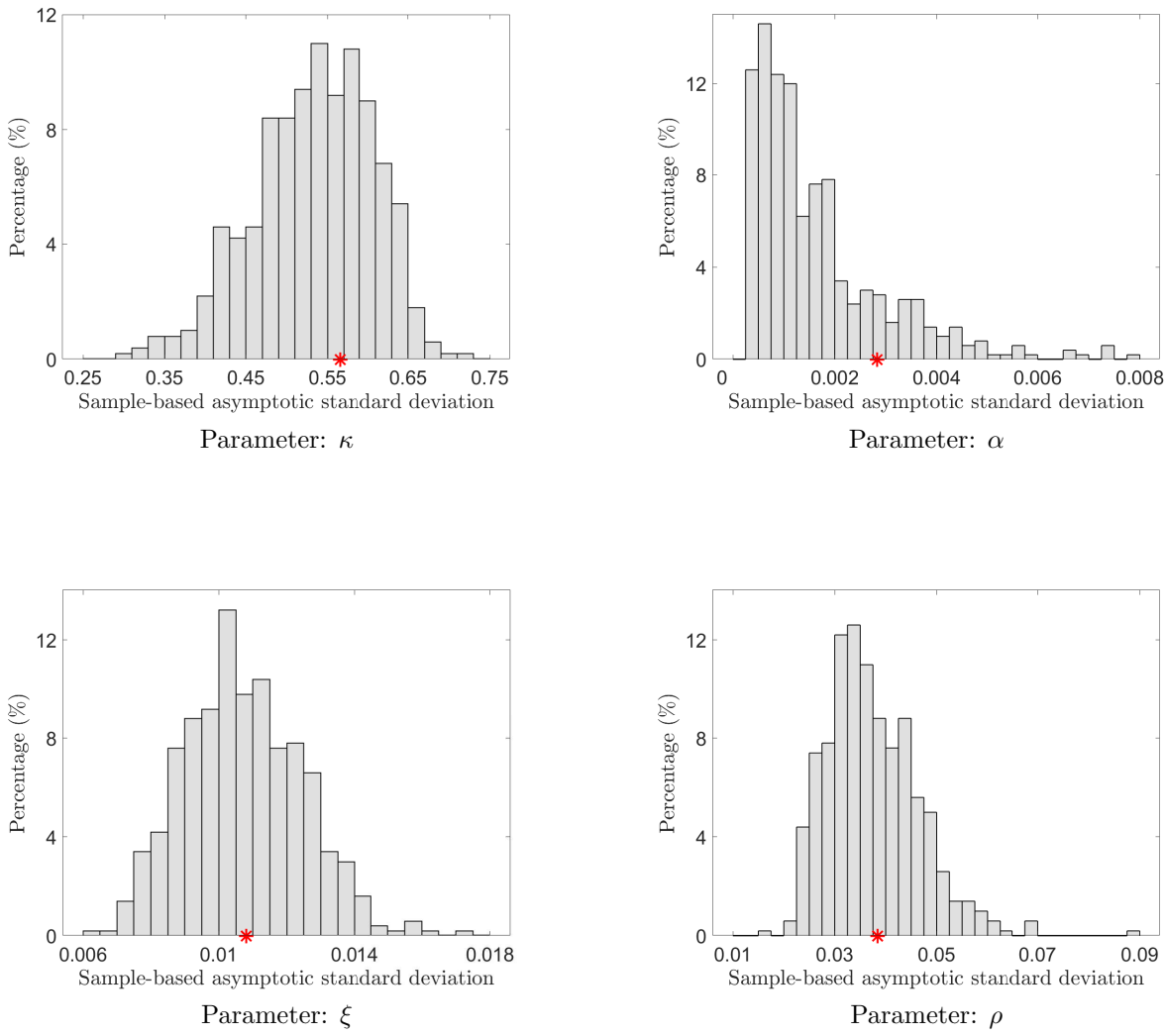

Figure 2: Histograms of the Newey-West estimators of asymptotic standard deviations for the exactly identified case

Note: In each panel, the histogram characterizes the distribution of 500 Newey-West (sample-based) estimators of asymptotic standard deviations. For each simulation trial, the sample-based asymptotic standard deviation is calculated according to (19) and (20). The red star marks the finite-sample standard deviation of the corresponding parameter as shown in the fourth column of Table 1.

