Estimating Stochastic Volatility Models using Prediction-based Estimating Functions

Asger Lunde and Anne Floor Brix

Estimating Stochastic Volatility Models using Prediction-based Estimating Functions[*]

Anne Floor Brix and **Asger Lunde**

Aarhus University, Department of Economics and Business,

Fuglesangs Allé 4, Aarhus V, Denmark & CREATES

July 5, 2013

Abstract

In this paper prediction-based estimating functions (PBEFs), introduced in Sørensen (2000), are reviewed and PBEFs for the Heston (1993) stochastic volatility model are derived. The finite sample performance of the PBEF based estimator is investigated in a Monte Carlo study, and compared to the performance of the GMM estimator based on conditional moments of integrated volatility from Bollerslev and Zhou (2002). The case where the observed log-price process is contaminated by i.i.d. market microstructure (MMS) noise is also investigated. First, the impact of MMS noise on the parameter estimates from the two estimation methods without noise correction are studied. Second, a noise robust GMM estimator is constructed by approximating integrated volatility by a realized kernel instead of realized variance. The PBEFs are also recalculated in the noise setting, and the two estimation methods ability to correctly account for the noise are investigated. Our Monte Carlo study shows that the estimator based on PBEFs outperforms the GMM estimator, both in the setting with and without MMS noise. Finally, an empirical application investigates the possible challenges and general performance of applying the PBEF based estimator in practice.

Keywords: GMM estimation, Heston model, high-frequency data, integrated volatility, market microstructure noise, prediction-based estimating functions, realized variance, realized kernel.

JEL Classification: C13, C22, C51.

1. **Introduction**

Continuous time stochastic volatility (SV) models are widely used in econometric and empirical finance for modeling prices of financial assets. Considerable efforts have been put into modeling and estimation of the latent volatility process, and 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 unobservable. This hidden Markov structure complicates inference, since the observed log-price process will not in it self 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. Likelihood inference is also not straightforward because an explicit expression for the transition density of the state vector is almost never available.

We will circumvent the above mentioned problems for conducting inference in SV models by using prediction-based estimating functions (PBEFs), introduced in Sørensen (2000), which are a generalization of martingale estimating functions. This generalization becomes especially useful when applied to observations from a non-Markovian model, such as a log-price process with stochastic volatility. PBEFs are estimating functions based on predictors of functions of the observed process. The structure of PBEFs are essentially a sum of weighted augmented prediction errors, and an estimator is found by making this sum zero.

In this paper, PBEFs will be reviewed, detailed and used for parameter estimation in the Heston model. The estimation method is easy to implement and fast to execute since the computation of PBEFs only rely on unconditional moments. When the Heston SV-model is considered no simulations are needed for constructing the relevant PBEFs.

As an alternative we also will consider the method suggested in Bollerslev and Zhou (2002). They derive a GMM-type estimation approach based on first and second order conditional moments of the integrated volatility (IV). Since IV is latent, the 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.

In this paper, we investigate the finite sample properties of the PBEF based estimator in a Monte Carlo study, and compare the performance to that of the GMM estimator from Bollerslev and Zhou (2002). When constructing the GMM estimator from Bollerslev and Zhou (2002), the intra-daily squared returns are transformed into daily realized measures, whereas in the

method using PBEFs the squared returns are used directly in the estimation procedure. We investigate whether the extra information contained in the PBEF based estimation procedure results in more accurate parameter estimates. The case when the efficient price is assumed to be directly observable, as well as the case when the price is observed with measurement error are considered. In particular, we contribute by extending the two competing methods to handle the latter case. To the best of our knowledge this is the first time the finite sample performance of PBEFs applied to SV-models are studied. In fact, this is in general 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 when a CIR process is observed with additive white noise, but the performance of PBEF estimation of SV-models has not previously been studied.

The paper also addresses implementation issues that arises when PBEFs are used in practice, and the link between the estimation method based on PBEFs and other well-known estimation methods such as GMM and linear regression are discussed. Especially, the connection between the optimal PBEF and the optimal choice of the weight matrix in GMM estimation is established. Lastly, an empirical application to the spot DM/$ exchange rate 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.

The paper is organized as follows: In the next section the PBEF estimation method is review 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 are derived. In section 3 we present our Monte Carlo study, which includes an investigation of how i.i.d. measurement errors impact the performances of the two methods, and shows how the methods can be extended to handle these errors. The fourth section contains an empirical application to the DM/$ spot exchange rate, that reveal the consequences of different choices made when constructing the PBEFs. The final section concludes and ideas on further research are outlined.

2. Estimating Stochastic Volatility Models

In this section the two estimation methods from Sørensen (2000) and Bollerslev and Zhou (2002) are reviewed. We detail the structure of the PBEFs and discuss the link between the optimal PBEF and a GMM estimator with the optimal choice of weight matrix.

Both estimation methods are fairly easy to implement and do not rely on simulation schemes.[1] When high-frequency data is 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), are that they are all based on time-series of daily realized measures, such as realized variance (RV) and bipower variation (BV). Instead of being transformed into daily realized measures, the squared intra-daily returns are used directly when constructing PBEFs. This means PBEFs have a potential informational advantage, the strength of which will be investigated throughout the paper.[2]

The focus of this paper is on how to use the two considered estimation methods for estimating SV models of the form

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

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

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

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 α. 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, which were introduced in Sørensen (2000) and further developed in Sørensen (2011b).

---

[1]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.

[2]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 Aït-Sahalia and Kimmel (2007).

Then, following Sørensen (2000), we derive the PBEFs for the Heston model which will be used in our Monte Carlo Study.

2.1.1. *The General Setup and Ideas*

The estimation method based on PBEFs are used for conducting parametric inference on observations $Y_1, Y_2, \ldots, Y_n$ 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 $Y_1, Y_2, \ldots, Y_n$ and θ. An estimator is then obtained by solving the p equations $G_n(\theta) = \underline{0}$ w.r.t. θ. When considering PBEFs, the task of solving $G_n(\theta) = \underline{0}$ is essentially the same as choosing θ to eliminate a sum of weighted augmented prediction errors.

Let $\mathcal{F}_i$ denote the σ-algebra generated by the observations $Y_1, Y_2, \ldots, Y_i$. When θ is the true parameter, we denote by $\mathcal{H}_i^\theta$ the $\mathcal{L}^2$-space of all square integrable, $\mathcal{F}_i$-measurable, 1-dimensional random variables. $\mathcal{H}_i^\theta$ is a Hilbert space of real-valued functions of the type $h(Y_1, \ldots, Y_i)$, with inner product given by $\langle h_1, h_2 \rangle = E_\theta[h_1(Y_1, \ldots, Y_i)h_2(Y_1, \ldots, Y_i)]$. For each i a closed linear subspace $\mathcal{P}_{i-1}^\theta$ of $\mathcal{H}_{i-1}^\theta$ can be chosen as the predictor space for predicting $f(Y_i)$, where f is some known 1-dimensional function[3]. The predictor space $\mathcal{P}_{i-1}^\theta$ can be chosen freely, but is often chosen to be finite dimensional in order to obtain tractable estimating functions.

In the setup of PBEFs, we study estimating functions of the form

$$G_n(\theta) = \sum_{i=1}^n \underbrace{\Pi^{(i-1)}(\theta)}_{p \times 1} \underbrace{[f(Y_i) - \underbrace{\hat{\pi}^{(i-1)}(\theta)}_{\in \mathcal{P}_{i-1}^\theta}]}_{1 \times 1}, \tag{3}$$

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 $E_\theta[f(Y_i)^2] < \infty$ for all $\theta \in \Theta$ and for $i = 1, \ldots, n$. Note that $E_\theta[\cdot]$ denotes expectation under the model with parameter θ. In (3), the p-dimensional stochastic vector $\Pi^{(i-1)}(\theta) = (\pi_1^{(i-1)}(\theta), \ldots, \pi_p^{(i-1)}(\theta))$ has elements that belong to the predictor space $\mathcal{P}_{i-1}^\theta$, and $\hat{\pi}^{(i-1)}(\theta)$ is the minimum mean square error (MMSE) predictor of $f(Y_i)$ in $\mathcal{P}_{i-1}^\theta$. That is, $\hat{\pi}^{(i-1)}(\theta)$ is the orthogonal projection of $f(Y_i)$ onto $\mathcal{P}_{i-1}^\theta$ w.r.t. the inner product in $\mathcal{H}_i^\theta$ defined above. Since the predictor space is both closed and linear, this

---

[3]One could also choose to predict functions of the type $f(Y_i, \ldots, Y_{i-s})$, see Sørensen (2011b), 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, $f_j(Y_i)$ $j = 1, \ldots, N$, are predicted, see Sørensen (2000) and Sørensen (2011b). This generalization will, however, not be necessary for estimating the SV-models we are considering.

orthogonal projection exists and is uniquely determined by the normal equations

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

see e.g. Thm. 3.1 in Karlin and Taylor (1975).[4] Note that (4) is just an orthogonality condition, stating that the prediction error $\big[f(Y_i) - \hat{\pi}^{(i-1)}(\theta)\big]$ is orthogonal to any element, π, from the predictor space $\mathcal{P}^\theta_{i-1}$. Using the normal equations (4) it easily follows that our PBEF, $G_n(\theta)$, is an unbiased estimating function, i.e. that $E_\theta\big[G_n(\theta)\big] = 0$ for all $\theta \in \Theta$. When we consider finite dimensional predictor spaces regularity conditions, such as stationarity and geometric α-mixing of the observed process, can be proved to ensure $\sqrt{n}$-consistency and asymptotic normality of the resulting estimator, (see Thm. 4.3 in Sørensen (2011b)).

A special class of PBEFs is the class of martingale estimating functions, which is obtained by choosing the predictor space to be the space of all square integrable $\mathcal{F}_{i-1}$-measurable functions, that is, by letting $\mathcal{P}^\theta_{i-1} = \mathcal{H}^\theta_{i-1}$. In this case, the MMSE predictor of $f(Y_i)$ in $\mathcal{P}^\theta_{i-1}$ is the conditional expectation

$$\hat{\pi}^{(i-1)}(\theta) = E_\theta\big[f(Y_i)|Y_1,\ldots,Y_{i-1}\big],$$

and $G_n(\theta)$ becomes a P_θ-martingale w.r.t. the filtration generated by the data process. Martingale estimating functions 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. Because continuous time SV-models are Non-Markovian, the idea is to use a smaller and tractable predictor space in place of $\mathcal{H}^\theta_{i-1}$. In fact, we will restrict our attention to finite dimensional predictor spaces and think of the resulting PBEF as an approximation to a martingale estimating function. The MMSE predictor obtained by projecting on a finite dimensional space can similarly be thought of as a linear approximation to the conditional expectation, needed to construct the martingale estimating function. The advantage of considering this linear approximation is that it only requires computing unconditional moments which, from a simulation point of view, are much easier to compute than conditional moments. Regarding efficiency issues, the conventional wisdom is that estimators based on conditional moments are more efficient than those based on unconditional moments. Likelihood inference is often complicated or infeasible when we consider non-Markovian models, such as SV-models, but since the score function is a martingale, we can obtain a close approximation to the score function by using martingale estimating functions. As mentioned, martingale estimating functions are constructed using conditional moments, but by suitably choosing the predictor space the MMSE predictors will be good approximations to the conditional moments, and the resulting PBEF will also be a good approximation

---

[4]Unique in the sense of mean square distance.

to the score function. This gives reason to believe that the estimators based on PBEFs can also obtain high efficiency.

