AARHUS
UNIVERSITY
BUSINESS AND SOCIAL SCIENCES
DEPARTMENT OF ECONOMICS AND BUSINESS

CREATES
Center for Research in Econometric Analysis of Time Series

ABC of SV: Limited Information Likelihood Inference in Stochastic Volatility Jump-Diffusion Models

Michael Creel and Dennis Kristensen

CREATES Research Paper 2014-30

Department of Economics and Business
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 8716 5515

ABC of SV: Limited Information Likelihood Inference in Stochastic Volatility Jump-Diffusion Models*

Michael Creel† and Dennis Kristensen‡

August 12, 2014

Abstract

We develop novel methods for estimation and filtering of continuous-time models with stochastic volatility and jumps using so-called Approximate Bayesian Computation which build likelihoods based on limited information. The proposed estimators and filters are computationally attractive relative to standard likelihood-based versions since they rely on low-dimensional auxiliary statistics and so avoid computation of high-dimensional integrals. Despite their computational simplicity, we find that estimators and filters perform well in practice and lead to precise estimates of model parameters and latent variables. We show how the methods can incorporate intra-daily information to improve on the estimation and filtering. In particular, the availability of realized volatility measures help us in learning about parameters and latent states. The method is employed in the estimation of a flexible stochastic volatility model for the dynamics of the S&P 500 equity index. We find evidence of the presence of a dynamic jump rate and in favor of a structural break in parameters at the time of the recent financial crisis. We find evidence that possible measurement error in log price is small and has little effect on parameter estimates. Smoothing shows that, recently, volatility and the jump rate have returned to the low levels of 2004-2006.

Keywords: Approximate Bayesian Computation; continuous-time processes; filtering; Indirect inference; jumps; realized volatility; stochastic volatility.

*We would like to thank the editors of the special issue and two anonymous referees for their comments and suggestions which greatly improved the manuscript. Earlier versions of this paper were presented at the BMRC-QASS conference on Macro and Financial Econometrics at Brunel University, and seminars at Lancaster University, University of Zürich, and University of Montreal. We thank all the participants for useful comments. All errors are ours. Creel acknowledges financial support from the Spanish Ministry of Economy and Competitiveness, through the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2011-0075), and from grants MICINN-ECO2009-11857 and SGR2009-578. Kristensen gratefully acknowledges research support by the Economic and Social Research Council through the ESRC Centre for Microdata Methods and Practice grant RES-589-28-0001, Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation, and the European Research Council (grant no. ERC-2012-StG 312474).

†Universitat Autònoma de Barcelona, BGSE (Barcelona Graduate School of Economics) and MOVE (Markets, Organizations and Votes in Economics). Dep. of Economics and Economic History, Edifici B, Universitat Autonoma de Barcelona, 08193 Bellaterra, Spain. E-mail: michael.creel@uab.es

‡University College London, CeMMaP (Centre for Microdata Methods and Practice, IFS) and CREATES (Center for Research in Econometric Analysis of Time Series, University of Aarhas). Dep. of Economics, UCL, Gower Street, London WC1E 6BT, United Kingdom. E-mail: d.kristensen@ucl.ac.uk.

JEL codes: C13, C14, C15, C33, G17.

1 Introduction

Continuous-time models are widely used in finance to describe the dynamics of asset prices since this framework gives researchers access to stochastic calculus as toolbox when deriving solutions to the models. In particular, this class of models allows for simple representations of prices of a variety of contingent claims. At the same time, densities and moments of most continuous-time processes used in theoretical finance cannot be expressed in closed-form thereby making implementation and estimation difficult. Additional complications arise from the fact that many of the important factors entering realistic asset pricing models are unobserved and so information about these have to be extracted using filtering methods.

One particular important class of asset pricing models is stochastic volatility (SV) diffusion models with jumps. These capture two important stylized facts of asset returns: First, because of changing beliefs and risk assessments amongst market participants, volatility of returns exhibit strong time-variation. Second, prices jump frequently due to news arrivals and other market surprises; see, for example, Barndorff-Nielsen and Shephard (2006) and Broadie, Chernov and Johannes (2009). The incorporation of these stylized facts into asset pricing models leads to the development of stochastic volatility jump-diffusion models in the pricing of options and other financial derivatives. An early example is the Heston (1993) model which has since been extended in a number of directions; see Bates (1996), Duffie et al. (2000) and Fang (2000). In econometrics, however, SV models tend to be formulated and estimated in a discrete-time setting at a daily frequency due to the aforementioned statistical and computational complexities associated with the continuous-time versions; see, e.g., Eraker, Johannes and Polson (2003), Kim, Shephard and Chib (1998). Unfortunately, it is unclear how estimated discrete-time models map into continuous-time equivalents, and so there seems to be a disconnect between the asset pricing and econometrics literature.

We propose a novel limited information method to estimate general continuous-time models and learn about latent states. The method allows for augmentation of daily returns data by so-called realized volatility (RV) measures in the estimation, leading to more precise estimates of parameters and better forecasts of future volatility and jumps. It also enables researchers to take into account first-step (intra-daily) estimation errors in the RV measures. The proposed method utilizes tools developed in the field of Approximate Bayesian Computation (ABC) to obtain computationally simple, yet precise estimators and filters. The main idea of ABC is to base inference on the likelihood of a set of sample statistics chosen by the researcher, thereby avoiding having to compute the full likelihood of all available data. This in general implies that not all information contained in data is used in the estimation, and so ABC estimators are limited information estimators (unless the researcher is so lucky as to choose a set of statistics that are sufficient for

the model in question).

ABC was originally developed in biostatistics for the Bayesian estimation of high-dimensional models as employed in, e.g., genetics; see Marin et al. (2012) for a review of this literature. It has since then been introduced to econometrics by, amongst others, Creel and Kristensen (2013) who referred to ABC estimators as indirect likelihood estimators. However, very little work has been done on the application of ABC in the estimation of dynamic models; some recent work in this direction include Calvet and Czellar (2012) and Dean, Singh, Jasra and Peters (2011). We complement these papers by showing how the ideas of ABC can be employed in the estimation and filtering of continuous-time models. We here focus on its implementation within the class of SV jump-diffusion models, but the methodology is applicable to many other classes of continuous-time models. By directly targeting continuous-time models and allowing for fast estimation of parameters and filtering and smoothing of spot volatility and jump intensity, our procedure can be used, for example, to perform on-line pricing of options and other derivatives.

There is a large, existing literature on estimation of continuous-time SV models. The early literature focused on estimation of these using only daily information (returns) and GMM-type or full likelihood-based estimators. GMM estimators were typically based on on Indirect Inference (II) methods and simulated method of moments (SMM); see Andersen, Benzoni, and Lund (2002), Carrasco, Chernov, Florens, and Ghysels (2007), Chernov, Gallant, Ghysels, and Tauchen (2003), and Monfardini (1998) for applications of these methods to asset pricing models. The advantages of ABC over II and SMM were demonstrated in Creel and Kristensen (2013) who showed that ABC estimators have, in general, smaller finite-sample variances when compared to the corresponding II and SMM estimators based on the same set of statistics. Moreover, ABC is computationally more efficient than these methods. Full likelihood inference were pursued by, amongst others, Johannes, Polson and Stroud (2009) and Jones (2003). The former employs particle filter algorithms to approximate the exact, unknown likelihood, while the latter advocates the use of MCMC methods. These algorithms require experience and fine-tuning, and are not guaranteed to deliver good estimates of the exact likelihood. This is particularly an issue for long data trajectories and high-dimensional state variables. Moreover, how to combine data observed at different frequencies in particle filters, e.g., daily and intra-daily observations, is not obvious. In contrast, our method is very simple to implement, can handle data at mixed frequencies, has no convergence issues, takes little time to run, and is fully parallelizable, so that it may easily be implemented using GPU computing (Creel and Zubair, 2012). This is of particular importance in the context of financial models, where there is a premium on obtaining an answer more or less in real time.

