Math. Commun. **30**(2025), 107–130

Asymptotic analysis for an optimal estimating function for Barndorff-Nielsen Shephard stochastic volatility models

Friedrich Hubalek[1] and Petra Posedel Šimović[2,*]

[1] *Institute of Statistics and Mathematical Methods in Economics, TU Wien, Wiedner Hauptstraße 8–10, 1040 Vienna, Austria*
[2] *Department of Information Science and Mathematics, Faculty of Agriculture, University of Zagreb, Svetošimunska cesta 25, HR-10 000 Zagreb, Croatia*

Received September 3, 2024; accepted January 15, 2025

Abstract. We provide and analyze optimal estimators from a fixed sample and asymptotic point of view for a class of discretely observed continuous-time stochastic volatility models with jumps. In particular, we consider a class of non-Gaussian Ornstein-Uhlenbeck-based models, as introduced by Barndorff-Nielsen and Shephard.
We develop in detail a martingale estimating function approach for this kind of processes, which are bivariate Markov processes that are not diffusions, but admit jumps. We assume that the bivariate process is observed on a discrete grid of fixed width, and the observation horizon tends to infinity.
We prove rigorously consistency and asymptotic normality of the optimal estimator based on a single assumption that all moments of the stationary distribution of the variance process are finite, and give explicit expressions for the asymptotic covariance matrix.
As an illustration, we provide a simulation study for daily increments, but the method applies unchanged to any time-scale, including high-frequency observations, without introducing any discretization error. Additionally, we compare the asymptotic covariance matrix of the optimal estimator with the one of the simple explicit estimators and investigate the improvement in variance reduction, even though this improvement is not significant.
This paper complements earlier works [24, 25].
AMS subject classifications: 60G51, 62F12, 62M05

Keywords: optimal martingale estimating functions, stochastic volatility models with jumps, consistency and asymptotic normality

1. Introduction

In [5], Barndorff-Nielsen and Shephard introduced a class of stochastic volatility models in continuous time, where the instantaneous variance follows an Ornstein-Uhlenbeck type process driven by an increasing Lévy process. BNS models, as we will refer to them from now on, allow flexible modelling, capture many stylized facts of financial time series, and yet are of great analytical tractability. This model class has been studied from various points of view in mathematical finance and econometrics, with MathSciNet listing, at the time of writing, over 400 citations of the seminal

*Corresponding author. *Email addresses:* `fhubalek@fam.tuwien.ac.at` (F. Hubalek), `pposedel@agr.hr` (P. Posedel Šimović)

https://www.mathos.unios.hr/mc

article underlining its relevance, see [35, 29, 34, 7, 42, 8, 30, 43, 4, 52, 21, 15, 2], and [6] for a multivariate setting. More recently, [19, 20] studied generalizations based on superpositions. In a comprehensive note reviewing some of the contributions of Ole Barndorff-Nielsen in financial econometrics, Neil Shephard discusses the model in [48, Section 2]. BNS models are also treated in the textbooks [12, 46, 47, 38, 3].

The literature on statistical estimation for BNS models focuses mostly on computationally intensive Bayesian Markov Chain Monte Carlo methods, see, for example, [43, 21, 16, 17, 15] and the references therein. Our approach is different, as we investigate how far we can get in the statistical estimation with explicit calculations, without introducing discretization errors and approximations. This is motivated by the analytical tractability of the BNS model class, which was exploited in other contexts, for example, by semi-closed form option pricing formulas, see [35].

Very often, estimating functions provide a tractable alternative when the likelihood function is not explicitly available for stochastic process models, when it is not tractable, or when it requires heavy computations [27, Section 1]. Indeed, in the case of stochastic volatility models, both without and with jumps, the likelihood function is often unknown, intractable or presents some maximization problems, or in semi-parametric cases, it does not exist. In that case, quasi-likelihood theory was developed as an alternative [23].

For a general discussion of and further references to the estimation of discretely observed stochastic volatility models we refer the reader to [24] and [25] and the references therein. In those papers, we found explicit estimators using the martingale estimating function approach and showed their consistency and asymptotic normality. However, the obtained estimators are not efficient. In this paper, we investigate the corresponding *optimal* martingale estimating functions in the sense of [23, Chapter 2], where the best element from a fixed sample size O_F and from an asymptotic point of view O_A is analyzed. We use heavily the general framework from [49] and [50]. Our contributions are as follows:

- We show that the framework in [49] can be applied in a setting with a bivariate process that is not a diffusion, but admits jumps.

- We derive explicit expressions for the optimal quadratic martingale estimating functions.

- We *prove* rigorously the assumptions required to show consistency and asymptotic normality of the optimal martingale estimating function for a particular example, the model with a stationary gamma distribution, the so-called Gamma-OU model.

- We show that the asymptotic covariance matrix of the estimator can be computed in closed form. In contrast to the estimators from [24, 25], the estimator itself, which is the solution of the estimating equations, must be computed numerically using a root-finding algorithm.

As a numerical illustration, we compared the optimal estimator to the simple estimator from [25]. We found the improvement due to using the optimal estimator instead of the simple one to be non-negligible in terms of the asymptotic variance-covariance matrix, but the gain is questionable considering the complexity of using the optimal estimator.

We would like to emphasize that the estimation problem for BNS models is very different from the estimation problem for Ornstein-Uhlenbeck type processes alone, treated in [9, 55, 10, 37]. The paper is organized as follows: In Section 2, we describe the BNS model studied and the optimal estimating functions, and give the results of consistency and asymptotic normality of the corresponding estimator. Section 3 deals with the computation of the optimal weights used for the optimal estimating functions, and we prove their limiting properties. The asymptotic performance of the optimal estimator and its comparison with the simple estimator is illustrated by a numerical example in Section 4. Section 5 concludes and gives directions for further developments. Longer proofs and calculations are presented in Appendices A and B.

2. Model setting and main results

2.1. The continuous time model: a general setting

For the derivation and study of the optimal estimator we use the same notation and framework as in [25], where a simple estimator was derived and studied. Therefore, we give only a brief summary and required details for introducing the model.

Assume we are given a probability space $(\Omega, \mathcal{F}, P)$ carrying a standard Brownian motion W and an independent subordinator Z. The Barndorff-Nielsen and Shephard model consists of a process modelling logarithmic returns denoted by X, and an instantaneous variance process is denoted by V satisfying

$$dX(t) = (\mu + \beta V(t-))dt + \sqrt{V(t-)}dW_\theta(t) + \rho dZ_\lambda(t), \quad X(0) = 0.$$

and

$$dV(t) = -\lambda V(t-)dt + dZ_\lambda(t), \quad V(0) = V_0.$$

The parameters μ, β, ρ and λ are real constants with $\lambda > 0$ and $Z_\lambda(t) = Z(\lambda t)$ for all $t \geq 0$. The random variable V_0 has a self-decomposable distribution D on $\mathbb{R}_+$ and Z is chosen such that the process V is strictly stationary, see the remark below for the construction and the required conditions. Typical examples for D used in finance are the gamma distribution, the (generalized) inverse Gaussian distribution, or tempered stable distributions.

$\mathbf{X} = (X, V)$ is called a BNS-D-OU model. The bivariate process $\mathbf{X}$ is clearly Markovian. For the statistical analysis we further assume throughout the paper that $E[V_0^n] < \infty$, for all $n \in \mathbf{N}$, denote the mean and variance by $\zeta = E[V_0]$ and $\eta = \mathrm{Var}[V_0]$, and introduce the parameter vector $\theta = (\lambda, \zeta, \eta, \mu, \beta, \rho)^\top$ for compactness.

Remark 1. *The construction and conditions for a stationary solution are discussed in [5, Sec.2.1, Thm.1], with reference to [56, 28, 1]. We require here the notion of self-decomposability, a concept from the study of infinite divisibility of distributions. For a definition and details, see [45, Sec. 3.15, P.90f]. All example distributions mentioned above are known to be self-decomposable.*

Now, given a self-decomposable distribution D, there is a probability space $(\Omega, \mathcal{F}, P)$ carrying a random variable V_0 with distribution D and a subordinator Z, such that (2.1) admits for every $\lambda > 0$ a stationary solution V with $V(0) = V_0$. The introduction of the time-changed subordinators Z_λ allows us to have D and Z that do not depend on λ with the intention of separating the stationary distribution from the dynamics. We are following this approach here.

Barndorff-Nielsen and Shephard discuss a second method, called later $OU - D$, that starts from a given subordinator Z and derives the stationary law D. This method requires a logarithmic moment condition for $Z(1)$ and yields slightly different dynamics and distributions for the model. As we do not adopt this approach here, we refer the interested reader to [5, Sec. 2.4] for further details.

For estimation purposes, we consider a parameterized family of probability measures, $(P_\theta : \theta \in \Theta)$, where $\Theta = \{\theta \in \mathbb{R}^6 : \theta^1 > 0, \theta^2 > 0, \theta^3 > 0\}$ is the parameter space. The expectation with respect to P_θ and P_{θ_0} is denoted by $E_\theta[.]$ and $E[.]$, respectively. We assume that there is a process $\mathbf{X}$ that is BNS-DOU(θ) under P_θ.

It is assumed that we observe the process (X, V) on a discrete grid of points in time, $0 = t_0 < t_1 < \ldots < t_n$, and denote the observations by $X_1, \ldots, X_n, V_1, \ldots, V_n$ with $X_i = X(t_i) - X(t_{i-1})$ for $i = 1, \ldots, n$ and $V_i = V(t_i)$ for $i = 0, \ldots, n$. We want to find an optimal estimator for θ_0 in the sense of [23, Def. 2.3] using observations $X_1, \ldots, X_n, V_1, \ldots, V_n$. We are interested in asymptotics as $n \to \infty$.

Let us briefly review the notions of optimality for martingale estimating functions. In [23], O_A and O_F optimality are distinguished, but the two notions agree on the class (in our setting) of estimating functions of the form

$$G_n(\theta) = \sum_{k=1}^{n} \alpha_k(\theta) \Delta M_k(\theta),$$

where $M(\theta)$ is a martingale and $\alpha(\theta)$ is a predictable process for all θ. An estimator of this form is O_F-optimal if it maximizes the information criterion $\mathcal{E}(G_n)$ which can be seen as a generalization of the Fisher information in the partial order of non-negative definite matrices, see [23, Section 2.2] for precise definitions and details.

