

Prediction-based estimating functions for stochastic volatility models with noisy data: comparison with a GMM alternative

Anne Floor Brix · Asger Lunde

Received: 20 January 2014 / Accepted: 13 February 2015 / Published online: 26 February 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract Prediction-based estimating functions (PBEFs), introduced in Sørensen (Econom J 3:123–147, 2000), are reviewed, and PBEFs for the Heston (Rev Financ Stud 6:327–343, 1993) stochastic volatility model are derived with and without the inclusion of noise in the data. The finite sample performance of the PBEF-based estimator is investigated in a Monte Carlo study and compared to the performance of the Generalized Method of Moments (GMM) estimator from Bollerslev and Zhou (J Econom 109:33–65, 2002) that is based on conditional moments of integrated variance. We derive new moment conditions in the presence of noise, but we also consider noise correcting the GMM estimator by basing it on a realized kernel instead of realized variance. Our Monte Carlo study reveals great promise for the estimator based on PBEFs. The study also shows that the PBEF-based estimator outperforms the GMM estimator, both in the setting with MMS noise and in the setting without MMS noise, especially for small sample sizes. Finally, in an empirical application we fit the Heston model to SPY data and investigate how the two methods handle real data and possible model misspecification. The empirical study also shows how the flexibility of the PBEF-based method can be used for robustness checks.

Keywords GMM estimation · Heston model · Market microstructure noise · Prediction-based estimating functions · Realized variance

Electronic supplementary material The online version of this article (doi:[10.1007/s10182-015-0248-6](https://doi.org/10.1007/s10182-015-0248-6)) contains supplementary material, which is available to authorized users.

We are grateful to Michael Sørensen for interesting discussions as well as many useful comments. We also appreciate the comments from the editor and two anonymous referees which have improved the quality of the paper significantly. We acknowledge financial support by the Center for Research in Econometric Analysis of Time Series, CREATES, funded by the Danish National Research Foundation.

A. F. Brix · A. Lunde (✉)

Department of Economics and Business and CREATES, Aarhus University, Aarhus, Denmark
e-mail: alunde@econ.au.dk

1 Introduction

Continuous time stochastic volatility (SV) models are widely used in econometrics and empirical finance for modeling prices of financial assets. Considerable efforts have been put into modeling and estimation of the latent volatility process. Most of this research is surveyed in part II of [Andersen et al. \(2009\)](#). Stochastic volatility diffusion models, such as the [Heston \(1993\)](#) model, represent a popular class of models within the continuous time framework. The Heston model will be the baseline model considered in this paper, since it is one of the most widely used models in financial institutions, due to its analytical tractability.

Parameter estimation in SV models is difficult because the volatility process is latent. The hidden Markov structure complicates inference, since the observed log-price process will not in itself be a Markov process, which implies that computing conditional expectations of functions of the observed process is practically infeasible. As a consequence, martingale estimating functions will not be a useful tool for conducting inference in SV models. Likewise, likelihood inference is not straightforward, because an analytical expression for the transition density is almost never available and methods based on extensive simulations are called for.

We will circumvent the above-mentioned problems for conducting inference in SV models using prediction-based estimating functions (PBEFs), introduced in [Sørensen \(2000\)](#), which are a generalization of martingale estimating functions. This generalization becomes particularly useful when applied to observations from a non-Markovian model. PBEFs are estimating functions based on predictors of functions of the observed process.

In this paper, we investigate and contrast two estimation approaches. First, PBEFs will be reviewed, detailed, and used for parameter estimation in the Heston model. The estimation method is fairly easy to implement and fast to execute, as the construction of PBEFs relies only on the computation of unconditional moments. When the Heston SV model is considered, no simulations are needed for constructing the PBEFs used in the paper.¹ As a benchmark, we consider the method suggested in [Bollerslev and Zhou \(2002\)](#). In [Bollerslev and Zhou \(2002\)](#), a Generalized Method of Moments (GMM) type estimator based on the first and second order conditional moments of the integrated variance (IV) is derived. Since IV is latent, realized variance (RV) is used as a proxy and the sample moments of RV are matched to the population moments of IV implied by the model. When high-frequency data are available, several other simulation-free methods have been suggested in the literature, see for instance [Barndorff-Nielsen and Shephard \(2002\)](#), [Corradi and Distaso \(2006\)](#), and [Todorov \(2009\)](#). Common to these methods, including the GMM-based estimator from [Bollerslev and Zhou \(2002\)](#), is that they are all based on time series of daily realized measures, such as realized variance (RV) and

¹ An implementable version of the optimal PBEF will however require simulation of a covariance matrix. Simulation based estimation methods for continuous time SV models, such as indirect inference, see [Gourieroux et al. \(1993\)](#), the efficient method of moments (EMM), see [Gallant and Tauchen \(1996\)](#), or Markov Chain Monte Carlo (MCMC), see [Eraker \(2001\)](#), are not as easily implemented, since many of them require substantial computational efforts. Another way of tackling the difficulties, that arise when considering parameter estimation in continuous time SV models, is based on approximations of the likelihood function, see for example [Ant-Sahalia and Kimmel \(2007\)](#).

bipower variation (BV). Instead of being transformed into daily realized measures, the squared intra-daily returns are used directly when constructing PBEFs. This means that PBEFs have a potential informational advantage, the strength of which will be investigated throughout this paper. More specifically, this paper investigates the finite sample properties of the PBEF-based estimator in a Monte Carlo study and compares its performance to that of the GMM estimator from [Bollerslev and Zhou \(2002\)](#).

Both the case where the efficient price is assumed to be directly observable and the case where noise is present in data are considered. In particular, we contribute by extending the two competing methods to handle the presence of noise. The usage of PBEFs for estimating SV models was suggested, among others, in [Barndorff-Nielsen and Shephard \(2001\)](#), but to the best of our knowledge this is the first time the finite sample performance of PBEFs applied to SV models is being studied. In fact, this is the most extensive Monte Carlo study of the finite sample performance of the PBEF-based estimation method. In [Nolsøe et al. \(2000\)](#) the authors conduct a small Monte Carlo study for the case where a Cox–Ingersoll–Ross (CIR) process is observed with additive white noise, but the potential of using PBEFs to estimate SV models has not previously been studied.

This paper also addresses the link between the estimation method based on PBEFs and GMM based on the moment conditions underlying the PBEF. Especially, the connection between the optimal PBEF and the optimal choice of the weight matrix in GMM estimation is established.

Lastly, an empirical application using SPY data is carried out, investigating how the two estimation methods handle real data characteristics and possible model misspecification. In the empirical application we also study how different choices in the flexible PBEF-based estimation method might impact the parameter estimates. In particular, we investigate how considering different choices of the predictor space might serve as a robustness check of whether there is a need for additional volatility factors in the model.

This paper is organized as follows: In the following section the PBEF estimation method is reviewed and detailed. The connection to GMM estimation is established, and a brief review of the GMM-based estimator from [Bollerslev and Zhou \(2002\)](#) is provided. For both methods, the estimator of the parameters in the Heston model is derived with and without the inclusion of noise in the data. In Sect. 3, we present our Monte Carlo study. This includes an investigation of how i.i.d. noise impacts the performances of the two methods and how the noise-corrected estimators perform. Section 4 contains an empirical application to SPY data that investigates how the methods handle real data and if and how the choice of estimation method impacts the parameter estimates. The final section concludes, and ideas on further research are outlined. Some details have been relegated to a separate appendix that can be obtained from the authors upon request.

2 Estimating stochastic volatility models

In this section, the two estimation methods from [Sørensen \(2000\)](#) and [Bollerslev and Zhou \(2002\)](#) are reviewed and extended to handle MMS noise. We also discuss the link

between the optimal PBEF and a GMM estimator with the optimal choice of weight matrix. The focus of this paper is on the performance of the two considered estimation methods used for estimating SV models of the form

$$dX_t = \sqrt{v_t}dW_t, \quad dv_t = b(v_t; \theta)dt + c(v_t; \theta)dB_t, \tag{1}$$

where W and B are independent standard Brownian motions. The independence assumption rules out the possibility of leverage effects, but it is only imposed for computational ease and could be relaxed in other applications. In particular, we are interested in studying inference for the Heston SV model

$$dX_t = \sqrt{v_t}dW_t, \quad dv_t = \kappa(\alpha - v_t)dt + \sigma\sqrt{v_t}dB_t, \tag{2}$$

where the spot volatility, v_t , is a CIR-process. The parameter, α , is the long run average variance of the observed process, $\{X_t\}$, and the other drift parameter, κ , is the rate at which v_t reverts to the long run average. The third parameter, σ , can be interpreted as the volatility of volatility. The Heston model is widely used in mathematical finance where the observed process, $\{X_t\}$, would be the logarithm of an asset price. The popularity of the Heston model in financial institutions is primarily due to the analytical tractability of the model, which allows for (quasi) closed form expressions for prices of financial derivatives, such as European options.

2.1 Estimation using prediction-based estimating functions

First, we explain the general setup and ideas underlying the estimation method based on PBEFs that was introduced in [Sørensen \(2000\)](#) and further developed in [Sørensen \(2011\)](#). Then, following [Sørensen \(2000\)](#), we derive the PBEFs for the Heston model without MMS noise. Finally, we add MMS noise to the observations and derive PBEFs in this setting.

2.1.1 The general setup and ideas

The estimation method based on PBEFs is used for conducting parametric inference on observations Y_1, Y_2, \dots, Y_n from a general stochastic process. The stochastic process is assumed to belong to a class of models parametrized by a p -dimensional vector, $\theta \in \Theta \subseteq \mathbb{R}^p$, that we wish to estimate. An estimating function is a p -dimensional function $G_n(\theta)$ that depends on the data Y_1, Y_2, \dots, Y_n and θ , and an estimator is obtained by solving the p equations $G_n(\theta) = \underline{0}$ w.r.t. θ . Specifically, the estimating functions we study have the form

$$G_n(\theta) = \sum_{i=1}^n \Pi^{(i-1)}(\theta) \left[f(Y_i) - \hat{\pi}^{(i-1)}(\theta) \right], \tag{3}$$

where the function to be predicted, $f(Y_i)$, is defined on the state space of the data generating process Y . The function f is assumed to satisfy the condition $E_\theta[f(Y_i)^2] <$

∞ for all $\theta \in \Theta$ and for $i = 1, \dots, n$. $\Pi^{(i-1)}(\theta) = (\pi_1^{(i-1)}(\theta), \dots, \pi_p^{(i-1)}(\theta))$ has elements that belong to \mathcal{P}_{i-1}^θ , the predictor space for predicting $f(Y_i)$, and $\hat{\pi}^{(i-1)}(\theta)$ is the minimum mean square error (MMSE) predictor of $f(Y_i)$ in \mathcal{P}_{i-1}^θ . For each i , the predictor space, \mathcal{P}_{i-1}^θ , can be chosen as closed linear subspace of the Hilbert space \mathcal{H}_{i-1}^θ . So $\hat{\pi}^{(i-1)}(\theta)$ is the orthogonal projection of $f(Y_i)$ onto \mathcal{P}_{i-1}^θ w.r.t. the inner product in \mathcal{H}_{i-1}^θ .² This orthogonal projection exists and is uniquely determined by the normal equations

$$E_\theta[\pi [f(Y_i) - \hat{\pi}^{(i-1)}(\theta)]] = 0, \quad \text{for all } \pi \in \mathcal{P}_{i-1}^\theta, \tag{4}$$

see e.g. Theorem 3.1 in [Karlin and Taylor \(1975\)](#).³

A special class of PBEFs is the class of martingale estimating functions (MGEFs), which is obtained by choosing $\mathcal{P}_{i-1}^\theta := \mathcal{H}_{i-1}^\theta$. In this case, the MMSE predictor of $f(Y_i)$ in \mathcal{P}_{i-1}^θ is the conditional expectation, $\hat{\pi}^{(i-1)}(\theta) = E_\theta[f(Y_i)|Y_1, \dots, Y_{i-1}]$, and $G_n(\theta)$ becomes a P_θ -martingale w.r.t. the filtration generated by the data process. MGEFs are however mainly useful when considering Markovian models, since for Non-Markovian models it is practically infeasible to calculate conditional expectations, conditioning on the entire past of observations. The idea underlying PBEFs is to use a smaller and more tractable predictor space in place of \mathcal{H}_{i-1}^θ and think of the resulting PBEF as an approximation of the MGEF. The advantage of considering this approximation is that it is only based on unconditional moments, which are much easier to compute, or simulate, than conditional moments.