Neither of the above cited papers use intra-daily information in the estimation. In recent years, so-called realized volatility (RV) measures have become increasingly popular in the analysis of asset return dynamics. Realized volatility measures are model-free measures of daily integrated volatility computed using intra-daily returns. These measures are noisy and potentially biased estimators of the actual daily volatility due to,

for example, finite intra-daily sampling, jumps, and market microstructure noise. These have been used for a number of different purposes such as learning about features of daily volatility, but also forecasting future volatility and estimation of continuous-time SV models; see, for example, Andersen and Bollerslev (1998), Andersen et al. (2003), and Barndorff-Nielsen and Shephard (2002, 2006). However, most estimation methods using RV tends to ignore or side-step the fact that RV measures are noisy estimates of exact integrated volatility; see, amongst others, Bollerslev and Zhou (2002), Corradi and Distaso (2006), Kanaya and Kristensen (2010) and Todorov (2009). To be specific, their asymptotic results are developed in a setting where the intra-daily sampling error is assumed to vanish sufficiently fast so that the integrated volatility can be treated as directly observed. In practice, only a finite number of intra-daily observations are available, and so the intra-daily sampling error in the RV measures should ideally be accounted for. Moreover, these papers do not provide estimators of the latent states of the models, only the parameters. Dobrev and Szerszen (2009), Hansen, Huang and Shek (2012), Koopman and Scharth (2013) and Takahashi, Omori and Watanabe (2009) are exceptions in that they incorporate measurement errors and allow for filtering of latent states. However, they formulate their models in discrete time, and their model specifications tend to be quite simple in order for them to be able to numerically compute the likelihood.

While ABC has been widely used in empirical work, a number of questions regarding its implementation remain: First, for a given model, how should one choose the set of auxiliary statistics used for estimation? Second, how should the bandwidth parameter used in the computation of the simulated version be chosen? Third, can ABC, or related ideas, be used in filtering of dynamic latent variable models? We contribute to all three issues. First, we show how the idea of auxiliary models as introduced in II can straightforwardly be carried over to ABC; this method for selecting statistics complements existing ones as proposed in Fernhead and Prangle (2012). Second, we demonstrate that standard bandwidth selection methods developed for nonparametric kernel regression can be employed in selecting the bandwidth parameter in ABC. Third, we develop a simple-to-implement, fast filter and smoother that can be used to track latent variables over time.

As an empirical application, we use the estimators to learn about the returns dynamics of the S&P 500 index and show how our filter allows us to track stochastic volatility and jump intensity over time. We find evidence of the presence of jumps and support for a structural break in parameters at the beginning of 2008. In particular, during the recent financial crisis, volatility has a higher mean and variance, and the probability of jumps more than doubles, compared to the pre-crisis level. We find evidence that possible measurement error in log price is small and has little effect on parameter estimates. Our filtering and smoothing methods identify well trends in volatility and jumps associated with historic events. Smoothing shows that, recently, volatility and the jump rate have returned to the low levels of 2004-2006.

The remainder of the paper is organized as follows: Section 2 introduces a general

class of jump-diffusion process with stochastic volatility together with the basic concepts and tools of realized volatility estimation. In Section 3, we present a general class of parametric jump-diffusion SV models and show how realized volatility measures can be used to augment the number of measurement equations at a daily frequency and thereby providing more information regarding the model parameters. The limited information likelihood estimator of the resulting model is introduced in Section 4 where we also discuss choice of statistics and bandwidth. The limited information filter is presented in Section 5. Section 6 discusses the practical implementation of the proposed method, while Section 7 and 8 contain a simulation study and the empirical application, respectively. We conclude in Section 9.

2 Jump-Diffusions and Realized Volatility Measures

We will throughout assume that the log-price of a given asset, $p_t = \log(P_t)$, can be described by the following continuous-time jump-diffusion model,

$$p_t = p_0 + \int_0^t \mu_t dt + \int_0^t \sigma_t dW_{1,t} + \sum_{i=1}^{N_t} J_i, \tag{1}$$

where μ_t is the time-varying drift, σ_t is the volatility process, $W_{1,t}$ is a standard Brownian motion, J_i is the size of jumps, and N_t is a Poisson process with jump intensity λ_t. This is a highly general model where the only real restriction is the assumption of finite jump activity. We could allow for more general specifications of the jump component and, for example, model it in terms of a Poisson random measure with compensator; see, e.g. Todorov (2009). All subsequent methods and results remain correct for this more general model. However, in the simulation study and empirical study we focus on the above Poisson specification and so maintain this for notational simplicity.

We measure time in days so that our unit of measurement will be one day. In the (unrealistic) case of continuous price record, we can directly observe the so-called quadratic variation (QV) of the diffusive and discontinuous components of the price process within a given day. The quadratic variation over day t is defined as the sum of squared log-returns within that day observed over an increasingly fine time grid,

$$QV_{t+1} = \lim_{M\to\infty} \sum_{i=1}^{M} r^2_{t+i/M}(i/M), \qquad r_t(\delta) := p_t - p_{t-\delta}.$$

Under general conditions, QV_{t+1} can be expressed as the sum of the quadratic variation of the diffusive and jump component,

$$QV_{t+1} = \int_t^{t+1} \sigma_t^2 dt + \sum_{i=N_t}^{N_{t+1}} J_i^2. \tag{2}$$

We will refer to the continuous component as the integrated volatility (IV), $IV_{t+1} = \int_t^{t+1} \sigma_t^2 dt$, and let $JV_{t+1} = \sum_{i=N_t}^{N_{t+1}} J_i^2$ denote the variation due to jumps.

In practice, we only observe prices at discrete time points within the day and so QV_{t+1} is not directly observable. However, we can estimate it from the discretely observed prices,

$$RV_{M,t+1} = \sum_{i=1}^{M} r_{t+i/M}^2 \,(i/M)\,, \tag{3}$$

where for notational simplicity we have assumed that $M \geq 1$ prices are observed at equidistant time points within each day. Under weak regularity conditions (see Barndorff-Nielsen et al, 2006), as the sampling frequency $M \to \infty$, RV_{t+1} is asymptotically normally distributed and centered around QV_{t+1},

$$\sqrt{M}(RV_{M,t+1} - QV_{t+1}) \to^d N\left(0, V_t^{QV}\right),$$

where V_t^{QV} is the asymptotic variance. If one is interested in separately estimating IV_{t+1}, a number of alternative estimators are available such as so-called bipower estimators (Barndorff-Nielsen and Shephard, 2004), threshold estimators (Mancini, 2009) and nearest neighbor truncation estimators (Andersen, Dobrev, and Schaumburg, 2011). Let $\hat{IV}_{M,t+1}$ be any of these estimators, and let $\hat{JV}_{M,t+1} = RV_{M,t+1} - \hat{IV}_{M,t+1}$ be the associated estimator of the jump component of the over-all realized volatility. We can then obtain consistent estimators of both the diffusive and jump component of QV,

$$\hat{IV}_{M,t+1} = IV_{t+1} + u_{M,t}^{IV}, \qquad \hat{JV}_{M,t+1} = JV_{M,t+1} + u_{M,t}^{JV}, \tag{4}$$

where $u_{M,t}^{IV}$ and $u_{M,t}^{JV}$ capture the estimation errors. As $M \to \infty$, $u_{M,t}^{IV}$ and $u_{M,t}^{JV}$ will vanish, but for finite number of intra-daily observations, they will affect the RV measures.