2.2. Description of the optimal estimating function

In this section, the notation from [50] is followed and the results are extended to bivariate and jump processes. Let us denote by $\mathcal{G}$ the class of zero mean, square integrable martingale estimating functions of the form $G_n = G_n\big(\{(X_i, V_i), i \le n\},\ \theta\big)$. We will consider the optimality within the class $\mathcal{M} \subset \mathcal{G}$ of quadratic estimating functions. We will find optimal quadratic martingale estimating functions for BNS models, following the general theory presented in [23].

For the sake of notational simplicity, let us introduce the abbreviations

$$h = h(x, v, w, \theta), \quad g = g(x, v, w, \theta)$$

with

$$v \mapsto V_{i-1}, \quad w \mapsto V_i, \quad x \mapsto X_i.$$

More precisely, an estimating function in $\mathcal{M}$ is of the form

$$G_n(\theta) = \sum_{i=1}^{n} g(X_i, V_i, V_{i-1}, \theta), \quad g = (g_1, g_2, \ldots, g_p)^T, \quad g_i = g_i(x, v, w, \theta), \tag{1}$$

where for $N = 5$ and $h = (h_1, h_2, \ldots h_5)^T$,

$$g(x_1, v_1, v_0, \theta) = \sum_{j=1}^{N} \alpha_j(v_0, \theta) h_j(x_1, v_1, v_0, \theta), \quad h_j = h_j(x, v, w, \theta), \tag{2}$$

$$\begin{array}{ll} h_1(x_1, v_1, v_0, \theta) = v_1 - f_1(v_0, \theta), & f_1(v, \theta) = E_\theta[V_1 | V_0 = v] \\ h_2(x_1, v_1, v_0, \theta) = x_1 - f_2(v_0, \theta), & f_2(v, \theta) = E_\theta[X_1 | V_0 = v] \\ h_3(x_1, v_1, v_0, \theta) = v_1^2 - f_3(v_0, \theta), & f_3(v, \theta) = E_\theta[V_1^2 | V_0 = v] \\ h_4(x_1, v_1, v_0, \theta) = x_1^2 - f_4(v_0, \theta), & f_4(v, \theta) = E_\theta[X_1^2 | V_0 = v] \\ h_5(x_1, v_1, v_0, \theta) = x_1 v_1 - f_5(v_0, \theta), & f_5(v, \theta) = E_\theta[X_1 V_1 | V_0 = v], \end{array} \tag{3}$$

and $\alpha_j(v, \theta)$ is a $p-$dimensional vector of measurable functions of v for each θ and $j = 1, \ldots, N$ such that $G_n(\theta)$ is square integrable.

We want to find the optimal estimating function for the class $\mathcal{M}$ distinguishing between fixed sample and asymptotic properties. In our case, $p = 6$.

Remark 2. *According to [23, Theorem 2.5], we will show that O_A-optimality will, in our case, imply the O_F-optimality.*

In the BNS setting, according to the general moment calculations given in [25, Appendix A], the conditional expectations given by (3) are polynomials in v. Namely, for every $j = 1, \ldots, 5$, we have

$$f_j(v, \theta) = E[X_1^{r_j} V_1^{s_j} | V_0 = v] = \sum_{l=0}^{p_j} \phi_l^j(\theta) \cdot v^l, \tag{4}$$

where the degree p_j and the coefficients $\phi_l^j(\cdot)$, which are smooth functions in θ, can be calculated explicitly. Based on these closed form expressions, the martingale estimating function $G_n(\theta)$ is still explicit, but the resulting estimator has to be solved numerically. The simple explicit estimator obtained in [25] might be a good starting point for solving equations $G_n(\theta) = 0$ numerically with general $\alpha_j(v, \theta)$.

A first step towards the computation of the optimal estimating function in a bivariate setting is given by the following result. In order to prove the following results, the theory from [23] is extended in the case of a bivariate Markov process that is not a diffusion, but admits jumps. We do not approximate transition probabilities, but can make use of exact expressions.

Theorem 1 (Optimality)**.** *Assume (X, V) is a BNS D-OU model with*

$$E[V_0^n] < \infty, \ \forall \ n \ \in \mathbf{N}, \quad \zeta = E[V_0], \quad \eta = \mathrm{Var}[V_0].$$

Denote the parameter vector $\theta = (\lambda, \zeta, \eta, \mu, \beta, \rho)^\top$, where μ, β, ρ and λ are real constants with $\lambda > 0$. Take observations $X_1, \ldots, X_n$ with $X_i = X(t_i) - X(t_{i-1})$ and

$V_0, V_1, \dots, V_n$ with $V_i = V(t_i)$, for $i = 1, \dots, n$ on an equidistant grid $t_0, \dots, t_n$ of fixed width $\Delta > 0$, and let $\mathcal{M}$ denote the square-integrable quadratic martingale estimating functions of the form (1)–(3), with

$$E_\theta \left[a_{ij}(V_0)^2 h_j(V_0, V_1, X_1)^2\right] < \infty, \quad \text{for all } 1 \le i \le 6, \quad 1 \le j \le 5.$$

Define the matrix functions

$$A^*(v;\theta) = \{a_{ij}(v;\theta)\}, \quad B(v;\theta) = \{b_{ij}(v;\theta)\}, \quad C(v;\theta) = \{c_{ij}(v;\theta)\} \tag{5}$$

$$b_{ij}(v,\theta) = \partial_i f_j(v,\theta), \quad c_{ij}(v,\theta) = E_\theta[h_i h_j | V_0 = v]. \tag{6}$$

Assume $C(v;\theta)$ is regular. In that case, define

$$A^*(v;\theta) = B(v;\theta)C(v;\theta)^{-1}$$

and

$$G_n^*(\theta) = \sum_{i=1}^n g^*(X_i, V_i, V_{i-1}, \theta),$$

where

$$g^* = A^*(v_0, \theta)h.$$

Then G_n^* is O_A-optimal in $\mathcal{M}$ and G_n^* is O_F-optimal in $\mathcal{M}$ for all $n \in \mathbb{N}$.

Proof. The proof is given in Appendix A. □

Remark 3. *The entries of B and C can be computed explicitly in terms of the parameters. In particular, $C(v;\theta)$ is a polynomial in v. Those expressions allow us to check in a mathematically rigorous way if*

$$\inf_{v>0} \det(C(v;\theta)) > 0$$

for concrete values of $\theta \in \Theta$, which is enough for estimation purposes, namely, that G_n^* is well-defined, for the conclusion of Theorem 1 and the asymptotic results in Section 2.3.

The simple estimator from [25, Eq. (3.1)] is derived from an estimating function with the same structure as (1), but with a function g° instead of g, that has entries g_i° that are very simple quadratic polynomials in x, v, w. The corresponding estimating equation system can be solved explicitly, and the solution is given in [25, (3.3)].

2.3. Consistency and asymptotic normality

In order to prove consistency and asymptotic normality of the optimal estimator, we use the general framework and results of [49]. In the BNS case, we need to extend the theory in the case of a bivariate Markov process. To apply [49, Corollary 2.6] and [49, Theorem 2.9] to consistency and asymptotic normality, respectively, it is necessary to show that [49, Condition 2.1] and [49, Condition 2.5] or [49, Condition 2.7] are satisfied.

Let

$$J(v,\theta) = -E[A^*(V_0,\theta)B(V_0,\theta)], \tag{7}$$

for matrices B and A^* given in Theorem 1, and

$$M_n^\alpha(\bar\theta) = \big\{\theta \in \Theta : \|\theta - \bar\theta\| \le \frac{\alpha}{\sqrt{n}}\big\}, \qquad \alpha > 0.$$

Theorem 2 (Consistency)**.** *Assume the conditions from Theorem 1. For every n, an estimator $\hat\theta_n$ exists that solves the optimal estimating equation $G_n(\hat\theta_n) = 0$ with a probability tending to one as $n \to \infty$, and that is weakly consistent.*

Proof. In order to prove weak consistency of the estimator, we can apply [49, Corollary 2.6]. This requires showing that [49, Condition 2.7], which contains five rather technical parts, is satisfied. This is shown in Appendix B. □

In order to show asymptotic normality of the optimal estimator, the central limit theorem for the optimal estimating function has to be proved.

Let $\Xi_k = \big(V_k, X_k, V_k^2, X_k^2, X_kV_k\big)^\top$ for $k = 1,\ldots,n$ and for ease of notation, $\Xi_1^m = X_1^{p_m}V_1^{q_m}$ for $m = 1,\ldots N$.

Proposition 1 (Asymptotic normality of the optimal estimating function)**.** *Assume the conditions from Theorem 1. We have*

$$\frac{1}{\sqrt{n}}G_n(\bar\theta) \xrightarrow{\mathcal{D}} N(0,\Phi), \quad \text{as} \quad n \to \infty,$$

where

$$\Phi = \sum_{m=1}^N \sum_{z=1}^N E\big[\alpha^*_{im}(V_0,\bar\theta)\alpha^*_{jz}(V_0,\bar\theta)Cov(\Xi_1^m, \Xi_1^z|V_0)\big].$$

Proof. To show the above result, we use the multivariate martingale central limit theorem. For that purpose, we introduce the vector martingale difference array

$$\chi_{n,k} = \frac{1}{\sqrt{n}}g^*(X_k,V_k,V_{k-1},\theta) = \frac{1}{\sqrt{n}}A^*(V_{k-1},\theta)h(X_k,V_k,V_{k-1},\theta), \tag{8}$$

where A^* and h are defined by (5) and (3). First we prove a multivariate Lyapunov condition which implies the Lindeberg condition. From (8) it follows that for $p = 1,\ldots,d$

$$\sqrt{n}\chi_{n,k}^p \le const(\theta)\sum_{j=1}^N V_{k-1}\big[X_k^{p_j}V_k^{q_j} - f_j(V_{k-1},\theta)\big]. \tag{9}$$

Expression (9) is of the form $p(X_1,V_1,V_0)$, where $p(x_1,v_1,v_0)$ is a polynomial in x_1,v_1,v_0, which does not depend on n. For verifying the second requirement needed for the martingale central limit theorem, we consider the (i,j)-th element of the matrix $\chi_{n,k}\chi_{n,k}^\top$ given by

$$\frac{1}{n}\sum_{m=1}^N\sum_{z=1}^N \alpha^*_{im}(V_{k-1},\theta)\alpha^*_{jz}(V_{k-1},\theta)\big[X_k^{p_m}V_k^{q_m} - f_m(V_{k-1},\theta)\big]\big[X_k^{p_z}V_k^{q_z} - f_z(V_{k-1},\theta)\big].$$

From Theorem 4 it follows that

$$\sum_{k=1}^{n} \left(\chi_{n,k}\chi_{n,k}^{\top}\right)_{i,j} \xrightarrow{a.s.} \sum_{m=1}^{N}\sum_{z=1}^{N} E\left\{\alpha_{im}^{*}(V_0,\theta)\alpha_{jz}^{*}(V_0,\theta)\left[X_1^{p_m}V_1^{q_m} - f_m(V_0,\theta)\right] \times \left[X_1^{p_z}V_1^{q_z} - f_z(V_0,\theta)\right]\right\}.$$

The expectation on the right-hand side is finite and explicit expressions can be given using explicit formulas as special cases of the general moment calculations given in [25]. □

Finally, we have the following result.