In the rest of the paper, we will restrict our attention to finite-dimensional predictor spaces, \mathcal{P}_{i-1}^θ , and assume that the observed process $\{Y_i\}$ is stationary. For asymptotic properties of the estimator in this setting, consult [Sørensen \(2000\)](#) and [Sørensen \(2011\)](#). In order to obtain even more tractable PBEFs, we will only consider $q + 1$ dimensional predictor spaces with basis elements of the form $Z_k^{(i-1)} = h_k(Y_{i-1}, \dots, Y_{i-s})$, $k = 0, \dots, q$, where $h_k : \mathbb{R}^s \mapsto \mathbb{R}$, $s \in \mathbb{N}$ and where the functions h_0, h_1, \dots, h_q are linearly independent and do not depend on θ . The predictor space used for predicting $f(Y_i)$ is then given by $\mathcal{P}_{i-1}^\theta = \text{span}\{Z_0^{(i-1)}, Z_1^{(i-1)}, \dots, Z_q^{(i-1)}\}$. The basis elements of the predictor space are no longer assumed to be functions of the entire past, but they are instead functions of the “most recent past” of a period of length s . The predictors in \mathcal{P}_{i-1}^θ will therefore be of the form $a_0 + a'Z^{(i-1)}$, where $a' = (a_1, \dots, a_q)$ and $Z^{(i-1)} = (Z_1^{(i-1)}, \dots, Z_q^{(i-1)})'$, imposing that $h_0 = 1$.⁴ The normal equations (4)

² One could also choose to predict functions of the type $f(Y_i, \dots, Y_{i-s})$, see [Sørensen \(2011\)](#), but for the purpose of this study $f(Y_i)$ will be general enough. The function f can be chosen freely but will often take the form $f(Y_i) = Y_i^\nu$, $\nu \in \mathbb{N}$, such that the moments needed to find the (optimal) PBEF are easier to calculate. PBEFs can in fact be further generalized to a setup where several functions of the data, $f_j(Y_i)$ $j = 1, \dots, N$, are predicted, see [Sørensen \(2000\)](#) and [Sørensen \(2011\)](#). This generalization will, however, not be necessary for estimating the SV model we are considering.

³ Unique in the sense of mean square distance.

⁴ In the case of the Heston model, we will consider basis elements of the form $Z_k^{(i-1)} = Y_{i-k}^2$, $k = 1, \dots, q$.

lead to the MMSE predictor

$$\hat{\pi}^{(i-1)}(\theta) = \hat{a}_0(\theta) + \hat{a}(\theta)^T Z^{(i-1)}, \tag{5}$$

where $\hat{a}(\theta) = C(\theta)^{-1}b(\theta)$ and $\hat{a}_0(\theta) = E_\theta[f(Y_i)] - \hat{a}(\theta)'E_\theta[Z^{(i-1)}]$. $C(\theta)$ denotes the $q \times q$ covariance matrix of $Z^{(i-1)}$ and $b(\theta) = [\text{Cov}_\theta(Z_1^{(i-1)}, f(Y_i)), \dots, \text{Cov}_\theta(Z_q^{(i-1)}, f(Y_i))]'$. Note that, since the observed process $\{Y_i\}$ is stationary, the coefficients of the MMSE predictor do not depend on i , but stay constant across time.⁵ For a formal derivation of the expressions for the coefficients in (5) see Appendix A. From (3)–(5) it follows that PBEFs can be calculated if we can calculate the first- and second-order moments of the random vector $(f(Y_i), Z_1^{(i-1)}, \dots, Z_q^{(i-1)})$.

Within the setup of the finite-dimensional predictor spaces considered above, we now turn to the specification of the $p \times 1$ -vector $\Pi^{(i-1)}(\theta)$ from (3). Since each element of the vector $\Pi^{(i-1)}(\theta)$ belongs to the predictor space \mathcal{P}_{i-1}^θ , the j^{th} element of $\Pi^{(i-1)}(\theta)$ is of the form $\pi_j^{(i-1)}(\theta) = \sum_{k=0}^q a_{jk}(\theta)Z_k^{(i-1)}$, where, as before, $Z_0^{(i-1)} = 1$. Note that the coefficients $a_{jk}(\theta)$ do not depend on i but are, like the coefficients of the MMSE, $\hat{a}(\theta)$ and $\hat{a}_0(\theta)$, constant over time. Therefore, in order to ease notation, we define

$$A(\theta)_{p \times (q+1)} = \begin{pmatrix} a_{10}(\theta) & \cdots & a_{1q}(\theta) \\ \vdots & & \vdots \\ a_{p0}(\theta) & \cdots & a_{pq}(\theta) \end{pmatrix}, \quad H^{(i)}(\theta)_{(q+1) \times 1} = \begin{pmatrix} Z_0^{(i-1)}[f(Y_i) - \hat{\pi}^{(i-1)}(\theta)] \\ \vdots \\ Z_q^{(i-1)}[f(Y_i) - \hat{\pi}^{(i-1)}(\theta)] \end{pmatrix},$$

for $i = 1, \dots, n$ and $F_n(\theta) := \sum_{i=s+1}^n H^{(i)}(\theta)$.⁶ With this notation at hand we are considering PBEFs of the form

$$G_n(\theta) = A(\theta)F_n(\theta), \tag{6}$$

where we need $p \leq q + 1$ to identify the p unknown parameters. Finding the optimal PBEF within a class of PBEFs of the type (6) is then a question of choosing an optimal weight matrix, $A^*(\theta)$. The optimal PBEF will then be the estimating function, within the considered class of estimating functions of type (6), that is closest to the score in an \mathcal{L}^2 -sense. For further details on the optimal PBEF see Sørensen (2000) or Appendix B.

2.1.2 Relating PBEFs to GMM estimation

The PBEFs and martingale estimating functions share many similarities with GMM estimation from the econometrics literature. In this subsection we will explain the link between the optimal PBEF and the optimal GMM estimator based on the moment conditions from the normal equations.

⁵ PBEFs with finite-dimensional predictor spaces can also be computed for non-stationary processes, but in this case computing the MMSE predictor, $\hat{\pi}(\theta)$, is a bit more complicated since the coefficients, $\hat{a}_0(\theta), \dots, \hat{a}_q(\theta)$, become time varying.

⁶ Note that the sum starts at $i = s + 1$ since $Z_k^{(i-1)}$ is only well defined for $i \geq s + 1$.

The PBEF-based estimator is obtained by solving $G_n(\theta) = \underline{0}$ for θ , but for numerical reasons it is often easier to minimize $G_n(\theta)'G_n(\theta)$ w.r.t. $\theta \in \Theta$. We employ this approach and find an estimator by solving

$$\min_{\theta \in \Theta} G_n(\theta)'G_n(\theta) = \min_{\theta \in \Theta} F_n(\theta)'A(\theta)'A(\theta)F_n(\theta).$$

This expression looks very similar to the GMM objective function that emerges if we perform GMM estimation using the $q + 1$ moment conditions $E_\theta[H(\theta)] = \underline{0}$. In this case the GMM objective function to be minimized is $\left(\frac{1}{n-s}F_n(\theta)\right)'W_n(\theta)\left(\frac{1}{n-s}F_n(\theta)\right)$, which is equivalent to minimizing $F_n(\theta)'W_n(\theta)F_n(\theta)$. In the latter case, the corresponding p first-order conditions are

$$2(\partial_\theta F_n(\theta))'W_n(\hat{\theta})F_n(\theta) = \underline{0}, \tag{7}$$

if we evaluate the GMM weight matrix, $W_n(\hat{\theta})$, at some consistent parameter estimate $\hat{\theta}$ such that the weight matrix does not depend on θ . The first-order conditions (7) have the same structure as the PBEFs in (6). The only difference is that the term in front of $F_n(\theta)$ in (7) becomes data dependent. However, it turns out that there is a strong link between (7) with $W_n(\hat{\theta})$ chosen optimally and the optimal PBEF of type (6). The optimal PBEF takes the form $G_n^*(\theta) = A_n^*(\theta)F_n(\theta)$, where $A_n^*(\theta) = U(\theta)'\overline{M}_n(\theta)^{-1}$ and the expression for $U(\theta)$ and $\overline{M}_n(\theta)^{-1}$ can be found in Sørensen (2000) or Appendix B. Straightforward calculations show that

$$-\frac{1}{n-s}\partial_\theta F_n(\theta)' \xrightarrow{p} U(\theta)', \quad \text{when } n \rightarrow \infty. \tag{8}$$

From the theory on GMM estimation, we know that the optimal choice of weight matrix, $W_n(\theta)$, is the inverse of the covariance matrix of $F_n(\theta)$ since the $H^{(i)}$'s are correlated. In the GMM setting this weight matrix will in practice be constructed using the sample version of the covariance matrix. When $W_n(\theta)$ is chosen optimally, $W_n(\theta)$ equals $\frac{1}{n-s}\overline{M}_n(\theta)^{-1}$ and (7) becomes the empirical analog of the optimal PBEF, $G_n^*(\theta) = A_n^*(\theta)F_n(\theta)$. Constructing the optimal PBEF is therefore the same as constructing the theoretical first-order conditions that emerge from the optimal GMM objective functions based on the moment conditions, $E_\theta[H(\theta)] = \underline{0}$, from the normal equations. The choice of f and predictor space then translates into which moment conditions to use in the GMM estimation. Hence, these choices will also impact the efficiency of the resulting PBEF-based estimator. Once these choices are made, the optimal PBEF-based estimator is linked to the optimal GMM estimator, based on the moment conditions from the normal equations, as described above. Note that a sub-optimal choice of the weight matrix $W_n(\theta)$ will lead to a sub-optimal PBEF, but the class of PBEFs is in general broader than the ones having the structure (7).

2.1.3 PBEFs for SV models without noise in the data

We now return to the setup from (1) and, following Sørensen (2000), review how to compute PBEFs for SV models without MMS noise.

Suppose the process X has been observed at discrete time points $X_0, X_\Delta, X_{2\Delta}, \dots, X_{n\Delta}$. In this setup, inference based on MGEFs becomes practically infeasible, since the conditional expectations appearing in the MGEFs, which are based on $f(Y_i) - E_\theta[f(Y_i)|Y_{i-1}, \dots, Y_1]$, are difficult to compute analytically, as well as numerically. One feasible approach for conducting inference is to use PBEFs instead. In fact, for many models, such as the Heston model, we are able to derive analytical expressions for the PBEFs. Then the continuous time returns from (1) are given by

$$Y_i = X_{i\Delta} - X_{(i-1)\Delta} = \int_{(i-1)\Delta}^{i\Delta} \sqrt{v_t} dW_t, \tag{9}$$

which allows for the decomposition $Y_i = \sqrt{S_i}Z_i$, where the Z_i 's are i.i.d. standard normal random variables independent of $\{S_i\}$, and where the process $\{S_i\}$ is given by $S_i = \int_{(i-1)\Delta}^{i\Delta} v_t dt$. We will assume v to be a positive, ergodic, diffusion process with invariant measure μ_θ and that $v_0 \sim \mu_\theta$ is independent of B , which implies that v is stationary.