In the rest of the paper we will restrict our attention to finite dimensional predictor spaces $\mathcal{P}_{i-1}^{\theta}$ and assume the observed process $\{Y_i\}$ to be stationary. In order to obtain even more tractable PBEFs, we will only consider predictor spaces with basis elements of the form $Z_k^{(i-1)} = h_k(Y_{i-1}, \ldots, Y_{i-s})$, $k = 0, \ldots, q$, where $h_k : \mathbb{R}^s \mapsto \mathbb{R}, s \in \mathbb{N}$ and where the functions $h_0, h_1, \ldots, h_q$ does not depend on θ. The predictor space used for predicting $f(Y_i)$ is then given by $\mathcal{P}_{i-1}^{\theta} = span\{Z_0^{(i-1)}, Z_1^{(i-1)}, \ldots, Z_q^{(i-1)}\}$. The functions $h_0, h_1, \ldots, h_q$ are assumed to be linearly independent in $\mathcal{H}_{i-1}^{\theta}$ such that the dimension of the predictor space $\mathcal{P}_{i-1}^{\theta}$ is $q + 1$. Note that the dimension does not depend on i. The basis elements of the predictor space are no longer assumed to be functions of the entire past, but are instead functions of the "most recent past" of a period of length s. To adapt to usual practice, and to ensure the resulting MMSE predictor of $f(Y_i)$ in $\mathcal{P}_{i-1}^{\theta}$ becomes unbiased, we will assume $h_0 = 1$. The predictors in $\mathcal{P}_{i-1}^{\theta}$ will therefore be of the form $a_0 + a'Z^{(i-1)}$, where $a' = (a_1, \ldots, a_q)$ and $Z^{(i-1)} = (Z_1^{(i-1)}, \ldots, Z_q^{(i-1)})'$.[5]

The normal equations (4) 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 $b(\theta) = [\text{Cov}_\theta(Z_1^{(i-1)}, f(Y_i)), \ldots, \text{Cov}_\theta(Z_q^{(i-1)}, f(Y_i))]'$. Note that, since the observed process $\{Y_i\}$ is stationary, the coefficients of the MMSE predictor does not depend on i, but stay constant across time.[6] The covariance matrix $C(\theta)$ will be invertible, because the functions $h_1, h_2, \ldots, h_q$ are linearly independent. For a formal derivation of the expressions for the coefficients in (5) see Appendix A of the online appendix.

From (3-5) it follows that PBEFs can be calculated provided we can calculate the first- and second-order moments of the random vector $(f(Y_i), Z_1^{(i-1)}, \ldots, Z_q^{(i-1)})$. In most models, we will have to use simulation schemes in order to obtain the unconditional moments we need for the computation of $\hat{\pi}^{(i-1)}(\theta)$. However, in special cases, such as in some affine stochastic volatility models, these unconditional moments can be found explicitly for certain choices of the function f. Again, it should be noted that computing unconditional moments using simulation is less demanding than simulating the corresponding conditional moments.

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

---

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

[6]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, $\hat{a}_0(\theta), \ldots, \hat{a}_q(\theta)$, become time-varying.

$\Pi^{(i-1)}(\theta)$ belongs to the predictor space $\mathcal{P}_{i-1}^{\theta}$, then the jth 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 $Z_0^{(i-1)} = 1$. Hence, each element of the vector $\Pi^{(i-1)}(\theta)$ is a linear combination of the basis elements spanning $\mathcal{P}_{i-1}^{\theta}$, where the coefficient $a_{jk}(\theta)$ denote the loading on $Z_k^{(i-1)}$ for the jth element $\pi_j^{(i-1)}(\theta)$. Note that the coefficients $a_{jk}(\theta)$ does 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

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

for $i = 1, \ldots, n$ and $F_n(\theta) := \sum_{i=s+1}^{n} H^{(i)}(\theta)$.[7] 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*

To obtain an estimator one should solve $G_n(\theta) = \underline{0}$ 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

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

The above 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 that case the GMM objective function is $\left(\frac{1}{n-s} F_n(\theta)\right)' W_n(\theta) \left(\frac{1}{n-s} F_n(\theta)\right)$, and the corresponding p first order conditions are

$$2\left(\frac{1}{n-s}\right)^2 \underbrace{\left(\partial_\theta F_n(\theta)\right)'}_{p \times (q+1)} \underbrace{\hat{W}_n(\hat{\theta})}_{(q+1) \times (q+1)} \underbrace{F_n(\theta)}_{(q+1) \times 1} = \underline{0}. \tag{7}$$

---

[7]Note that the sum now runs from $i = s + 1$ since $Z_k^{(i-1)}$ is only well-defined for $i \geq s + 1$.

Here $\hat{W}_n(\hat{\theta})$ is the GMM weight matrix evaluated at some consistent parameter estimate $\hat{\theta}$. 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, which we do not allow $A(\theta)$ to be. However, it turns out that there is strong link between (7) with $\hat{W}_n(\hat{\theta})$ 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 of the online appendix. Straightforward calculations show that

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

From the theory for 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 co-variance matrix. When $W_n(\theta)$ is chosen optimally, $\frac{1}{n-s}W_n(\theta)$ equals $\overline{M}_n(\theta)^{-1}$ 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 emerges from the optimal GMM objective functions based on the moment conditions, $E_\theta[H(\theta)] = \underline{0}$, from the normal equations. The choice of which function of the data to predict and which predictor space to use, then translate into which moment conditions to use in the GMM estimation.

2.1.3. *Prediction-based Estimating Functions for SV-Models*

We now return to the setup from (1) and consider how to compute PBEFs for SV-models. Suppose the process X has been observed at discrete time points $X_0, X_\Delta, X_{2\Delta}, \ldots, X_{n\Delta}$. It is more convenient to base the statistical inference on the differences $Y_i = X_{i\Delta} - X_{(i-1)\Delta}$ since the process $\{Y_i\}$, in contrast to $\{X_{i\Delta}\}$, will be stationary since $\{v_t\}$ is assumed sta-tionary. In this setup, inference based on martingale estimating functions becomes practi-cally infeasible, since the conditional expectations appearing in estimation functions based on $f(Y_i) - E_\theta[f(Y_i)|Y_{i-1}, \ldots, Y_1]$ are difficult to compute analytically, as well as numerically. One feasible approach for conducting inference is to use PBEFs instead, since these are not based on conditional expectations and do not require extensive simulation. In fact, for many mod-els, such as the Heston model, we are able to find explicit expressions for the PBEFs. The continuous time returns from (1) are given by

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

which allows for the following decomposition

$$Y_i = \sqrt{S_i}Z_i, \tag{9}$$

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

$$S_i = \int_{(i-1)\Delta}^{i\Delta} v_t \, dt.$$

The distribution of v_t is the same on all intervals $[0, \Delta), \ldots, [(n-1)\Delta, n\Delta)$ because $\{v_t\}$ is stationary, hence it follows that $\{S_i\}$ and $\{Y_i\}$ are stationary processes. It is easy to see that the $Y_i's$ have zero mean and are uncorrelated, but not independent.

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 $Y_{i-1}, Y_{i-2}, \ldots Y_{i-q}$ will not work. As is the case with empirical data where we have volatility clustering, squared returns from the SV-model are often correlated and a natural choice for f would therefore be $f(x) = x^2$. The decomposition in (9) also reveals that $f(x) = x^2$ is convenient as it eases the computation of the moments required to construct the PBEFs. As our predictor spaces we choose

$$\mathcal{P}_{i-1}^{\theta} = \{a_0 + a_1 Y_{i-1}^2 + \cdots + a_q Y_{i-q}^2 | a_k \in \mathbb{R} \ k = 0, 1, \ldots, q\}.$$

This means that aside from the constant, $Z_0^{(i-1)} = 1$, the predictor variables $Z_k^{(i-1)} = Y_{i-k}^2$ for $k = 1, 2, \ldots, q$ have the same functional form as the function of the data to predict.[8] Notice, that in this case $s = q$ since $\mathcal{P}_{i-1}^{\theta}$ 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

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

and

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

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 we are considering PBEFs of the form

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

with $\Pi^{(i-1)}(\theta) = A(\theta) \tilde{Z}^{(i-1)}$, where $\tilde{Z}^{(i-1)} = (1, Y_{i-1}^2, \ldots, Y_{i-q}^2)'$ are the basis elements of the predictor space $\mathcal{P}_{i-1}^{\theta}$ and $A(\theta)$ is a $p \times (q+1)$ weight matrix.

---

[8]It should be noted, that one does not have to choose a predictor space spanned by variables of the same 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 and q need not be very large, (see thm. 3.3 in Bradley (2005)). Note that if the volatility process $\{v_t\}$ is α-mixing then the observed process $\{Y_i\}$ inherits this property and is also α-mixing, (see lemma 6.3 in Sørensen (2000)).

In our Monte Carlo study we will use the following sub-optimal, yet simple weight matrix

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

since computing the optimal weight matrix $A^*(\theta)$ involves analytically computing the covariance matrix of the sum of weighted augmented predition errors, $(F_n(\theta))$. The resulting sub-optimal PBEF is

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

Equating (12) to zero and solving for θ gives a $\sqrt{n}$-consistent estimator, but we may loose some efficiency for not using the optimal weight matrix $A^*(\theta)$.

From the above expressions, the econometrician might notice the link to OLS regression. In a standard linear regression $y = \beta X + \varepsilon$ the fitted value vector $\hat{y} = \hat{\beta}X$ is exactly the MMSE predictor of y, arising from the projection of y onto the linear subspace spanned by the columns in X. The residuals from the regression, $\hat{\varepsilon}$, are then the prediction errors. Another way to see the link to regression, is to consider the simple regression $y = \alpha + \beta x + \varepsilon$. The formulas for the regression coefficients are

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}, \quad \hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\text{sample covariance of x and y}}{\text{sample variance of x}},$$

where $\bar{x}$ denotes the sample mean of x and $\bar{y}$ denotes the sample mean of y. It is evident that the formulas for $\hat{\beta}$ and $\hat{\alpha}$ are just the sample versions of the analytical formulas in (10), if $y_i = Y_i^2$, $x_i = Y_{i-1}^2$ and $q = 1$. What we are doing, when estimating SV-models using PBEFs, are in some sense to fit an $AR(q)$ process, including a constant, to our series of squared returns. But instead of carrying out a regression to find the coefficients determining the orthogonal projection, we compute these coefficients, $\hat{a}(\theta)$, analytically. The coefficients will be expressed in terms of the parameters of interest to us, namely θ. Our estimate $\hat{\theta}$ is then the parameter vector for which the chosen weighted sum of augmented prediction errors, from fitting the $AR(q)$-process and constant, are eliminated.

The MMSE predictor $\hat{\pi}^{(i-1)}(\theta)$ can be computed if we can calculate $\hat{a}_0(\theta), \hat{a}_1(\theta), \ldots, \hat{a}_q(\theta)$. 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 $j = 1, \ldots, q$, so we have to assume $E_\theta[Y_i^4] < \infty$ for the MMSE predictor to be well-defined.[10] The requested moments can be

---

[10]From Jensen's inequality it follows that $E_\theta[v_t^{\beta/2}] < \infty$ implies $E_\theta[Y_i^\beta] < \infty$ for $\beta \geq 2$. For $\beta \leq 2$, $E_\theta[v_t] < \infty$ implies $E_\theta[Y_i^\beta] < \infty$.

calculated from the moments of the volatility process $\{v_t\}$. By the Itô isometry and recalling that $\{S_i\}$ is stationary we have that $E_\theta[Y_i^2] = E_\theta[S_1]$. Straightforward calculations gives $\text{Var}_\theta(Y_i^2) = 3\text{Var}_\theta(S_1) + 2E_\theta(S_1)^2$ and $\text{Cov}_\theta(Y_i^2, Y_{i+j}^2) = \text{Cov}_\theta(S_1, S_{1+j})$.

Now, define the mean, variance and autocorrelation function of the volatility process using the following notation $\xi(\theta) = E_\theta[v_t]$, $\omega(\theta) = \text{Var}_\theta(v_t)$, and $r(u;\theta) = \text{Cov}_\theta(v_t, v_{t+u})/\omega(\theta)$. From (Barndorff-Nielsen and Shephard, 2001, pp. 179-181) it follows that

$$E_\theta[Y_i^2] = \Delta\xi(\theta), \tag{13}$$

$$\text{Var}_\theta(Y_i^2) = 6\omega(\theta)R^*(\Delta;\theta) + 2\Delta^2\xi(\theta)^2, \tag{14}$$

$$\text{Cov}_\theta(Y_i^2, Y_{i+j}^2) = \omega(\theta)[R^*(\Delta(j+1);\theta) - 2R^*(\Delta j;\theta) + R^*(\Delta(j-1);\theta)], \tag{15}$$

where $R^*(t;\theta) = \int_0^t \int_0^s r(u;\theta)\,\mathrm{d}u\mathrm{d}s$. So all we need is now the first- and second-order moments of the unobserved volatility process $\{v_t\}$. This is, for instance, possible if the volatility process belongs to the class of Pearson diffusions, see Forman and Sørensen (2008). Amongst others, the CIR process is a Pearson diffusion so explicit expressions for PBEFs exist when the Heston model is considered.