Theorem 3 (Asymptotic normality of the optimal estimator). *Assume the conditions from Theorem 1. The estimator $\hat{\theta}_n$ obtained by solving the equation $G_n^*(\theta) = 0$ is asymptotically normal, namely*

$$\sqrt{n}(\hat{\theta}_n - \bar{\theta}) \xrightarrow{\mathcal{D}} N\left(0, J(\bar{\theta})^{-1}\Phi(J(\bar{\theta})^{-1})^{\top}\right),$$

as $n \to \infty$, where Φ is given in Proposition 1 and the matrix $J(\theta)$ is given by (7), with $A^*(v,\theta)$ being the weight matrix having entries that are rational functions in v, and $\dfrac{\partial f}{\partial \theta}(V_0,\theta)$ being the Jacobian matrix of the vector function $f(v,\theta)$ with respect to θ, whose entries are polynomials in v.

Proof. Directly from [49, Theorem 2.9]. □

Remark 4. *A self-decomposable distribution has a density [45, Thm. 27.13], thus the entries of J can be computed by one-dimensional numerical integration using the corresponding density. In the case of the BNS-Γ-OU, we can evaluate the integrals explicitly in terms of the error function. Namely, we have*

$$\int_0^{\infty} \frac{v^k}{v-r} \cdot \frac{\alpha^{\nu}}{\Gamma(\nu)} v^{\nu-1} e^{-\alpha v}\, dv = (\nu)_k \alpha^{1-k} e^{-\alpha r} E_{k+\nu}(-r\alpha),$$

where $(\nu)_k$ and $E_k(z)$ denote the Pochhammer symbol and the exponential integral function, respectively, see [36, §8.19]. Alternatively it can be expressed by an incomplete gamma function. The integral is evaluated for real negative r by an elementary substitution and a quick proof for complex r can be done using an analytic continuation argument.

3. Properties of the optimal weights

3.1. Computation of the optimal weights

The optimal weight α is obtained as a projection. In our setting, it can be described as follows. Let $q(v,\theta) = \det(C(v,\theta))$, where the matrix C is given by (6). From Theorem 1 we know that the optimal A solves

$$A^*(v,\theta)C(v,\theta) = B(v,\theta), \tag{10}$$

where B is given by (5).

Remark 5. *We know that $C(v,\theta)$ is a covariance matrix and is thus positive semi-definite, so $q(v,\theta)=\det(C(v,\theta))\geq 0$.*

From the definition of the determinant, we have

$$\det(C(v,\theta))=\sum_{\pi\in\gamma_6}\mathtt{sgn}(\pi)c_{1\pi(1)}\dots c_{5\pi(5)},$$

where π denotes the permutation function and $|\gamma_6|=6!$. It follows that $q(v,\theta)$ is a polynomial in v, and its coefficients are smooth functions in θ as they are given by, sums of products of entries of $C(v,\theta)$. Since the entries of C are polynomials in v and can be computed explicitly, we can also compute the determinant of C, which is again a polynomial in V. The resulting expression is rather lengthy, but with the help of Mathematica, it can be shown that $q(v,\theta)$, as a polynomial in v has in fact degree 4, i.e., $q(v,\theta)=q_0(\theta)+q_1(\theta)v+q_2(\theta)v^2+q_3(\theta)v^3+q_4(\theta)v^4$ for every θ in some neighborhood of θ_0.

Conjecture 1. *$q(v,\theta)$ as a polynomial in v, for fixed θ, has 4 distinct roots, none of which is real and nonnegative.*

In order to illustrate the optimal estimating function, its performance will be compared with respect to the simple explicit estimator studied in a numerical example in [25]. The conjecture can be numerically checked in practice for the example, but using arbitrary precision arithmetic and respectively interval arithmetic, it can be verified mathematically rigorously for concrete values with the help of a computer, see [54] for more details. But we have not been able to verify the conjecture for *all* values $\theta\in\Theta$. If Conjecture 1 is true, we could use the Cardano-Ferrari formula [51, 11] to write an explicit expression for the roots.

Assuming Conjecture 1, let the roots $r_1(\theta),\dots,r_4(\theta)$. It follows that $r_1(\theta),\dots,r_4(\theta)$ are smooth functions in θ if the coefficients $q_i(\theta)$ are smooth.
If we use Cramer's rule to solve (10), a similar argument shows that $\alpha^*_{ij}(v,\theta)=p_{ij}(v,\theta)/q(v,\theta)$, where $p_{ij}(v,\theta)$ is the *determinant* of the matrix obtained from $C(v,\theta)$, replacing the i-th column by the j-th column of the matrix $B(v,\theta)$. Hence, it follows that $p_{ij}(v,\theta)$ is a polynomial in v and its coefficients are smooth functions in θ as they are given by the sums of products of the entries of $C(v,\theta)$ and $B(v,\theta)$. Similarly, it can be shown that $p_{ij}(v,\theta)$ as a polynomial in v has in fact at most degree 6, i.e.,

$$p_{ij}(v,\theta)=p_{ij0}(\theta)+p_{ij1}(\theta)v+p_{ij2}(\theta)v^2+p_{ij3}(\theta)v^3+p_{ij4}(\theta)v^4+p_{ij5}(\theta)v^5+p_{ij6}(\theta)v^6$$

for every θ in some neighborhood of θ_0. Thus, $\alpha_{ij}(v,\theta)$ are rational functions in v and if $q(v,\theta)\neq 0$ for all $v\geq 0$, then $|\alpha_{ij}(v,\theta)|\leq K(\theta)(1+v^2)$ for some constant $K(\theta)$. A partial fraction decomposition yields

$$\begin{aligned}\alpha^*_{ij}(v,\theta)&=\alpha^{*(0)}_{ij}(\theta)+\alpha^{*(1)}_{ij}(\theta)v+\alpha^{*(2)}_{ij}(\theta)v^2+\frac{\kappa^{(1)}_{ij}(\theta)}{v-r_1(\theta)}+\frac{\kappa^{(2)}_{ij}(\theta)}{v-r_2(\theta)}\\&\quad+\frac{\kappa^{(3)}_{ij}(\theta)}{v-r_3(\theta)}+\frac{\kappa^{(4)}_{ij}(\theta)}{v-r_4(\theta)},\end{aligned}\tag{11}$$

where $\kappa_{ij}^{(l)}(\theta),\ l=1,\dots,4$ can be calculated explicitly. The calculation of the coefficients in a partial fraction decomposition is an elementary and well-known issue that can be found in many textbooks. A particularly thorough reference is [22, Section 7.1].

3.2. Strong laws for the optimal weights

For the purpose of investigating the optimality of the estimating function, we first need to prove some results of the limiting properties on mixed moments of some special functions of processes (V) and (X).

Lemma 1. *Assume the conditions from Theorem 1. Then:*

1. *For every $p,q,r\geq 0$, we have*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}X_i^qV_i^p(1+V_{i-1})^r=E[X_1^qV_1^p(1+V_0)^r];$$

2. *For every $\theta\leq 0$ and every $p,r\geq 0$, we have*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}e^{\theta(1+V_{i-1})}X_i^pV_i^q=E\big[e^{\theta(1+V_0)}X_1^pV_1^q\big]; \tag{12}$$

3. *For every $p,q\geq 0$ and $r\in\mathbb{N}$, we have*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\frac{X_i^pV_i^q}{(1+V_{i-1})^r}=E\left[\frac{X_1^pV_1^q}{(1+V_0)^r}\right]; \tag{13}$$

4. *For every $\theta\geq 0$ and every $p,q,r\geq 0$, we have*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}X_i^pV_i^qe^{-\theta(1+V_{i-1})}V_{i-1}^r=E\big[X_1^pV_1^qe^{-\theta(1+V_0)}V_0^r\big]. \tag{14}$$

Proof. The first claim is proven directly using the binomial theorem and [25, Lemma 3.6]. For the second claim, we extend the strong law of large numbers from [25, Lemma 3.6] to exponential functions. Expanding the exponential function and using [25, Lemma 3.6], the result follows. Thirdly, for every $x,v\in\mathbb{R}$, we have

$$\int_0^\infty x^pv^q\frac{\theta^r}{r!}e^{-\theta(1+v)}d\theta=\frac{x^pv^q}{(1+v)^r}.$$

Now the result follows using (12) that has just been proven and the convergence property of the sequence. The fourth result follows by expanding the exponential function in series, using the convergency property of the series just obtained and the result from [25, Lemma 3.6]. □

Finally, we are able to state the most general limiting result for mixed moment terms of processes (V) and (X). The proofs of auxiliary results for the strong laws of weight functions are lengthy, but of the same structure as those given in [24, 25].