To construct PBEFs, we have to decide on which function of the data to predict. Since the Y_i 's are uncorrelated, trying to predict Y_i using $Y_{i-1}, Y_{i-2}, \dots, Y_{i-q}$ will not work and $f(y) = y$ is a bad choice. To match empirical data, where we have volatility clustering, squared returns from the considered SV models are often correlated, and a natural choice for f would therefore be $f(y) = y^2$. The decomposition $Y_i = \sqrt{S_i}Z_i$ also reveals that $f(y) = y^2$ is a convenient choice, as it eases the computation of the moments required to construct the PBEFs. Other choices of f might result in efficiency gains, but without further knowledge of the intractable score functions that we aim to estimate, we will stick to the class of polynomial PBEFs as they offer computational ease.⁷ We choose $\mathcal{P}_{i-1}^\theta = \{a_0 + a_1Y_{i-1}^2 + \dots + a_qY_{i-q}^2 | a_k \in \mathbb{R} \ k = 0, 1, \dots, q\}$ as our predictor space. Hence, $Z_k^{(i-1)} = Y_{i-k}^2$ for $k = 1, 2, \dots, q$ having the same functional form as the function of the data to predict.⁸ Notice that in this case $s = q$ since \mathcal{P}_{i-1}^θ is spanned by the ‘‘most recent past of squared returns of length q ’’.⁹ With the above choice of f and predictor space, the MMSE predictor is given by

$$\hat{\pi}^{(i-1)}(\theta) = \hat{a}_0(\theta) + \hat{a}(\theta)'Z^{(i-1)}, \quad \text{with } Z^{(i-1)} = (Y_{i-1}^2, \dots, Y_{i-q}^2),$$

$$\hat{a}(\theta) = C^{-1}(\theta)b(\theta), \quad \hat{a}_0(\theta) = E_\theta(Y_1^2)[1 - (\hat{a}_1(\theta) + \dots + \hat{a}_q(\theta))]. \tag{10}$$

As before, C denotes the covariance matrix of $Z^{(i-1)}$, and b is the $q \times 1$ -vector with j^{th} element given by $\text{Cov}_\theta(Y_{i-j}^2, Y_i^2)$. Together with (6), this means that we are considering PBEFs of the form

⁷ Higher power of y could also have been considered, but very high moments are often not reliable for empirical investigations, and we choose to stick with $f(y) = y^2$ as suggested in Sørensen (2000).

⁸ It should be noted that one does not have to choose a predictor space spanned by variables of the same functional form as f , even though it seems like the most natural choice.

⁹ When the volatility process $\{v_t\}$ is ρ -mixing the coefficient $\hat{a}_k(\theta)$ decreases exponentially with k . Then q is not required to be very large since q represents the ‘‘required’’ information for predicting $f(Y)$, [see Theorem 3.3 in Bradley (2005)].

$$G_n(\theta) = \sum_{i=q+1}^n \underbrace{\Pi^{(i-1)}(\theta)}_{p \times 1} [Y_i^2 - (\hat{\alpha}_0(\theta) + \hat{\alpha}_1(\theta)Y_{i-1}^2 + \dots + \hat{\alpha}_q(\theta)Y_{i-q}^2)], \quad (11)$$

with $\Pi^{(i-1)}(\theta) = A(\theta)\tilde{Z}^{(i-1)}$, where $\tilde{Z}^{(i-1)} = (1, Y_{i-1}^2, \dots, Y_{i-q}^2)'$. In our Monte Carlo study we will use the following sub-optimal, yet simple weight matrix

$$A(\theta) = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \end{pmatrix},$$

since computing the optimal weight matrix $A^*(\theta)$ involves computing the covariance matrix of $F_n(\theta)$.¹⁰ The resulting sub-optimal PBEF is

$$G_n(\theta) = \sum_{i=q+1}^n \begin{pmatrix} 1 \\ Y_{i-1}^2 \\ Y_{i-2}^2 \end{pmatrix} [Y_i^2 - (\hat{\alpha}_0(\theta) + \hat{\alpha}_1(\theta)Y_{i-1}^2 + \dots + \hat{\alpha}_q(\theta)Y_{i-q}^2)]. \quad (12)$$

Equating (12) to zero and solving for θ gives a \sqrt{n} -consistent estimator, but we may lose some efficiency for not using the optimal weight matrix $A^*(\theta)$. However, the aim of the paper is to study the finite sample performance of an easy implementable and simulation-free PBEF. As we shall see in our Monte Carlo study, the estimator based on the sub-optimal PBEF performs well in finite samples, and a study of the possible further improvements from using the optimal PBEF is left for future research.

The sub-optimal PBEF from (12) can now be computed if we can calculate $\hat{\alpha}_0(\theta), \hat{\alpha}_1(\theta), \dots, \hat{\alpha}_q(\theta)$. For this we need $E_\theta[Y_i^2]$, $\text{Var}_\theta(Y_i^2)$ and $\text{Cov}_\theta(Y_i^2, Y_{i+j}^2)$ for $j = 1, \dots, q$. Following (Barndorff-Nielsen and Shephard 2001, pp. 179–181), the required moments can be calculated from the moments of the volatility process $\{v_i\}$. For the Heston they are given by

$$E_\theta[Y_i^2] = \Delta\alpha, \quad \text{Var}_\theta(Y_i^2) = \frac{6\alpha\sigma^2}{2\kappa^3}(e^{-\kappa\Delta} + \kappa\Delta - 1) + 2\Delta^2\alpha^2, \quad \text{and}$$

$$\text{Cov}_\theta(Y_i^2, Y_{i+j}^2) = \frac{\alpha\sigma^2}{2\kappa^3}(e^{-\kappa\Delta j}[e^{-\kappa\Delta} - 2 + e^{\kappa\Delta}]).$$

2.1.4 PBEFs for SV models with noisy data

We now add noise to the observation scheme from (9). More specifically, we add i.i.d. Gaussian noise to the log-price process

$$X_i = X_i^* + U_i, \quad U_i \text{ i.i.d. } N(0, \omega^2), \quad (13)$$

¹⁰ A task that involves computing $E_\theta[Y_i^2 Y_j^2 Y_k^2 Y_l^2]$ for $i \geq j \geq k$. For further details on how to compute optimal PBEFs for stochastic volatility models, see Sørensen (2000). In Sørensen (2000) an analytical formula for the optimal PBEF for an affine SV model, such as the Heston model, is also given. Even though an analytical expression for $A^*(\theta)$ is in principle available, it is a very complicated expression and not easily implementable. In practice, a feasible strategy could be to simulate $A^*(\theta)$.

where the efficient log-price process X^* comes from the Heston model and X^* and U are assumed to be independent. The additive error term U_i will be interpreted as market microstructure (MMS) noise due to market frictions such as bid-ask bounce, liquidity changes, and discreteness of prices. When MMS noise is present, the observed returns have the following structure

$$Y_i = X_i - X_{i-1} = (X_i^* - X_{i-1}^*) + (U_i - U_{i-1}) = Y_i^* + \epsilon_i, \tag{14}$$

where the MA(1) process, ϵ , is normally distributed, $N(0, 2\omega^2)$, and independent of the efficient return process Y^* .

To correct for MMS noise in the PBEFs, the moments used to construct the MMSE predictor have to be recalculated. That is, $E_\theta[Y_i^2]$, $\text{Var}_\theta(Y_i^2)$ and $\text{Cov}_\theta(Y_i^2, Y_{i+j}^2)$ need to be computed in the setting from (14).

Straightforward calculations give $E_\theta[Y_i^2] = \Delta\alpha + 2\omega^2$, since Y^* and ϵ are independent and have mean zero. We can now derive the bias in α that can be expected to occur, when performing the PBEF-based estimation procedure without correcting for MMS noise. If the MMS noise is not taken into account, then the equation, $E_\theta[Y_i^2] = \Delta\alpha$, is erroneously used for constructing the PBEF. Therefore, the expected bias in α is given by $\frac{2\omega^2}{\Delta}$, and as we shall see, this quantity matches with the bias found in our Monte Carlo Study. Since we do not have an analytical expression for the estimators, we will not attempt to derive the bias encountered in κ and σ .

As for the variance of the squared returns, it follows from (14) that $Y_i^2 = Y_i^{*2} + \epsilon_i^2 + 2Y_i^*\epsilon_i$, and since the three terms are uncorrelated, we have that

$$\text{Var}_\theta(Y_i^2) = \text{Var}_\theta(Y_i^{*2}) + \text{Var}_\theta(\epsilon_i^2) + 4\text{Var}_\theta(Y_i^*\epsilon_i). \tag{15}$$

Given the structure of the noise process, and since the noise process is normally distributed with mean zero and a variance of $2\omega^2$, we find that $\text{Var}_\theta(\epsilon_i^2) = 8\omega^4$. The efficient return process and the noise process are independent, and both have mean zero, so $\text{Var}_\theta(Y_i^*\epsilon_i) = E_\theta[\epsilon_i^2]E_\theta[Y_i^{*2}] = 2\omega^2\Delta\alpha$. Plugging this into (15) yields

$$\text{Var}_\theta(Y_i^2) = \text{Var}_\theta(Y_i^{*2}) + 8\omega^4 + 8\omega^2\Delta\alpha. \tag{16}$$

Regarding the covariance structure of the squared returns, only the first-order covariance will change due to the MA(1) structure in the return errors ϵ . By, once again, exploiting that Y^* and ϵ are independent and both have mean zero, we obtain the following expression for the first-order covariance of the observed squared return series

$$\text{Cov}_\theta(Y_i^2, Y_{i+1}^2) = \text{Cov}_\theta(Y_i^{*2}, Y_{i+1}^{*2}) + \text{Cov}_\theta(\epsilon_i^2, \epsilon_{i+1}^2) = \text{Cov}_\theta(Y_i^{*2}, Y_{i+1}^{*2}) + 2\omega^4. \tag{17}$$

We can now compute the noise corrected version of the PBEF previously described. Note that we can choose to estimate the variance of the noise, ω^2 , in a first step, by for instance, plugging a non-parametric estimator into the noise corrected PBEF used for estimating κ , α and σ . Another approach would be to expand the parameter vector to

$\theta = (\kappa, \alpha, \sigma, \omega^2)$ and use the noise corrected PBEF to estimate all four parameters. In the last approach one would have to choose a $4 \times (q + 1)$ weight matrix, $A(\theta)$. This will result in a 4×1 estimating function $G_n(\theta)$, such that our estimator $\hat{\theta}$ is obtained by solving four equations in four unknowns. We will follow the last approach and estimate all four parameters in one step. Since we have chosen $q = 3$ and wish to estimate ω^2 , the weight matrix will be a 4×4 matrix. This means, that the weight matrix can be ignored when solving $G_n(\theta) = \underline{0}$, under the assumption that $A(\theta)$ is invertible and the sub-optimal PBEF we have considered so far will, in this setting, be optimal.

2.2 A GMM estimator based on moments of integrated volatility

In this subsection, the GMM-based estimation procedure from [Bollerslev and Zhou \(2002\)](#) is reviewed and extended to handle MMS noise. In [Bollerslev and Zhou \(2002\)](#), the moment conditions for constructing the GMM estimator arise from the analytical derivations of the conditional first- and second-order moments of the daily integrated variance (IV) process. We will consider both a parametric and a non-parametric way of accounting for the presence of MMS noise in the data used for constructing the GMM estimator. In the parametric approach, the moment conditions are adjusted to hold in the MMS noise setting, and in the non-parametric approach we use a noise robust estimate of IV, namely the realized kernel (RK) from [Barndorff-Nielsen et al. \(2008a\)](#).

2.2.1 The GMM estimator without noise in the data

We now review the GMM estimator from [Bollerslev and Zhou \(2002\)](#) in the case where we have observations from the Heston model without MMS noise. Since the daily IV is latent, the realization of this time series is approximated by the daily realized variance (RV). Replacing population moments of IV with sample moments of RV results in an easy-to-implement GMM estimator. The GMM-based estimation method crucially depends on the availability of high-frequency data, since high-frequency data will ensure that RV is a good approximation of IV the moment conditions will hold approximately for RV.