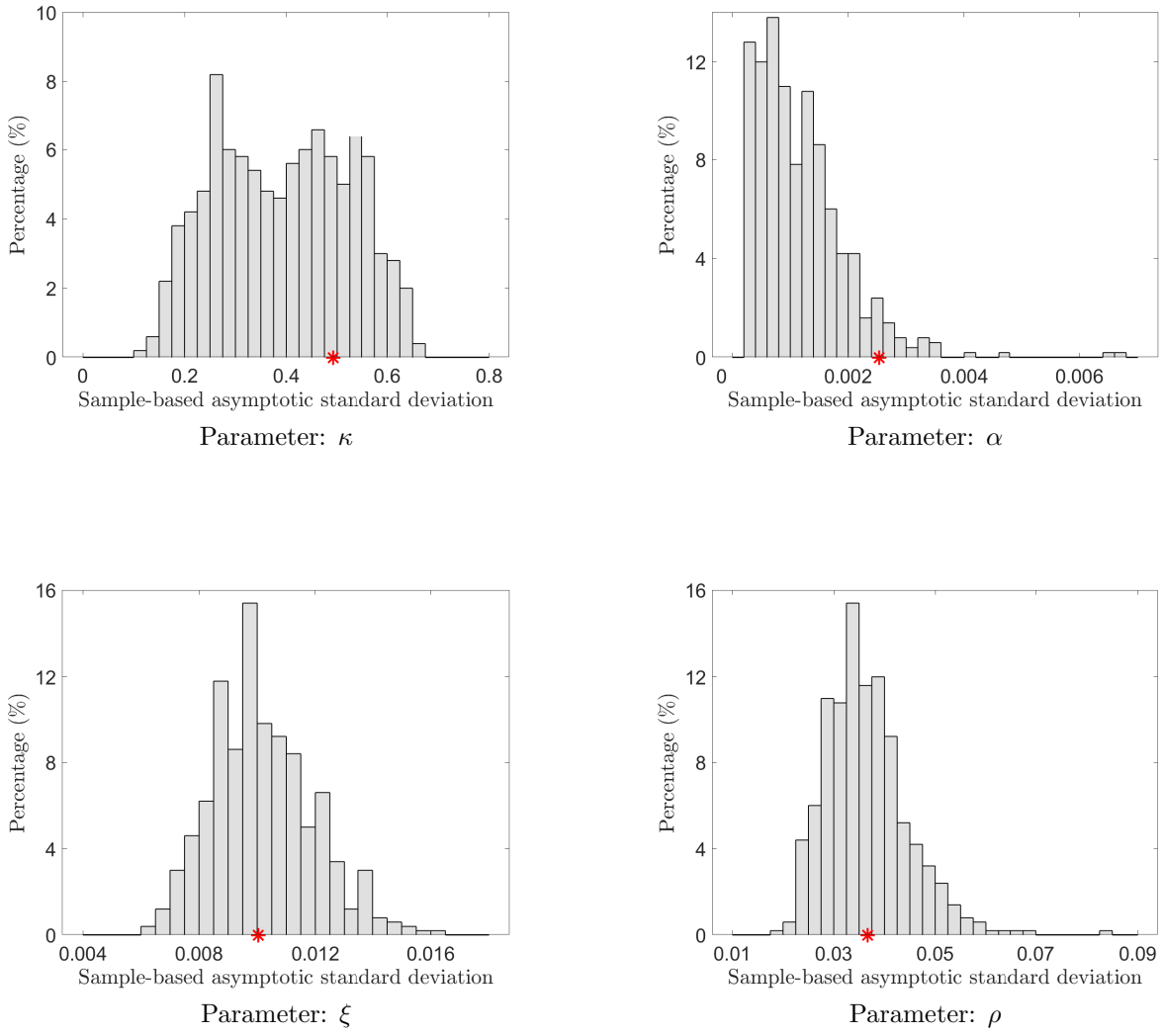

Figure 3: Histograms of the Newey-West estimators of asymptotic standard deviations for the over-identified case

Note: Except for switching to the over-identified case, all the other settings for these four panels remain the same as those for producing Figure 2.