In the Heston model the stationary distribution of $\{v_t\}$ is the Gamma distribution with shape parameter $2\kappa\alpha\sigma^{-2}$ and scale parameter $2\kappa\sigma^{-2}$, provided that $\sigma > 0, \alpha > 0$ (non-negativity), $\kappa > 0$ (stationary in mean), and $2\kappa\alpha \geq \sigma^2$ (stationary in volatility). Thus, we have

$$\xi(\theta) = \alpha, \quad \omega(\theta) = \frac{\alpha\sigma^2}{2\kappa}, \quad r(u;\theta) = e^{-\kappa u} \quad \text{and} \quad R^*(t;\theta) = \frac{1}{\kappa^2}\left(e^{-\kappa t} + \kappa t - 1\right).$$

Using the formulas (13) - (15), we can now calculate the following moments needed to compute $\hat{a}_0(\theta)$, $\hat{a}(\theta)$ and hence the MMSE predictor $\hat{\pi}^{(i-1)}(\theta)$. The moments we need are given by

$$E_\theta[Y_i^2] = \Delta\alpha, \tag{16}$$

$$\text{Var}_\theta(Y_i^2) = \frac{6\alpha\sigma^2}{2\kappa^3}\left(e^{-\kappa\Delta} + \kappa\Delta - 1\right) + 2\Delta^2\alpha^2, \tag{17}$$

$$\text{Cov}_\theta(Y_i^2, Y_{i+j}^2) = \frac{\alpha\sigma^2}{2\kappa^3}\left(e^{-\kappa\Delta j}\left[e^{-\kappa\Delta} - 2 + e^{\kappa\Delta}\right]\right), \tag{18}$$

and the sub-optimal PBEF can now be constructed.[11]

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. In Bollerslev and Zhou (2002) the moment conditions for constructing the GMM

---

[11]For 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 easy implementable. In practice, a feasible strategy could be to simulate $A^*(\theta)$.

estimator arise from the analytical derivations of the conditional first- and second-order moments of the daily integrated volatility (IV) process. 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 result in an easy-to-implement GMM estimator for the model parameters. Once again, the statistical inference will be based on the discretely sampled returns $Y_i^\Delta = X_{i\Delta} - X_{(i-1)\Delta}$, which we will assume to be available at high frequencies. 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 to IV, and hence ensure that the moment conditions underlying the GMM estimator holds approximately for RV.

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

$$E[IV_{t+1,t+2} - \delta IV_{t,t+1} - \beta | \mathcal{G}_t] = 0,$$

$$E[IV_{t+1,t+2}^2 - H(IV_{t,t+1}^2) - I(IV_{t,t+1}) - J | \mathcal{G}_t] = 0,$$

where $\mathcal{G}_t = \sigma\{IV_{t-s-1,t-s}; s = 0, 1, 2, \ldots \infty\}$ and δ, β, H, I and J are functions of the parameters κ, α and σ. 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 in the online appendix. 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 restrictions. By replacing daily IV by daily RV and using the unconditional versions of these six moment conditions, we are now able to construct a feasible GMM estimator for the parameters of interest $\theta = (\kappa, \alpha, \sigma)$. The feasible GMM estimator is given by

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

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 where $f_t(\theta)$ is given by

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

13

The number of trading days and hence the number of daily *RV* in our sample is denoted by
T. The finite sample performance of this estimator for the parameters in the Heston model, as
well as the one based on the sub-optimal PBEF, will be investigated in a Monte Carlo study in
the following section.

3. **A Monte Carlo Study of the Finite Sample Performances**

3.1. **The Setup and the Case with no Measurement Error**

In this section the finite sample performances of the GMM estimator from Bollerslev and Zhou
(2002) and the PBEF based estimator from Sørensen (2000) are investigated in a Monte Carlo
study. The data used for estimation is simulated realizations from the Heston model (2). We
use a first-order Euler scheme for simulating the volatility- and log-price process. The log-
price is sampled every 30 seconds in the artificial 6.5 hours of daily trading, for sample sizes
of $T = 1000, 4000$ trading days. Using the simulated data, daily realized variances based on
the artificial five-minute returns are constructed. We think of the five-minute returns as our
available data and investigate the performance of the corresponding GMM based and PBEF
based estimator. Since we are using five-minute returns over 6.5 hours of trading we put
$\Delta = 1/78$.

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), as the autocorrelation is given by $r(u, \theta) = e^{-\kappa u}$ the correlation between the
  volatility process sampled five minutes apart equals $e^{-0.03*1/78} = 0.9996$.

- 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.

- 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 where considered in the Monte Carlo study conducted in Bollerslev and
Zhou (2002). In each of the three scenarios we simulate data from the Heston model and use
it as input in the two estimation procedures. We impose strict positivity of the parameter
estimates $\hat{\kappa}, \hat{\alpha}$ and $\hat{\sigma}$ and require that the Feller condition, $2\hat{\kappa}\hat{\alpha} \geq \hat{\sigma}^2$, be satisfied. In each case

the number of Monte Carlo replications is 1000 and we use the true values of θ as starting values.

3.1.1. *Finite Sample Performance of the GMM Estimator based on* IV

When estimating the asymptotic covariance matrix of $g_T(\theta)$ we use the heteroscedasticity and autocorrelation consistent estimator from Newey and West (1987) . When implementing the GMM estimation procedure, we use continuously updated GMM, where the weight matrix is estimated simultaneously with the parameters θ. Regarding the lag length in the Bartlett kernel we follow Bollerslev and Zhou (2002) and choose a lag length of five. The results from the GMM estimations are reported in Table 1.

The finite sample performance of the GMM estimator, using realized variance based on five-minute returns, is quite good. As in Bollerslev and Zhou (2002), we find the mean-reversion parameter κ to exhibit a small upwards bias, while the long-run mean parameter α is slightly downwards biased. The biases of the two drift parameters seem to worsen when the volatility of volatility is high (as in scenario 3), but in all cases decrease when the sample size increases. The volatility of volatility parameter σ has a small, yet systematic, upwards bias, that actually seems to be more pronounced when the sample size increases from 1000 to 4000. Below we will try to account for the bias in σ by adjusting the moment conditions used for constructing the GMM estimator. Turning our attention to the RMSE, we observe that the RMSE for the two drift parameters is roughly halved when the sample size increases from 1000 to 4000, as one would expect. However, the decay rate in RMSE for σ, when the sample grows from 1000 to 4000 is not always 2, ($\sqrt{4000/1000}$).

To better understand the bias in σ, we consider the discretization error $u_{t,t+1} = RV_{t,t+1} - IV_{t,t+1}$. From Barndorff-Nielsen and Shephard (2002) we know that, in the case of zero drift, the error made by replacing $IV_{t,t+1}$ with $RV_{t,t+1}$ is averaged out in the first moment condition and the corresponding augmented ones, since these depend linearly on the discretization error. However, Barndorff-Nielsen and Shephard (2002) also show that $RV_{t,t+1}^2$, for any fixed sampling frequency, is an upwards biased estimator of $IV_{t,t+1}^2$. To account for this discretization error Bollerslev and Zhou (2002) therefore introduce a nuisance parameter, γ, and approximate $IV_{t+1,t+2}^2$ by $RV_{t+1,t+2}^2 - \gamma$. We follow the same strategy and implement the discretization error correction (DEC) described above in the three second order moment conditions. The results on σ from using the GMM estimation procedure with DEC are reported in Table 1. For the results on the drift parameters consult Table 1 in Appendix D of the online appendix. The results show that the drift parameters κ and α are not as affected by the discretization error correction as σ is. This is not surprising since the drift parameters enters both the first and second order

Table 1: Performance of the GMM estimator based on *RV*.

	Mean		Bias		Variance		RMSE	
	$T = 1000$	$T = 4000$	$T = 1000$	$T = 4000$	$T = 1000$	$T = 4000$	$T = 1000$	$T = 4000$
Scenario 1								
GMM with daily realized variance from five-minute returns								
$\kappa = 0.03$	0.0342	0.0312	0.0042	0.0012	1.162e-04	2.495e-05	0.0116	0.0051
$\alpha = 0.25$	0.2422	0.2458	-0.0078	-0.0042	0.0029	6.758e-04	0.0541	0.0263
$\sigma = 0.10$	0.1015	0.1023	0.0015	0.0023	6.814e-05	1.802e-05	0.0084	0.0048
GMM with daily realized variance using DEC								
$\sigma = 0.10$	0.0934	0.0994	-0.0066	-0.0006	2.773e-04	7.100e-05	0.0179	0.0084
Scenario 2								
GMM with daily realized variance from five-minute returns								
$\kappa = 0.10$	0.1043	0.1012	0.0043	0.0012	4.273e-04	9.367e-05	0.0211	0.0098
$\alpha = 0.25$	0.2445	0.2461	-0.0055	-0.0039	2.481e-04	5.795e-05	0.0167	0.0086
$\sigma = 0.10$	0.1061	0.1071	0.0061	0.0071	5.874e-05	1.413e-05	0.0098	0.0080
GMM with daily realized variance using DEC								
$\sigma = 0.10$	0.0992	0.1013	-0.0008	0.0013	2.861e-04	5.492e-05	0.0169	0.0075
Scenario 3								
GMM with daily realized variance from five-minute returns								
$\kappa = 0.10$	0.1101	0.1035	0.0101	0.0035	4.844e-04	1.075e-04	0.0242	0.0109
$\alpha = 0.25$	0.2381	0.2447	-0.0119	-0.0053	8.873e-04	2.263e-04	0.0321	0.0160
$\sigma = 0.20$	0.2029	0.2044	0.0029	0.0044	1.271e-04	3.422e-05	0.0116	0.0073
GMM with daily realized variance using DEC								
$\sigma = 0.20$	0.1890	0.1991	-0.0110	-0.0009	5.509e-04	1.988e-04	0.0259	0.0141

moment conditions. But for σ, that only the enters the second order moment conditions, the results change. The DEC seems to remove the systematic upwards bias in the estimates of σ, in fact there is now a downwards bias when $T = 1000$ and practically no bias when $T = 4000$. Unfortunately, the variance of σ has now increased and the RMSE has roughly doubled. The size of the RMSE is still very small, but of course this is a bias-variance trade-off to consider. There is definitely still room for improvement, which could be filled by using more sophisticated methods for discretization error correction. We will not pursue any other DEC methods here since this is not the aim of our study.

3.1.2. *Finite Sample Performance of the Estimator based on PBEFs*

We now investigate the finite sample performance of the estimator associated with the following sub-optimal PBEF from (12)

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

An interesting question, that arises when considering PBEFs, is how to optimally choose q. One approach for answering this question, could be to consider the partial-autocorrelation function for the squared returns and then choose q as the cut-off point where the function dies out. Instead, we start at the smallest interesting choice, $q = 3$, and then later on investigate the choice of q.[12]

Table 2: Performance of the sub-optimal PBEF based estimator with $q = 3$.