Asset prices in general suffer from market microstructure noise due to discreteness of the price, and properties of the trading mechanism. In this case, we only observe a noisy measure $\hat{p}_{t,i}$ of the true, efficient price,

$$\hat{p}_{t,i} = p_{t+i/M} + \epsilon_{t,i}, \tag{5}$$

where $\epsilon_{t,i}$ captures the market microstructure noise. In this case, the above measures will be biased such that

$$\hat{IV}_{M,t+1} = B_{M,t}^{IV} + IV_{t+1} + u_{M,t}^{IV}, \qquad \hat{JV}_{M,t+1} = B_{M,t}^{JV} + JV_{t+1} + u_{M,t}^{JV}, \tag{6}$$

where $B_{M,t}^{IV}$ and $B_{M,t}^{JV}$ capture the biases while $u_{M,t}^{IV}$ and $u_{M,t}^{JV}$ contain the remaining intra-daily sampling errors, c.f. Zhang (2006). If the market microstructure noise is i.i.d. and independent of the efficient price then, for large M, $B_{M,t}^{IV} \simeq B_M^{IV}$ and $B_{M,t}^{JV} \simeq B_M^{JV}$ are time-invariant and only depend on the intra-daily sampling frequency and so are approximately constant across different days. One can remove these biases using noise-robust

estimators as developed in the literature; see, for example, Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008), Podolskij and Vetter (2009), and Zhang (2006).

All of the above-cited realized volatility and jump-size measures are model-free since they impose no parametric assumptions on the drift, volatility and jump process. They are also robust to intra-daily seasonalities in the volatility and jump processes which are integrated out since we work with daily measures. Thus, supposing we are not willing to take a stand on intra-daily seasonalities and the particular nature of the market microstructure noise, these provide useful daily statistics carrying information about the diffusive and the jump component. On the other hand, these measures are backward-looking and are on their own not informative about future volatility and (size and probability of) jumps. For the purpose of of forecasting volatility, jumps or other risk measures and for computing derivative prices, we have to impose additional structure on the price process. The next section introduces a general class of parametric jump-diffusion models that achieves this goal.

3 A Class of Parametric Jump-Diffusion Models

We here impose additional parametric restrictions on the volatility and jump components. This will allow us to forecast volatility and jumps. First, we assume that the jump size sequence J_i is independent of past price and volatility movements and follows some distribution known up to some unknown parameter θ, $J_i \sim F(J;\theta)$; this could be weakened to allow for the jump distribution to depend on s_t and p_t as in, for example, Chan and Maheu (2002). We assume that the Poisson process N_t has jump intensity λ_t which is allowed to be time-varying. The drift, volatility and intensity processes are driven by some underlying process s_t, $(\mu_t, \sigma_t, \lambda_t) = v(s_t;\theta)$ for some function $v(s;\theta)$; for example, $s_t = (h_t, \lambda_t)$ and$(\mu_t, \sigma_t, \lambda_t) = (\mu_0 + \mu_1 \exp(h_t), \exp(h_t), \lambda_t)$, where h_t is the log-spot volatility. To close the model, we model s_t as a Markov diffusion process solving

$$ds_t = a(s_t;\theta)\,dt + b(s_t;\theta)\,dW_{2,t}, \tag{7}$$

for some (vector) Brownian motion $W_{2,t}$ which is potentially correlated with $W_{1,t}$ to allow for leverage effects. Here, the drift, $a(s_t;\theta)$, and the volatility, $b(s_t;\theta)$, of s_t are known up to some parameter θ. We could in principle also have jumps in s_t but again maintain eq. (7) for simplicity. The complete model is fully parametric with all components specified up to the unknown parameter $\theta \in \Theta$. This is a very general set-up and covers most known asset price models found in the asset pricing literature, including the ones found in Broadie, Chernov and Johannes (2009), Bates (1996), Duffie et al (2000) and Heston (1993) which focus on affine versions.

The aim is to estimate the unknown parameter θ from observed log-prices. We will consider two sampling scenarios: (i) Only low-frequency, daily observations of p_t is available, $p_1, p_2, ..., p_n$, or (ii) high-frequency, intra-daily, observations are available from

which we construct daily measures of realized quadratic variation and integrated volatility $\left(RV_t, I\hat{V}_t\right)$, $i = 1, ..., n$. In the first case (i), we have only one measurement equation,

$$p_t \sim f_p\left(p_t | p_{t-1}, s_{t-1}; \theta\right),$$

where $f_p\left(p_t | p_{t-1}, s_{t-1}; \theta\right)$ is the transition density induced by eq. (1) together with the parametric restrictions above, and the state equation, $s_t \sim f_s\left(s_t | s_{t-1}; \theta\right)$, where $f_s\left(s_t | s_{t-1}; \theta\right)$ is the transition density induced by eqs. (7). In the second case (ii), the measurement equation for p_t is augmented by the two additional ones for the realized measures as given in either eq. (4), if noise is not present or noise-robust estimates are used, or eq. (6) if noise is present and not controlled for. These two additional measurement equations provide information about s_t and J_i and so can be used to obtain more precise estimates of the components of θ that concern the features of s_t and J_i.

Whether we are under observation scheme (i) or (ii), the estimation of the SV model is complicated by two facts: First, the conditional densities f_p and f_s are, in general, not available on closed form. Second, we do not observe s_t and so face a dynamic latent variable problem. This means that full-information likelihood inference is computationally very challenging. Finally, under observation scheme (ii), full likelihood-based methods require precise specification of the measurement equation for the realized volatility measures that leads to a Markov structure in order to employ existing methods for filtering and computation of likelihood. We instead develop a simpler estimation method in the next section that avoids these complications.

Given access to intra-daily returns, one could try to utilize all the information contained in these instead of, as we do, summarize this in terms of the chosen RV measures. This should in principle provide better estimates of the model parameters. However, at the same time, this would require us to specify intra-daily seasonalities in the volatility (see, e.g., Hasbrouck, 1999), the precise nature of the market microstructure noise, and other intra-daily features in data. By only using RV measures at daily frequencies, we to some extent can sidestep these issues: First, we avoid having to model intra-daily seasonalities (this does however create biases in the RV measures; see, .e.g., Andersen, Dobrev and Schaumburg, 2011). Second, we expect that the correct specification of the market microstructure noise is less important since this is averaged out in the computation of the RV measures and so, by the central limit theorem, will be approximately normally distributed.