When considering the Heston model, the conditional moment conditions used for constructing the GMM estimator are given by

$$\begin{aligned} E_\theta[IV_{t+1,t+2} - \delta IV_{t,t+1} - \beta \mathcal{G}_t] &= 0, \\ E_\theta[IV_{t+1,t+2}^2 - H(IV_{t,t+1}^2) - I(IV_{t,t+1}) - J \mathcal{G}_t] &= 0, \end{aligned} \tag{18}$$

where $IV_{t,t+1}$ denotes the integrated variance from day t to day $t - 1$ and $\mathcal{G}_t = \sigma(\{IV_{t-s-1,t-s} | s = 0, 1, 2, \dots, \infty\})$. The functions $\delta, \beta, H, I,$ and J are functions of the parameters $\kappa, \alpha,$ and σ and can be found in Appendix C. The functions δ and β only depend on the drift parameters κ and α which is why the second moment condition is needed. For further details on the derivation of the two conditional moments conditions see [Bollerslev and Zhou \(2002\)](#) or Appendix C. To get enough moment conditions to

identify θ , the two moment conditions are augmented by $IV_{t-1,t}$ and $IV_{t-1,t}^2$, yielding a total of six moment conditions. By replacing daily IV with daily RV, and using the unconditional versions of the moment conditions, a feasible GMM estimator is obtained from

$$\hat{\theta}_T = \underset{\theta}{\operatorname{argmin}} \left(\frac{1}{T-2} \sum_{t=1}^{T-2} f_t(\theta) \right)' \hat{W} \left(\frac{1}{T-2} \sum_{t=1}^{T-2} f_t(\theta) \right), \tag{19}$$

with $\hat{W} = \hat{S}^{-1}$, where \hat{S} is a consistent estimate of the asymptotic covariance matrix of $g_T(\theta) = \frac{1}{T-2} \sum_{t=1}^{T-2} f_t(\theta)$ and $f_t(\theta)$ is given by

$$f_t(\theta) = \begin{pmatrix} RV_{t+1,t+2} - \delta RV_{t,t+1} - \beta \\ RV_{t+1,t+2}^2 - H(RV_{t,t+1}^2) - I(RV_{t,t+1}) - J \\ [RV_{t+1,t+2} - \delta RV_{t,t+1} - \beta] RV_{t-1,t} \\ [RV_{t+1,t+2}^2 - H(RV_{t,t+1}^2) - I(RV_{t,t+1}) - J] RV_{t-1,t} \\ [RV_{t+1,t+2} - \delta RV_{t,t+1} - \beta] RV_{t-1,t}^2 \\ [RV_{t+1,t+2}^2 - H(RV_{t,t+1}^2) - I(RV_{t,t+1}) - J] RV_{t-1,t}^2 \end{pmatrix}. \tag{20}$$

2.2.2 Parametrically correcting for noisy data in the GMM approach

Recall, when MMS noise is present, the observed returns have the following structure

$$Y_i = X_i - X_{i-1} = (X_i^* - X_{i-1}^*) + (U_i - U_{i-1}) = Y_i^* + \epsilon_i.$$

If we denote the realized variance based on the MMS noise contaminated returns by RV^{MMS} and the realized variance based on the efficient return process by RV^* , we can rewrite RV^{MMS} over day t as

$$RV_{t,t+1}^{\text{MMS}} = RV_{t,t+1}^* + \sum_{i=1}^m \epsilon_{i,t}^2 + 2 \sum_{i=1}^m \epsilon_{i,t} Y_{i,t}^*, \tag{21}$$

where the number of intra-day observations is given by $m := \Delta^{-1}$.

The idea is to noise correct the GMM estimation approach from [Bollerslev and Zhou \(2002\)](#) by adjusting the moment condition from (20) such that they hold for RV^{MMS} . In order to do so, we have to extend the filtration we condition on to a larger filtration, making RV^{MMS} measurable w.r.t that filtration. The moment conditions from [Bollerslev and Zhou \(2002\)](#) were derived using the sigma-algebra $\mathcal{G}_t = \sigma(\{IV_{t-s-1,t-s} | s = 0, 1, 2, \dots\})$, that was approximated by $\hat{\mathcal{G}}_t = \sigma(\{RV_{t-s-1,t-s}^* | s = 0, 1, 2, \dots\})$. Instead of $\hat{\mathcal{G}}_t$, we will now consider the larger filtration, $\hat{\mathcal{H}}_t$, generated by RV^* , the efficient returns process, Y^* , and the noise process, ϵ , up until the beginning of day t . Define

$$\hat{\mathcal{H}}_t := \sigma(\{RV_{t-s-1,t-s}^*, Y_{i,t-s-1}^*, \epsilon_{i,t-s-1} | s = 0, 1, 2, \dots \text{ and } i = 1, 2, \dots, m\}),$$

where $Y_{i,t-1}^*$ and $\epsilon_{i,t-1}$ for $i = 1, 2, \dots, m$ denote the intra-day returns of the efficient price process and the MMS noise process during day $t - 1$ respectively. We now consider how to extend the first conditional moment condition from (18), using the decomposition from (21)

$$\begin{aligned}
 & E_\theta \left[\text{RV}_{t+1,t+2}^{\text{MMS}} - \delta \text{RV}_{t,t+1}^{\text{MMS}} - \beta | \hat{\mathcal{H}}_t \right] \\
 &= E_\theta \left[\text{RV}_{t+1,t+2}^* - \delta \text{RV}_{t,t+1}^* - \beta | \hat{\mathcal{H}}_t \right] \\
 &+ E_\theta \left[\sum_{i=1}^m \epsilon_{i,t+1}^2 - \delta \sum_{i=1}^m \epsilon_{i,t}^2 | \hat{\mathcal{H}}_t \right] \\
 &+ 2E_\theta \left[\sum_{i=1}^m \epsilon_{i,t+1} Y_{i,t+1}^{*2} - \delta \sum_{i=1}^m \epsilon_{i,t} Y_{i,t}^{*2} | \hat{\mathcal{H}}_t \right].
 \end{aligned}
 \tag{22}$$

Let us first consider (22). From the moment conditions used in the no noise case we know that we approximately get a zero when conditioning on $\hat{\mathcal{G}}_t$ in (22) instead. Under the assumption that the efficient return series, $\{Y^*\}$, up till time t do not add significant explanatory power compared to the information contained in $\hat{\mathcal{G}}_t$, (22) will also equal zero approximately, since both the $\{\text{RV}^*\}$ and $\{Y^*\}$ series are independent of the noise process. We do not use overnight returns, so the MA(1) structure in the noise process only holds within the trading day. This means that it will not impact the calculation of the conditional expectation, which will just equal the unconditional expectation $(1 - \delta)2\omega^2 m$. Since, as just discussed, the noise process is independent of any realizations from previous days and has mean zero, the conditional expectation will just equal zero. All in all, this leaves us with the noise adjusted moments condition

$$E_\theta \left[\text{RV}_{t+1,t+2}^{\text{MMS}} - \delta \text{RV}_{t,t+1}^{\text{MMS}} - \beta - (1 - \delta)2\omega^2 m | \hat{\mathcal{H}}_t \right] \approx 0.
 \tag{23}$$

As in the no MMS noise case we will augment the conditional moment condition (23) by $\text{RV}_{t-1,t}^{\text{MMS}}$ and $(\text{RV}_{t-1,t}^{\text{MMS}})^2$ to get two additional moment conditions.

Turning our attention to the second conditional moment condition from (18), we wish to compute

$$E_\theta \left[(\text{RV}_{t+1,t+2}^{\text{MMS}})^2 - H(\text{RV}_{t,t+1}^{\text{MMS}})^2 - I(\text{RV}_{t,t+1}^{\text{MMS}}) - J | \hat{\mathcal{H}}_t \right].
 \tag{24}$$

This task is however not feasible since it involves computing $E_\theta[Y_{i,t+1}^{*2} | \mathcal{H}_t]$ and $E_\theta[Y_{i,t}^{*2} | \mathcal{H}_t]$. If this was possible, we could use these expressions to form martingale estimating functions and would not need to use PBEFs for the estimation of our SV model. The problem is that we do not have an analytical expression for the conditional expectation of the squared returns during day $t + 1$ and day t given the filtration generation by the return series up until time t . Instead we will settle for four moment conditions and simply use the unconditional expectation of (24) given by

$$\begin{aligned}
 E_{\theta}[(RV_{i+1,t+2}^{\text{MMS}})^2 - H(RV_{i,t+1}^{\text{MMS}})^2 - I(RV_{i,t+1}^{\text{MMS}}) - J - K - L] &\approx 0, \\
 K &= (1 - H)(4m^2\omega^4 + 4m\omega^2\alpha + 12m\omega^4 - 4m\omega^4 + 8\omega^2\alpha), \\
 L &= -2m\omega^2 I
 \end{aligned} \tag{25}$$

The derivation of the moment condition above can be found in Appendix C.

2.2.3 Non-parametrically correcting for noisy data in the GMM approach

In the presence of i.i.d. MMS noise that is independent of the efficient log-price process, [Hansen and Lunde \(2006\)](#) show that the bias in RV equals $2\Delta^{-1}\omega^2$.

In fact, the variance of RV also diverges to infinity as the sampling frequency increases. In the setting with MMS noise, RV is no longer a consistent estimator of IV. We will therefore use a noise robust estimate of IV when constructing the GMM estimator. Instead of basing the estimation procedure on the time series of daily RV, we will use the time series of daily realized kernels (RK) from [Barndorff-Nielsen et al. \(2008a\)](#). The estimator is then constructed using the moment conditions (20), replacing RV with RK. We use the flat-top Tukey–Hanning₂ kernel, since the resulting RK is closest to being efficient in the setting of i.i.d. noise that is independent of the observed process. As for the bandwidth, H , we follow the asymptotic derivations from [Barndorff-Nielsen et al. \(2008a\)](#) and let $H \propto (1/\Delta)^{1/2}$, in order to obtain the optimal rate of convergence, $(1/\Delta)^{1/4}$, of RK to IV.¹¹

3 A Monte Carlo study of the finite sample performances

3.1 The setup and the case without noisy data

In this subsection and the following, the finite sample performances of the PBEF-based estimator from [Sørensen \(2000\)](#) and the GMM estimator from [Bollerslev and Zhou \(2002\)](#) are investigated in a Monte Carlo study. This subsection investigates the potential of using the intra-day returns directly in the PBEF-based estimator in a setting without MMS noise. The benchmark used for evaluating the performance of the PBEF-based method is the GMM approach from the previous subsection. In the next subsection, we first consider a setup with mild model misspecification, in the sense that we now add MMS noise to the simulated data and investigate how this impacts the two estimation methods. Afterwards, the performance of the noise corrected estimation methods are studied.

The data used for constructing the estimators are simulated realizations from the Heston model (2). We use a first-order Euler scheme to simulate the volatility- and log-price processes. The log-price is sampled every 30 s in the artificial 6.5 h of daily trading, for sample sizes of $T = 100, 400, 1000$ and 4000 trading days. Using the simulated data, daily realized variances based on the artificial 5-min returns are con-

¹¹ For further details on how the bandwidth is chosen, consult section 4 of [Barndorff-Nielsen et al. \(2008a\)](#).

structured. We will think of the five-minute returns as our available data. Since we are using 5-min returns over 6.5 h of trading, we have $\Delta = 1/78$.

To get a better grasp of the finite sample performance of the estimator based on PBEFs, as well as the GMM estimator, we conduct our Monte Carlo experiment in three different scenarios of parameter configurations.

- Scenario 1: $(\kappa, \alpha, \sigma) = (0.03, 0.25, 0.10)$. The volatility process is highly persistent (near unit-root). The autocorrelation is given by $r(u, \theta) = e^{-\kappa u}$ and the correlation between the volatility processes sampled five minutes apart equals $e^{-0.03 \cdot 1/78} = 0.9996$. The half-life of the volatility process equals 23.1 days.
- Scenario 2: $(\kappa, \alpha, \sigma) = (0.10, 0.25, 0.10)$. Here we have a slightly less persistent volatility process due to the increase in the mean-reversion parameter. The half-life now equals 6.93 days.
- Scenario 3: $(\kappa, \alpha, \sigma) = (0.10, 0.25, 0.20)$. The local variance of volatility is now increased. This process is also close to the non-stationary region since the CIR process is stationary if and only if $2\kappa\alpha \geq \sigma^2$, and here $2\kappa\alpha - \sigma^2 = 0.01$ (compared to 0.04 in scenario 2).

The same scenarios were considered in the Monte Carlo study conducted in [Bollerslev and Zhou \(2002\)](#). In [Bollerslev and Zhou \(2002\)](#), the authors only consider the rather large sample sizes $T = 1000$ and $T = 4000$, corresponding to 4 and 16 years of data. Our Monte Carlo study therefore also contributes by investigating the usability of this method when less data are available. We impose strict positivity of the parameter estimates $\hat{\kappa}$, $\hat{\alpha}$ and $\hat{\sigma}$ and use the true values of θ as starting values in the numerical routines. In each case the number of Monte Carlo replications is 1000.

An interesting question that arises when considering PBEFs is how to optimally choose q . No theory exists for this choice, as this would require knowledge of the intractable conditional expectation $E_\theta[f(Y_i)|Y_1, \dots, Y_{i-1}]$ that we wish to approximate. One approach for answering this question could be to consider the partial-autocorrelation function for $f(Y_i)$ and the functions used for predicting $f(Y_i)$ and then choose q as the cutoff point where the function dies out. In our setting this corresponds to considering the partial-autocorrelation function for the squared returns. However, inverting the covariance matrix $C(\theta)$ can cause numerical challenges and inaccuracies for large values of q .