	Mean		Bias		Variance		RMSE	
	$T = 1000$	$T = 4000$	$T = 1000$	$T = 4000$	$T = 1000$	$T = 4000$	$T = 1000$	$T = 4000$
Scenario 1								
Sub-optimal PBEF estimation based on five-minute returns								
$\kappa = 0.03$	0.0303	0.0301	0.0003	0.0001	2.621e-06	7.525e-07	0.0016	0.0009
$\alpha = 0.25$	0.2500	0.2500	-4.076e-05	3.317e-05	0.0027	6.799e-04	0.0517	0.0261
$\sigma = 0.10$	0.0989	0.0993	-0.0011	-0.0007	1.063e-04	3.247e-05	0.0104	0.0057
Scenario 2								
Sub-optimal PBEF estimation based on five-minute returns								
$\kappa = 0.10$	0.1003	0.1002	0.0003	0.0002	5.536e-06	1.162e-06	0.0024	0.0011
$\alpha = 0.25$	0.2500	0.2500	-3.823e-06	-1.163e-05	2.436e-04	5.952e-05	0.0156	0.0077
$\sigma = 0.10$	0.0995	0.0996	-0.0005	-0.0004	2.159e-05	4.454e-06	0.0047	0.0021
Scenario 3								
Sub-optimal PBEF estimation based on five-minute returns								
$\kappa = 0.10$	0.1002	0.1003	0.0002	0.0003	1.039e-05	2.732e-06	0.0032	0.0017
$\alpha = 0.25$	0.2501	0.2506	-0.0001	0.0006	8.894e-04	2.316e-04	0.0298	0.0152
$\sigma = 0.20$	0.1996	0.1990	-0.0004	-0.0010	1.583e-04	4.255e-05	0.0126	0.0066

When the parameters $\theta = (\kappa, \alpha, \sigma)$ are estimated we minimize $G_n(\theta)'G_n(\theta)$ instead of solving $G_n(\theta) = \underline{0}$. The finite sample performance of the sub-optimal PBEF based on the series, $\{Y_i^2\}$, of squared five-minute returns are reported in Table 2. Compared to the results for the GMM estimator based on IV, using PBEF to estimate the Heston model looks promising. The results show indicate that the parameter estimates becomes almost unbiased when using this choice of PBEF. Comparing the RMSEs for the drift parameters it seems that the mean reversion parameter κ is easier to estimate when using PBEFs, whereas the RMSEs for α are similar to those obtained using the GMM procedure. In scenario 2, the RMSE for σ is smaller when the PBEF is used, but for the other two scenarios the RMSE is similar to the those from the GMM procedure. In contrast to what we observed for the GMM estimator based on IV, the RMSE for σ in the PBEF case is actually halved when the sample size doubles. For κ the variance is smaller when the PBEF is used, but for α and σ the variances are of the same magnitude as for the GMM procedure.

---

[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 $G_n(\theta) = \underline{0}$.

These results were to be expected, because the estimation method based on PBEFs use the intra-daily squared returns directly, whereas the squared returns are transformed into daily realized measures before the GMM estimator is constructed. Based on our results, it seems that the PBEF based estimator is able to exploit the extra information contained in the intra-daily returns.
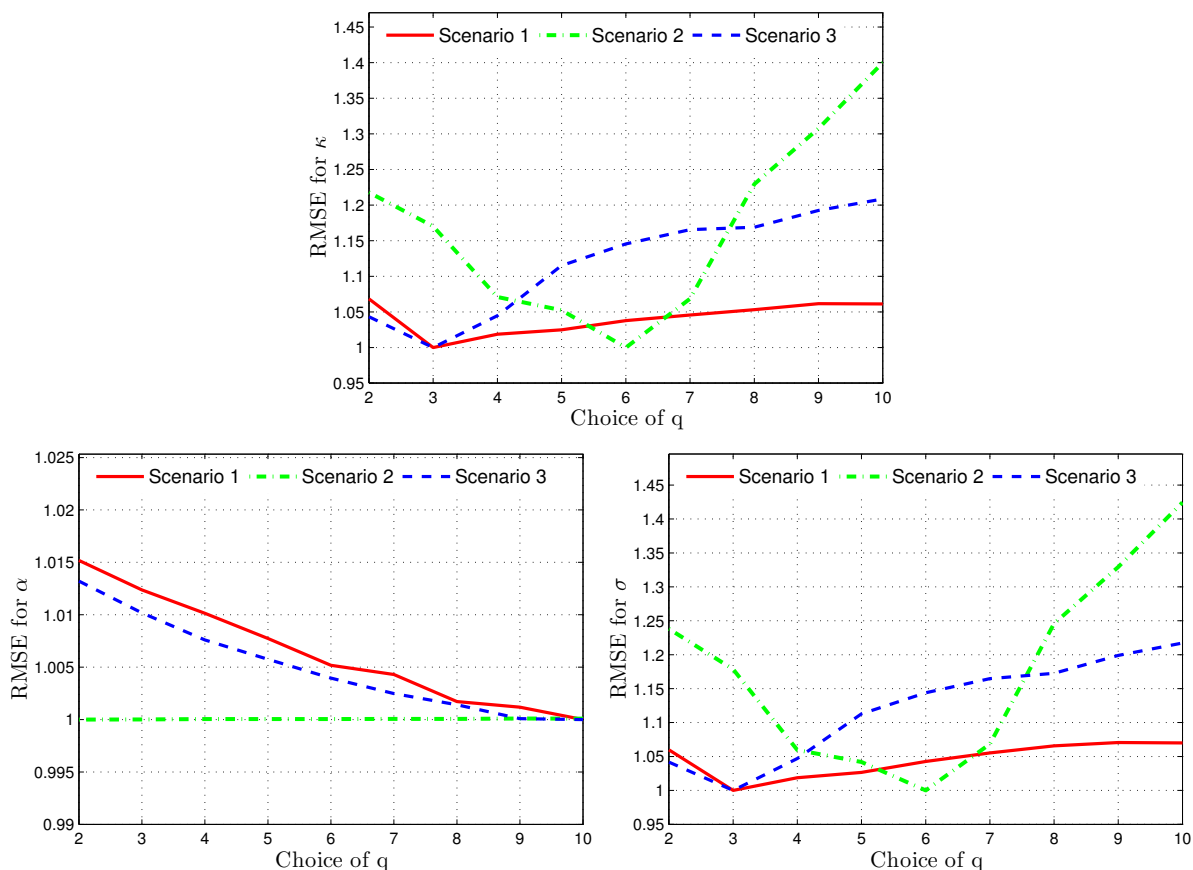

Figure 1: Normalized RMSE for the three parameter estimates in scenario 1-3, with $T = 1000$, plotted as a function of the number of predictor variables q. The RMSE are in each case normalized to have a minimum of 1.

To investigate the optimal choice of q in the PBEF, the impact on the RMSEs from increasing q is examined. In Figure 1 the three different scenarios are considered for $T = 1000$ and the RMSEs for the three parameters are plotted against the choice of q. The shape of the plots for κ and σ look almost identical. In the two 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 choice for both parameters. Looking at the three plots for α there does not seem to be much variation in the RMSE across the choice of q. In in scenario 1 and 3, where we are close to the non-stationary region, the RMSE decreases when q increases. Therefore, we chose $q = 3$ in our Monte Carlo Study even though the variations in the RMSEs

are small. It should also be noted that the optimal choice of q might be depend on the PBEF under consideration.

In the next section we investigate how adding i.i.d. measurement errors to the observed process effects the finite sample performance of the two estimation methods. We also extend the methods to take account of these measurement errors and compare the performance of the two resulting estimators. In the simulations we fix q at 3.

3.2. Including Measurement Errors in the Observed Log-price Process

We now consider the following observation equation

$$X_t = X_t^* + U_t, \quad U_t \text{ i.i.d } N(0, \omega^2), \tag{21}$$

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_t will be interpreted as market microstructure (MMS) noise due to market frictions such as bid-ask bounce, liquidity changes and discreteness of prices. In this section, the impact of MMS noise on the parameter estimates from the two estimation procedures is investigated. In the next section we consider how to change the two estimation methods to incorporate the MMS noise. We now simulate data with MMS noise and then perform parameter estimation ignoring the noise. We consider two levels of noise, $\omega^2 = 0.001$ which we present in the paper. Results for $\omega^2 = 0.0005$ can be found in Appendix D of the online appendix.

Table 3: Performance of the GMM estimator based on RV in the presence of noise, $\omega^2 = 0.001$.

	Mean		Bias		Variance		RMSE	
	$T = 1000$	$T = 4000$	$T = 1000$	$T = 4000$	$T = 1000$	$T = 4000$	$T = 1000$	$T = 4000$
Scenario 1								
GMM with daily realized variance from five-minute returns								
$\kappa = 0.03$	0.0374	0.0336	0.0074	0.0036	2.325e-04	4.266e-05	0.0169	0.0075
$\alpha = 0.25$	0.3976	0.3981	0.1476	0.1481	0.0031	6.808e-04	0.1578	0.1503
$\sigma = 0.10$	0.0797	0.0806	-0.0203	-0.0194	1.425e-04	3.626e-05	0.0235	0.0203
Scenario 2								
GMM with daily realized variance from five-minute returns								
$\kappa = 0.10$	0.1069	0.1043	0.0069	0.0043	0.0010	2.304e-04	0.0329	0.0158
$\alpha = 0.25$	0.3987	0.3999	0.1487	0.1499	2.532e-04	5.915e-05	0.1495	0.1501
$\sigma = 0.10$	0.0915	0.0927	-0.0085	-0.0073	1.760e-04	4.135e-05	0.0157	0.0097
Scenario 3								
GMM with daily realized variance from five-minute returns								
$\kappa = 0.10$	0.1211	0.1133	0.0211	0.0133	7.786e-04	1.512e-04	0.0350	0.0181
$\alpha = 0.25$	0.3929	0.3967	0.1429	0.1467	0.0010	2.458e-04	0.1464	0.1475
$\sigma = 0.20$	0.1611	0.1619	-0.0389	-0.0381	1.451e-04	3.4253e-05	0.0407	0.0386

The results for GMM estimation based on RV are reported in Table 3. From the results it follows that the inclusion of MMS noise in the observed process leads to biases in the parameters, when the GMM estimation procedure without noise correction is carried out. Compared to the no noise case, we note that the upwards bias in the drift parameter κ has worsened and the small downwards bias in the other drift parameter α has been turned into a severe upwards bias. Regarding the parameter σ, the small upwards bias has now been replaced by a notable downwards bias. The biases in κ and σ seems to worsen in scenario 3 where σ is increased. For all three parameters, the biases become more pronounced when ω^2 increases, but for κ and σ the biases become less pronounced when the sample size increases. In all three cases, and for both choices of sample size, the RMSE for the parameters estimates goes up when ω^2 increases. For κ and σ, the RMSE goes down when the sample size increases, because the variance decrease. For α the RMSE remains unchanged.

To understand how the MMS noise effects the estimator from the PBEF estimation method, we start by investigating the impact of noise on the autocorrelation function for the squared return series, in each of the three parameter scenarios.
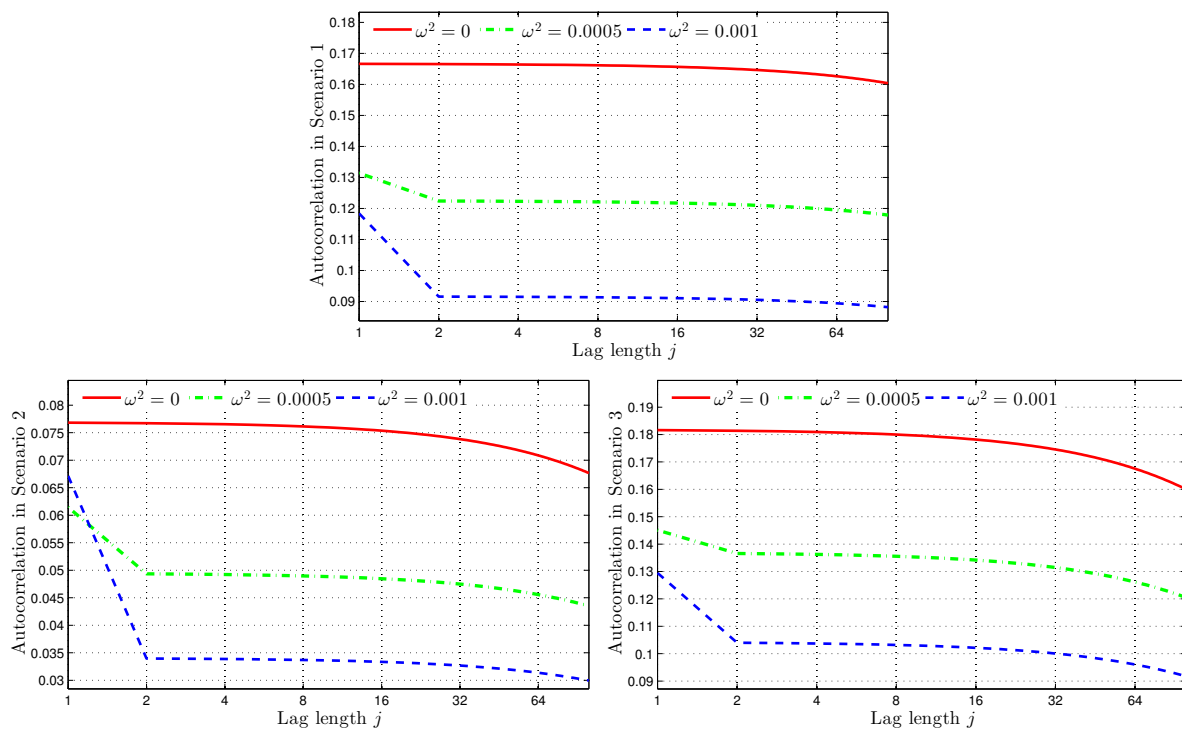

Figure 2: Autocorrelation functions for the squared returns in scenario 1-3.