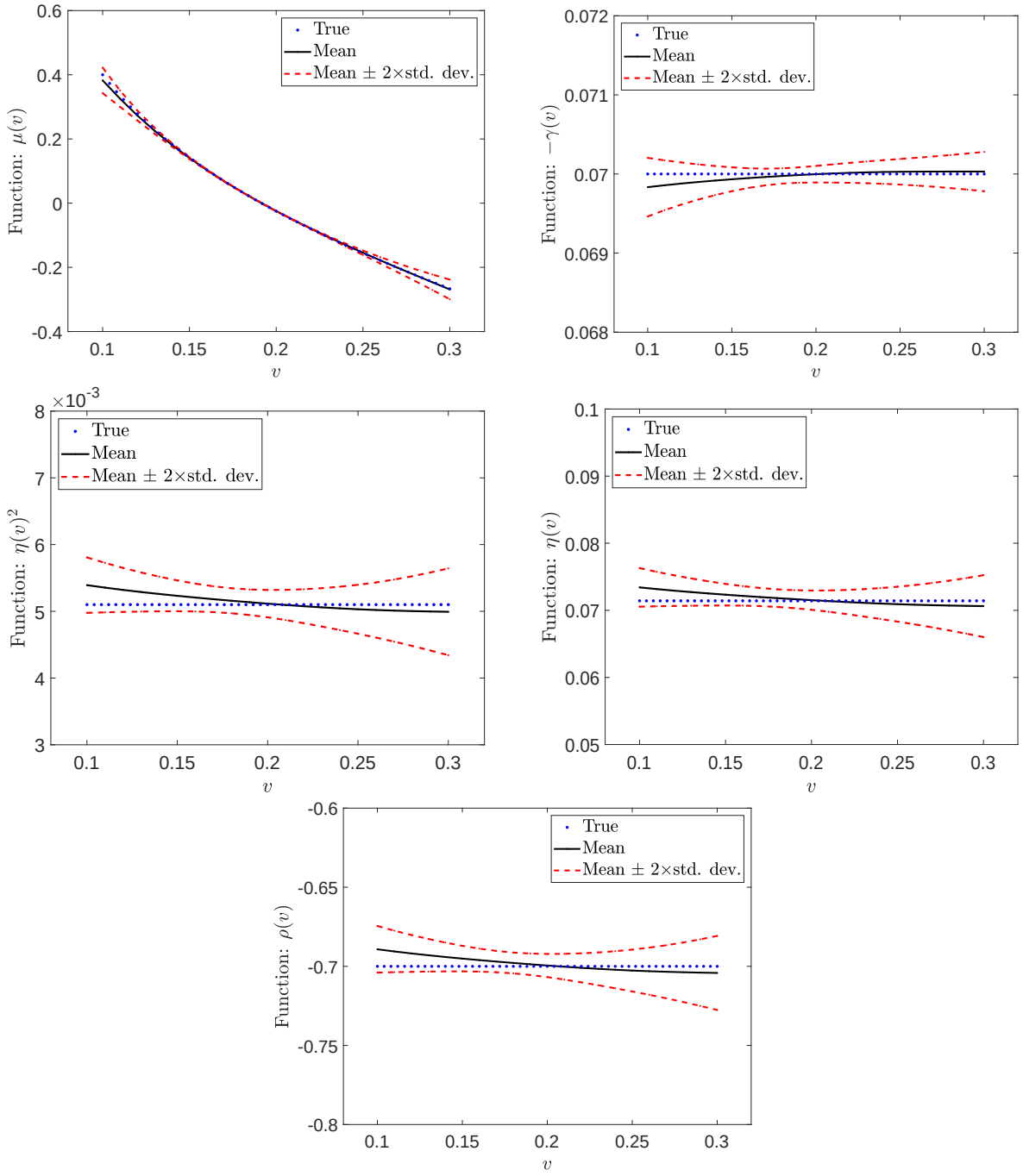

Figure 4: Monte Carlo results for nonparametric implied stochastic volatility model (1a)–(1b)

Note: In each panel, the true function is determined or calculated according to (28). The black solid curve represents the mean of nonparametric estimators corresponding to the 500 simulation trials. Each point on the upper (resp. lower) red dashed curve is plotted by vertically upward (resp. downward) shifting the corresponding one on the black mean curve by a distance equal to twice of the corresponding finite-sample standard deviation.