Instead, we start at the smallest interesting choice, $q = 3$, and then later on investigate the sensitivity w.r.t the choice of q .¹² When the parameters $\theta = (\kappa, \alpha, \sigma)$ are estimated, we minimize $G_n(\theta)'G_n(\theta)$ instead of solving $G_n(\theta) = \underline{0}$. In the implementation of the GMM estimation procedure, we use continuously updated GMM, where the weight matrix is estimated simultaneously with the parameters θ . The asymptotic covariance matrix of $g_T(\theta)$ is estimated using the heteroscedasticity and autocorrelation consistent estimator from [Newey and West \(1987\)](#). Regarding the lag length in the Bartlett kernel we follow a rule-of-thumb from [Newey and West \(1987\)](#), that is $\lfloor 4(T/100)^{2/9} \rfloor$. The results on the finite sample performance of the two estimation methods in the absence of noise are summarized in Table 1.

¹² $q = 2$ would automatically result in an optimal PBEF because the weight matrix $A(\theta)$ would then be a 3×3 matrix and could be disregarded when solving $G_n(\theta) = \underline{0}$.

Table 1 Performance of estimators in absence of noise

	Relative Bias (%)				Root MSRE (%)			
	100	400	1000	4000	100	400	1000	4000
	<i>Panel A: PBEF-based estimator with $q = 3$</i>							
Scenario 1								
$\kappa = 0.03$	-6.924[0.295]	-3.386[0.173]	-1.467[0.111]	-0.279[0.044]	11.97	6.648	3.975	1.499
$\alpha = 0.25$	-1.983[1.590]	-2.150[0.893]	-0.929[0.594]	-0.256[0.311]	52.74	29.67	19.70	10.30
$\sigma = 0.10$	19.97 [0.974]	8.368[0.434]	3.486[0.260]	0.631[0.093]	37.96	16.64	9.309	3.151
Scenario 2								
$\kappa = 0.10$	-1.207[0.113]	-0.165[0.025]	-0.062[0.018]	-0.012[0.007]	3.934	0.842	0.584	0.245
$\alpha = 0.25$	0.185[0.570]	-0.021[0.293]	-0.265[0.186]	-0.096[0.096]	18.91	9.726	6.172	3.175
$\sigma = 0.10$	2.940[0.275]	0.348[0.048]	0.132[0.034]	0.026[0.015]	9.594	1.623	1.146	0.486
Scenario 3								
$\kappa = 0.10$	-4.418[0.192]	-1.436[0.111]	-0.373[0.058]	-0.129[0.025]	7.750	3.942	1.956	0.837
$\alpha = 0.25$	-2.285[1.101]	-1.244[0.567]	-0.337[0.364]	0.298[0.184]	36.59	18.83	12.06	6.104
$\sigma = 0.20$	10.92[0.527]	3.347[0.254]	0.853[0.122]	0.275[0.051]	20.61	9.065	4.148	1.719

Table 1 continued

	Relative Bias (%)				Root MSRE (%)			
	100	400	1000	4000	100	400	1000	4000
	<i>Panel B: GMM with daily realized variance from 5-min returns</i>							
Scenario 1								
$\kappa = 0.03$	180.7[10.34]	36.88[2.127]	15.26[1.110]	4.925[0.518]	386.5	79.59	39.85	17.85
$\alpha = 0.25$	226.5[47.73]	1.956[2.677]	-3.437[0.620]	-2.567[0.311]	1594	88.76	20.84	10.64
$\sigma = 0.10$	-3.781[0.892]	-0.762[0.367]	0.900[0.232]	2.053[0.127]	29.73	12.18	7.743	4.690
Scenario 2								
$\kappa = 0.10$	54.14[3.840]	13.60[1.172]	7.114[0.644]	2.370[0.291]	138.3	41.15	22.51	9.948
$\alpha = 0.25$	51.57[24.82]	-2.793[0.303]	-2.640[0.187]	-1.885[0.095]	824.0	10.43	6.747	3.668
$\sigma = 0.10$	3.609[1.036]	5.573[0.390]	6.584[0.234]	6.977[0.115]	34.52	14.07	10.16	7.946
Scenario 3								
$\kappa = 0.10$	74.58[4.084]	22.22[1.215]	10.67[0.673]	4.222[0.323]	154.5	45.99	24.73	11.51
$\alpha = 0.25$	33.65[11.96]	-7.543[0.585]	-5.095[0.364]	-2.806[0.181]	397.6	20.82	13.10	6.629
$\sigma = 0.20$	-0.839[0.643]	0.402[0.278]	1.301[0.170]	2.038[0.086]	21.31	9.234	5.796	3.504

The table reports the relative bias (with Monte Carlo std. error in brackets) and the root mean squared relative error (root MSRE) of the estimates from the sub-optimal PBEF and the RV-based GMM estimation procedure in the three different scenarios of parameter configurations. The number of Monte Carlo replications is 1000

The potential of PBEFs is clear from Panel A in Table 1. The estimators are practically unbiased for the large sample sizes, $T = 1000, 4000$, and the bias for the small sample size is of an acceptable size. For the small sample sizes, there is a small downwards bias in κ and a more pronounced upwards bias in σ . The root MSREs behave as expected, decaying with T and are roughly halved when the sample size grows from $T = 100$ to $T = 400$ and from $T = 1000$ to $T = 4000$. The mean-reversion rate, κ , is most accurately estimated in terms of root MSRE, whereas the other drift parameter, α , has the highest root MSRE. All three parameters are easiest to estimate in Scenario 2, where the volatility process is less persistent and less volatile. This could be because the other two scenarios are closer to the non-stationary region where the Feller condition is violated. In Scenario 1, with a highly persistent volatility process, the root MSREs are higher compared to the other two scenarios.

Turning our attention to Panel B of Table 1, we see that the GMM estimator from Bollerslev and Zhou (2002) performs poorly when the sample size is small. In fact, the results for $T = 100$ indicate that the method is not working for sample sizes this small. For the larger sample sizes, our results match those found in Bollerslev and Zhou (2002). The table reveals an upwards bias in κ and a smaller downwards bias in α . The volatility of volatility parameter σ has a small, yet systematic, upwards bias that actually seems to worsen when the sample size increases.¹³ The drift parameters again appear to be easiest to estimate in Scenario 2. In contrast to the results for the PBEF-based estimator, the most accurate estimates of σ are now found in Scenario 3, where the volatility of volatility is high.

If we compare Panel A and B of Table 1, it is clear that the PBEF-based method outperforms the GMM approach, especially when the sample size is small. The informational content of 100 observations of daily realized variance is too small to fully extract the dynamics of the underlying volatility process compared to 7800 observations of intra-day squared returns, and in general it seems that the PBEF-based estimator is able to exploit the extra information contained in the intra-daily returns. Furthermore, a GMM estimator based on 100 observations will, in general, often result in inaccurate estimates. The gains from using PBEFs are most prominent for the mean-reversion rate, κ , whereas the root MSREs for α are similar across the two estimation methods for the larger sample sizes.

To investigate the optimal choice of q in the PBEF, the impact on the root MSREs from increasing q is examined. In Fig. 1 the three different scenarios are considered for $T = 1000$, and the root MSREs for the three parameters are plotted against the choice of q . The shape of the plots for κ and σ looks almost identical. In Scenarios 1 and 3, where the volatility process is close to the non-stationary region, $q = 3$ seems to be the optimal choice for κ and σ . In Scenario 2, $q = 6$ appears to be the optimal

¹³ This bias in σ can, as found in Bollerslev and Zhou (2002), be explained by the variance of the discretization error $u_{t,t+1} := RV_{t,t+1} - IV_{t,t+1}$, since Barndorff-Nielsen and Shephard (2002) show that $RV_{t,t+1}^2$ is for any fixed sampling frequency, an upwards biased estimator of $IV_{t,t+1}^2$. To account for this discretization error, Bollerslev and Zhou (2002) introduce a nuisance parameter, γ , and approximate $IV_{t+1,t+2}^2$ by $RV_{t+1,t+2}^2 - \gamma$. We also implemented this simple discretization error correction and found, in line with Bollerslev and Zhou (2002), that it helps remove the systematic bias in σ , but also roughly doubles the root MSRE. We will therefore proceed without this correction in the rest of our Monte Carlo study. The results with discretization error correction are available upon request.

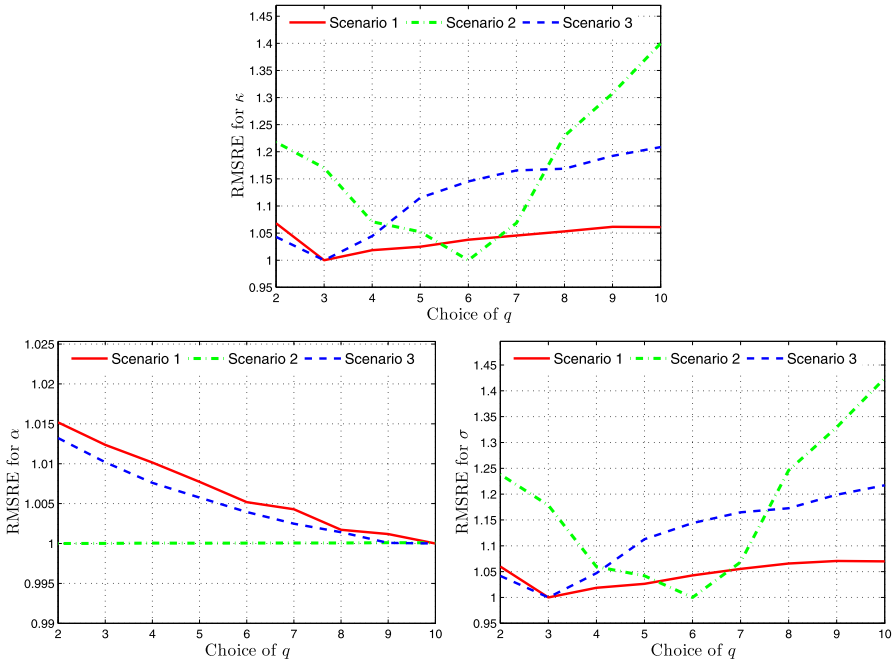


Fig. 1 Normalized root MSRE for the three parameter estimates in scenario 1–3, with $T = 1000$, plotted as a function of the number of predictor variables q . The root MSREs are in each case normalized to have a minimum of 1

choice for both parameters. Looking at the three plots for α there does not seem to be much variation in the root MSRE across the choice of q . In Scenario 1 and 3, where we are close to the non-stationary region, the root MSRE decreases when q increases. Therefore, we chose $q = 3$ in our Monte Carlo Study even though the variations in the root MSREs are small.

3.2 Including noise in the observations

In this section, the impact of MMS noise on the parameter estimates from the two estimation procedures is investigated. We consider the noise level, $\omega^2 = 0.001$, as this choice is in line with the empirical estimates found for stock returns in Hansen and Lunde (2006).¹⁴ First, we will simulate data from the Heston model with the inclusion of MMS noise and then perform parameter estimation ignoring the presence of noise. The resulting estimates will be analyzed in the next subsection. Then, in the following subsection, the finite sample performance of the noise corrected estimators will be investigated.

¹⁴ We also considered the noise level, $\omega^2 = 0.0005$, and the results are available upon request.

3.2.1 The impact of failing to correct for the presence of noise

The finite sample performances of the two estimation methods without noise correction are summarized in Table 2. Panel A of the table reports the results for the PBEF-based estimator, and Panel B reports the results for the GMM estimator based on RV.

From the results, it follows that the inclusion of MMS noise in the observed process leads to biases in the parameters. Panel A of the table shows that the downwards bias in κ has worsened in the presence of noise and the small downwards bias in α has turned into a severe upwards bias. The bias in α matches exactly the expected bias, $\frac{2\omega^2}{\Delta}$, which in relative terms equals 62.4%. The upwards bias in σ has also worsened, but α is the parameter that is most affected by ignoring the presence of noise. For all three parameters, the highest root MSREs still occur in Scenario 1, and the lowest in Scenario 2. In Panel B, the results for the performance of the GMM estimator based on RV reveal that the bias in κ has roughly doubled compared to the no noise setting, and it is approximately twice the size of the bias in κ in Panel A. As for the other drift parameter, the downwards bias in α has been turned into a severe upwards bias of a similar size as the one found in Panel A. The sign of the bias in σ has also changed, and the table now reports a downwards bias around the same size as the upwards bias in Panel A.