From Figure 2, we see how the theoretical autocorrelation function drops as the variance of the noise increases. It also becomes evident that the $MA(1)$ structure in the errors from

the return series causes the autocorrelation function to drastically drop from the first to the second lag. This feature becomes more pronounced as the variance of the noise increases. The low level of the autocorrelation function, especially in scenario 2, means that our predictors $Y_{i-1}^2, \ldots, Y_{i-q}^2$ have little predictive power and the PBEF estimation method might not produce as accurate parameter estimates as in the no noise case. The results from the estimation method based on the sub-optimal PBEF without noise correction are reported in Table 4 in the $\omega^2 = 0.001$ case.

Table 4: Performance of the PBEF based estimator in the presence of noise, $\omega^2 = 0.001$ and $q = 3$.

	Mean		Bias		Variance		RMSE	
	$T = 1000$	$T = 4000$	$T = 1000$	$T = 4000$	$T = 1000$	$T = 4000$	$T = 1000$	$T = 4000$
Scenario 1								
Sub-optimal PBEF estimation based on five-minute returns								
$\kappa = 0.03$	0.0329	0.0322	0.0029	0.0022	3.929e-06	1.192e-06	0.0035	0.0025
$\alpha = 0.25$	0.4050	0.4061	0.1550	0.1561	0.0028	6.829e-04	0.1638	0.1582
$\sigma = 0.10$	0.0843	0.0870	-0.0157	-0.0130	1.134e-04	3.538e-05	0.0190	0.0143
Scenario 2								
Sub-optimal PBEF estimation based on five-minute returns								
$\kappa = 0.10$	0.0968	0.0964	-0.0032	-0.0036	1.133e-05	9.147e-06	0.0047	0.0047
$\alpha = 0.25$	0.4060	0.4060	0.1560	0.1560	2.470e-04	6.0034e-05	0.1568	0.1562
$\sigma = 0.10$	0.1071	0.1078	0.0071	0.0078	4.878e-05	4.109e-05	0.0100	0.0101
Scenario 3								
Sub-optimal PBEF estimation based on five-minute returns								
$\kappa = 0.10$	0.1082	0.1073	0.0082	0.0073	2.116e-05	5.272e-06	0.0094	0.0077
$\alpha = 0.25$	0.4054	0.4066	0.1554	0.1566	9.565e-04	2.331e-04	0.1585	0.1573
$\sigma = 0.20$	0.1714	0.1737	-0.0286	-0.0263	2.163e-04	5.445e-05	0.0322	0.0273

From the tables we see that the drift parameter κ becomes upwards biased, but not as much as for the GMM method. The bias also increases when then noise variance, ω^2, goes from 0.0005 to 0.001, except for scenario 2 where the bias becomes negative when the noise variance doubles. This is a bit odd, but can be explained by the fact that the autocorrelation function in scenario 2 is extremely low, especially when $\omega^2 = 0.001$, which means that the PBEF estimation procedure might not work as well as in the no noise case. The long run average of variance parameter, α, becomes severely upwards biased when we include MMS in our log-price process and fail to correct for it in the PBEF. In all the considered cases, the positive bias in α roughly doubles when the variance of the noise doubles. The bias in α is bigger in the PBEF setting, compared to the GMM setting, but the increase in the bias from the no noise case to the MMS noise case is roughly the same for the two estimation methods. As for the volatility of volatility parameter, σ, it now becomes downwards biased. The bias in σ worsen when ω^2 increases but decreases when the sample size decreases, except for scenario

21

2 where the bias again changes sign when the noise variance is increased. When comparing the results to those for the GMM estimator based on RV, we find that the variance of κ has gone down, the variance of α is roughly unaltered and the variance of σ is bigger in scenario 1 and 3 and smaller in scenario 2. The results on the behavior of the bias and variance of the parameter estimates means that the RMSE for kappa as gone significantly down, the RMSE for α has increased and the RMSE for σ is the same, when compared to the GMM results. Compared to the no noise results from Table 2, the RMSE has gone up in all the considered scenarios. In all cases, the RMSE goes down when the sample size increases, but it is not close to being halved as otherwise expected.

By looking at the autocorrelation functions in the three scenarios, some of the above results for the sub-optimal PBEF could have been anticipated. From the expression for the autocorrelation function for the squared returns in the no noise case, (obtained using equation (17) and (18))

$$\text{Corr}(Y_i^2, Y_{i+j}^2) = \frac{\text{Cov}(Y_i^2, Y_{i+j}^2)}{\text{Var}(Y_i^2)} = \frac{e^{-\kappa\Delta j}[e^{-\kappa\Delta} - 2 + e^{\kappa\Delta}]}{6(e^{-\kappa\Delta} + \kappa\Delta - 1) + \frac{4\Delta^2\alpha\kappa^3}{\sigma^2}}, \tag{22}$$

we see that the decay rate is completely determined by κ. When the presence of MMS noise is ignored, the above autocorrelation structure is fitted to the ones from the MMS noise case depicted in Figure 2. But due to rapid drop from the first to the second autocorrelation when we have MMS noise, κ will increase such that the correlation structure (22) from the no noise case better fits the autocorrelation structure from the MMS noise case. This could be one explanation for the upwards bias in κ. It was also noted that the level of the autocorrelation function drops when MMS noise is added to the observed log-price process. From (22) we see that, assuming κ has been determined to fit the decay rate, one way of getting the level of the autocorrelation function to drop would be to increase α and/or decrease σ. The inspection of the autocorrelation functions offers some explanation of the biases in the parameters. In the next subsection, where the two estimation procedures are extended to handle MMS noise, it will become clear why α becomes significantly upwards biased when we do not correct for the MMS noise. Furthermore, an analytical expression for the bias will be derived.

3.3. **Correcting for MMS Noise in the two Estimation Procedures**

The two estimation methods are now extended to handle MMS noise. Since RV is no longer a consistent estimator of IV when MMS noise present, we will extend the GMM estimation procedure from Bollerslev and Zhou (2002) to handle this situation by replacing daily RV with a noise robust estimate of IV. As our estimate of daily IV we will use the realized kernel (RK) estimator from Barndorff-Nielsen et al. (2008a). The PBEF based method can also be

extended to handle MMS noise, simply by recalculating the moments $E_\theta[Y_i^2]$, $\text{Var}_\theta(Y_i^2)$ and $\text{Cov}_\theta(Y_i^2, Y_{i+j}^2)$ needed for constructing $\hat{\pi}(\theta)$, and hence the PBEFs, in this new setting.

3.3.1. *Correcting for MMS noise 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

$$E_\theta[RV_{t,t+1}] = \int_t^{t+1} v_s \, \mathrm{d}s + 2\Delta^{-1}\omega^2. \tag{23}$$

In fact, also the variance of *RV* diverges to infinity as the sample frequency increases. In the setting with MMS noise, *RV* is no longer a consistent estimator of *IV* and we 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 instead use the time series of daily realized kernels (*RK*) from Barndorff-Nielsen et al. (2008a). We use the flat-top Tukey-Hanning$_2$ 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*. For further details on how the bandwidth is chosen consult section 4 of Barndorff-Nielsen et al. (2008a).

The finite sample performances of the GMM estimator based on the time series of daily *RK* are reported in Table 5 when $\omega^2 = 0.001$.

Table 5: Finite sample behavior of the GMM estimator based on *RK*, $\omega^2 = 0.001$.

	Mean		Bias		Variance		RMSE	
	$T = 1000$	$T = 4000$	$T = 1000$	$T = 4000$	$T = 1000$	$T = 4000$	$T = 1000$	$T = 4000$
Scenario 1								
GMM with daily realized kernels from five-minute returns								
$\kappa = 0.03$	0.0329	0.0306	0.0029	0.0006	2.037e-04	5.507e-05	0.0146	0.0074
$\alpha = 0.25$	0.2405	0.2414	-0.0095	-0.0086	0.0029	5.831e-04	0.0548	0.0256
$\sigma = 0.10$	0.1058	0.1093	0.0058	0.0093	4.954e-04	1.673e-04	0.0230	0.0159
Scenario 2								
GMM with daily realized kernels from five-minute returns								
$\kappa = 0.10$	0.1024	0.1005	0.0024	0.0005	0.0018	4.540e-04	0.0428	0.0223
$\alpha = 0.25$	0.2410	0.2417	-0.0090	-0.0083	5.075e-04	5.810e-05	0.0242	0.0113
$\sigma = 0.10$	0.1259	0.1291	0.0259	0.0291	7.654-04	2.035e-04	0.0379	0.0324
Scenario 3								
GMM with daily realized kernels from five-minute returns								
$\kappa = 0.10$	0.1040	0.1009	0.0040	0.0009	4.412e-04	1.031e-04	0.0214	0.0102
$\alpha = 0.25$	0.2386	0.2422	-0.0114	-0.0078	5.866e-04	1.467e-04	0.0267	0.0144
$\sigma = 0.20$	0.2116	0.2163	0.0116	0.0163	2.452e-04	7.909e-05	0.0195	0.0185

From the tables it is clear that the upwards bias in κ has gone down, and is in fact a bit smaller than the upwards bias from the no noise case in Table 1. Regarding the RMSE for κ, it is almost the same for both levels of noise and is roughly halved when the sample size is doubled. Compared to the no noise case from Table 1 the RMSE has gone up in the first two scenarios, especially in scenario 2. When compared to the noise case without noise correction from Table 3, the RMSE is unaltered in scenario 1, has gone up in scenario 2 and gone down in scenario 3.

Turning our attention to the other drift parameter α, we see that the huge upward biases from Table 3 have now been turned into a small but noticeable downwards bias, that is more pronounced than in the no noise case from Table 1. Looking at the RMSE, the values are now almost the same as in the no noise case of Table 1 when scenario 1 is considered and a bit higher and a bit lower in scenario 2 and 3 respectively.

Regarding the volatility of volatility parameter σ, the situation is a bit different. The downwards bias, caused by the MMS noise, has now been replaced by an upwards bias from the use of RK. This upwards bias is most pronounced in scenario 2, and in all three scenarios more severe than the one we observed in the no noise case. As in the no noise case, the bias seems to worsen when the sample size increases. As a consequence, the RMSE has also gone significantly up, compared to the no noise case, especially in scenario 2. In scenario 2, the RMSE for σ is actually larger than in the case without noise correction. In scenario 1 and 3 the results are a bit mixed. When compared to the noise case without noise correction, the RMSE for σ has gone slightly up in scenario 1 when $\omega^2 = 0.0005$, whereas in scenario 3 the RMSE has gone down for both levels of the noise variance. When the sample size doubles, the RMSE is again far from being halved. Common to all three parameter estimates is the fact that when the series of RK is used as input in the GMM estimation, the RMSE seems to be almost unchanged when the noise level increases. The severe upwards bias in σ might be, as discussed earlier in the no noise case, due to the discretization error since RK^2 is an upwards biased estimator of IV^2. The slower convergence rate of RK might explain why the upwards bias in σ is more pronounced than in the no noise case where RV is used to construct the GMM estimator.

In conclusion, the use of RK in the presence of MMS noise is not perfect in the setting we are considering. It helps reduce the bias in κ and almost removes the huge bias in α, but it also causes a severe upwards bias in σ. This way of correcting for the noise also causes the RMSE for κ and σ to increase compared to the no noise case, and in some cases it even increased when compared to the case with MMS noise and no noise correction. However, this increase is not dramatic and there is still much to be gained from using RK to correct for the MMS noise.

3.3.2. *Correcting for MMS noise in the Estimation Method based on PBEFs*

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 (21)

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

where the $MA(1)$ process ϵ is normally distributed, $N(0, 2\omega^2)$, and independent of the efficient return process Y^*. The moments needed to compute PBEFs in the presence of i.i.d MMS noise are given below. Straight forward 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 $\frac{2\omega^2}{\Delta}$. The bias equals 0.0780 and 0.1560 when the noise variance is 0.0005 and 0.001 respectively, and in fact these numbers exactly match the actual biases reported in Table 4.