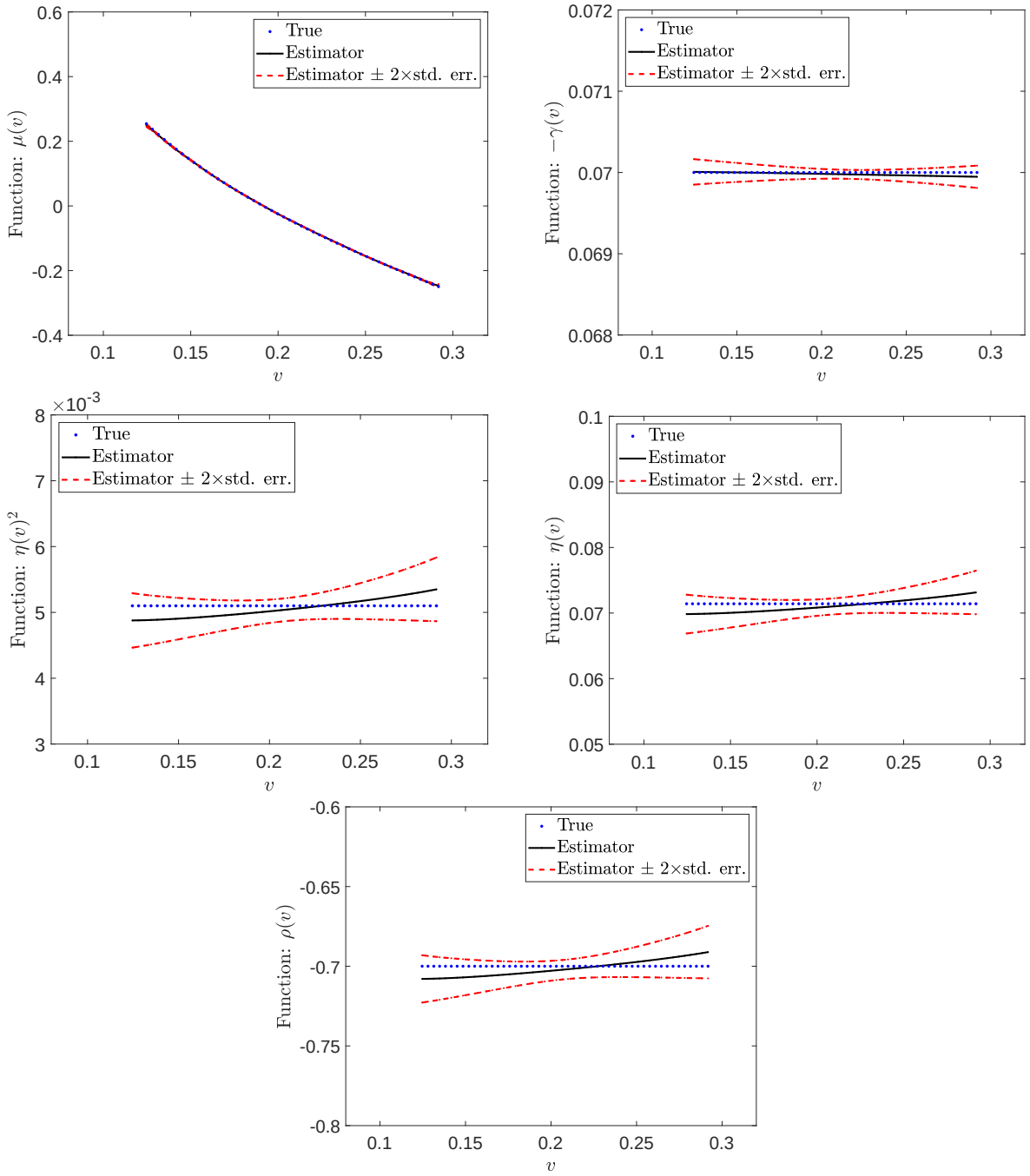

Figure 5: Nonparametric implied stochastic volatility model (1a)–(1b) built from one-trial simulation

Note: In each panel, the true function is determined or calculated according to (28). The black solid curve represents the one-trial nonparametric estimator. Each point on the upper (resp. lower) red dashed curve is plotted from vertically upward (resp. downward) shifting the corresponding one on black curve by a distance equal to twice of the corresponding standard error. Here, the standard error is calculated by the bootstrap strategy introduced in Section 5.2.