The root MSREs reported in Table 2 have all gone up compared to the no noise case from Table 1, and the results show that failing to correct for noise strongly impacts the parameter estimates. The long run mean of volatility, α , appears to be affected the most. As in the no noise setting of Table 1, the performance of the GMM-based estimator is quite poor for small sample sizes, and it only produces trustworthy results when $T = 400$ and we are in Scenario 2 or 3. When the larger sample sizes $T = 1000$ and $T = 4000$ are considered, the root MSREs of κ are higher in Panel B, but the impact of noise on the root MSREs of α and σ appears to be the same across the two estimation methods.

3.2.2 Finite sample performances of the noise corrected estimation procedures

The performances of the estimator based on the noise corrected PBEF and the GMM estimator where noise is corrected for parametrically are reported in Table 3. Table 4 presents the results for the GMM estimator based on RK.

The results in Panel A of Table 3 show that the noise corrected PBEF-based estimation procedure is able to correctly account for the presence of noise and produce unbiased estimates for the larger sample sizes. In fact, the small upwards bias in σ found in Table 1 has decreased significantly for the smaller sample sizes, and it disappears as the sample size grows. The variance of the noise process, ω^2 , is also accurately estimated. The root MSREs for κ are in general very similar to those reported in Table 1. For the other drift parameter, α , the root MSREs have now gone down, compared to the no noise setting. Due to the bias reduction in σ , for the smaller sample sizes $T = 100$ and $T = 400$, the root MSREs are also lower than what was reported in Table 1. For the larger sample sizes $T = 1000$ and $T = 4000$, the root MSREs for σ

Table 2 Performance of estimators in presence of noise, $\omega^2 = 0.001$

		Relative Bias (%)				Root MSRE (%)			
		100	400	1000	4000	100	400	1000	4000
<i>Panel A: PBEF-based estimator with $q = 3$</i>									
Scenario 1									
$\kappa = 0.03$	-9.624[0.436]	-8.276[0.239]	-7.992[0.148]	-7.544[0.055]	17.37	11.46	9.383	7.764	
$\alpha = 0.25$	60.40[1.589]	60.22[0.893]	61.46[0.594]	62.14[0.311]	80.14	67.11	64.53	62.99	
$\sigma = 0.10$	36.08[1.915]	22.65[0.907]	19.58[0.504]	17.29[0.147]	73.02	37.63	25.74	17.96	
Scenario 2									
$\kappa = 0.10$	-5.750[0.091]	-5.485[0.036]	-5.490[0.022]	-5.563[0.012]	6.492	5.614	5.539	5.577	
$\alpha = 0.25$	62.66[0.572]	62.40[0.295]	62.13[0.186]	62.31[0.096]	65.46	63.16	62.44	62.39	
$\sigma = 0.10$	12.88[0.249]	11.92[0.087]	11.90[0.052]	12.06[0.028]	15.30	12.27	12.03	12.09	
Scenario 3									
$\kappa = 0.10$	-8.494[0.291]	-7.842[0.132]	-7.647[0.068]	-7.717[0.031]	12.85	8.983	7.976	7.784	
$\alpha = 0.25$	60.10[1.107]	61.18[0.567]	62.12[0.364]	62.71[0.184]	70.42	64.00	63.28	63.01	
$\sigma = 0.20$	24.43[1.154]	18.38[0.422]	17.16[0.180]	17.16[0.077]	45.39	23.10	18.17	17.35	

Table 2 continued

	Relative Bias (%)			Root MSRE (%)				
	100	400	1000	4000	100	400	1000	4000
	<i>Panel B: GMM with daily realized variance from 5-min returns</i>							
Scenario 1								
$\kappa = 0.03$	251.2[28.73]	48.49[2.902]	24.77[1.479]	13.07[0.678]	982.2	107.7	54.93	26.00
$\alpha = 0.25$	466.1[55.44]	82.07[11.53]	57.64[0.615]	58.53[0.310]	1890	390.9	61.13	59.42
$\sigma = 0.10$	-21.71[1.531]	-23.15[0.589]	-21.23[0.373]	-19.48[0.180]	55.06	30.28	24.56	20.37
Scenario 2								
$\kappa = 0.10$	70.58[7.012]	14.78[1.829]	10.29[1.020]	5.459[0.459]	242.6	62.42	35.35	16.17
$\alpha = 0.25$	212.4[29.47]	59.98[0.471]	59.10[0.191]	59.67[0.096]	998.5	61.98	59.44	59.76
$\sigma = 0.10$	-5.983[1.735]	-10.01[0.728]	-7.537[0.409]	-7.136[0.192]	57.75	26.13	15.51	9.559
Scenario 3								
$\kappa = 0.10$	99.39[9.789]	35.58[1.518]	20.86[0.797]	14.58[0.395]	339.4	61.63	33.67	19.59
$\alpha = 0.25$	210.8[41.62]	54.90[0.607]	56.99[0.372]	58.11[0.190]	1396	58.47	58.31	58.45
$\sigma = 0.20$	-21.13[0.974]	-20.20[0.291]	-19.61[0.176]	-18.99[0.091]	38.58	22.38	20.46	19.23

The table reports the relative bias (with Monte Carlo std. error in brackets) and the root mean squared relative error (root MSRE) of the estimates from the sub-optimal PBEF and the RV-based GMM estimation procedure in the three different scenarios of parameter configurations. The number of Monte Carlo replications is 1000

Table 3 Performance of noise corrected estimators, $\omega^2 = 0.001$

		Relative Bias (%)				Root MSRE (%)			
		100	400	1000	4000	100	400	1000	4000
<i>Panel A: PBEF-based estimator with $q = 3$</i>									
Scenario 1									
$\kappa = 0.03$	-2.119[0.398]	-1.157[0.190]	-0.302[0.106]	-0.191[0.060]	13.35	6.388	3.535	1.986	
$\alpha = 0.25$	-0.945[1.166]	-1.326[0.657]	-0.549[0.432]	-0.130[0.225]	38.66	21.81	14.35	7.475	
$\sigma = 0.10$	8.567[0.676]	3.595[0.414]	0.985[0.221]	0.505[0.123]	24.00	14.17	7.389	4.097	
$\omega^2 = 0.001$	-1.680[0.699]	-1.359[0.386]	-0.624[0.263]	-0.200[0.139]	23.22	12.87	8.742	4.614	
Scenario 2									
$\kappa = 0.10$	0.230[0.156]	0.090[0.047]	0.019[0.030]	0.020[0.014]	5.176	1.567	0.993	0.469	
$\alpha = 0.25$	0.185[0.413]	0.016[0.213]	-0.185[0.135]	-0.060[0.069]	13.71	7.068	4.464	2.300	
$\sigma = 0.10$	0.023[0.189]	-0.114[0.085]	-0.009[0.058]	-0.033[0.029]	6.278	2.805	1.928	0.947	
$\omega^2 = 0.001$	0.121[0.261]	-0.023[0.134]	-0.130[0.084]	-0.052[0.044]	8.652	4.456	2.796	1.456	
Scenario 3									
$\kappa = 0.10$	-1.743[0.247]	-0.297[0.110]	-0.134[0.077]	-0.088[0.042]	8.356	3.649	2.554	1.403	
$\alpha = 0.25$	-1.350[0.817]	-0.742[0.416]	-0.163[0.267]	0.251[0.134]	27.13	13.79	8.865	4.454	
$\sigma = 0.20$	5.617[0.520]	0.991[0.226]	0.466[0.159]	0.235[0.086]	18.13	7.545	5.290	2.864	
$\omega^2 = 0.001$	-1.512[0.483]	-0.761[0.260]	-0.178[0.170]	0.096[0.082]	16.07	8.661	5.648	2.731	

Table 3 continued

		Relative Bias (%)				Root MSRE (%)			
		100	400	1000	4000	100	400	1000	4000
<i>Panel B: GMM with daily realized variance from 5-min returns</i>									
Scenario 1									
$\kappa = 0.03$		227.0[36.85]	37.05[2.709]	13.99[1.452]	3.166[0.683]	1226	97.15	50.13	22.88
$\alpha = 0.25$		504.5[94.12]	1.871[2.559]	-2.444[0.468]	-1.674[0.228]	3118	84.86	15.71	7.732
$\sigma = 0.10$		2.073[5.774]	-5.076[0.816]	-1.670[0.440]	0.909[0.216]	188.8	27.52	14.70	7.209
$\omega^2 = 0.001$		22.76[3.409]	5.935[1.124]	-0.732[0.300]	-1.308[0.143]	113.8	37.74	9.987	4.912
Scenario 2									
$\kappa = 0.10$		78.30[6.307]	11.78[1.773]	6.247[1.000]	1.541[0.459]	223.1	59.96	33.75	15.29
$\alpha = 0.25$		209.6[59.22]	-2.932[0.511]	-1.869[0.147]	-1.248[0.071]	1973	17.21	5.219	2.662
$\sigma = 0.10$		6.528[2.338]	3.196[0.832]	6.355[0.451]	6.846[0.214]	77.73	27.78	16.25	9.855
$\omega^2 = 0.001$		18.27[3.274]	1.651[0.392]	-1.276[0.118]	-1.736[0.050]	110.0	13.11	4.109	2.401
Scenario 3									
$\kappa = 0.10$		69.43[4.119]	17.70[1.287]	7.345[0.721]	2.263[0.373]	153.1	46.20	25.01	12.57
$\alpha = 0.25$		136.3[54.52]	-2.967[0.430]	-2.182[0.266]	-1.422[0.134]	1811	14.57	9.088	4.674
$\sigma = 0.20$		-1.539[1.072]	-0.614[0.342]	0.546[0.208]	1.784[0.109]	35.54	11.35	6.919	4.020
$\omega^2 = 0.001$		15.14[4.856]	-1.739[0.270]	-1.611[0.162]	-1.402[0.081]	161.5	9.123	5.597	3.035

The table reports the relative bias (with Monte Carlo std. error in brackets) and the root mean squared relative error (root MSRE) of the estimates from the noise corrected versions of the sub-optimal PBEF and the RV-based GMM estimation procedure in the three different scenarios of parameter configurations. The number of Monte Carlo replications is 1000

Table 4 Performance of the GMM estimator based on RK, $\omega^2 = 0.001$

	Relative Bias (%)			Root MSRE (%)				
	100	400	1000	4000	100	400	1000	4000
Scenario 1								
$\kappa = 0.03$	282.0[26.76]	36.34[3.053]	13.81[1.716]	3.484[0.834]	928.8	107.5	58.55	27.85
$\alpha = 0.25$	398.7[62.41]	13.14[5.635]	-4.466[0.719]	-4.056[0.308]	2103	187.3	24.25	10.98
$\sigma = 0.10$	15.83[2.544]	3.289[1.138]	6.062[0.802]	9.292[0.423]	85.63	37.85	27.26	16.83
Scenario 2								
$\kappa = 0.10$	179.6[30.65]	12.80[2.324]	5.320[1.375]	1.527[0.655]	1031	78.10	45.88	21.75
$\alpha = 0.25$	206.5[32.43]	4.226[2.124]	-3.348[0.407]	-3.674[0.095]	1094	70.54	13.89	4.840
$\sigma = 0.10$	50.92[6.523]	23.22[1.409]	26.45[0.899]	28.85[0.425]	222.1	52.15	39.85	32.10
Scenario 3								
$\kappa = 0.10$	75.02[4.856]	19.33[1.403]	8.028[0.779]	2.546[0.389]	177.5	50.37	27.03	13.14
$\alpha = 0.25$	166.0[35.48]	-7.895[0.584]	-5.964[0.356]	-4.141[0.179]	1187	20.90	13.23	7.238
$\sigma = 0.20$	2.936[1.112]	6.264[0.499]	7.654[0.327]	9.132[0.178]	36.95	17.69	13.26	10.87

The table reports the relative bias (with Monte Carlo std. error in brackets) and the root mean squared relative error (root MSRE) of the estimates from the RK-based GMM estimation procedure in the three different scenarios of parameter configurations. The number of Monte Carlo replications is 1000

are a bit higher or similar to those found in Table 1. For all three parameters, the root MSREs are smallest in Scenario 2, as was also the case in the no noise setting.