4 Estimation by ABC

Due to the aforementioned challenges of full likelihood inference in continuous-time SV models, we choose to base inference on ABC. For convenience, we let $y_{M,t}$ denote the tth observation given M intra-daily observations in both of the two sampling scenarios introduced in the previous section. That is, either (i) $y_{M,t} := \hat{p}_t$ or (ii) $y_{M,t} = \left(\hat{p}_t, RV_{M,t}, I\hat{V}_{M,t}\right)$

for $t = 1, ..., n$, where $n \geq 1$ is the number of days that we have followed the given asset. Given data for, we choose a set of statistics for inferential purposes; these should summarize the relevant information in data regarding the jump-diffusion model. At the most general level, the set of statistics is a vector mapping taking data into a set, say $d \geq 1$, new variables, $Z_{M,n} = Z_n(y_{M,1}, ..., y_{M,n}) \in \mathbb{R}^d$. Given the chosen set of sample statistics, we then estimate the parameters through the posterior mean obtained from the likelihood of $Z_{M,n}$ as implied by the model, $f_{M,n}(Z_{M,n}|\theta)$, together with some prior $\pi(\theta)$ on the parameter space Θ. Formally, the estimator takes the form

$$\hat{\theta} = \int_{\Theta} \theta f_{M,n}(\theta|Z_{M,n})\, d\theta, \tag{8}$$

where $f_{M,n}(\theta|Z_{M,n})$ is the posterior distribution,

$$f_{M,n}(\theta|Z_{M,n}) = \frac{f_{M,n}(Z_{M,n}, \theta)}{f_{M,n}(Z_{M,n})} = \frac{f_{M,n}(Z_{M,n}|\theta)\pi(\theta)}{\int_{\Theta} f_{M,n}(Z_{M,n}|\theta)\pi(\theta)\, d\theta}.$$

One can choose to interpret $\hat{\theta}$ as a Bayesian estimator, as is frequently done in the ABC literature, or as a frequentist estimator, as done in Creel and Kristensen (2013). In the latter case, Creel and Kristensen (2013) show that if the chosen statistic satisfies a central limit theorem, $\sqrt{n}(Z_{M,n} - Z_M(\theta)) \to^d N(0, \Omega_M(\theta))$, for some limit $Z_M(\theta)$ that uniquely identifies θ and some asymptotic variance $\Omega_M(\theta)$, both depending on the true data-generating parameter value $\theta \in \Theta$, then, under additional, weak regularity conditions, $\hat{\theta}$ is consistent and asymptotically normally distributed:

$$\sqrt{n}(\hat{\theta} - \theta) \to^d N\left(0, \mathcal{I}_M^{-1}(\theta)\right), \tag{9}$$

where $\mathcal{I}_M(\theta) = \dot{Z}_M(\theta)' \Omega_M^{-1}(\theta) \dot{Z}_M(\theta)$ and $\dot{Z}_M(\theta) = \partial Z_M(\theta) / (\partial \theta')$. Observe here that we here keep the number of intradaily observations (M) fixed, and only let the number of long-span observations (n) diverge. This reflects the empirical application in the paper, where $M = 96$ is relatively small compared to $n = 3027$. In particular, the above asymptotics take into account the particular intra-daily sampling used to compute the RV measures. This is in contrast to the existing literature, where inference is based on the assumption that M diverges fast enough relative to n so that intra-daily sampling errors can be ignored; see, for example, Corradi and Distaso (2006) and Todorov (2009).

The two main assumptions for eq. (9) to hold is that (i) $Z_{M,n}$ is $\sqrt{n}$-asymptotically normally distributed, and (ii) $Z_M(\theta)$ identifies θ. One sufficient condition for (i) to hold is that $(R_{M,t}, E_{M,t})$, where $R_{M,t} = (r_{t+1/M}(1/M), ..., r_{t+1}(1/M))$ and $E_{M,t} = (\epsilon_{t,1}, ..., \epsilon_{t,M})$ contain the noise-free intra-daily returns and measurement errors during day t, respectively, is a stationary and mixing sequence with suitable moments, and that $Z_{M,n}$ are functions of sample moments of $(R_{M,t}, E_{M,t})$. If this is the case, we can appeal to CLT's for stationary and mixing sequences to obtain the desired result. Stationarity and mixing of $(R_{M,t}, E_{M,t})$ is in turn implied by, for example, the continuous-time stochastic

process$\{s_t : t \geq 0\}$ and the discrete-time process $\{E_{M,t} : t = 1, 2, ...\}$ being mutually independent and each of them being stationary and mixing. Most standard specifications of σ_t and λ_t have stationary solutions under suitable restrictions on the parameters; these restrictions are normally satisfied in data. The assumption of market microstructure noise being stationary also seems plausible. To formally model dependence between s_t and the market microstructure noise seems difficult since the former is a continuous time process while the latter is a discrete time one. Regarding the identification condition (ii), we show through simulations that for the model estimated in the empirical application our chosen statistics indeed do identify model parameters.

In the following, we very often suppress any dependence on M for notational simplicity since we treat it as given by the particular intra-daily sampling scheme that generated our sample, and so is considered fixed.

Similar to full-likelihood based Bayesian estimators, $\hat{\theta}$ is only available on closed form in a few special cases and in general we have to compute a simulation-based version of it. However, in contrast to full likelihood estimators, the computation of a simulated version of the BIL estimator is extremely simple due to the fact that it is based on a low-dimensional statistic, Z_n, instead of the full sample. We here follow the ABC literature and Creel and Kristensen (2013) and combine simulations and nonparametric regression techniques to obtain an approximate version of $\hat{\theta}$: First, obtain a swarm of i.i.d. draws (θ^s, Z_n^s), $s = 1, ...S$, from $f_n(Z_n, \theta)$, and then compute the following kernel regression estimator,

$$\hat{\theta}_S = \hat{E}_S[\theta | Z_n] = \frac{\sum_{s=1}^{S} \theta^s K_h(Z_n^s - Z_n)}{\sum_{s=1}^{S} K_h(Z_n^s - Z_n)}, \tag{10}$$

where $K_h(z) := K(z/h)/h$, $K : \mathbb{R}^d \mapsto \mathbb{R}$ is a so-called kernel function, for example, a standard normal density, and $h > 0$ is a bandwidth; see Li and Racine (2007) for an introduction. The draws can be obtained by following three steps:

1. Draw θ from $\pi(\theta)$.
2. Given the draw θ, simulate a trajectory $\{y_1(\theta), ..., y_n(\theta)\}$ from the jump-diffusion model evaluated at θ.
3. Compute $Z_n(\theta) = Z_n(y_1(\theta), ..., y_n(\theta))$.

Repeating above steps 1-3 S times, we obtain the desired swarm (θ^s, Z_n^s), $s = 1, ...S$, where $Z_n^s = Z_n(\theta^s)$. The main challenge is the second step of the above algorithm since this requires simulating from a continuous-time model. We resolve this issue by simulating from the corresponding Euler discretization: Choose a discretization step size $\overline{M} \geq 1$, and then compute iteratively, for any given draw of θ from $\pi(\theta)$, intra-daily log-prices

and latent states recursively using the Euler scheme,

$$\begin{aligned} p_{t+(i+1)/\bar{M}} &= p_{t+i/\bar{M}} + \mu\frac{1}{\bar{M}} + v\left(s_{t+i/\bar{M}}\right)\frac{1}{\sqrt{\bar{M}}}\varepsilon_{1,t,i} + J_{t,i}\Delta N_{t+i/M}, \\ s_{t+(i+1)/\bar{M}} &= s_{t+i/\bar{M}} + a\left(s_{t+i/\bar{M}};\theta\right)\frac{1}{\bar{M}} + b\left(s_{t+i/\bar{M};\theta}\right)\frac{1}{\sqrt{\bar{M}}}\varepsilon_{2,t,i}, \end{aligned}$$