Theorem 4 (Strong laws for the weights)**.** *Assume the conditions from Theorem 1. For every $\theta \in \Theta$ and $p, q, r \geq 0$ and every $j = 1, \ldots, N,\ p = 1, \ldots, d$ we have*

$$\frac{1}{n} \sum_{i=1}^{n} \alpha^*_{pj}(V_{i-1}, \theta) X_i^p V_i^q V_{i-1}^r \xrightarrow{a.s.} E\big[\alpha^*_{pj}(V_0, \theta) X_1^p V_1^q V_0^r\big], \quad \text{as} \quad n \to \infty.$$

Proof. Rewriting (11) we have

$$\alpha^*_{pj}(v, \theta) = \alpha^{*(0)}_{pj}(\theta) + \alpha^{*(1)}_{pj} v + \alpha^{*(2)}_{pj} v^2 + \sum_{l=1}^{4} \frac{\kappa^{(l)}_{pj}(\theta)}{v - r_l(\theta)}. \tag{15}$$

Thus, it follows that for every $j = 1, \ldots, N,\ p = 1, \ldots, d$ and $p, q, r \geq 0$ we have

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^{n} \alpha^*_{pj}(V_{i-1}, \theta) X_i^p V_i^q V_{i-1}^r \\ &\quad = \alpha^{*(0)}_{pj}(\theta) \frac{1}{n} \sum_{i=1}^{n} X_i^p V_i^q V_{i-1}^r + \alpha^{*(1)}_{pj} \frac{1}{n} \sum_{i=1}^{n} X_i^p V_i^q V_{i-1}^{r+1} \\ &\qquad + \alpha^{*(2)}_{pj} \frac{1}{n} \sum_{i=1}^{n} X_i^p V_i^q V_{i-1}^{r+2} + \sum_{l=1}^{4} \kappa^{(l)}_{pj}(\theta) \frac{1}{n} \sum_{i=1}^{n} \frac{X_i^p V_i^q V_{i-1}^r}{V_{i-1} - r_l(\theta)}. \end{aligned} \tag{16}$$

Using Lemma 1, (13) and [25, Lemma 3.6] from relation (16) we have

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^{n} \alpha^*_{pj}(V_{i-1}, \theta) X_i^p V_i^q V_{i-1}^r &\xrightarrow{a.s.} \alpha^{*(0)}_{pj}(\theta) E\big[X_1^p V_1^q V_0^r\big] + \alpha^{*(1)}_{pj} E\big[X_1^p V_1^q V_0^{r+1}\big] \\ + \alpha^{*(2)}_{pj} E\big[X_1^p V_1^q V_0^{r+2}\big] &+ \sum_{l=1}^{4} \kappa^{(l)}_{pj}(\theta) E\left[\frac{X_1^p V_1^q V_0^r}{V_0 - r_l(\theta)}\right] = E\big[\alpha^*_{pj}(V_0, \theta) X_1^p V_1^q V_0^r\big], \end{aligned}$$

as $n \to \infty$, since the last equality follows from (15). That completes the proof. □

For a concise notation, let

$$\alpha^*_{pjk}(v, \theta) = \partial_k \alpha^*_{pj}(v, \theta), \quad \kappa^{(l)}_{pjk}(\theta) = \partial_k \kappa^{(l)}_{pj}(\theta) \tag{17}$$

for $l = 1, \ldots, 4, \quad p, k = 1, \ldots, d, \quad j = 1, \ldots, N$. We have the following result.

Theorem 5 (Strong laws for the derivatives of the weights)**.** *Assume the conditions from Theorem 1. For every $\theta \in \Theta$ and $p, q, r \geq 0$, for every $j = 1, \ldots, N$ and $p, k = 1, \ldots, d$ we have*

$$\frac{1}{n} \sum_{i=1}^{n} \alpha^*_{pjk}(V_{i-1}, \theta) X_i^p V_i^q V_{i-1}^r \xrightarrow{a.s.} E\big[\alpha^*_{pjk}(V_0, \theta) X_1^p V_1^q V_0^r\big], \quad \text{as} \quad n \to \infty.$$

Proof. Using the notation from (17), the statement follows by applying [25, Lemma 3.6] and Lemma 1, (13). □

4. Numerical illustration

To illustrate the performance of the optimal estimator obtained by solving $G_n^*(\theta) = 0$ and compare it with the performance of the simple estimator, we consider the Γ-OU model, where the variance V has a stationary gamma distribution.[‡] The background-driving Lévy process is a compound Poisson process with exponential jumps and it was simulated using the iid sequence of exponentially distributed interarrival times.[§] The variance process and the integrated variance process are then computed pathwise as deterministic functions of the jump times and jump heights of the background driving Lévy process and the initial variance, which is simply a gamma variable, according to the model equations. Finally, for the simulation of log returns, we use the fact that they are conditionally Gaussian given the simulated background-driving Lévy process and the conditional mean and variance are given in terms of the previously simulated quantities.

This simulation method is elementary and exact. It relies substantially on the fact that the background-driving Lévy process in the Γ-OU model is a compound Poisson process, thus it has finite activity, that is, finitely many jumps in any finite interval.[¶]

We use one year as the time unit, consisting of $n = 250$ trading days. The true parameters are

$$\nu = 2.56; \quad \alpha = 64; \quad \lambda = 256; \quad \beta = -0.5; \quad \rho = -0.1; \quad \mu = 1.2$$

The parameters for the variance process have the following meaning: The parameter λ determines the speed of mean-reversion in the OU-model, the parameters ζ and η are the mean and variance of the stationary variance process in general, and in the concrete Γ-OU model ν is the degree of freedom and α is the exponential parameter of the stationary gamma distribution with $\zeta = \dfrac{\nu}{\alpha}, \eta = \dfrac{\nu}{\alpha^2}$. The parameters for the log returns are as follows: The parameter μ is a linear drift component, the parameter β could be seen as a parameter for the market price of risk, as it is related to certain structure-preserving measure changes, see [35, 26], and ρ is the parameter governing the leverage effect. All parameters are estimated by the estimating function approach. More details about the statistical properties of the instantaneous variance process and log returns in this specific case can be found in [25].

[‡]Simulations were done in C using the GNU scientific library [18]. From that library, we selected the Mersenne Twister as the basic generator for uniform random numbers[33]. For gamma, exponential and Gaussian random numbers, we used the default routines, which implement the Marsaglia-Tsang fast gamma method [32] and the elementary inversion method [14, Sec. II.2.2], while for Gaussian random numbers we used the polar (Box-Mueller) method [14, Sec.V.4.4].

[§]This is called the exponential spacings method in [14, VI.1.2].

[¶]For infinite-activity processes, such as in the OU-Γ model, the IG-OU and the OU-IG model, or models based on tempered stable processes, more advanced simulation methods must be used, for example the methods based on the characteristic function, as in [13], or approximative methods, such as the truncated inverse Lévy measure method, also related to shot-noise representations, see [40, 44, 58, 53, 57].

4.1. The asymptotic covariance matrix of the simple and optimal estimator

In this numerical analysis, we do not estimate the asymptotic covariance, but evaluate the explicit expression using the true parameters. Denoting the vector of asymptotic standard deviations of the estimates and the correlation matrix by $s/\sqrt{n}$ and r, respectively, for the simple estimator, we have:

$$s = \begin{bmatrix} 4.86 \\ 125 \\ 650 \\ 7.36 \\ 253 \\ 0.526 \end{bmatrix}, \quad r = \begin{bmatrix} 1 & 0.89 & 0.41 & 0.03 & 0.09 & -0.02 \\ 0.89 & 1 & 0.4 & 0.03 & 0.09 & -0.03 \\ 0.41 & 0.4 & 1 & 0.06 & 0.22 & 0 \\ 0.03 & 0.03 & 0.06 & 1 & -0.75 & 0.06 \\ 0.09 & 0.09 & 0.22 & -0.75 & 1 & -0.57 \\ -0.02 & -0.03 & 0 & 0.06 & -0.57 & 1 \end{bmatrix}.$$

Even though the optimal α's can be calculated explicitly and thus the optimal estimating function $G_n^*(\theta)$ is explicit, the estimator $\bar{\theta}$ has to be solved numerically. Nevertheless, Conjecture 1 can be checked for the true parameter, coefficients $\kappa_{ij}^{(k)}$ and the asymptotic variance-covariance matrix $J(\bar{\theta})^{-1}\Phi(J(\bar{\theta})^{-1})^\top$ can be calculated explicitly. First we calculate the coefficients $\alpha_{ij}^{*(0)}(\bar{\theta})$, $\alpha_{ij}^{*(1)}(\bar{\theta})$, $\alpha_{ij}^{*(2)}(\bar{\theta})$ and $\kappa_{ij}^{(1)}(\bar{\theta})$, $\kappa_{ij}^{(2)}(\bar{\theta})$, $\kappa_{ij}^{(3)}(\bar{\theta})$ and $\kappa_{ij}^{(4)}(\bar{\theta})$. Furthermore, explicit expressions for the optimal estimating function are given. Since the variance-covariance matrix Φ from Proposition 1 can be given explicity, from Theorem 3 it follows that the variance-covariance matrix of the optimal estimator is T^*/n, where

$$T^* = J(\bar{\theta})^{-1}\Phi\big(J(\bar{\theta})^{-1}\big)^T.$$

Denoting the vector of asymptotic standard deviations of the estimates and the correlation matrix by $s^*/\sqrt{n}$ and r^*, respectively, for the optimal estimator, we have:

$$s^* = \begin{bmatrix} 4.686 \\ 120.23 \\ 491.038 \\ 6.179 \\ 219.193 \\ 0.504 \end{bmatrix}, \quad r^* = \begin{bmatrix} 1. & 0.88 & 0.32 & 0.03 & 0.06 & -0.02 \\ 0.88 & 1. & 0.31 & 0.03 & 0.06 & -0.03 \\ 0.32 & 0.31 & 1. & 0.04 & 0.2 & -0.01 \\ 0.03 & 0.03 & 0.04 & 1. & -0.69 & -0.02 \\ 0.06 & 0.06 & 0.2 & -0.69 & 1. & -0.58 \\ -0.02 & -0.03 & -0.01 & -0.02 & -0.58 & 1. \end{bmatrix}.$$

4.2. A numerical illustration of the actual performance of the optimal estimator

In this subsection, we report the actual estimation in the setting described in the previous subsections. Namely, 10 000 calculations with the optimal estimating functions were performed. In Figure 1, histograms together with the theoretically calculated Gaussian limiting distributions are displayed. We may notice an improvement of the optimal estimator versus the simple one, even though this improvement is not significant. More specifically, the biggest improvement is achieved for the λ parameter denoting the speed of the mean-reversion in the corresponding OU-model, and the μ parameter describing the linear drift component of the returns.

The multivariate root-finding was done in C using the the `hybrids` solve routine from the GSL library [18], which is, according to the documentation, based on a modified version of Powell's hybrid method, see [39], which replaces calls to the Jacobian function by its finite difference approximation. As the optimal quadratic estimating function is explicit and elementary, the Jacobian can be implemented by a very patient coder or with the help of an automatic symbolic differentiation system, but this was not required for the illustration here [41, 31]. The histograms are from the empirical distribution of this illustration, the Gaussian densities are plotted using the theoretical results of the asymptotic limit distributions given the known true parameter values.