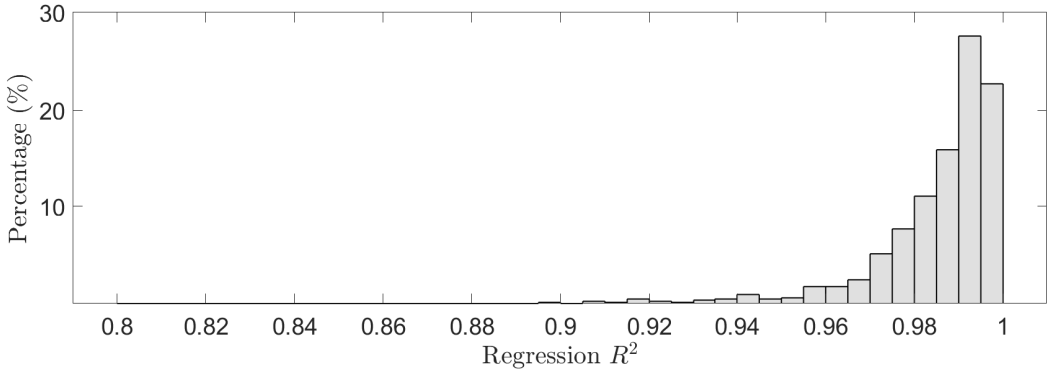

Figure 6: Histogram of R^2 for parametric regressions (32) for individual days across the whole sample covering the period of January 2, 2013 to December 29, 2017.

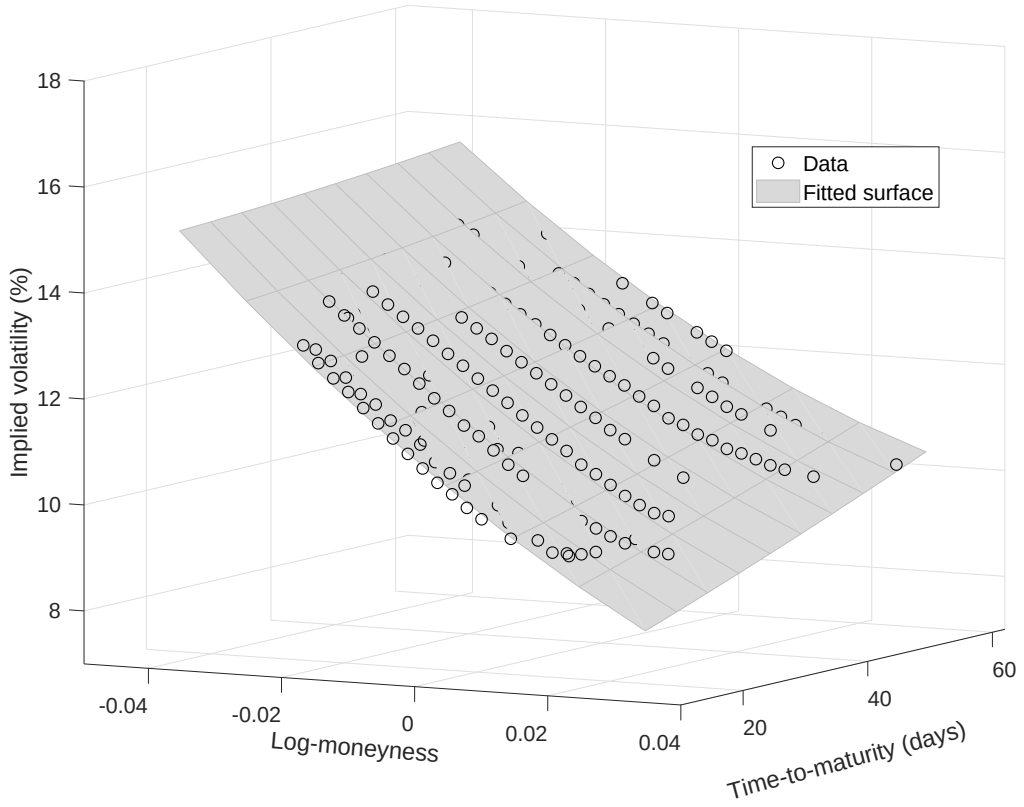

Figure 7: Implied volatility data on January 3, 2017 and the corresponding parametric fitted surface with regression $R^2 = 0.9868$

Note: The parametric fitted surface is calculated according to bivariate regression (32).