where $\varepsilon_{t,i} = (\varepsilon_{1,t,i}, \varepsilon_{2,t,i})$, $i = 1, ..., \bar{M}$ and $t = 1, ..., n$, is a double-indexed sequence of i.i.d. random variables drawn from a bivariate Normal distribution; the pair $(\varepsilon_{1,t,i}, \varepsilon_{2,t,i})$ is potentially mutually correlated in order to capture leverage effects, J_i is i.i.d. and drawn from the jump-size distribution , and $\Delta N_{t+i/M}$ is drawn from a Poisson distribution with intensity $\lambda_{t+i/M}$. We have here suppressed the simulated values' dependence on θ for notational simplicity. Under sampling scheme (ii), once we have computed an approximate trajectory, $p_{t+i/\bar{M}}$, , $i = 1, ..., \bar{M}$, we obtain the corresponding realized measures using the formulae for RV_t and $I\hat{V}_t$ based on the intra-daily sampling frequency M that the actual data was observed at. The Euler discretization implies a bias, but this can be controlled by choosing the step size $\bar{M}$ large enough. Note that $\bar{M}$ should be chosen at least as large as M under sampling scheme (ii) in order for the simulated versions of RV_t and $I\hat{V}_t$ to match the sampling frequency used to compute their sample versions.

Using standard results for kernel regression estimators, we find that the simulated version satisfies, conditional on Z_n,

$$E\left[\hat{\theta}_S|Z_n\right] = \hat{\theta} + h^2 B\left(Z_n\right) + o\left(h^2\right), \tag{11}$$

$$Var\left[\hat{\theta}_S|Z_n\right] = \frac{1}{Sh^d}\frac{\|K\|^2 V\left(Z_n\right)}{f\left(Z_n\right)} + o\left(\frac{1}{Sh^d}\right), \tag{12}$$

where $d = \dim\left(Z_n\right)$, $V\left(Z_n\right) = Var\left[\theta|Z_n\right]$, $f\left(Z_n\right)$ is the density of Z_n, and $B\left(Z_n\right) = \sum_{j=1}^{d} B_j\left(Z_n\right)$ with

$$B_j\left(Z_n\right) = \frac{1}{2}\partial^2 E\left[\theta|Z_n\right] / \left(\partial^2 Z_{n,j}\right) + f\left(Z_n\right)^{-1}\partial E\left[\theta|Z_n\right] / \left(\partial Z_{n,j}\right)\partial f\left(Z_n\right) / \left(\partial Z_{n,j}\right);$$

see, e.g. Li and Racine (2007, Section 2.1). This highlights two important features of the ABC (SBIL) estimator: First, the bandwidth h has to be chosen to balance the bias and variance due to simulations and smoothing. Second, the simulated version suffers from a curse-of-dimensionality with, for a given number of simulations S, the variance due to simulations blowing up as d increases. In particular, the simple simulator proposed here does not work when Z_n is chosen as the full sample since in this case d is of the same order as n and so the curse of dimensionality becomes very severe. Instead, we wish to choose a moderate number of informative statistics in the estimation. One consequence of the above bias-variance expansion is that, as $h \to 0$ and $Sh^{d+2} \to \infty$ and conditional on Z_n,

$$\sqrt{Sh^d}(\hat{\theta}_S - \hat{\theta}) \to^d N\left(0, \frac{\|K\|^2 V\left(Z_n\right)}{f\left(Z_n\right)}\right), \tag{13}$$

That is, the simulated version converges towards the exact BIL estimator as the number of simulations diverges and the bandwidth sequence converges to zero at a suitable rate. The performance of the above estimator obviously depends on the choice of Z_n and the bandwidth h. Below, we discuss in turn methods for choosing these two.

4.1 Choice of statistics

Ideally one would like to choose Z_n as a sufficient statistic that fully summarizes all the relevant information contained in the sample. If such is used in the estimation, there is no efficiency loss and the ABC estimator is fully efficient. Unfortunately, to our knowledge, such are only available on closed form when data comes from a distribution within the exponential family (see, e.g., Grelaud, Robert, Marin, Rodolphe, and Taly, 2009). Outside this class, one can at most hope to find a set of statistics that approximate the sufficient statistic. The search for a suitable statistic can be done either in a model-specific manner or in a non-model based way. One can potentially mix the approaches, using some statistics motivated by the features of the model supplemented with Fourier-type terms.

4.1.1 Non-model based statistics

In the non-model based approach, the researcher uses (a relatively large number of) test functions that (approximately) span the unknown score function. Examples of test functions within this approach are Hermite polynomials (Bansal et al, 1994) and Fourier series (Carrasco et al, 2007); see also Fernhead and Prangle (2012) for some results on how this approach works together with ABC. We here discuss how this approach can be employed within the framework of continuous-time SV models. First note that the stochastic differential equation for the log-price process at a daily frequency can be rewritten as

$$r_t = \mu + \sqrt{IV_t}\varepsilon_{1,t} + \sum_{i=N_{t-1}}^{N_t} J_i, \tag{14}$$

where, $r_t = r_t(1) = p_t - p_{t-1}$ is the daily log-return, IV_t is the integrated volatility leading up to day t, and $N_t - N_{t-1}$ is the number of jumps within the $t-1$th day. In particular, it should be possible to learn about the volatility dynamics and jump-dynamics from the sequence of realized volatility measures, c.f. eq. (4). We therefore distinguish between the two sampling cases where realized volatility measures are available or not.

Consider first the case of sampling scheme (i): We here only have available daily returns to learn about the volatility dynamics. If jumps are not present, $E\left[r_t^2|\mathcal{I}_{t-1}\right] = IV_t$ where in turn IV_t should be informative about the dynamics of the spot volatility process. Thus, a natural choice is therefore to base the ABC on the autocorrelation structure that we observe in squared daily returns, r_t^2. When jumps are present, we need to decompose this autocorrelation function into its continuous and jump component. One could follow Barndorff-Nielsen and Shephard (2004) and choose as test functions differ-

ent powers of absolute returns, $|r_t|^p |r_{t-1}|^q$ for different values of $p, q > 0$ since these will allow us to get a rough measure of the diffusive and jump components. To formally show that these power moments identify the jump component, Barndorff-Nielsen and Shephard (2004) consider in-fill asymptotics where the time distance between observed prices shrinks. We here look at a fixed daily frequency, where to our knowledge there is no formal result showing that bipower moments of the form $E\left[|r_t|^p |r_{t-1}|^q\right]$ identify relevant jump characteristics for suitable choices of p and q. However, Barndorff-Nielsen and Shephard's asymptotic results indicate that, even if time distance between observations is fixed, bipower moments should be informative about jumps. Alternatively, one can use cosine functions as in, amongst others, Creel and Kristensen (2012).

In the case (ii) when, in addition to daily returns, we also have access to realized volatility measures, we can directly use these to learn about the volatility dynamics. In the case of no jumps, we can use sample moments of $\log RV_t$ to learn about the parameters governing the dynamics of s_t. When jumps are present, we combine sample moments of RV_t and $I\hat{V}_t$.

4.1.2 Model-based statistics

In the model-specific procedure for choosing the statistics, which is used in the simulation and empirical application, the researcher bases the choice of Z_n on the particular features of the model in question. For a given model, one chooses (a small number of) test functions that are believed to identify the parameters of interest. These should reflect the particular features of model and data. One particularly fruitful method for finding a suitable statistic originates from the literature on II. We here describe how this method can be employed within our setting.