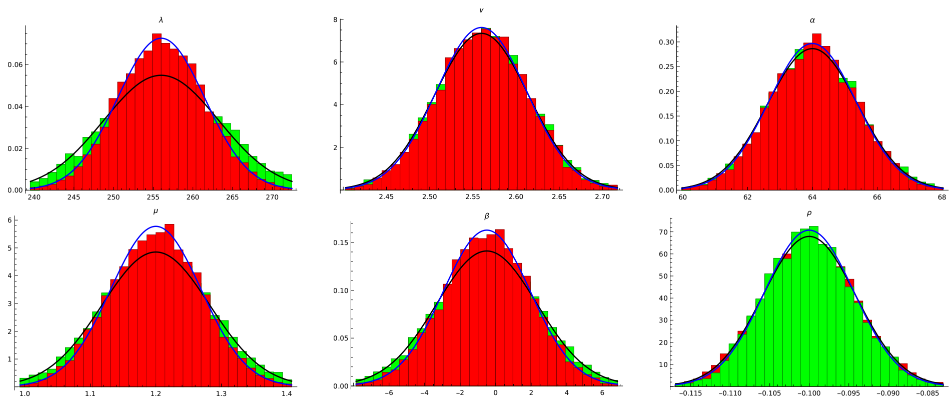

Figure 1: *Comparison of simple and optimal estimates*

5. Conclusions and further research

In this paper, we provide and analyze the optimal quadratic estimator from a fixed and asymptotic sample point of view for a class of discretely observed continuous-time stochastic volatility models with jumps. In particular, we consider a class of non-Gaussian Ornstein-Uhlenbeck-based models. First, we developed in detail a martingale estimating function approach for this kind of processes, which are bivariate Markov processes that are not diffusions, but admit jumps. Second, we proved rigorously consistency and asymptotic normality of the optimal estimator based on a single assumption that all moments of the stationary distribution of the variance process are finite, and showed that the optimal martingale estimating function and the asymptotic covariance matrix of the estimator can be computed in closed form. Third, in a numerical illustration we found a non-negligible improvement using the optimal instead of the simple estimator in terms of the asymptotic variance-covariance matrix.

As a natural continuation of this work, we should study numerical aspects of solving the optimal estimating equations, and illustrate and analyze the small sample performance of the optimal estimator in comparison to the simple explicit estimator, which requires no numerical root-finding.

Finally, we address the issue that volatility, or instantaneous variance, is not directly observable in discrete time. A potential practical solution to the problem is to use our estimation framework with the unobservable instantaneous variance replaced by a suitable observable substitute, see the suggestions in [25, Section 5] and the references given therein. Another possibility would be to extend the present analysis for the optimal estimating functions to the joint analysis of stock returns and some measure of trading intensity, such as trading volume, for example, and compare its performance to the simple estimator developed in [24].

Acknowledgements

The authors would like to thank the referees for the helpful suggestions.

References

[1] O. E. Barndorff-Nielsen, J. L. Jensen, M. Sørensen, *Some stationary processes in discrete and continuous time*, Adv. in Appl. Probab. **30**(1998), 989–1007.

[2] O. E. Barndorff-Nielsen, A. E. D. Veraart, *Stochastic volatility of volatility and variance risk premia*, J. Financ. Econom. **11**(2013), 1–46.

[3] O. E. Barndorff-Nielsen, F. E. Benth, A. E. D. Veraart, *Ambit stochastics*, Springer, Cham, 2018.

[4] O. E. Barndorff-Nielsen, E. Nicolato, N. Shephard, *Some recent developments in stochastic volatility modelling*, Quant. Finance **2**(2002), 11–23.

[5] O. E. Barndorff-Nielsen, N. Shephard, *Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics*, J. R. Stat. Soc. Ser. B Stat. Methodol. **63**(2001), 167–241.

[6] O. E. Barndorff-Nielsen, R. Stelzer, *The multivariate supOU stochastic volatility model*, Math. Finance **23**(2013), 275–296.

[7] F. E. Benth, M. Groth, *The minimal entropy martingale measure and numerical option pricing for the Barndorff-Nielsen-Shephard stochastic volatility model*, Stoch. Anal. Appl. **27**(2009), 875–896.

[8] F. E. Benth, K. H. Karlsen, K. Reikvam, *Merton's portfolio optimization problem in a Black and Scholes market with non-Gaussian stochastic volatility of Ornstein-Uhlenbeck type*, Math. Finance **13**(2003), 215–244.

[9] P. J. Brockwell, R. A. Davis, Y. Yang, *Estimation for nonnegative Lévy-driven Ornstein-Uhlenbeck processes*, J. Appl. Probab. **44**(2007), 977–989.

[10] P. J. Brockwell, A. Lindner, *Ornstein-Uhlenbeck related models driven by Lévy processes*, in: *Statistical Methods for Stochastic Differential Equations*, (M. Kessler, A. Lindner and M. Sørensen, Eds.), CRC Press, Boca Raton, 2012, 383–427.

[11] M. Chávez-Pichardo, M. A. Martínez-Cruz, A. Trejo-Martínez, D. Martínez-Carbajal, T. Arenas-Resendiz, *A complete review of the general quartic equation with real coefficients and multiple roots*, Mathematics **10**(2022), 2377.

[12] R. Cont, P. Tankov, *Financial modelling with jump processes*. Chapman & Hall/CRC, Boca Raton, 2004.

[13] L. Devroye, *On the computer generation of random variables with a given characteristic function*, Comput. Math. Appl. **7**(1981), 547–552.

[14] L. DEVROYE, *Nonuniform random variate generation*, Springer-Verlag, New York, 1986.

[15] S. FRÜHWIRTH-SCHNATTER, L. SÖGNER, *Bayesian estimation of stochastic volatility models based on OU processes with marginal Gamma law*, Ann. Inst. Statist. Math. **61**(2009), 159–179.

[16] M. P. S. GANDER, D. A. STEPHENS, *Simulation and inference for stochastic volatility models driven by Lévy processes*, Biometrika **94**(2007), 627–646.

[17] M. P. S. GANDER, D. A. STEPHENS, *Stochastic volatility modelling in continuous time with general marginal distributions: inference, prediction and model selection*, J. Statist. Plann. Inference **137**(2007), 3068–3081.

[18] B. GOUGH, *GNU scientific library reference manual*, Network Theory Ltd., Godalming, 2009.

[19] D. GRAHOVAC, N. N. LEONENKO, M. S. TAQQU, *Limit theorems, scaling of moments and intermittency for integrated finite variance supOU processes*, Stochastic Process. Appl. **129**(2019), 5113–5150.

[20] D. GRAHOVAC, N. N. LEONENKO, M. S. TAQQU, *The multifaceted behavior of integrated supOU processes: the infinite variance case*, J. Theoret. Probab. **33**(2020), 1801–1831.

[21] J. E. GRIFFIN, M. F. J. STEEL, *Inference with non-Gaussian Ornstein-Uhlenbeck processes for stochastic volatility*, J. Econometrics, **134**(2006), 605–644.

[22] P. HENRICI, *Applied and computational complex analysis. Vol. 1.* John Wiley & Sons, New York, 1988.

[23] C. C. HEYDE, *Quasi-likelihood and its application*, Springer-Verlag, New York, 1997.

[24] F. HUBALEK, P. POSEDEL, *Joint analysis and estimation of stock prices and trading volume in Barndorff-Nielsen and Shephard stochastic volatility models*, Quant. Finance **11**(2011), 917–932.

[25] F. HUBALEK, P. POSEDEL, *Asymptotic analysis and explicit estimation of a class of stochastic volatility models with jumps using the martingale estimating function approach*, Glas. Mat. Ser. III **48**(2013), 185–210.

[26] F. HUBALEK, C. SGARRA, *On the Esscher transforms and other equivalent martingale measures for Barndorff-Nielsen and Shephard stochastic volatility models with jumps*, Stochastic Process. Appl. **119**(2009), 2137–2157.

[27] J. JACOD, M. SØRENSEN, *A review of asymptotic theory of estimating functions*, Stat. Inference Stoch. Process. **21**(2018), 415–434.

[28] Z. J. JUREK, W. VERVAAT, *An integral representation for self-decomposable Banach space valued random variables*, Z. Wahrsch. Verw. Gebiete **62**(1983), 247–262.

[29] J. KALLSEN, J. MUHLE-KARBE, M. VOSS, *Pricing options on variance in affine stochastic volatility models*, Math. Finance **21**(2011), 627–641.

[30] C. LINDBERG, *News-generated dependence and optimal portfolios for n stocks in a market of Barndorff-Nielsen and Shephard type*, Math. Finance **16**(2006), 549–568.

[31] C. C. MARGOSSIAN, *A review of automatic differentiation and its efficient implementation*, CoRR **abs/1811.05031**(2018). Available at `http://arxiv.org/abs/1811.05031`.

[32] G. MARSAGLIA, W. W. TSANG, *A simple method for generating gamma variables*, ACM Trans. Math. Softw. **26**(2000), 363–372.

[33] M. MATSUMOTO, T. NISHIMURA, *Mersenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator*, ACM Trans. Model. Comput. Simul. **8**(1998), 3–30.

[34] J. MUHLE-KARBE, O. PFAFFEL, R. STELZER, *Option pricing in multivariate stochastic volatility models of OU type*, SIAM J. Financial Math. **3**(2012), 66–94.

[35] E. Nicolato, E. Venardos, *Option pricing in stochastic volatility models of the Ornstein-Uhlenbeck type*, Math. Finance **13**(2003), 445–466.

[36] F. W. J. Olver, D. W. Lozier, R. F. Boisvert, C. W. Clark, Eds., *NIST handbook of mathematical functions*, Cambridge University Press, Cambridge, 2010.

[37] Y. Peng, M. C. Fu, J. Q. Hu, *Gradient-based simulated maximum likelihood estimation for Lévy-driven Ornstein-Uhlenbeck stochastic volatility models*, Quant. Finance **14**(2014), 1399–1414.

[38] M. Podolskij, R. Stelzer, S. Thorbjørnsen, A. E. D. Veraart, Eds., *The fascination of probability, statistics and their applications*, Springer, Cham, 2016.

[39] M. J. D. Powell, *A hybrid method for nonlinear equations*, in: *Numerical Methods for Nonlinear Algebraic Equations*, (P. Rabinowitz, Ed.), Gordon and Breach, London-New York-Paris, 1970, 87–114.