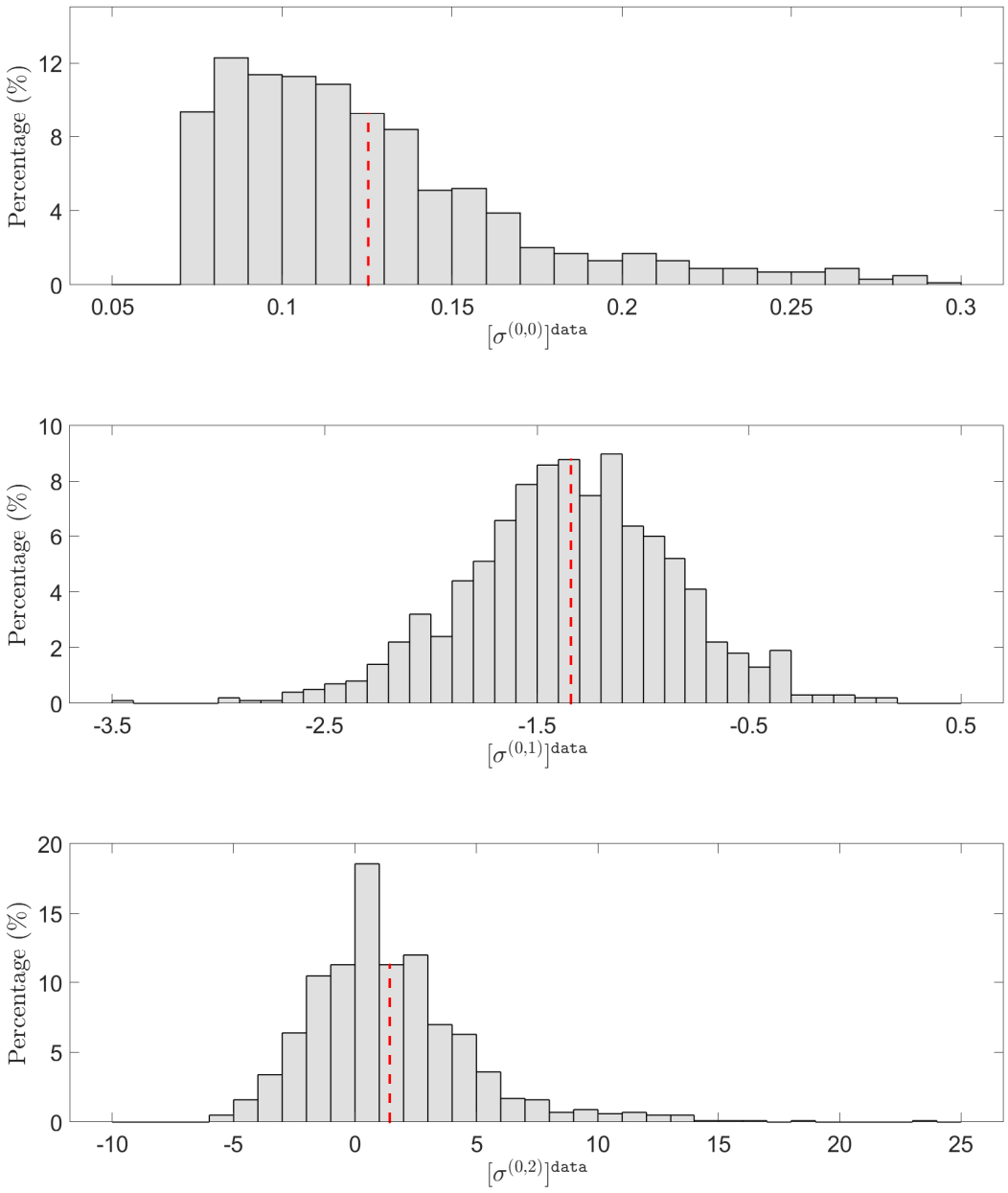

Figure 8: Histograms for the data of $[\sigma^{(0,0)}]^{\text{data}}$, $[\sigma^{(0,1)}]^{\text{data}}$, and $[\sigma^{(0,2)}]^{\text{data}}$

Note: $[\sigma^{(0,0)}]^{\text{data}}$, $[\sigma^{(0,1)}]^{\text{data}}$, and $[\sigma^{(0,2)}]^{\text{data}}$ are the data of expansion terms $\sigma^{(0,0)}$, $\sigma^{(0,1)}$, and $\sigma^{(0,2)}$, respectively. They are prepared from the bivariate regression (32) across the whole sample. In each panel, we plot a red dashed vertical bar to represent the mean of the corresponding histogram.