As for the variance of the squared returns we have that

$$\text{Var}_\theta(Y_i^2) = E_\theta[Y_i^4] + E_\theta[Y_i^2]^2, \tag{24}$$

where the last term can be computed from the above calculations. By using the binomial formula we find

$$E_\theta[Y_i^4] = E_\theta[(Y_i^*)^4] + E_\theta[\epsilon_i^4] + 6E_\theta[(Y_i^*)^2]E_\theta[\epsilon_i^2],$$

again due to Y^* and ϵ being independent and having mean zero. From previously derived results we get

$$E_\theta[Y_i^4] = 3E_\theta[S_1^2] + 3(2\omega^2)^2 + 6\Delta\alpha2\omega^2 = 3E_\theta[S_1^2] + 12\omega^4 + 12\Delta\alpha\omega^2, \tag{25}$$

and by using the derivation of equation (14) as well as equation (17) we obtain

$$E_\theta[S_1^2] = 2\text{Var}_\theta(v_t)R^*(\Delta, \theta) + E_\theta[S_1]^2 = \frac{\alpha\sigma^2}{\kappa^3}(e^{-\kappa\Delta} + \kappa\Delta - 1) + \Delta^2\alpha^2. \tag{26}$$

Now, by plugging (26) into (25) we can obtain an expression for $\text{Var}_\theta(Y_i^2)$ by using (24)

$$\text{Var}_\theta(Y_i^2) = \frac{3\alpha\sigma^2}{\kappa^3}(e^{-\kappa\Delta} + \kappa\Delta - 1) + 3\Delta^2\alpha^2 + 12\omega^4 + 12\Delta\alpha\omega^2 - (\Delta\alpha + 2\omega^2)^2$$

$$= \text{Var}_\theta(Y_i^{*2}) + 8\omega^4 + 8\Delta\alpha\omega^2. \tag{27}$$

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

$$\text{Cov}_\theta(Y_i^2, Y_{i+1}^2) = \text{Cov}_\theta(Y_i^{*2}, Y_{i+1}^{*2}) + \text{Cov}_\theta(\epsilon_i^2, \epsilon_{i+1}^2) = \text{Cov}_\theta(Y_i^{*2}, Y_{i+1}^{*2}) + 2\omega^4. \qquad (28)$$

We can now compute the noise corrected version of the PBEF we previously used for estimation. Note that we can choose to estimate the variance of the noise, ω^2, in a first step, by for instance using a non-parametric estimator based on *RV* and *RK*, and then plug it 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, $A(\theta)$, that will results 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 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 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 now be optimal. The finite sample performances of the (now optimal) PBEF corrected for MMS noise are reported in Table 6 when $\omega^2 = 0.001$.

Table 6: Performance of the PBEF based estimator corrected for noise, $\omega^2 = 0.001$ and $q = 3$.

	Mean		Bias		Variance		RMSE	
	$T = 1000$	$T = 4000$	$T = 1000$	$T = 4000$	$T = 1000$	$T = 4000$	$T = 1000$	$T = 4000$
Scenario 1								
Optimal PBEF estimation based on five-minute returns								
$\kappa = 0.03$	0.0308	0.0301	0.0008	0.0001	3.911e-06	1.031e-06	0.0021	0.0010
$\alpha = 0.25$	0.2525	0.2512	0.0025	0.0012	0.0023	5.107e-04	0.0477	0.0226
$\sigma = 0.10$	0.0958	0.0995	-0.0042	-0.0005	1.506e-04	4.398e-05	0.0130	0.0066
$\omega^2 = 0.001$	9.776e-04	9.932e-04	-2.236e-05	-6.853e-06	1.0153e-08	2.517e-09	1.032e-04	5.061e-05
Scenario 2								
Optimal PBEF estimation based on five-minute returns								
$\kappa = 0.10$	0.0999	0.0999	-0.0001	-0.0001	6.904e-06	1.225e-06	0.0026	0.0011
$\alpha = 0.25$	0.2513	0.2503	0.0013	0.0003	1.890e-04	3.903e-05	0.0138	0.0063
$\sigma = 0.10$	0.1005	0.1003	0.0005	0.0003	2.796e-05	5.073e-06	0.0053	0.0023
$\omega^2 = 0.001$	9.915e-04	9.981e-04	-8.455e-06	-1.934e-06	1.765e-09	2.613e-10	4.283e-05	1.627e-05
Scenario 3								
Optimal PBEF estimation based on five-minute returns								
$\kappa = 0.10$	0.1014	0.1002	0.0014	0.0002	1.721e-05	4.346e-06	0.0044	0.0021
$\alpha = 0.25$	0.2520	0.2509	0.0020	0.0009	0.0010	2.104e-04	0.0320	0.0145
$\sigma = 0.20$	0.1955	0.1994	-0.0045	-0.0006	2.483e-04	6.667e-05	0.0164	0.0082
$\omega^2 = 0.001$	9.841e-04	9.977e-04	-1.592e-05	-2.289e-06	1.028e-08	2.096e-09	1.026e-04	4.582e-05

In Table 6 we see that by noise correcting the PBEF estimation procedure the bias in κ

vanishes when $T = 4000$ and only a small bias is left when $T = 1000$. As for the other drift parameter, α, only a small negligible upwards bias is left. Compared with the results on noise corrected GMM estimator based on RK, the bias in α is smaller both when $\omega^2 = 0.0005$ and when $\omega^2 = 0.001$. In all the considered cases, the bias in α is reduced when the sample size is increased. The downwards bias in σ is now much smaller than in the MMS noise case without noise correction, and the bias almost disappears when the sample size increases to 4000 trading days. The RMSE is once again halved when the sample size increases from 1000 to 4000 trading days. Compared to the MMS noise case without noise correction, the RMSE has dropped drastically for all three parameter estimates and for both levels of the noise variance. The RMSEs have also roughly reached the same level as in the no noise case from Table 2. The variances of the parameter estimates are also smaller than in the MMS noise case without noise correction. The noise variance is also estimated, and the results show that the noise corrected PBEF produces unbiased estimates of this quantity. The RMSEs coming from the noise corrected PBEF estimation method are in all cases, except for α in scenario 3, smaller than the RMSEs obtained using the noise corrected GMM estimation method, with the difference being most remarkable for the parameters κ and σ. In conclusion, the noise corrected PBEF estimation method produces (almost) unbiased parameter estimates with RMSEs comparable to those from the no noise case. In our Monte Carlo setup where the Heston model with additive MMS noise is considered, the estimation method based on PBEFs seems to outperform the GMM estimation method from Bollerslev and Zhou (2002).

4. Empirical Application

In this section we use actual five-minute returns from a widely used data set on the DM/$ spot exchange rate as input in the two estimation methods analyzed in our Monte Carlo study. First, completely ignoring MMS noise effects, we fit a Heston model to the data using both estimation methods and compare the results. Then, we try to estimate the MMS noise level in the data and see how the parameter estimates changes when we account for MMS noise effects. We are well aware that the Heston model might not fit the chosen data well. This is however not the purpose of this exercise, the application to data should rather be seen as a check of what happens when the estimation methods are used to fit a (possibly misspecified) model to real data. The empirical application is also an investigation of how different choices, such as MMS noise correction and the choice of predictor space in the flexible PBEF based method, might affect the parameter estimates.

4.1. **Data description**

For our empirical illustration we use five-minute returns for the DM/ $ spot exchange rate covering the period from October 1, 1992 through September 30, 1993. The data consist of 288 daily five-minute returns for each of the 260 trading days in our sample, yielding a total of 74880 five-minute returns.

As a first inspection of the data characteristics, we consider the empirical autocorrelation functions for the squared five-minute returns, reported in Figure 3. 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 if the two estimation methods will yield similar parameter estimates or 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).
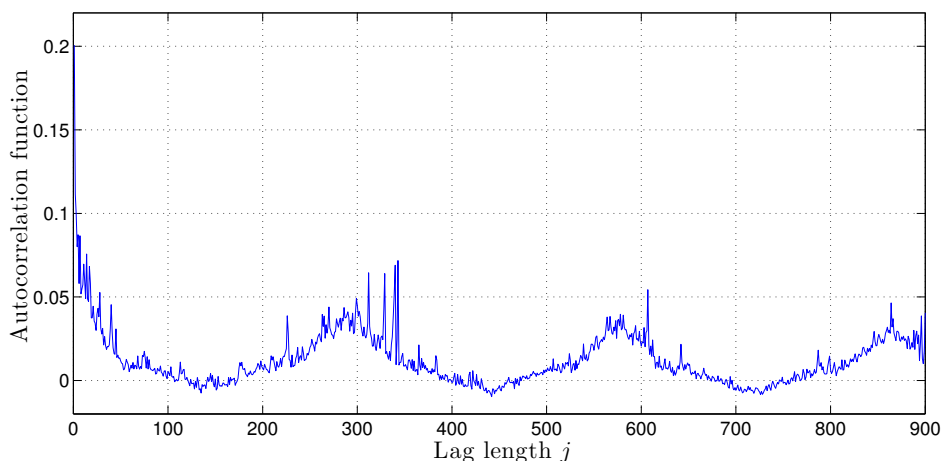

Figure 3: Autocorrelation function for the squared five-minute returns on the DM/$ the exchange rate.

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 five-minute returns are aggregated into the daily realized measures used for constructing the estimator. This same does not apply for the estimator based on PBEFs, since this estimator is based directly on the squared five-minute 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 five-minute returns using a non-parametric kernel regression[13].

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 five-minute interval in which the observation falls. Finally, the squared returns are normalized such that the overall variance of the squared returns remain unchanged. Figure 4 displays the autocorrelation function for the adjusted data. From the figure it is clear that
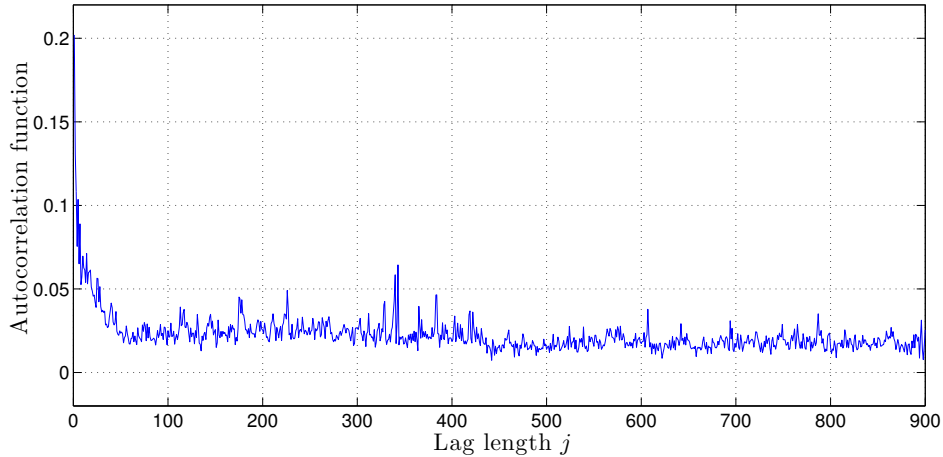

Figure 4: Autocorrelation function for the adjusted squared five-minute returns on the DM/$ the exchange rate.

the intra-daily periodicity has been removed. It is however still evident that the autocorrelation function is not exponentially decaying, rendering the Heston model a poor model choice. There also seems to be a need for at least a two factor SV model in order to properly capture the dynamics of 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 prediction 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 three different choices of basis elements spanning the predictor space. The three cases we consider corresponds to having 30 min., 12 hours and 24 hours between each of the basis

---

[13]We use the fit function in MATLAB with a smoothing spline, setting the smoothing parameter equal to 0.001

elements.[14] The same simple choice of weight matrix, as used in the Monte Carlo study, is employed when constructing the PBEFs. As in our Monte Carlo study, both the case where the possible presence of MMS noise is ignored and the case where it is corrected for are considered.