[40] Y. Qu, A. Dassios, H. Zhao, *Exact simulation of Ornstein-Uhlenbeck tempered stable processes*, J. Appl. Probab. **58**(2021), 347–371.

[41] L. B. Rall, G. F. Corliss, *An introduction to automatic differentiation*, in: *Computational differentiation*, (M. Berz, C. Bischof, G. Corliss and A. Griewank, Eds.), SIAM, Philadelphia, 1996, 1–18.

[42] T. Rheinländer, G. Steiger, *The minimal entropy martingale measure for general Barndorff-Nielsen/Shephard models*, Ann. Appl. Probab. **16**(2006), 1319–1351.

[43] G. O. Roberts, O. Papaspiliopoulos, P. Dellaportas, *Bayesian inference for non-Gaussian Ornstein-Uhlenbeck stochastic volatility processes*, J. R. Stat. Soc. Ser. B Stat. Methodol. **66**(2004), 369–393.

[44] J. Rosiński, *Series representations of Lévy processes from the perspective of point processes*, in: *Lévy processes*, (O. E. Barndorff-Nielsen, T. Mikosch and S. I. Resnick, Eds.), Birkhäuser, Boston, 2001, 401–415.

[45] K. -I. Sato, *Lévy processes and infinitely divisible distributions*, Cambridge University Press, Cambridge, 2013.

[46] W. Schoutens, *Lévy processes in finance*, John Wiley & Sons, Inc., New York-London-Sydney, 2003.

[47] N. Shephard, Ed., *Stochastic volatility*, Oxford University Press, Oxford, 2005.

[48] N. Shephard, *Ole Eiler Barndorff-Nielsen and financial econometrics*, working paper, Harvard University, 2024.

[49] M. Sørensen, *On asymptotics of estimating functions*, Braz. J. Probab. Stat. **13**(1999), 111–136.

[50] M. Sørensen, *Statistical inference for discretely observed diffusion*, lectures at Berliner Graduiertenkolleg "Stochastische Prozesse und Probabilistische Analysis", 1997.

[51] I. Stewart, *Galois theory*. CRC Press, Boca Raton, 2015.

[52] E. ter Horst, *A Levy generalization of compound Poisson processes in finance: Theory and applications*, Thesis (Ph.D.), Duke University, Durham, 2003.

[53] V. Todorov, G. Tauchen, *Simulation methods for Lévy-driven continuous-time autoregressive moving average (CARMA) stochastic volatility models*, J. Bus. Econom. Statist. **24**(2006), 455–469.

[54] W. Tucker, *Validated numerics*, Princeton University Press, Princeton, 2011.

[55] L. Valdivieso, W. Schoutens, F. Tuerlinckx, *Maximum likelihood estimation in processes of Ornstein-Uhlenbeck type*, Stat. Inference Stoch. Process. **12**(2009), 1–19.

[56] S. J. Wolfe, *On a continuous analogue of the stochastic difference equation* $X_n = \rho X_{n-1} + B_n$, Stochastic Process. Appl. **12**(1982), 301–312.

[57] S. Zhang, X. Zhang, *Exact simulation of IG-OU processes*, Methodol. Comput. Appl. Probab. **10**(2008), 337–355.

[58] S. ZHANG, *Exact simulation of tempered stable Ornstein-Uhlenbeck processes*, J. Stat. Comput. Simul. **81**(2011), 1533–1544,.

Appendix A. Proof of Theorem 1

Proof. For $k = 1, \dots, d$ and $j = 1, \dots, N$, define two $d \times N$ matrices by

$$A(v_{i-1};\theta) = \{\alpha_{k,j}(v_{i-1};\theta)\}, \quad \tilde{A}(v_{i-1};\theta) = \{\tilde{\alpha}_{k,j}(v_{i-1};\theta)\}, \qquad i \le n,$$

and two $d-$dimensional estimating functions

$$G_n(\theta) = \sum_{i=1}^{n} A(V_{i-1};\theta)h(V_i, X_i|V_{i-1};\theta), \quad \tilde{G}_n(\theta) = \sum_{i=1}^{n} \tilde{A}(V_{i-1};\theta)h(V_i, X_i|V_{i-1};\theta),$$

where $h = (h_1, \dots, h_N)^T$. Denote the predictable quadratic covariation $(d\times d)$ matrix by $\langle G(\theta)\rangle_n = \{\langle G^i(\theta), G^j(\theta)\rangle_n,\ i,j = 1,\dots,d\}$. Then, using definition (6) we obtain

$$\begin{aligned}\langle G(\theta), \tilde{G}(\theta)\rangle_n &= \sum_{i=1}^{n} E_\theta\big[A(V_{i-1};\theta)h(V_i, X_i|V_{i-1};\theta)\big(\tilde{A}(V_{i-1};\theta)h(V_i, X_i|V_{i-1};\theta)\big)^T|V_{i-1}\big] \\ &= \sum_{i=1}^{n} A(V_{i-1};\theta)E_\theta\big[h(V_i, X_i|V_{i-1};\theta)h^T(V_i, X_i|V_{i-1};\theta)|V_{i-1}\big]\tilde{A}(V_{i-1};\theta)^T \\ &= \sum_{i=1}^{n} A(V_{i-1};\theta)C(V_{i-1};\theta)\tilde{A}(V_{i-1};\theta)^T. \end{aligned} \tag{18}$$

Since the matrix $C(v_0;\theta)$ is assumed to be regular, under Conjecture 1 from (5) and (6) we have

$$A^*(v_{i-1};\theta) = B(v_{i-1};\theta)C(v_{i-1};\theta)^{-1}, \tag{19}$$

where $A^*(v_{i-1};\theta) = \{\alpha_{k,j}(v_{i-1};\theta), k = 1,\dots,d,\ j = 1,\dots N\}$. In particular, from (18) and (19) it follows that

$$\langle G^*(\theta)\rangle_n = \sum_{i=1}^{n} B(V_{i-1};\theta)C(V_{i-1};\theta)^{-1}B(V_{i-1};\theta)^T$$

and

$$\begin{aligned}\langle G(\theta), G^*(\theta)\rangle_n &= \sum_{i=1}^{n} A(V_{i-1};\theta)C(V_{i-1};\theta)A^*(V_{i-1};\theta)^T \\ &= \sum_{i=1}^{n} A(V_{i-1};\theta)C(V_{i-1};\theta)C(V_{i-1};\theta)^{-1}B(V_{i-1};\theta)^T \\ &= \sum_{i=1}^{n} A(V_{i-1};\theta)B(V_{i-1};\theta)^T. \end{aligned} \tag{20}$$

Let $\bar{G}_n(\theta)$ denote the compensator of $\partial_{\theta^T}G_n(\theta)$ under P_θ, i.e. $\bar{G}_n(\theta)$ is $\mathcal{F}_{n-1}$ measurable for every $n \in \mathbb{N}$ and each component of the matrix $\partial_{\theta^T}G_n(\theta) - \bar{G}_n(\theta)$ is a P_θ martingale w.r.t. $\{\mathcal{F}_n\}$. In order to find $\bar{G}_n(\theta)$ we differentiate $G_n(\theta)$ w.r.t. θ_k, where $k = 1, \dots, d$.

$$\partial_{\theta_k} G_n(\theta) = \sum_{i=1}^{n} \left(\frac{\partial}{\partial_{\theta_k}} A(V_{i-1};\theta) \right) h(V_i, X_i | V_{i-1};\theta) + \sum_{i=1}^{n} A(V_{i-1};\theta) \frac{\partial}{\partial_{\theta_k}} h(V_i, X_i | V_{i-1};\theta). \quad (21)$$

But, for every $j = 1, \dots, N$, it holds that $E_\theta \left[\sum_{i=1}^{n} h_j(V_i, X_i | V_{i-1};\theta) \right] = 0$, so calculating the conditional expectation of both sides of equation (21), it follows that

$$E_\theta \left[\partial_{\theta_k} G_n(\theta) | V_{i-1} \right] = \sum_{i=1}^{n} A(V_{i-1};\theta) E_\theta \left[\frac{\partial}{\partial_{\theta_k}} h(V_i, X_i | V_{i-1};\theta) \middle| V_{i-1} \right].$$

Hence, using definition (5) and relation (20) we derive

$$\begin{aligned} \bar{G}_n(\theta) &= \sum_{i=1}^{n} A(V_{i-1};\theta) E_\theta \left[\frac{\partial}{\partial_{\theta_k}} h(V_i, X_i | V_{i-1};\theta) | V_{i-1} \right] \\ &= - \sum_{i=1}^{n} A(V_{i-1};\theta) B(V_{i-1};\theta) = \left\langle G(\theta), G^*(\theta) \right\rangle_n. \end{aligned}$$

Especially, $\bar{G^*}_n(\theta) = -\left\langle G^*(\theta) \right\rangle_n$. Finally, we have that

$$\bar{G}_n(\theta)^{-1} \left\langle G(\theta), G^*(\theta) \right\rangle_n = -I_d = \bar{G^*}_n(\theta)^{-1} \left\langle G^*(\theta) \right\rangle_n, \quad (22)$$

where I_d is the identity matrix of dimension d and this quantity is obviously non-random. Since by (22) we have $\bar{G}_n(\theta)^{-1} \left\langle G(\theta), G^*(\theta) \right\rangle_n = \bar{G^*}_n(\theta)^{-1} \left\langle G^*(\theta) \right\rangle_n$, applying [23, Theorem 2.4], it follows that $G_n^*(\theta) \in \mathcal{M}$ is an O_A-optimal estimating function within $\mathcal{M}$. Now, we have the O_A-optimality of $G_n^*(\theta)$, and from (22) it follows that the quantity $\bar{G^*}_n(\theta)^{-1} \left\langle G^*(\theta) \right\rangle_n$ is non-random. Finally, by [23, Theorem 2.5], it thus follows that $G_n^*(\theta)$ is O_F-optimal within $\mathcal{M}$. This completes the proof. □

Appendix B. Fulfillment of [49, Condition 2.7]

Proposition 2. *Assume the conditions from Theorem 1. Condition* 2.7 *of [49] is satisfied, namely,*

(*i*) *the mapping $\theta \longmapsto G_n(\theta)$ is twice continuously differentiable.*

(*ii*) *There exist a $\bar{\theta} \in$ int Θ and an invertible non-random $d \times d$ matrix $J(\bar{\theta})$ such that*

$$\sup_{\theta^{(i)} \in M_n^\alpha(\bar{\theta})} \left\| \frac{1}{n} J_n(\theta^{(1)}, \dots, \theta^{(d)}) - J(\bar{\theta}) \right\| \to 0$$

in probability as $n \to \infty$ for all $\alpha > 0$.