The performance of the parametrically noise corrected GMM estimator is summarized in Panel B of Table 3. An inspection of the results shows that although the GMM estimator does not produce unbiased estimates it succeeds at incorporating the noise and produce estimates with small biases comparable to those from Table 1. In fact, the biases are now a bit lower than in Table 1, with the difference being most apparent in Scenario 3. The results for the smallest sample size, $T = 100$, appear unreliable, in the sense that 100 observations are simply not enough to infer the dynamics of the underlying volatility process and produce estimates with an acceptable level of bias and root MSREs. Comparing the root MSREs to those reported in Table 1, we find that the root MSREs are in general similar to the no noise setting, with the root MSREs for κ and σ being a bit bigger than in Table 1 and the root MSREs for α a bit smaller. The patterns, found in Panel B of Table 1, across the different scenarios are repeated, the drift parameters have lower root MSREs in Scenario 2, whereas σ is most accurately estimated in Scenario 3.

When comparing Panel A and B of Table 3, it is clear that the PBEF-based estimation method produces more accurate parameter estimates. The root MSREs of α are quite similar for the two methods, but the root MSREs are lower in Panel A, for the other two parameters and the noise variance. The difference between the two methods is most prominent for the mean-reversion parameter, which is extremely well estimated with the PBEF-based method.

We also considered estimating the Heston model using noisy data, by simply replacing RV with a realized kernel, RK, in the original moment conditions from [Bollerslev and Zhou \(2002\)](#). The finite sample performance of this estimator is summarized in Table 4. Even though the estimator is based on six moment conditions, compared to the four moment conditions used for constructing the parametrically noise corrected estimator, the performance is not nearly as good. The biases in the parameters are larger, but have the same signs as in Panel B of Table 3. The small systematic bias in σ due to the discretization error and noisy data has also increased compared to Table 1. This could be explained by the slower rate of convergence of RK to IV, compared to the convergence rate of RV. The biases in the drift parameters are however comparable to those reported in the no noise setting, and for κ the bias is in fact a bit lower. The two ways of correcting for noise in the GMM approach give rise to somewhat similar results for the mean-reversion rate κ , but the root MSREs are lower for α and σ when the noise corrected estimator based on RV is employed. The investigation of robustness towards misspecification of the noise process is outside the scope of this paper. In our setting, without model misspecification, the parametric approach of correcting for the noise outperforms the non-parametric approach.

In conclusion, the PBEF-based estimation method produced promising results, and it appears that the method is able to exploit the extra information contained in directly using the dynamics of the intra-day returns, instead of aggregating them into realized measures. What remains to be investigated is how the two estimation methods perform empirically.

4 Empirical application

In this section, we use actual 5-min returns as input in the two estimation methods analyzed in our Monte Carlo study. We are aware that the Heston model might not fit the chosen data. As such our application should be seen as a check of what happens when the estimation methods are used to fit a (possibly misspecified) model to real data.

4.1 Data description

For our empirical illustration we use 5-min returns for SPDR S&P 500 (SPY). SPY is an exchange traded fund (ETF) that tracks the S&P 500. The sample covers the period from January 4, 2010 through December 31, 2013. We sample the first price each day at 9:35 and then every 5 minute until the close at 16:00. Thus, we have 77 daily 5-min returns for each of the 1006 trading days in our sample, yielding a total of 77,462 5-min returns. We cleaned the data following the step-by-step cleaning procedure used in [Barndorff-Nielsen et al. \(2008b\)](#).

As a first inspection of the data characteristics, we consider the empirical autocorrelation functions for the squared 5-min returns, reported in the top panel of Fig. 2. The autocorrelation function does not seem to be exponentially decaying, revealing that the Heston model will not be able to properly account for the dynamics of the data. However, our main interest lies in investigating whether the two estimation methods will yield similar parameter estimates or whether they perform differently. The autocorrelation function also exhibits cyclical patterns corresponding to a lag length of one trading day. This is due to the well-documented intra-day periodicity in volatility in foreign exchange and equity markets, see for instance [Andersen and Bollerslev \(1997\)](#) or [Dacorogna et al. \(1993\)](#).

The intra-daily periodicity in the volatility might cause the two estimation methods to perform differently. The intra-daily periodicity should not affect the GMM-based estimator much, since the intra-daily pattern in volatility will be “smoothed out” when the 5-min returns are aggregated into the daily realized measures. The same does not apply for the estimator based on PBEFs, since this estimator is based directly on the squared 5-min returns. Hence, the intra-daily periodicity might effect the parameter estimates when the PBEF-based estimation method is carried out. In order to avoid this, the intra-daily volatility pattern is captured by fitting a spline function to the intra-daily averages of 5-min returns using a non-parametric kernel regression.

The data are then adjusted for periodicity in intra-daily volatility by dividing the squared returns through by the fitted values from the spline function, matched according to the intra-daily 5-min interval in which the observation falls. Finally, the squared returns are normalized such that the overall variance of the squared returns remains unchanged. The bottom panel of Fig. 2 displays the autocorrelation function for the adjusted data. From the figure, it is clear that the intra-daily periodicity has been removed. It is however also evident that the autocorrelation function is not exponentially decaying, rendering the Heston model a poor model choice. There seems to be a need for at least a two factor SV model, in order to properly capture the dynamics of

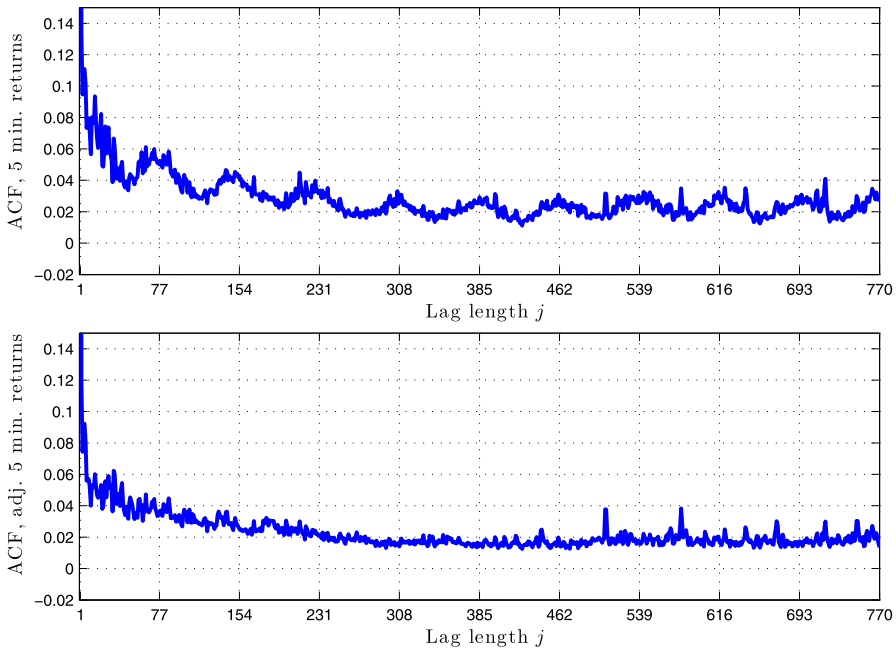


Fig. 2 Autocorrelation function for the squared 5-min returns (*top*) and the adjusted squared 5-min returns (*bottom*) on SPY

the autocorrelation function. One factor is needed for capturing the fast decay in the autocorrelation function at the short end, whereas the other factor should be slowly mean reverting and thereby account for the persistent or long memory-like factor in the variance.

4.2 Estimation results

In the Heston model, the decay rate of the autocorrelation function for the squared returns is uniquely governed by the mean-reversion parameter κ . Due to the dynamic structure of the autocorrelation function, discussed above, the choice of predictor space might heavily influence the estimated value of κ . Depending on the largest time lag of past squared returns included in the predictor space, different dynamics might be captured. When fitting the Heston model to the adjusted data, we hold the dimension of the predictor space fixed at 4, ($q = 3$), but consider four different choices of basis elements, spanning the predictor space. The four cases correspond to having 1 h, 1, 2, and 4 days between each of the basis elements.¹⁵ The time lag between the basis elements will be denoted by the l variable. The simple choice of weight matrix used

¹⁵ That is, in the first case we let the predictor space be spanned by $Y_{i-1}^2, Y_{i-13}^2, Y_{i-25}^2$ and a constant. In the second case we choose $Y_{i-1}^2, Y_{i-78}, Y_{i-155}$ and a constant as the basis elements of the predictor space and so on.

in the Monte Carlo study is also employed when constructing the PBEFs. We will only consider the noise corrected estimators, as MMS noise is a stylized feature of the present data. The variance of the noise process is not estimated in the GMM approach based on RK, so instead we report the non-parametric estimate of the daily MMS noise variance, using the formula favored in [Barndorff-Nielsen et al. \(2008a\)](#),

$$\check{\omega}^2 = \exp[\log(\hat{\omega}^2) - \text{RK}/\text{RV}],$$

with RK and RV constructed using the intra-daily adjusted 5-min returns and where $\hat{\omega}^2 = \text{RV}/2m$. In our case $m = 77$, and the overall estimate of the MMS noise variance is found by averaging the 1006 daily estimates. For the SPY data we find $\check{\omega}_{\text{avg}}^2 = 0.0015629$. The realized kernel is now computed using the parzen kernel with $H \propto (1/\Delta)^{3/5}$ as recommended in [Barndorff-Nielsen et al. \(2008b\)](#) for empirical applications. The obtained convergence rate of RK to IV is now $(1/\Delta)^{1/5}$. The choice of bandwidth H , that resulted in the convergence rate of $(1/\Delta)^{1/4}$ in our Monte Carlo study, relies heavily on the assumption of i.i.d. MMS noise which might not hold in practice. The results from fitting the Heston model to the data, using the various estimators, are reported in [Table 5](#). Computation of asymptotic standard errors for the PBEF-based estimator is challenging. It involves computation of the matrix $\bar{M}_n(\theta)$, that also enters the expression for the optimal weight matrix $A^*(\theta)$, and was in fact the main reason why we focused on the sub-optimal PBEF in our Monte Carlo study. Therefore, we resort to bootstrap methods for computing standard errors. The standard errors and 95% confidence intervals (CI) reported in [Table 5](#) are computed using the moving block bootstrap, as recommended in [Lahiri \(1999\)](#). The confidence intervals are equal tail intervals, constructed using the percentile method. As for the block length in the bootstrap method, then $T^{1/3} (\approx 10)$ days is the rule-of-thumb advocated in [Hall et al. \(1995\)](#) for constructing standard errors. Due to the strong persistence in the data, we choose to be conservative and use a block length of 20 days.

[Table 5](#) also reports the fit to the moments of the adjusted squared returns implied by the parameter estimates from using the various estimation methods. The obtained parameter estimates vary across the different estimation methods, but all are meaningful and within the same range. From [Table 5](#), we see that the two different ways of noise correcting the GMM estimator impact the parameter estimates, especially σ . The estimated noise variance in the parametrically noise corrected method is only about half of the non-parametric estimate $\check{\omega}_{\text{avg}}^2$. It also becomes evident that the choice of the predictor space, represented by the variable l , highly impacts the parameter estimates. When l is low, κ is high, and more emphasis is put, by the PBEF estimator, on capturing the fast decaying part in the short end of the autocorrelation function. As the l variable increases, κ drastically drops and reveals the need for several volatility factors in order to fully capture the dynamics of the data. The split between how much of the mean of the squared adjusted returns is due to the long run mean of the volatility process, α , and how much is due to the variance of MMS noise, ω^2 , also varies with the l variable. An exception is when $l = 77$, where the estimated noise variance is similar to the estimate obtained with the parametrically noise corrected GMM method, otherwise the noise variance is either severely underestimated or potentially overestimated.