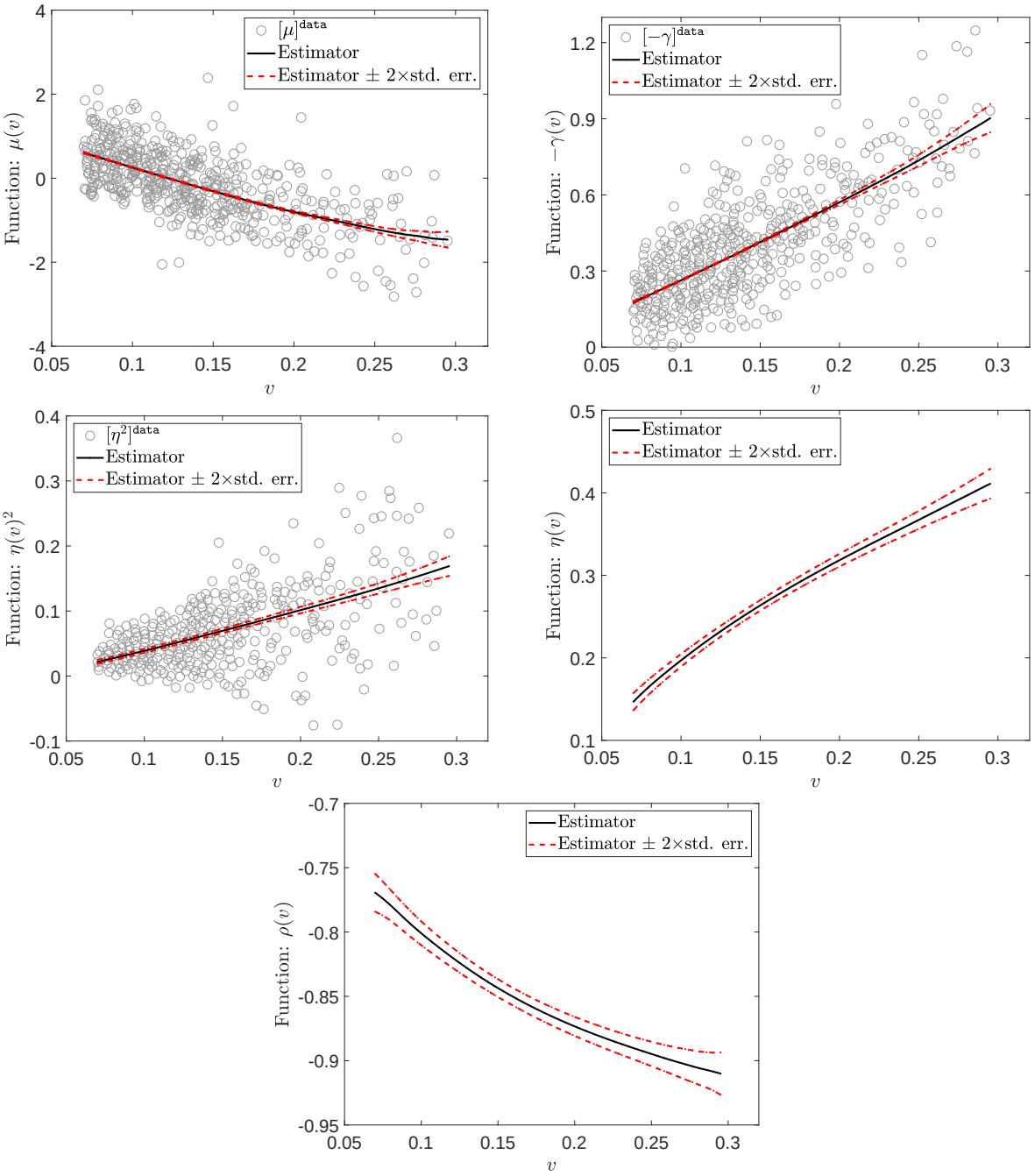

Figure 9: Nonparametric implied stochastic volatility model (1a)–(1b)

Note: In the upper left and middle left panels, the data $[\mu]^{\text{data}}$ and $[\eta^2]^{\text{data}}$ are calculated according to (26) and (25), respectively. In the upper right panel, the data $[-\gamma]^{\text{data}}$ are simply the opposite numbers of the data $[\gamma]^{\text{data}}$, which are calculated according to (22). In all these three panels, the nonparametric estimators are obtained by local linear regressions according to the method proposed in Section 4. In the middle right panel, the nonparametric estimator of η follows by taking square root of the estimator of η^2. In the lowest panel, the nonparametric estimator of ρ follows from (31). In all the panels, the standard errors of estimators are calculated by the bootstrap strategy introduced in Section 5.2.