4.2.1. *Results for estimation methods without noise correction*

The parameter estimates from the two estimation methods without noise correction are reported in Table 7. For the PBEF based estimation method the *lag* variable equals 6 (6x5min lag between the basis elements) if we are in the first case and the *laq* variable equals 144 or 288 depending on whether we are in the second or the third case. Table 7 also shows how the various estimated Heston models fit the mean, variance and first order autocorrelation of the squared adjusted returns and daily IV. Since daily IV is not observable the sample moments of IV are approximated by $E[RV]$, $\text{Var}(RV) - \gamma$ and $\text{Corr}(RV, 1) * \left(\text{Var}(RV)/(\text{Var}(RV) - \gamma)\right)$ respectively, where γ is the bias correction term insuring that $E[IV^2] = E[RV^2] - \gamma$.

The estimated values for the long-run mean of volatility, reported in the second column of Table 7, do not vary much with the choice of estimation method. Not surprisingly, the fourth column in the lower part of Table 7 reveals that the estimated value of α are, in all cases, close to the unconditional mean of daily RV. The best fit to the unconditional mean of daily RV, and hence also to the unconditional mean of the squared five-minute returns, are obtained by the PBEF based estimation methods. When the *lag* variable equals 6 we obtain the closest fit, and as the lag value increases the parameter estimate for α decreases. For the other two parameters there are large variation in the parameter estimates obtained using the different estimation methods. For the GMM based approaches, the estimate of the mean reversion parameter κ decreases and the estimate for the volatility of volatility parameter increases when the naive DEC is employed. The use of DEC also yields results more in line with those reported in Bollerslev and Zhou (2002) for the bigger data set spanning the period December 1, 1986 through December 1 1996. Regarding the PBEF based estimation method, the need for a multifactor volatility seems to be apparent from the estimated values for the mean-reversion parameter reported in Table 7. When the predictor space only covers a short period ,($lag = 6$), we capture the fast decaying part in the beginning of the autocorrelation function from Figure 4, hence the estimated value of κ is large. In contrast, as the lag length between the predictor variables increase the estimated value of κ decreases drastically, since more emphasis is put on fitting the more

---

[14]That is, in the first case we let the predictor space we spanned by $Y_{i-1}^2, Y_{i-7}^2, Y_{i-13}^2$ and a constant. In the second case we choose $Y_{i-1}^2, Y_{i-145}, Y_{i-289}$ and a constant as the basis elements of the predictor space and finally in the third case we try to capture some of the more persistent dynamics by letting the predictor space span a period of two days by using $Y_{i-1}^2, Y_{i-289}^2, Y_{i-577}^2$ and a constant as basis elements.

persistent dynamics of the autocorrelation function. In all three cases, the estimates for κ are much larger than those obtained by the GMM estimation procedure. The same holds true for σ, and the estimated value of σ also decreases as the *lag* variable increases. As a robustness check we separate out the parameter (κ) that controls the dynamic behavior of the data from those (α and σ) that govern the stationary distribution, by fitting the Heston model with the following time-changed CIR specification for the variance process to the adjusted data

$$\mathrm{d}v_t = \kappa(\alpha - v_t)\mathrm{d}t + \sigma\sqrt{v_t}\mathrm{d}B_{\kappa t}, \text{ where } \mathrm{d}B_{\kappa t} \overset{\mathcal{L}}{=} \sqrt{\kappa}\mathrm{d}B_t.$$

The advantage of using this specification is that the stationary distribution of the variance process becomes the Gamma distribution with shape parameter $2\alpha\sigma^{-2}$ and scale parameter $2\sigma^{-2}$, hence the distribution no longer depends on κ. This also mean that the second order moments of the returns that we use for constructing the PBEFs change since we now have $\xi(\theta) = \alpha$, $\omega(\theta) = \frac{\alpha\sigma^2}{2}$, and $r(u;\theta) = e^{-\kappa u}$. The obtained parameter estimates from fitting the time-changed model exactly match those in Table 7. Changing the dimension of the predictor space to 6, $(q = 5)$, did not influence the parameter estimates either.

Table 7: Estimation results for the estimation methods without noise correction.

Estimation method	κ	α	σ	γ	FC	ω^2
Methods without noise correction						
GMM with *RV*	0.3651	0.8032	0.3937	-	-0.4314	-
GMM with *RV* and DEC	0.2621	0.8285	0.5833	-0.1217	-0.0941	-
PBEF with $q = 3$ and $lag = 6$	26.922	0.9070	8.5680	-	24.573	-
PBEF with $q = 3$ and $lag = 144$	2.1550	0.8984	2.0557	-	0.3537	-
PBEF with $q = 3$ and $lag = 288$	1.1689	0.8888	1.5081	-	0.1967	-

Estimation method	$E[Y_i^2]$	$Var(Y_i^2)$	$Corr(Y_i^2,1)$	$E[IV]$	$Var(IV)$	$Corr(IV,1)$
Methods without noise correction						
Sample moments, *RV*	0.003149	1.0e-04	0.2020	0.9070	0.4452	0.4452
GMM with *RV*	0.002789	2.2e-05	0.0945	0.8032	0.1515	0.7898
GMM with *RV* and DEC	0.002877	3.6e-05	0.1799	0.8285	0.4937	0.8429
PBEF with $q = 3$ and $lag = 6$	0.003149	6.3e-05	0.2150	0.9070	0.0885	0.0193
PBEF with $q = 3$ and $lag = 144$	0.003119	5.1e-05	0.2057	0.8984	0.4821	0.3075
PBEF with $q = 3$ and $lag = 288$	0.003086	5.0e-05	0.2065	0.8888	0.6070	0.4953

The table reports the parameter estimates from fitting the Heston model to the DM/$ spot exchange rate data from October 1, 1992 to September 30, 1993. The variable *FC* denotes the Feller condition, $\sigma^2 - 2\kappa\alpha$, and 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. Since *IV* is not observable the sample moments of *IV* is constructed using *RV*.

For all of the three considered cases, the parameter estimates from the PBEF based estimation methods violate the Feller condition. This is not surprising since we expected the Heston

model to provide a poor fit to the exchange rate returns. The Feller condition actually holds when the GMM estimation method is employed, but only barely when DEC is used. The model misspecification is also highlighted by the negative estimate of the correction factor γ, indicating that $RV_{t,t+1}^2$ is a downwards biased estimate of $IV_{t,t+1}$, which is not the case. The presence of MMS noise could also influence the estimated values of γ. This will be investigated in the next subsection. From the first three rows in the lower part of Table 7 the usefulness of DEC is again highlighted, as the obtained parameter estimates provide better fit to the sample moments, except when it comes to fitting $\text{Corr}(IV, 1)$. Not surprisingly, it is also evident from Table 7 that the PBEF based estimation procedures yields parameter estimates that produce better fits to the mean, variance and first order autocorrelation of the squared adjusted returns. The mean and variance is best matched if we consider the short time horizon, $lag = 6$, whereas the first order autocorrelation is slightly better matches if longer horizons, $(lag = 144, 228)$, are used.

If the aim is to match the moments of daily IV, then the results vary drastically with the choice of time span in the predictor space. For all three considered choices of the lag variable the mean is matched fairly well. Letting $lag = 6$ results in a poor fit to the sample variance and first order autocorrelation, but with $lag = 144, 288$ the fits improve drastically and are in fact better than when the GMM estimation method is used.

4.2.2. *Results for estimation methods with noise correction*

As in our Monte Carlo study, we now consider how noise correcting the estimation procedures affects the parameter estimates. The GMM estimation procedure is once again noise corrected by using the time series of RK as input for the estimation method. 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.

After computing the daily realized kernels we compute a non-parametric estimate of the daily MMS noise variance using the formula favored in Barndorff-Nielsen et al. (2008a)

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

with RK and RV constructed using the intra-daily adjusted five-minute returns and where $\hat{\omega}^2 = RV/2n$. In our case $n = 288$ and the overall estimate of the MMS noise variance is found by averaging over the 260 daily estimates. For the adjusted exchange rate data we find

$\breve{\omega}^2_{avg} = 7.3629e - 04$. This means the MMS noise has a standard deviation of 0.0271%, which are well in line with the values found in Bessembinder (1994) and those used in Aït-Sahalia et al. (2005).

The PBEF estimation procedure is noise corrected according to the assumption of Gaussian i.i.d noise. It may not hold in practice and this highlight a possible pitfall of using PBEFs - one has to specify the structure of the MMS noise in order to compute the PBEFs. We consider both including the noise variance ω^2 as a fourth parameter to be estimated (referred to as 1 step PBEF) and the case where it is held fixed at the non-parametric estimate found above (referred to as 2 step PBEF). The results of these procedures, as well as the parameter estimates obtain by using the GMM estimator based on the time series of RK, are reported in Table 8.

Table 8: Estimation results for the estimation methods with noise correction.

Estimation method	κ	α	σ	γ	FC	ω^2
Methods with noise correction						
GMM with RK	0.2097	0.6949	0.3799	-	-0.1470	-
GMM with RK and DEC	0.2360	0.6673	0.5167	-0.0569	-0.0481	-
PBEF with $q = 3$ and $lag = 6$, 1 step	25.490	0.9070	8.3020	-	22.678	0.0000
PBEF with $q = 3$ and $lag = 144$, 1 step	0.9776	0.2374	1.5707	-	2.0029	0.0011
PBEF with $q = 3$ and $lag = 288$, 1 step	0.8940	0.3527	1.5552	-	1.7881	9.30e-04
PBEF with $q = 3$ and $lag = 6$, 2 step	21.539	0.4829	8.9102	-	58.587	7.36e-04
PBEF with $q = 3$ and $lag = 144$, 2 step	1.8741	0.4743	2.2533	-	3.2996	7.36e-04
PBEF with $q = 3$ and $lag = 288$, 2 step	1.0261	0.4646	1.6632	-	1.8130	7.36e-04

Estimation method	$E[Y_i^2]$	Var(Y_i^2)	Corr$(Y_i^2, 1)$	$E[IV]$	Var(IV)	Corr$(IV, 1)$
Methods with noise correction						
Sample moments, RK	0.003149	1.0e-04	0.2020	0.7137	0.3380	0.4471
GMM with RK	0.003885	3.9e-05	0.1021	0.6949	0.2233	0.8717
GMM with RK and DEC	0.003789	4.3e-05	0.1300	0.6673	0.3493	0.8571
PBEF with $q = 3$ and $lag = 6$, 1 step	0.003149	6.3e-05	0.2152	0.9070	0.0924	0.0204
PBEF with $q = 3$ and $lag = 144$, 1 step	0.003024	3.0e-05	0.2058	0.2374	0.2218	0.5498
PBEF with $q = 3$ and $lag = 288$, 1 step	0.003085	3.6e-05	0.2058	0.3527	0.3380	0.5763
PBEF with $q = 3$ and $lag = 6$, 2 step	0.003149	5.1e-05	0.2156	0.4829	0.0788	0.0243
PBEF with $q = 3$ and $lag = 144$, 2 step	0.003119	4.3e-05	0.2059	0.4743	0.3759	0.3487
PBEF with $q = 3$ and $lag = 288$, 2 step	0.003086	4.2e-05	0.2066	0.4646	0.4576	0.5354

The table reports the parameter estimates from fitting the Heston model to the DM/$ spot exchange rate data from October 1, 1992 to September 30, 1993. The variable *FC* denotes the Feller condition, $\sigma^2 - 2\kappa\alpha$, and 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. Since *IV* is not observable the sample moments of *IV* is constructed using *RK*.

As expected from the results in the Monte Carlo study, the long run mean of volatility parameter, α, drops significantly when we correct for MMS noise in all of the considered cases. However, the parameter estimate of α now varies a lot across the different estimation methods employed. One explanation for this might be that the i.i.d Gaussian noise assumption un-

derlying the construction of the PBEFs does not hold. The mean reversion parameter, κ, also behaves like in the Monte Carlo study and drops when we try to correct for the presence of MMS noise. As before the value of the *lag* variable influence the estimate of κ dramatically. The drops in the two drift parameters are most pronounced when PBEFs is used for the estimation, especially when ω^2 is included in the parameter space. Only when the 2 step PBEF based method is used does σ increase compared to the case without noise correction. In the other cases σ drops, in contrast to what was expected from the results in our Monte Carlo study. As for the estimate of ω^2, it is either driven to zero when $lag = 6$ or takes on values higher than $\breve{\omega}_{avg}^2$ when $lag = 144, 288$. When the noise variance, ω^2, is included as a fourth parameter to be estimated the minimization of $G_n(\theta)'G_n(\theta)$ becomes extremely sensitive to the choice of starting values and often a local minimum is reached. In contrast, the optimization routine becomes much more stable when the noise variance is fixed at the non-parametric estimate.