Table 5 Estimation results for the estimation methods with noise correction

Estimation method	κ	α	σ	FC	$100 \times \omega^2$	$100 \times E[Y_t^2]$	$100 \times V(Y_t^2)$	$100 \times ACF_1(Y_t^2)$
Sample moments								
GMM with RK	0.0776 (0.0561) [0.0469;0.2770]	0.500 (0.111) [0.347;0.692]	0.485 (0.120) [0.299;0.745]	0.157 (0.115) [-0.012;0.386]	0.1563 (0.0250) [0.1119;0.2103]	0.684	0.0442	0.240
GMM noiseCorr.	0.1013 (0.0528) [0.0317;0.2586]	0.426 (0.186) [0.269;0.661]	0.775 (0.223) [0.456;1.284]	0.514 (0.356) [0.127;1.460]	0.0880 (0.0094) [0.0768;0.1059]	0.866	0.0788	0.276
PBEF 1 step								
$q = 3, l = 12$	0.3874 (0.2085) [0.1036;0.9379]	0.403 (0.140) [0.144;0.641]	1.017 (0.520) [0.413;2.319]	0.721 (1.897) [-0.075;5.160]	0.0803 (0.0745) [0.0004;0.2450]	0.836	0.0412	0.231
$q = 3, l = 77$	0.2593 (0.0532) [0.2300;0.4461]	0.514 (0.078) [0.375;0.680]	0.786 (0.350) [0.437;1.642]	0.351 (0.908) [-0.075;2.313]	0.0085 (0.0111) [0.0004;0.0086]	0.981	0.0502	0.215
$q = 3, l = 154$	0.0956 (0.0506) [0.0376;0.2474]	0.128 (0.059) [0.034;0.268]	0.630 (0.617) [0.195;2.376]	0.373 (2.587) [0.020;5.635]	0.2598 (0.0631) [0.1259;0.3741]	0.479	0.0180	0.275
$q = 3, l = 308$	0.0677 (0.0330) [0.0539;0.1807]	0.139 (0.132) [0.053;0.547]	0.535 (0.405) [0.251;1.684]	0.267 (1.185) [0.021;2.824]	0.2534 (0.0899) [0.0087;0.3493]	0.493	0.0197	0.276
PBEF 2 step								
$q = 3, l = 12$	0.2356 (0.2232) [0.0085;0.8752]	0.286 (0.043) [0.214;0.379]	0.876 (0.632) [0.118;2.372]	0.632 (2.032) [0.007;5.301]	0.1563 (0.0250) [0.1119;0.2103]	0.684	0.0329	0.253
$q = 3, l = 77$	0.2270 (0.0498) [0.1667;0.3747]	0.287 (0.043) [0.214;0.379]	0.894 (0.466) [0.411;2.008]	0.668 (1.491) [0.064;3.853]	0.1563 (0.0250) [0.1119;0.2103]	0.685	0.0349	0.257
$q = 3, l = 154$	0.1585 (0.0386) [0.1255;0.2813]	0.287 (0.043) [0.213;0.378]	0.709 (0.376) [0.325;1.669]	0.412 (0.966) [0.035;2.658]	0.1563 (0.0251) [0.1115;0.2103]	0.686	0.0325	0.251
$q = 3, l = 308$	0.0963 (0.0253) [0.0892;0.1912]	0.289 (0.043) [0.213;0.379]	0.569 (0.301) [0.295;1.355]	0.269 (0.561) [0.034;1.746]	0.1563 (0.0250) [0.1119;0.2103]	0.687	0.0340	0.255

The table reports the parameter estimates from fitting the Heston model to the SPY data. The variable FC denotes the Feller condition, $\sigma^2 - 2k\alpha$, and it is positive if the parameter constraint is violated. The table also reports sample moments as well as theoretical moments implied by the various obtained parameter estimates. The std. errors and 95% equal tail CIs are computed using the moving block bootstrap with a block length of 20 days and $B = 999$

This could be a consequence of the i.i.d. Gaussian noise assumption, which might not hold. It could also be that the PBEF-based method has problems identifying ω^2 and α in the data at hand. We therefore also consider fixing the noise variance at the estimate $\hat{\omega}_{\text{avg}}^2$ and only estimate the three parameters from the Heston model. This procedure is denoted by PBEF 2 step in the Table. The results from the PBEF 2 step method reveal that fixing the noise variance mainly affects the estimate of α , which is now constant across the choice of predictor space. The PBEF 2 step method also appears more stable, with the relative standard errors¹⁶ being constant across the values of the time lag in the predictor space, except for the intra-day time lag ($l = 12$). Checking for parameter stability across the different predictor spaces employed could serve as a general robustness check that might reveal model misspecification.

The last three columns of Table 5 report the model-implied fit to the sample moments of the data. The mean of the adjusted squared returns is extremely well fitted by the PBEF 2 step procedure, whereas the other methods do not give as good fits. The variance is however reasonably well matched when the lower values of l are used in the PBEF 1 step procedure, but is poorly matched when l is high. The PBEF 2 step procedure and the GMM estimator based on RK also produce reasonable fits to the variance. The first-order autocorrelation of the squared returns is best matched by the GMM method based on RV and the PBEF 1 step method with $l = 12$, with the PBEF 2 step approach being the runner-up. Not surprisingly, the overall best fit is obtained by the PBEF-based method, and the fit appears more stable when the PBEF 2 step approach is used.

From Table 5 we also observe that the Feller condition is violated in all the estimation procedures, indicating that the Heston model provides a poor fit to the data.¹⁷ However, this has no influence on the specific aim of this section which was to investigate how the two different estimation methods handle real data with possible model misspecification. The problem seems to be that the dynamic structure implied by the Heston model is not flexible enough to adequately model the observed dynamics. The need for allowing for several volatility factors is best highlighted by the PBEF-based estimation method. The flexibility of the PBEF-based method can in general serve as a robustness check of the specified model, including the noise specification.

5 Conclusion and final remarks

The general theory underlying PBEFs was reviewed and detailed. We explicitly constructed PBEFs for parameter estimation in the Heston model with and without the inclusion of noise in the data. The link between optimal GMM estimation and the optimal PBEF was also derived. As a benchmark, we considered the GMM estimator from [Bollerslev and Zhou \(2002\)](#), and we extended the method to handle noisy data in a parametric and non-parametric way. The finite sample performance of the estimator based on PBEFs was investigated in a Monte Carlo study and compared to that of

¹⁶ The standard errors normalized by the parameter estimate.

¹⁷ 0 is actually contained in the CI's for the FC variable when the GMM method based on RK and the PBEF 1 step method with $l = 12$ and $l = 77$ are employed, but only barely.

the GMM estimator. In the no MMS noise setting, there are gains to be made from using PBEFs, both in terms of bias and root MSRE, especially when the sample size is small. The gain from using the PBEF-based method was most prominent for the mean-reversion parameter, κ , that was extremely well estimated. The PBEF-based method produced promising results in all three parameter configurations, but the root MSREs were lower when the volatility process was less persistent and less volatile.

Including MMS noise in the observation equation, but neglecting to correct for it, produced biased estimates with the upwards bias in the long run average variance, α , being most severe.

We then considered the performance of the noise corrected estimation methods. The PBEF-based estimator and the parametrically noise corrected GMM estimator produced results similar to those found in the no MMS noise setting. The non-parametric approach, where the GMM estimator is based on RK, did not perform as well. Hence, the GMM estimator could not compete with the results obtained using the noise corrected PBEF.

In our empirical application, by fitting the Heston model to SPY data, we investigated how the two different approaches handle real data. The empirical application revealed that the choice of estimation approach impacts the parameter estimates. The study also made it clear, how the great flexibility of the PBEF-based estimation method could serve as a way of conducting robustness checks, for instance by checking for parameter stability across different time spans of the predictor space.

It would be of interest to see how the estimation method based on PBEFs performs if we extend the Monte Carlo setup by leaving the assumption of i.i.d. noise. This would however complicate the construction of the MMS noise corrected PBEF and the recalculation of the moments used for constructing the GMM estimator. A solution to this potential problem could be to filter out the noise in a first step using the method of pre-averaging introduced by [Jacod et al. \(2009\)](#), instead of modeling the noise directly. The performance of this approach is still to be investigated. Since the PBEF-based estimation method is quite general, an important contribution to the existing literature would be to consider PBEFs in a setting where the driving sources of randomness are general Lévy processes, like the models considered in [Brockwell \(2001\)](#), [Barndorff-Nielsen and Shephard \(2001\)](#), and in [Todorov and Tauchen \(2006\)](#). Finally, quantifying the gain from using the optimal PBEF in different settings, as well as how to simulate or approximate the optimal weight matrix, would also be a topic for future research.

Acknowledgments The authors acknowledge financial support by the Center for Research in Econometric Analysis of Time Series, CREATES, funded by the Danish National Research Foundation.

References

- Andersen, T.G., Bollerslev, T.: Intraday periodicity and volatility persistence in financial markets. *J. Empir. Financ.* **4**, 115–158 (1997)
- Andersen, T.G., Davis, R., Kreiss, J.-P., Mikosch, T.: *Handbook of Financial Time Series*. Springer, Berlin (2009)
- Ant-Sahalia, Y., Kimmel, R.: Maximum likelihood estimation of stochastic volatility models. *J. Financ. Econ.* **83**, 413–452 (2007)

- Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N.: Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica* **76**, 1481–1536 (2008a)
- Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N.: Realised kernels in practice: trades and quotes. *Econom. J.* **04**, 1–32 (2008b)
- Barndorff-Nielsen, O.E., Shephard, N.: Non-Gaussian OU-based models and some of their uses in financial economics (with discussion). *J. R. Stat. Soc. B* **63**, 167–241 (2001)
- Barndorff-Nielsen, O.E., Shephard, N.: Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J. R. Stat. Soc. B* **64**, 253–280 (2002)
- Bollerslev, T., Zhou, H.: Estimating stochastic volatility diffusions using conditional moments of integrated volatility. *J. Econom.* **109**, 33–65 (2002)
- Bradley, R.C.: Basic properties of strong mixing conditions: a survey and some open questions. *Probab. Surv.* **2**, 107–144 (2005)
- Brockwell, P.J.: LTVy-driven CARMA processes. *Ann. Inst. Statist. Math.* **53**, 113–124 (2001)
- Corradi, V., Distaso, W.: Semi-parametric comparison of stochastic volatility models using realized measures. *Rev. Econ. Stud.* **73**, 635–667 (2006)
- Dacorogna, M., Mnller, U., Nagler, R., Olsen, R., Pictet, O.: A geographical model for the daily and weekly seasonal volatility in the foreign exchange market. *J. Int. Money Financ.* **12**, 413–438 (1993)
- Eraker, B.: Markov Chain Monte Carlo analysis of diffusion models with application to finance. *J. Bus. Econ. Stat.* **19–2**, 177–191 (2001)
- Gallant, A.R., Tauchen, G.: Which moments to match? *Econom. Theory* **12**, 657–681 (1996)
- Gourieroux, C., Monfort, A., Renault, E.: Indirect inference. *J. Appl. Econom.* **8**, S85–S118 (1993)
- Hall, P., Horowitz, J., Jing, B.: On blocking rules for the bootstrap with dependent data. *Biometrika* **82**, 561–574 (1995)
- Hansen, P.R., Lunde, A.: Realized variance and market microstructure noise. *J. Bus. Econ. Stat.* **24**, 127–218 (2006)
- Heston, S.L.: A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Rev. Financ. Stud.* **6**, 327–343 (1993)
- Jacod, J., Li, Y., Mykland, P., Podolskij, M., Vetter, M.: Microstructure noise in the continuous case: the pre-averaging approach. *Stoch. Process. Appl.* **119**, 2249–2276 (2009)
- Karlin, S., Taylor, H.M.: *A First Course in Stochastic Processes*. Academic Press, New York (1975)
- Lahiri, S.N.: Theoretical comparison of block bootstrap methods. *Ann. Stat.* **27**, 386–404 (1999)
- Newey, W.K., West, K.D.: A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55**, 703–708 (1987)
- Nolsøe, K., Nielsen, J.N., Madsen, H.: Prediction-based estimating functions for diffusion processes with measurement noise. In: Technical Reports no. 10, Informatics and mathematical modelling. Technical University of Denmark (2000)
- Sørensen, M.: Prediction-based estimating functions. *Econom. J.* **3**, 123–147 (2000)
- Sørensen, M.: Prediction-based estimating functions: review and new developments. *Braz. J. Probab. Stat.* **25**, 362–391 (2011)
- Todorov, V.: Estimation of continuous-time stochastic volatility models with jumps using high-frequency data. *J. Econom.* **148**, 131–148 (2009)
- Todorov, V., Tauchen, G.: Simulation methods for LTVy-driven CARMA stochastic volatility models. *J. Bus. Econ. Stat.* **24**, 455–469 (2006)