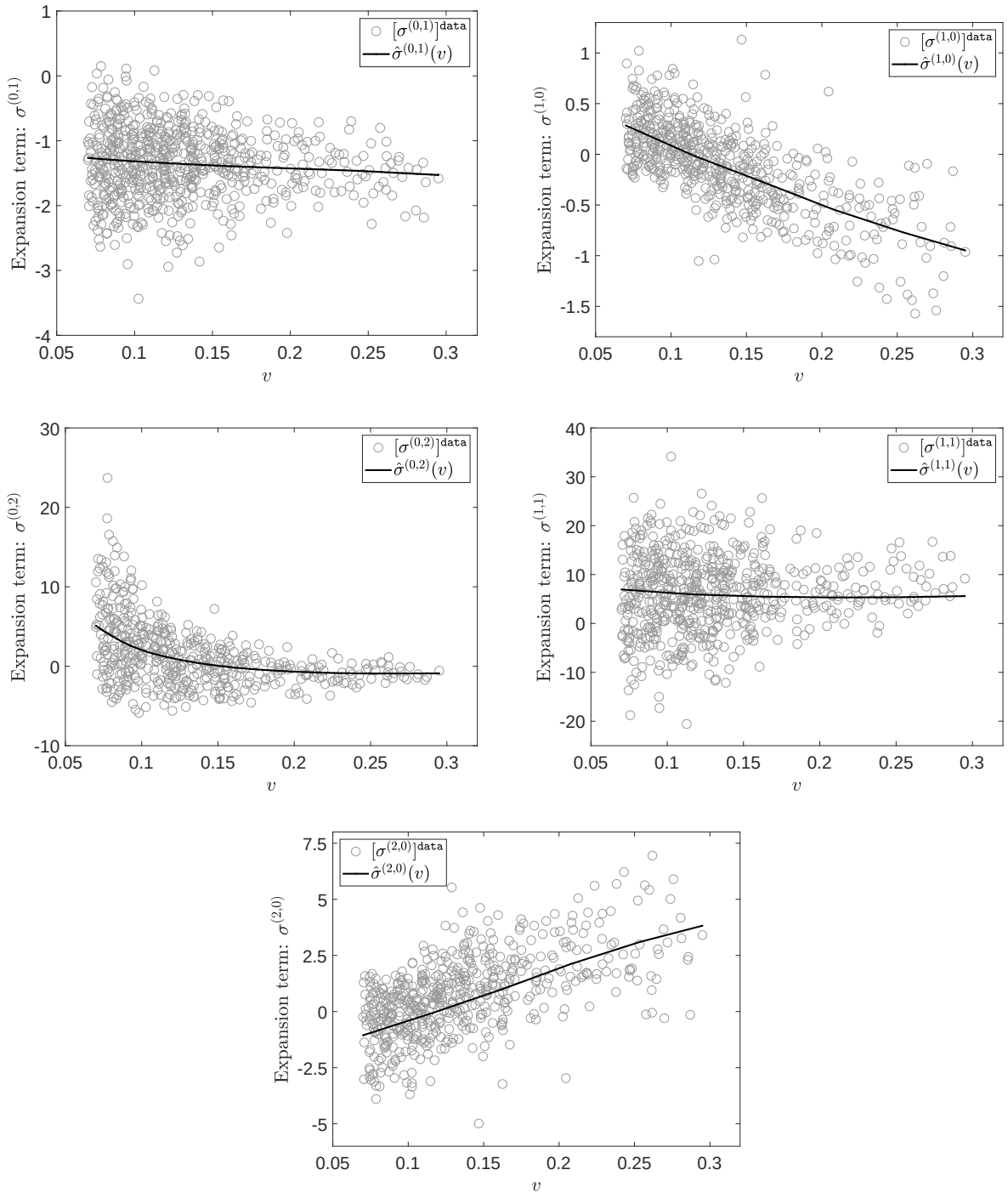

Figure 10: Back-check of the fitting performances on expansion terms

Note: In each panel, the data $[\sigma^{(i,j)}]^{\text{data}}$ are obtained from bivariate regression (32), while the fitted expansion terms $\hat{\sigma}^{(i,j)}$ are obtained from replacing the functions μ, γ, and η, as well as their derivatives by their nonparametric estimators in the formula of $\sigma^{(i,j)}$.