Table 8 also report the implied fit to the moments of the adjusted squared returns and *IV* obtained using the different estimation methods. The moments of the adjusted squared returns implied by the fitted models are computed using the formulas in section 3.3 with $\omega^2 = \breve{\omega}_{avg}^2$ when the GMM estimation method and the 2 step PBEF are used, and estimated within the model when the 1 step PBEF is used. The model implied moments of *IV* are constructed as in the case without noise correction, only this time *RK* is used as a proxy for *IV*. Even though the assumption of i.i.d. Gaussian noise underlying the construction of the noise corrected PBEF might not hold, then the parameters estimates from the PBEF based estimation methods provide the best overall fit to the moments of the adjusted squared returns. Using *RK* as input int the GMM based estimation procedure improved the fit to $\text{Var}(Y_i^2)$ both with and without DEC, but for the other two moments the fit is better without noise correction. Except for the fit to $\text{Var}(Y_i^2)$ when $lag = 6$, the 2 step PBEF provides the best overall fit to the moments of Y_i^2. The best fit to $E[IV]$, is obtained using the GMM estimation method, but as in the case without noise correction the PBEF based method with $lag = 144, 288$ provides fits to the variance of *IV* comparable with those from the GMM based methods and are actually better at fitting the first order autocorrelation of *IV*.

Over all, the choice of estimation method depends on whether you want to fit the intra-day squared returns well or if the aim is fitting the moments of the daily integrated volatility. For the considered data set, where there seems to be noise present, the 2 step PBEF based method with $lag = 144, 288$ produces quite good fits to both the moments of Y_i^2 and *IV*. Whereas, the moments implied by the GMM based method only reasonably matches the mean and variance, and provide poor fits to the first order autocorrelation of the the two time series.

The aim of this section was never to find the best model for the exchange rate data, but in-

stead to investigate how the two different estimation methods handle real data with possible model misspecification. Here 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 relevance of allowing for several volatility factor is best highlighted by the PBEF based estimation method.

5. **Conclusion and Final Remarks**

We reviewed and detailed the general theory underlying PBEFs and explicitly constructed PBEFs for parameter estimation in the Heston model. Implementation issues and the link to other estimation methods, such as the link between optimal GMM estimation and the optimal PBEF, were discussed. The finite sample performance of the estimator based on PBEFs was investigated in a Monte Carlo study and compared to that of the GMM estimator from Bollerslev and Zhou (2002). Both the case with and without the inclusion of additive i.i.d. MMS noise were considered. In the no MMS noise setting, there are gains to be made from using PBEFs in terms of bias and the RMSE is also lower for the mean reversion parameter, κ.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. The two estimation methods were then extended to handle i.i.d. MMS noise by basing the GMM estimator on a time series of realized kernels from Barndorff-Nielsen et al. (2008a) and recalculating the moments needed for constructing the PBEFs. Only the PBEF based estimator had a performance comparable with the no MMS noise setup and once again seemed to outperform the GMM based estimator in the considered setting.

However, one concern regarding the application of the PBEF based method to real data is the sensitivity towards intra-daily dynamics, like intra-daily periodicity in volatility. Of course, one could adjust for this periodicity as we do in our empirical application, but if modeling the observed data is the aim of the study the GMM based estimator might be preferred. In the empirical application we investigated how the two estimation methods handle possible model misspecifaction and how the parameter estimates change when MMS noise correction is employed. The parameter estimates from the two estimation methods were quite different. The choice of predictor space influenced the estimates of κ and σ significantly, revealing the need for a multifactor volatility model which was not apparent from the use of the GMM based estimator. The study confirmed the presence of MMS noise in the DM/\$ exchange rate data and highlighted a possible pitfall when using PBEFs. When the data is contaminated with MMS noise, the structure of the noise has to be specified for the PBEFs to be computable. This is not necessary when the GMM estimator is used. Not surprisingly, our empirical study

showed that the model implied fit to the moments of the adjusted squared returns were best when the PBEF based estimator was used. Actually, when the predictor space spanned longer horizons, the fit to moments of daily *IV* were better than those obtained using the GMM based estimator.

It would be of interest to see how the estimation method based on PBEFs perform if we leave the assumption of i.i.d noise. This would however complicate the construction of the MMS noise corrected PBEFs. A solution to this potential problem could be to try 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, the importance of using the optimal weight matrix when basing inference on PBEFs in different settings, as well as how to simulate or approximate the optimal weight, could also be a topic for further research.

References

Andersen, T. G. and T. Bollerslev (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance 4*, 115–158.

Andersen, T. G., H.-J. Chung, and B. E. Sørensen (1999). Efficient method of moments estimation of a stochastic volatility model: A monte carlo study. *Journal of Econometrics 91*, 61–87.

Andersen, T. G., R. Davis, J.-P. Kreiss, and T. Mikosch (2009). *Handbook of Financial Time Series*. Springer.

Andersen, T. G. and B. E. Sørensen (1996). GMM estimation of a stochastic volatility model: A monte carlo study. *Journal of Business & Economic Statistics 14*, 328–352.

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

Aït-Sahalia, Y., P. A. Mykland, and L. Zhang (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies 18*, 351–416.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008a). Designing realized kernels to meausure the ex post variantion of equity prices in the presence of noise. *Econometrica 76*, 1481–1536.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008b). Realised kernels in practice: Trades and quotes. *Econometrics Journal 04*, 1–32.

Barndorff-Nielsen, O. E. and N. Shephard (2001). Non-Gaussian OU-based models and some of their uses in financial economics (with dicussion). *Journal of the Royal Statistical Society B 63*, 167–241.

Barndorff-Nielsen, O. E. and N. Shephard (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society B 64*, 253–280.

Bessembinder, H. (1994). Bid-ask spreads in the interbank foreign exchange markets. *Journal of Financial Econometrics 35*, 317–348.

Bollerslev, T. and H. Zhou (2002). Estimating stochastic volatility diffusions using conditional moments of integrated volatility. *Journal of Econometrics 109*, 33–65.

Bradley, R. C. (2005). Basic properties of strong mixing conditions: A survey and some open questions. *Probability Surveys 2*, 107–144.

Brockwell, P. J. (2001). Lévy-driven CARMA processes. *Ann.Inst.Statist.Math 53*, 113–124.

Corradi, V. and W. Distaso (2006). Semi-parametric comparison of stochastic volatility models using realized measures. *Review of Economic Studies 73*, 635–667.

Cox, J. C., J. E. Ingersoll, and S. A. Ross (1985). A theory of the term structure of interest rates. *Econometrica 53*, 385–408.

Dacorogna, M., U. Müller, R. Nagler, R. Olsen, and O. Pictet (1993). A geographical model for the daily and weekly seasonal volatility in the foreign exchange market. *Journal of International Money and Finance 12*, 413–438.

Eraker, B. (2001). Markov Chain Monte Carlo analysis of diffusion models with application to finance. *Journal of Business and Economic Statistics 19-2*, 177–191.

Forman, J. L. and M. Sørensen (2008). The Pearson diffusions: A class of statistically tractable diffusion processes. *Scandinavian Journal of Statistics 35*, 438–465.

Gallant, A. R. and G. Tauchen (1996). Which moments to match? *Econometric Theory 12*, 657–681.

Gourieroux, C., A. Monfort, and E. Renault (1993). Indirect inference. *Journal of Applied Econometrics 8*, S85–S118.

Hansen, P. R. and A. Lunde (2006). Realized variance and market microstructure noise. *Journal of Business and Economic Statistics 24*, 127–218.

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

Jacod, J., Y. Li, P. Mykland, M. Podolskij, and M. Vetter (2009). Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Processes and Their Applications 119*, 2249–2276.

Karlin, S. and H. M. Taylor (1975). *A First Course in Stochastic Processes*. Academic Press, New York.

Meddahi, N. (2002). A theoretical comparison between integrated and realized volatility. *Journal of Applied Econometrics 17*, 475–508.

Meddahi, N. (2003). ARMA representation of integrated and realized variances. *Econometrics Journal 6*, 335–356.

Melino, A. and S. M. Turnbull (1990). Pricing foreign currency options with stochastic volatility. *Journal of Econometrics 45*, 239–265.

Newey, W. K. and K. D. West (1987). A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*, 703–708.

Nolsøe, K., J. N. Nielsen, and H. Madsen (2000). Prediction-based estimating functions for diffusion processes with measurement noise. Technical reports no. 10, Informatics and mathematical modelling, Technical University of Denmark.

Renault, E. (2009). Moment-based estimation of stochastic volatiliy models. In T. G. Andersen, R. Davis, J.-P. Kreiss, and T. Mikosch (Eds.), *Handbook of Financial Time Series*, pp. 269–311. Springer.

Sørensen, M. (2000). Prediction-based estimating functions. *Econometrics Journal 3*, 123–147.

Sørensen, M. (2011a). Estimating functions for diffusion-type processes. *In Kessler, Lindner and Søerensen editors, Statistical Methods for Stochastic Differential Equations. Chapmann and Hall.*

Sørensen, M. (2011b). Prediction-based estimating functions: Review and new developments. *Brazilian Journal of Probability and Statistics 25*, 362–391.

Todorov, V. (2009). Estimation of continuous-time stochastic volatility models with jumps using high-frequency data. *Journal of Econometrics 148*, 131–148.

Todorov, V. and G. Tauchen (2006). Simulation methods for Lévy-driven CARMA stochastic volatility models. *Journal of Business and Economic Statistics 24*, 455–469.

Research Papers
2013

CREATES
Center for Research in Econometric
Analysis of Time Series

2013-06:   Debopam Bhattacharya, Pascaline Dupasand Shin Kanaya: Estimating the Impact of Means-tested Subsidies under Treatment Externalities with Application to Anti-Malarial Bednets

2013-07:   Sílvia Gonçalves, Ulrich Hounyo and Nour Meddahi: Bootstrap inference for pre-averaged realized volatility based on non-overlapping returns

2013-08:   Katarzyna Lasak and Carlos Velasco: Fractional cointegration rank estimation

2013-09:   Roberto Casarin, Stefano Grassi, Francesco Ravazzolo and Herman K. van Dijk: Parallel Sequential Monte Carlo for Efficient Density Combination: The Deco Matlab Toolbox

2013-10:   Hendrik Kaufmann and Robinson Kruse: Bias-corrected estimation in potentially mildly explosive autoregressive models

2013-11:   Robinson Kruse, Daniel Ventosa-Santaulària and Antonio E. Noriega: Changes in persistence, spurious regressions and the Fisher hypothesis

2013-12:   Martin M. Andreasen, Jesús Fernández-Villaverde and Juan F. Rubio-Ramírez: The Pruned State-Space System for Non-Linear DSGE Models: Theory and Empirical Applications

2013-13:   Tom Engsted, Stig V. Møller and Magnus Sander: Bond return predictability in expansions and recessions

2013-14:   Charlotte Christiansen, Jonas Nygaard Eriksen and Stig V. Møller: Forecasting US Recessions: The Role of Sentiments

2013-15:   Ole E. Barndorff-Nielsen, Mikko S. Pakkanen and Jürgen Schmiegel: Assessing Relative Volatility/Intermittency/Energy Dissipation

2013-16:   Peter Exterkate, Patrick J.F. Groenen, Christiaan Heij and Dick van Dijk: Nonlinear Forecasting With Many Predictors Using Kernel Ridge Regression

2013-17:   Daniela Osterrieder: Interest Rates with Long Memory: A Generalized Affine Term-Structure Model

2013-18:   Kirstin Hubrich and Timo Teräsvirta: Thresholds and Smooth Transitions in Vector Autoregressive Models

2013-19:   Asger Lunde and Kasper V. Olesen: Modeling and Forecasting the Volatility of Energy Forward Returns - Evidence from the Nordic Power Exchange

2013-20:   Anders Bredahl Kock: Oracle inequalities for high-dimensional panel data models

2013-21:   Malene Kallestrup-Lamb, Anders Bredahl Kock and Johannes Tang Kristensen: Lassoing the Determinants of Retirement

2013-22:   Johannes Tang Kristensen: Diffusion Indexes with Sparse Loadings

2013-23:   Asger Lunde and Anne Floor Brix: Estimating Stochastic Volatility Models using Prediction-based Estimating Functions