We take as starting point a so-called auxiliary model. This should generate data that have the same main features as the model of interest, but at the same time should be simpler to work with from a computational point of view. Suppose that such a model has been chosen and, given data, can be represented by a (quasi-)likelihood function $g_n(Y_n|\psi) = g_n(y_1, ..., y_n|\psi)$ where ψ contains the parameters of the auxiliary model. We can then choose Z_n as the MLE of the auxiliary model,

$$Z_n = \arg\max_{\psi \in \Psi} g_n(Y_n|\psi), \tag{15}$$

with Y_n being the data generated by the original model. The auxiliary model should be chosen with an eye on the model of interest. In our case, we wish to choose an auxiliary model that generates time series data with features that are similar to those of a SV jump-diffusion model. Take as starting point the equation describing the evolution of prices at a daily frequency as given in eq. (14): In the case of no jumps, a natural approximation of this continuous-time model is a GARCH-type model,

$$r_t = \mu + \sqrt{v_t}\varepsilon_{1,t}, \qquad \varepsilon_{1,t} \sim N(0,1), \tag{16}$$

where v_t plays the role of IV_t. We would then like to build a simple model for v_t that mimics the exact one for IV_t, but is simpler. The precise approximate model depends on the specification of the spot volatility dynamics implied by eq. (7) and its implications for the dynamics of IV_t. Here, Bollerslev and Zhou (2002) provide some helpful guidelines. For example, in the case of a Heston model, a reasonable auxiliary specification would be a GARCH-type model,

$$v_t = \omega + \alpha\varepsilon_{1,t-1}^2 + \beta v_{t-1},$$

while in the case of log-volatility being a Gaussian process, an EGARCH model would be a good auxiliary model,

$$\log v_t = \omega + \alpha\left|\varepsilon_{1,t-1}\right| + \gamma\varepsilon_{1,t-1} + \beta\log v_{t-1}. \tag{17}$$

If jumps are included in the model, this could be captured by adding a jump component to eq. (16):

$$r_t = \mu_r + \sqrt{v_t}\varepsilon_{1,t} + J_t N_t, \tag{18}$$

where J_t has the distribution as given by the jump-diffusion model and N_t is a Bernoulli variable with jump probability λ_t. Again, the specification of the dynamics of the auxiliary jump probability λ_t should be guided by the specified dynamics of the jump intensity. For example, if it is specified as a CIR-type process, a good auxiliary model would be

$$\lambda_t = a + br_{t-1}^2 + c\lambda_{t-1}. \tag{19}$$

Finally, if we are under sampling scheme (ii), we can add equations to the above auxiliary model to describe the realized volatility measures. One would of doing so would be to use the realized GARCH model of Hansen et al (2012). For example,

$$\begin{aligned} v_t &= \omega + \alpha\varepsilon_{1,t-1}^2 + \beta v_{t-1} + \gamma RV_{t-1}, \\ RV_t &= a + bRV_{t-1} + cv_{t-1} + \varepsilon_{2,t}. \end{aligned} \tag{20}$$

Similarly, we can extend eq. (19) to incorporate the information contained in JV_t about the jump sizes and their probabilities. Collecting the parameters appearing in the auxiliary model outlined in eqs. (18)-(20) in ψ, we can compute the pseudo-MLE of ψ given data of daily returns and RV measures and use this as Z_n, c.f. eq. (15).

The above choice of Z_n involves numerical optimization. Recall that the simulated version of BIL takes the form given in eq. (10). For the computation of the ABC estimator, this requires computation of $Z_n^s = Z_n\left(\theta^s\right)$ for each draw θ^s from the posterior, where

$$Z_n\left(\theta\right) = \arg\max_{\psi\in\Psi} g_n\left(Y_n\left(\theta\right)|\psi\right).$$

This can be computationally burdensome. Instead, one may replace the maximization step with an integration step and choose $Z_n(\theta)$ as the Bayesian posterior mean of the

auxiliary model,

$$Z_n(\theta) = \int_\Psi \psi \hat{g}_n(\psi | Y_n(\theta))\, d\psi = \frac{\sum_{i=1}^N \psi^i g_n\left(Y_n(\theta) | \psi^i\right)}{\sum_{i=1}^N g_n\left(Y_n(\theta) | \psi^i\right)},$$

where ψ^i are i.i.d. draws from the chosen auxiliary prior on Ψ.

4.1.3 Reducing the number of statistics

Whether the statistics have been chosen using the model-based or non-model based approach, or a combination of the two, it may be the case that the initial collection of statistics is high-dimensional, which may lead to difficulties when performing the nonparametric fit upon which the ABC estimator is based. As shown in Creel and Kristensen (2013), the ABC estimator is first-order equivalent to the efficient GMM estimator based on Z_n. Thus, when the dimension of Z_n is large, the ABC estimator suffers from the same issues as GMM estimators with a large number of moment restrictions: The use of (too) many moment conditions can lead to poor finite sample performance of the GMM estimator. In addition, recognizing that nonparametric regression is needed to compute the ABC estimator, it is well known that nonparametric regression has biases and variances, the latter increasing with the dimension of the conditioning variables, c.f. equation (12). In the context of ABC, the number of simulations, S, serves as the sample size for the nonparametric fitting step, and so these additional biases and variances can be controlled by increasing the number of simulations. However, to control the computational requirements, it may be desirable to limit the dimension of Z_n, so that good results may be obtained with a reasonably small number of simulations. This again entails finding a means of selecting a "good" set of auxiliary statistics out of a larger body of possible statistics.

Given the first-order equivalence to GMM, the same methods that are used to select moment conditions for GMM estimation can be used to select statistics for ABC estimation. We here follow the recent literature on shrinkage-type GMM estimators (see, e.g., Caner and Zhang, 2014) and propose to use shrinkage to reduce the dimension of the initial set of statistics: Let δ be $d \times 1$ vector of zeros and ones, where a zero indicates that the corresponding auxiliary statistic is not used, and a one indicates that it is, and let $Z_n(\delta)$ be the corresponding vector of selected statistics. Let Δ be the set of all possible values of δ. This set as 2^d elements, a number which can be very large when a number of statistics are under consideration. The cross-validated set of statistics, with a penalty function to further encourage a parsimonious choice, is given by

$$\hat{\delta} = \arg\min_{\delta \in \Delta} CV(\delta), \qquad CV(\delta) = \sum_{s=1}^S \left\| \theta^s - \hat{E}_{-s}\left[\theta | Z_n^s(\delta)\right] \right\|^2 + w \|\delta\| \tag{21}$$

where $w > 0$ determines the penalty associated with an increase in the dimension of the

selected set, and $\hat{E}_{-s}\left[\theta|Z_n^s(\delta)\right]$ is the so-called leave-one-out estimator

$$\hat{E}_{-s}\left[\theta|Z_n^s(\delta)\right] := \frac{\sum_{t\neq s}\theta^t K_h\left(Z_n^t(\delta) - Z_n^s(\delta)\right)}{\sum_{t\neq s} K_h\left(Z_n^t(\delta) - Z_n^s(\delta)\right)},$$

using only the selected statistics. This is a non-differentiable minimization problem over many possible combinations. As simulated annealing is known to be able to solve the conceptually similar traveling salesman problem (Černý, 1985), we obtain an approximately optimal set of statistics by applying simulated annealing to approximately minimize the cross validation score.