(*iii*) *There exist* d *non-random* $d \times d$ *matrices* $B^i(\bar\theta)$, $i = 1, \dots, d$, *such that*

$$\sup_{\theta^{(i)} \in M_n^\alpha(\bar\theta)} \Big\| \frac{1}{n} Q_n^{(i)}(\theta^{(1)}, \dots, \theta^{(d)}) - B^i(\bar\theta) \Big\| \to 0$$

in probability as $n \to \infty$ *for all* $\alpha > 0$ *and all* $i = 1, \dots, d$, *where* $Q_n^{(i)}(\theta) = \partial_\theta^2 G_n^i(\theta)$.

(*iv*) $\left\{ \dfrac{G_n(\bar\theta)}{n} : n \in \mathbb{N} \right\}$ *is stochastically bounded.*

(*v*) $\sup_{\theta \in M_n^\alpha(\bar\theta)} \left\| \dfrac{G_n(\theta)}{n} \right\| \to 0$ *in probability as* $n \to \infty$ *for all* $\alpha > 0$.

Proof. In our case, the number of parameters is $d = 6$ and $N = 5$. From the definitions above it immediately follows that the mapping $\theta \mapsto G_n^p(\theta)$ is twice continuously differentiable with respect to θ. Let us consider the matrix $J_n = \big(J_n^{p,k}\big)_{p,k=1,\dots,d}$ componentwise. For $j = 1, \dots, N$, let

$$\alpha^*_{pjk}(v;\theta) = \frac{\partial \alpha^*_{pj}(v;\theta)}{\partial \theta_k}, \quad f_j^k(v;\theta) = \frac{\partial f_j(v;\theta)}{\partial \theta_k},$$

$$\phi_{jl}^k(\theta) = \frac{\partial \phi_{jl}(\theta)}{\partial \theta_k}, \qquad \phi_{jl}^{k,z}(\theta) = \frac{\partial \phi_{jl}^k(\theta)}{\partial \theta_z} \qquad z = 1, \dots, d.$$

From (4) it follows that

$$f_j(v;\theta) = \sum_{l=0}^{p_j+q_j} \phi_{jl}(\theta) v^l, \tag{23}$$

where ϕ_{jl} are explicit in terms of the parameter θ. From (23) it follows that

$$f_j^k(v;\theta) = \sum_{l=0}^{p_j+q_j} \phi_{jl}^k(\theta) v^l.$$

Thus we have

$$\begin{aligned} J_n^{p,k}(\theta) &= \sum_{i=1}^n \sum_{j=1}^N \alpha^*_{pjk}(V_{i-1},\theta)\big[X_i^{p_j} V_i^{q_j} - f_j(V_{i-1},\theta)\big] - \sum_{i=1}^n \sum_{j=1}^N \alpha^*_{pj}(V_{i-1},\theta) f_j^k(V_{i-1},\theta) \\ &= \sum_{i=1}^n \sum_{j=1}^N \alpha^*_{pjk}(V_{i-1},\theta)\big[X_i^{p_j} V_i^{q_j} - f_j(V_{i-1},\theta)\big] - \sum_{i=1}^n \sum_{j=1}^N \alpha^*_{pj}(V_{i-1},\theta) \sum_{l=0}^{p_j+q_j} \phi_{jl}^k(\theta) V_{i-1}^l. \end{aligned} \tag{24}$$

Let us define

$$J^{p,k}(\bar\theta) = -\sum_{j=1}^N \sum_{l=0}^{p_j+q_j} \phi_{jl}^k(\bar\theta) E\big[\alpha^*_{pj}(V_0,\bar\theta) V_0^l\big].$$

Using the expressions for $f_j(v,\theta)$ and (24), $m=1,\dots,d$, we have

$$\begin{aligned}
&\sup_{\theta^{(m)}\in M_n(\bar\theta)}\left|\frac{1}{n}J_n^{p,k}(\theta^{(m)})-J^{p,k}(\bar\theta)\right|\\
&\quad=\sup_{\theta^{(m)}\in M_n(\bar\theta)}\left|\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{N}\alpha^*_{pjk}(V_{i-1},\theta^{(m)})\big[X_i^{p_j}V_i^{q_j}-f_j(V_{i-1};\theta^{(p)})\big]\right.\\
&\qquad\left.-\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{N}\alpha^*_{pj}(V_{i-1},\theta^{(m)})\sum_{l=0}^{p_j+q_j}\phi^k_{jl}(\theta^{(m)})V_{i-1}^l+\sum_{j=1}^{N}\sum_{l=0}^{p_j+q_j}\phi^k_{jl}(\bar\theta)E\big[\alpha^*_{pj}(V_0,\bar\theta)V_0^l\big]\right|\\
&\quad\le\sup_{\theta^{(m)}\in M_n(\bar\theta)}\left|\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{N}\alpha^*_{pjk}(V_{i-1},\theta^{(m)})\big[X_i^{p_j}V_i^{q_j}-f_j(V_{i-1},\theta^{(m)})\big]\right|\\
&\qquad+\sup_{\theta^{(m)}\in M_n(\bar\theta)}\left|-\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{N}\alpha^*_{pj}(V_{i-1},\theta^{(m)})\sum_{l=0}^{p_j+q_j}\phi^k_{jl}(\theta^{(m)})V_{i-1}^l\right.\\
&\qquad\left.+\sum_{j=1}^{N}\sum_{l=0}^{p_j+q_j}\phi^k_{jl}(\bar\theta)E\big[\alpha^*_{pj}(V_0,\bar\theta)V_0^l\big]\right|. \qquad (25)
\end{aligned}$$

Adding and subtracting expressions

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{N}\alpha^*_{pj}(V_{i-1},\bar\theta)\sum_{l=0}^{p_j+q_j}\phi^k_{jl}(\theta^{(m)})V_{i-1}^l$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{N}\alpha^*_{pj}(V_{i-1},\bar\theta)\sum_{l=0}^{p_j+q_j}\phi^k_{jl}(\bar\theta)V_{i-1}^l$$

in (25) it follows that

$$\begin{aligned}
&\sup_{\theta^{(m)}\in M_n(\bar\theta)}\left|\frac{1}{n}J_n^{p,k}(\theta^{(p)})-J^{p,k}(\bar\theta)\right|\\
&\quad\le\sup_{\theta^{(m)}\in M_n(\bar\theta)}\left|\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{N}\alpha^*_{pjk}(V_{i-1},\theta^{(m)})\big[X_i^{p_j}V_i^{q_j}-f_j(V_{i-1};\theta^{(p)})\big]\right|\\
&\qquad+\sum_{j=1}^{N}\sup_{\theta^{(m)}\in M_n(\bar\theta)}\left|\frac{1}{n}\sum_{i=1}^{n}\big[\alpha^*_{pj}(V_{i-1},\bar\theta)-\alpha^*_{pj}(V_{i-1},\theta^{(m)})\big]\sum_{l=0}^{p_j+q_j}\phi^k_{jl}(\theta^{(m)})V_{i-1}^l\right|\\
&\qquad+\sum_{j=1}^{N}\sup_{\theta^{(m)}\in M_n(\bar\theta)}\left|\sum_{l=0}^{p_j+q_j}\big[\phi^k_{jl}(\bar\theta)-\phi^k_{jl}(\theta^{(m)})\big]\frac{1}{n}\sum_{i=1}^{n}\alpha^*_{pj}(V_{i-1},\bar\theta)V_{i-1}^l\right|\\
&\qquad+\sum_{j=1}^{N}\sum_{l=0}^{p_j+q_j}\left|\phi^k_{jl}(\bar\theta)\left[E[\alpha^*_{pj}(V_0,\bar\theta)V_0^l]-\frac{1}{n}\sum_{i=1}^{n}\alpha^*_{pj}(V_{i-1},\bar\theta)V_{i-1}^l\right]\right|. \qquad (26)
\end{aligned}$$

Since a function belonging to $\mathcal{C}^\infty$ is bounded on a compact set, we have

$$\sup_{\theta^*\in M_n^\alpha(\bar\theta)}\left|\frac{\partial\phi_l^{j,k}}{\partial\theta_m^{(p)}}(\theta^*)\right|\le M,$$

for some finite constant M. Using the definition of the set M_n we thus have

$$\begin{aligned}\sup_{\theta^{(m)}\in M_n(\bar\theta)}&\left|\sum_{l=0}^{p_j+q_j}\left[\phi_{jl}^k(\bar\theta)-\phi_{jl}^k(\theta^{(m)})\right]\right|\\ &\le\sum_{l=0}^{p_j+q_j}\sup_{\theta^{(p)}\in M_n(\bar\theta)}\sum_{d=1}^{p}\sup_{\theta^*\in M_n^\alpha(\bar\theta)}\left|\frac{\partial\phi_{jl}^k}{\partial\theta_d^{(m)}}(\theta^*)\right|\left|\bar\theta_m-\theta_d^{(m)}\right|\\ &\le\sum_{l=0}^{p_j+q_j}\sum_{d=1}^{p}M\frac{\alpha}{\sqrt n}=:\frac{C_{1,j}}{\sqrt n},\end{aligned}$$

where $C_{1,j}=M\alpha(p_j+q_j)p$. For ease of notation, let us define

$$K_n^{p,k}(\theta^{(m)})=\frac1n\sum_{i=1}^n\alpha_{pjk}^*(V_{i-1};\theta^{(m)})\big[X_i^{p_j}V_i^{q_j}-\phi^j(V_{i-1};\theta^{(m)})\big].$$