Alternative methods for selecting statistics has received a good amount of attention in the ABC literature. Blum et al. (2013) provides a review of this work, with some new results based on use of regularization methods. The methods that have been used fall into two classes: best subset selection, and projection methods that define new statistics by combining statistics in some way to reduce dimension. Among papers that use projection, Fearnhead and Prangle (2012) explore using a training set of simulated observations, regressing parameters drawn from the prior on statistics computed using the simulated data, and then using the fitted values (a weighted index of the initial statistics that is an approximation of the posterior mean of the parameter) as a single statistic for ABC estimation, for each parameter in turn. This approach reduces to one the dimension of the initial large set of statistics. This may lead to an excessive loss of information, in that there may exist no single linear index that captures the information in the large body of initial statistics. A further limitation is the use of linear regression, which may perform poorly if the true relationship between the parameter and the informative statistics is nonlinear. Fearnhead and Prangle (2012) address this by using methods to endogenously choose the training set from a region of reasonably high posterior mass. Representative of papers that select a subset of statistics, Joyce and Marjoram (2008) propose an algorithm to determine whether a new statistic should be added to a previously selected set of statistics, based on approximations of the posterior odds ratio, computed using rejection-based ABC. If the new statistic is added, then removal of the previously selected statistics is contemplated. This proceeds until all statistics have been considered. This proposal appears to be complex and costly to implement if the number of initial statistics is large.

Some comments: First, it is not necessary to achieve the exact solution to the above optimization problem, a solution close to the minimum is adequate for the purpose of achieving dimension reduction. For this reason, rapid cooling and a limited number of trials may be used. Second, for the purpose of choosing which statistics to retain, we may use a relatively small subset of the simulations to do cross validation, as the informativeness of the statistics does not depend on the number of simulations. Third, cross validation uses ABC estimation in the same way as it is used for actual estimation, using nonparametric regression. This approach does not suffer from a potential loss of information due to use of linear regressions. Fourth, the output of the algorithm is the vector

of indices of the selected statistics. We retain the statistics for which the corresponding element of δ is one. We do not form a weighted index of the statistics. One can potentially run the above cross-validation for each parameter so that the selected statistics are specific to each parameter. In this case, the final selected set for estimation of the entire parameter vector can be defined as the union of the selected statistics for each parameter. Last, this algorithm is somewhat computationally demanding, but is is conceptually very simple and easy to program. It simply involves doing many ABC fits, and minimizing out of sample RMSE over the set of statistics, using simulated annealing.

4.2 Bandwidth selection

Standard implementations of ABC-type estimators tend to choose the bandwidth in a rather ad hoc fashion. However, as is well-known from the literature on nonparametric regression, kernel regression estimators are quite sensitive to the choice of h . In the context of computing the simulated BIL estimator given in eq. (10), we saw in eqs. (11)-(12) that there is a bias-variance trade-off in choosing h; choosing h too large or too small will lead to large biases and/or variances in $\hat{\theta}_S$ relative to the exact, but unknown ABC estimator, $\hat{\theta}$. Thus, there may be benefits to choosing h carefully.

Given that our implementation of the BIL estimator is simply a kernel regression estimator, we can use existing methods developed in this literature for selecting the bandwidth. Again, we propose to use cross-validation, where h is chosen as

$$\hat{h} = \arg\min_{h>0} CV(h), \qquad CV(h) = \sum_{s=1}^{S} \left\| \theta^s - \hat{E}_{-s}\left[\theta | Z_n^s\right] \right\|^2$$

where $\hat{E}_{-s}\left[\theta | Z_n^s\right]$ is the leave-one-out estimator,

$$\hat{E}_{-s}\left[\theta | Z_n\right] := \frac{\sum_{t \neq s} \theta^t K_h\left(Z_n^t - Z_n^s\right)}{\sum_{t \neq s} K_h\left(Z_n^t - Z_n^s\right)}.$$

It can be shown that, as $S \to \infty$, $\hat{h} \simeq h_{opt}$ where h_{opt} is the bandwidth sequence that minimizes the mean-square error (MSE) of the kernel regression estimator, see Li and Racine (2007, Theorem 2.3). One may use different bandwidths for the different components of $\hat{\theta}_{SBIL}$. This can be done by computing a cross-validated bandwidth for each individual parameter; that is, we minimize $CV_k(h) = \sum_{s=1}^{S} \left(\theta_k^s - \hat{E}_{-s}\left[\theta_k | Z_n^s\right]\right)^2$, for $k = 1, ..., \dim(\theta)$.

The above selection method is global in the sense that it minimizes the integrated MSE (IMSE). The IMSE is defined as the average MSE over all potential outcomes of Z_n, $IMSE = \int MSE(Z_n) f(Z_n) dZ_n$, where

$$MSE(Z_n) = E\left[\left\| \hat{\theta}_S - \hat{\theta} \right\|^2 \middle| Z_n \right].$$

Since we are only interested in obtaining a good estimate of $\hat{\theta} = E\left[\theta | Z_n\right]$, as evaluated at the particular value of Z_n that we have observed in data, there are potential gains from

using so-called local bandwidth selection methods. These are designed to choose the bandwidth to minimize the pointwise MSE at any given value of Z_n, $MSE(Z_n)$, and so should provide a better estimate of $\hat{\theta}$; see, e.g., Fan and Gijbels (1992, 1995), Schucany (1995), Fan et al. (1996), and Prewitt and Lohr (2006).

5 Limited Information Filtering and Smoothing

Once the model parameters θ have been estimated, it is often of interest to learn about the realized in-sample trajectories of the latent variables and to do forecasting. We propose a novel, simple method that, similar to the estimation procedure, relies on limited information. Collecting the state variables in $w_t = (p_t, s_t, \lambda_t)$, suppose we wish to learn about $g_{t+m} := g(w_{t+m})$ for some given function $g(w)$ and some forecast/filtering horizon $m \geq 0$. For example, if the goal is to forecast spot volatility m periods ahead, we choose $g_{t+m} = \sigma_{t+m}$. Alternatively, one may be interested in forecasting daily integrated volatility m periods ahead in which case $g_{t+m} = IV_{t+m}$. We wish to do so given the information available to us at time t, $\{y_1, ..., y_t\}$. Suppose we know the data-generating parameter value θ. We then propose to forecast g_{t+m} by

$$g_{t+m|t} := E_\theta [g_{t+m}|F_t],$$

where $F_t = F(y_t, ..., y_1) \in \mathbb{R}^{d_F}$ is some function of data and $E_\theta[\cdot|F_t]$ denotes expectations under the model evaluated at θ. Ideally, one would use all information and so choose $F_t = (y_t, ..., y_1)$, but, as $t \to \infty$, this becomes a high-dimensional computational problem. Instead, we propose to use only a smaller fixed subset, e.g., $F_t = (y_t, ..., y_{t-q})$, for some finite, fixed number of lags $q \geq 1$. Here, q should be chosen such that the dimension of the forecast information variable remains moderate, but at the same time large enough that the most relevant information in data is contained in F_t. This is similar to our limited information estimator, where a moderate sized statistic was used in place of the full data set, but this should still convey most of the relevant information in data. We discuss below "data-driven" method for choosing q.