Since $\alpha_{pj}^*(v,\theta)$ defined by (11) is Lipschitz if Conjecture 1 holds, from definition (1) and relation (26) it follows that

$$\begin{aligned}\sup_{\theta^{(m)}\in M_n(\bar\theta)}&\left|\frac1n J_n^{p,k}(\theta^{(p)})-J^{p,k}(\bar\theta)\right|\\ \le&\sum_{j=1}^N\sup_{\theta^{(m)}\in M_n(\bar\theta)}\left|K_n^{p,k}(\theta^{(m)})\right|+\frac{\alpha}{\sqrt n}\sum_{j=1}^N\sup_{\theta^{(m)}\in M_n(\bar\theta)}\left|\sum_{l=0}^{p_j+q_j}\phi_{jl}^k(\theta^{(m)})\right|\left|\frac1n\sum_{i=1}^nV_{i-1}^l\right|\\ &+\sum_{j=1}^N\frac{C_{1,j}}{\sqrt n}\left|\frac1n\sum_{i=1}^n\alpha_{pj}^*(V_{i-1};\bar\theta)V_{i-1}^l\right|\\ &+\sum_{j=1}^N\sum_{l=0}^{p_j+q_j}\left|\phi_{jl}^k(\bar\theta)\left[\frac1n\sum_{i=1}^n\alpha_{pj}^*(V_{i-1};\bar\theta)V_{i-1}^l-E[\alpha_{pj}^*(V_0;\bar\theta)V_0^l]\right]\right|.\end{aligned}\tag{27}$$

Adding and subtracting expressions

$$\frac1n\sum_{i=1}^n\alpha_{pjk}^*(V_{i-1};\bar\theta)\big[X_i^{p_j}V_i^{q_j}-f_j(V_{i-1};\theta^{(m)})\big]$$

and

$$\frac1n\sum_{i=1}^n\big[\alpha_{pjk}^*(V_{i-1};\theta^{(p)})-\alpha_{pjk}^*(V_{i-1};\bar\theta)\big]f_j(V_{i-1};\bar\theta)$$

in $K_n^{p,k}(\theta^{(m)})$ in relation (27) and since functions $\alpha^*_{pjk}(v,\theta)$ are Lipschitz if Conjecture 1 holds, for every $j=1,\ldots,N$, we obtain

$$
\begin{aligned}
\sup_{\theta^{(m)}\in M_n(\bar\theta)} & \left|K_n^{p,k}(\theta^{(m)})\right| \\
\le & \sup_{\theta^{(m)}\in M_n(\bar\theta)}\left|\frac{1}{n}\sum_{i=1}^n\big[\alpha^*_{pjk}(V_{i-1};\theta^{(m)})-\alpha^*_{pjk}(V_{i-1};\bar\theta)\big]\big[X_i^{p_j}V_i^{q_j}-f_j(V_{i-1};\bar\theta)\big]\right| \\
& + \sup_{\theta^{(m)}\in M_n(\bar\theta)}\left|\frac{1}{n}\sum_{i=1}^n\big[\alpha^*_{pjk}(V_{i-1};\theta^{(m)})-\alpha^*_{pjk}(V_{i-1};\bar\theta)\big]\big[f_j(V_{i-1};\bar\theta)-f_j(V_{i-1};\theta^{(m)})\big]\right| \\
& + \sup_{\theta^{(m)}\in M_n(\bar\theta)}\left|\frac{1}{n}\sum_{i=1}^n\alpha^*_{pjk}(V_{i-1};\bar\theta)\big[f_j(V_{i-1};\bar\theta)-f_j(V_{i-1};\theta^{(m)})\big]\right| \\
& + \left|\frac{1}{n}\sum_{i=1}^n\alpha^*_{pjk}(V_{i-1};\bar\theta)\big[X_i^{p_j}V_i^{q_j}-f_j(\bar\theta)\big]\right| \\
\le & \frac{\alpha}{\sqrt{n}}\left\{\left|\frac{1}{n}\sum_{i=1}^n\big[X_i^{p_j}V_i^{q_j}-f_j(V_{i-1};\bar\theta)\big]\right|\right. \\
& \left. + \sup_{\theta^{(m)}\in M_n(\bar\theta)}\frac{1}{n}\sum_{i=1}^n\big[f_j(V_{i-1};\bar\theta)-f_j(V_{i-1};\theta^{(m)})\big]\right|\right\} \\
& + \sup_{\theta^{(m)}\in M_n(\bar\theta)}\frac{1}{n}\sum_{i=1}^n\alpha^*_{pjk}(V_{i-1};\bar\theta)\big[f_j(V_{i-1};\bar\theta)-f_j(V_{i-1};\theta^{(m)})\big] \\
& + \left|\frac{1}{n}\sum_{i=1}^n\alpha^*_{pjk}(V_{i-1};\bar\theta)\big[X_i^{p_j}V_i^{q_j}-f_j(\bar\theta)\big]\right|.
\end{aligned}
\tag{28}
$$

Furthermore, we have

$$
\begin{aligned}
\left|f_j(V_{i-1};\bar\theta)-f_j(V_{i-1};\theta^{(m)})\right| &\le \sum_{l=0}^{p_j+q_j}\left|\phi_{jl}(\bar\theta)-\phi_{jl}(\theta^{(m)})\right|V_{i-i}^l \\
&\le \sum_{l=0}^{p_j+q_j}\sum_{d=1}^{p}\sup_{\theta^*\in M_n^\alpha(\bar\theta)}\left|\frac{\partial\phi_{jl}^k}{\partial\theta_d^{(m)}}(\theta^*)\right|\cdot\left|\theta_d^{(m)}-\bar\theta_d\right|V_{i-1}^l \\
&\le \frac{C_{1,j}}{\sqrt{n}}V_{i-1}^l,
\end{aligned}
$$

where $C_{1,j}$ as before. Thus, from relation (28) using Theorem 4 it follows that

$$\sup_{\theta^{(p)}\in M_n(\bar\theta)}\left|K_n^{p,k}(\theta^{(m)})\right|\xrightarrow{a.s.}0,$$

when $n\to\infty$. Using the result just obtained in expression (27) and using again (14), it follows that

$$\sup_{\theta^{(m)}\in M_n(\bar\theta)}\left|\frac{1}{n}J_n^{p,k}(\theta^{(m)})-J^{p,k}(\bar\theta)\right|\xrightarrow{a.s.}0, \tag{29}$$

as $n \to \infty$. This proves part (ii).

Additionally, for $z = 1, \dots, d$, let us define

$$Q_n^{p,k,z}(\theta) = \frac{\partial J_n^{p,k}(\theta)}{\partial \theta_z}, \quad \alpha^*_{pjkz}(v,\theta) = \frac{\partial \alpha^*_{pjk}(v,\theta)}{\partial \theta_z}, \quad f_j^{k,z}(v,\theta) = \frac{\partial f_j^k(v,\theta)}{\partial \theta_z},$$

$$B^{p,k,z}(\theta) = -\sum_{j=1}^{N} \sum_{l=0}^{p_j+q_j} \big\{ \phi^z_{jl}(\theta) E\big[\alpha^*_{pjk}(V_0,\theta)V_0^l\big] + \phi^k_{jl}(\theta) E\big[\alpha^*_{pjz}(V_0,\theta)V_0^l\big] + \phi^{k,z}_{jl}(\theta) E\big[\alpha^*_{pj}(V_0,\theta)V_0^l\big]\big\}.$$

Thus, part (iii) can be proved following the lines of the proof of part (ii) considering $\phi^k_{jl}, \phi^{k,z}_{jl}$ instead of ϕ_{jl} and ϕ^k_{jl}, taking into consideration that

$$\alpha^*_{pj}(v,\theta) \le cons_1(\theta) v^2, \quad \alpha^*_{pjk}(v,\theta) \le cons_2(\theta) v^2 \quad \alpha^*_{pjkz}(v,\theta) \le cons_3(\theta) v^2$$

and using Theorems 4 and 5 and (14).

Let $K_\epsilon^2 := \frac{1}{\epsilon} \sum_{j=1}^{N}\sum_{k=1}^{N} E\big[\alpha^*_{pj}(V_0,\theta)\alpha^*_{pk}(V_0,\theta)\big(X_1^{p_j}V_1^{q_j} - f_j(V_0,\theta)\big(X_1^{p_k}V_1^{q_k} - f_k(V_0,\theta)\big]$.
For an arbitrary $\epsilon > 0$ and $K_\epsilon > 0$, using the Chebyshev inequality and the stationarity of the volatility process we have

$$\begin{aligned} P\left[\left|\frac{G_n^p(\bar\theta)}{\sqrt{n}}\right| > K_\epsilon\right] &\le \frac{E\big[|G_n^p(\bar\theta)|^2\big]}{nK_\epsilon^2} \\ &= \frac{1}{nK_\epsilon^2}\sum_{i=1}^{n} E\Big\{\big[\sum_{j=1}^{N} \alpha^*_{pj}(V_{i-1},\theta)(X_i^{p_j}V_i^{q_j} - f_j(V_{i-1},\theta))\big]^2\Big\} \\ &= \frac{1}{K_\epsilon^2}\sum_{j=1}^{N}\sum_{k=1}^{N} E\Big[\alpha^*_{pj}(V_0,\theta)\alpha^*_{pk}(V_0,\theta)\big(X_1^{p_j}V_1^{q_j} - f_j(V_0,\theta)\big(X_1^{p_k}V_1^{q_k} - f_k(V_0,\theta)\big)\Big] \\ &< \epsilon. \end{aligned}$$

Thus, it follows that for every $\epsilon > 0$ there exists a $K_\epsilon > 0$ such that $\sup_{n\in\mathbb{N}} P\left[\left|\frac{G_n^j(\bar\theta)}{\sqrt{n}}\right| > K_\epsilon\right] < \epsilon$, which proves part (iv). Finally, using the result that has just been proven, the definition of $M_n^\alpha(\bar\theta)$ and (29), we have

$$\begin{aligned} \sup_{\theta\in M_n^\alpha(\bar\theta)} \big|\frac{1}{n}G_n^p(\theta)\big| &= \sup_{\theta\in M_n^\alpha(\bar\theta)} \big|\frac{1}{n}G_n^p(\bar\theta) + \frac{1}{n}J_n^{p,k}(\theta^*)(\theta - \bar\theta)\big| \\ &\le \big|\frac{1}{n}G_n^p(\bar\theta)\big| + \sup_{\theta^*\in M_n^\alpha(\bar\theta)} \big|\frac{1}{n}J_n^{p,k}(\theta^*)\big| \sup_{\theta\in M_n^\alpha(\bar\theta)} |\theta - \bar\theta| \\ &\le \frac{1}{\sqrt{n}}K_\epsilon + \sup_{\theta^*\in M_n^\alpha(\bar\theta)} \big|\frac{1}{n}J_n^{p,k}(\theta^*)\big| \frac{\alpha}{\sqrt{n}} \xrightarrow{P} 0, \end{aligned}$$

as $n \to \infty$. This completes the proof. □