The computation of $g_{t+m|t}$ is done by combining simulations and kernel regression methods, similar to the implementation of the ABC estimator: First, obtain a swarm of i.i.d. draws $(p_t^r(\theta), s_t^r(\theta), \lambda_t^r(\theta), RV_t^r(\theta), IV_t^r(\theta))$, $r = 1, ...R$, from the model evaluated at the estimated parameter value, $\theta = \hat{\theta}$ - these draws are obtained in a similar fashion as for the computation of the BIL - compute the corresponding simulated values of the forecast variable and conditioning variable, $(g_{t+m}^r(\hat{\theta}), F_t^r(\hat{\theta}))$, $r = 1, ...R$, and feed these into a kernel regression,

$$\hat{g}_{t+m|t} = \hat{E}_{\hat{\theta}}[g_{t+m}|F_t] = \frac{\sum_{r=1}^R g_{t+m}^r(\hat{\theta}) K_b\left(F_t^r(\hat{\theta}) - F_t\right)}{\sum_{r=1}^R K_b\left(F_t^r(\hat{\theta}) - F_t\right)}. \tag{22}$$

for some bandwidth $b > 0$; this should in general be chosen different from the one, h,

used in the computation of the SBIL. Under standard regularity conditions for kernel regression estimators, as $h \to 0$ and $Rh^{d_F+2} \to \infty$, we have, conditional on F_t and $\hat{\theta}_{BIL}$,

$$\sqrt{Rb^{d_F}}(\hat{g}_{t+m|t} - g_{t+m|t}) \to^d N\left(0, \frac{\|K\|^2 V\left(F_t|\hat{\theta}\right)}{f\left(F_t|\hat{\theta}\right)}\right), \tag{23}$$

where $d_F = \dim(F_t)$, $V(F_t|\theta) = Var_\theta[g_{t+h}|F_t]$ and $f(F_t|\theta)$ is the density of F_t when θ is the true parameter value; see Li and Racine (2007, Theorem 2.2). Instead of simulating R trajectories of length n, one could simulate one "long" trajectory of length R in order to construct the forecast $g_{t+m|t}$. If the fitted model is stationary and mixing, then the simulated forecast based on this long trajectory will still satisfy eq. (23), see Li and Racine (2007, Sec. 18.2). This method has some conceptual similarities to the reprojection method proposed by Gallant and Tauchen (1998), which involves evaluation of the moments of an estimated semi-nonparametric density. We believe that our proposed method is simpler to use in practice, as it is simply a kernel regression estimator.

The above method can also be used to construct confidence bands for the forecast (or provide median-based point forecasts). This can, for example, be done using kernel quantile regression,

$$\hat{q}_{t+m|t}(\tau) = \arg\min_g \sum_{r=1}^{R} K_b\left(F_t^r(\hat{\theta}) - F_t\right) \rho_\tau\left(g_{t+m}^r(\hat{\theta}) - g\right), \tag{24}$$

where $\rho_\tau(\cdot)$ is the so-called check-function as known from quantile regression, see Koenker and Bassett (1978). In particular, $\left[\hat{q}_{t+m|t}(0.025), \hat{q}_{t+m|t}(0.975)\right]$ will provide a consistent 95% forecast interval, while $\hat{q}_{t+m|t}(0.5)$ will be the median forecast of g_{t+h}. This corresponds to doing density forecasting since there is a one-to-one correspondence between the quantile range and the cumulative distribution function. The forecast interval can be adjusted to take into account the simulation error by using standard errors for nonparametric quantile regression estimators found in . One can also take into account additional errors contained in the forecast $\hat{g}_{t+m|t}$ due to the sampling and simulation error in $\hat{\theta}_{BIL}$ along the lines of Hansen (2006). Finally, one can also generalize the above idea to generate density forecasts of $g_{t+m} = g$:

$$\hat{f}(g|F_t) = \frac{\sum_{r=1}^{R} K_b\left(g_{t+m}^r(\hat{\theta}) - g\right) K_b\left(F_t^r(\hat{\theta}) - F_t\right)}{\sum_{r=1}^{R} K_b\left(F_t^r(\hat{\theta}) - F_t\right)}.$$

This could in turn be used for Value-at-Risk computations.

The method can be adjusted to do smoothing. In this case, we simply choose $F_t = F(y_1, ..., y_n)$ as a function of all data, not just observations up to time t. For example, we could choose F_t to include the first q leads and lags together with the concurrent observation, $F_t = (y_{t-q}, ..., y_t, ..., y_{t+q})$.

An important user choice for the above filter is the set of predictors F_t. One "data-driven" approach for choosing F_t is to use variable selection methods as used in nonpara-

metric regression; for example, Zhang (1991) advocates the use of cross-validation to inform the researcher about the choice of predictors: Consider for the sake of clarity the case where the set of potential candidate predictors takes the form $F_{t,q} = (y_t, ..., y_{t-q})$, $q = 0, 1, 2, ...$ and so we wish to choose q. To this end, for a given choice of forecast variable g_t, define the simulated cross-validation criterion $CV_g(b,q) = \sum_{r=1}^{R} \left\| g_{t+m}^r - \hat{E}_{-r}\left[g_{t+m}||F_{q,t}^r\right] \right\|^2$, where $\hat{E}_{-r}\left[g_{t+m}||F_{q,t}^r\right]$ is the leave-one-out version of the kernel regression estimator in eq. (22), which implicitly depends on the bandwidth b and Zhang (1991) then shows that the following method for choosing (b,q) is consistent

$$\left(\hat{b}, \hat{q}\right) = \arg \min_{h>0, q\geq 0} CV_g(b,q).$$

As a computationally simpler alternative to the above kernel filter and smoother, we also consider a linear forecasting model,

$$g_{t+m|t}^{LIN} = \beta_t' F_t,$$

where $\beta_t \in \mathbb{R}^{d_F}$ is chosen to minimize the mean-square error. Similar to before, β_t can be learned through simulations: Simulate $(g_{t+m}^r(\hat{\theta}), F_t^r(\hat{\theta}))$, $r = 1, ...R$, and compute

$$\hat{\beta} = \left[\sum_{r=1}^{R} F_t^r(\hat{\theta}) F_t^r(\hat{\theta})'\right]^{-1} \sum_{r=1}^{R} F_t^r(\hat{\theta}) g_{t+m}^r(\hat{\theta}).$$

As in linear regression models estimated by OLS, $g_{t+m|t}^{LIN}$ is the best linear predictor in the mean-square-error sense.

In the case where g_t is chosen as $g_t = \log QV_t$, $g_{t+m|t}^{LIN}$ is similar to the HAR-RV-J forecasting model proposed by Andersen, Bollerslev and Diebold (2007) for forecasting volatility when jump variation measures are available. The HAR-RV-J model takes the form

$$\log RV_{t,t+h} = \beta_0 + \beta_1 \log RV_t + \beta_W \log RV_{t,t-5} + \beta_M \log RV_{t,t-22} + \beta_J \log(jumps_t + 1) + e_t, \tag{25}$$

where $RV_{t,t+h} := (RV_{t+1} + RV_{t+2} + \cdots + RV_{t+h})/h$. Here, β_0, β_1, β_W, β_M and β_J can be learned directly from data without the need of simulations and so the HAR-RV-J model allows for reduced-form forecasting without the need of specifying the dynamics of the the underlying latent states. On the other hand, the HAR-RV-J model only allows forecasting of observed variables, while the proposed non-linear and linear smoothers and filters can be used to learn about latent states. In particular, note that the HAR-RV-J model in general only provides a forecast of the log-quadratic variation over the chosen time horizon h and ignores that RV_t is a noisy measure of the quadratic variation.The HAR_RV-J model can also be used to forecast the diffusive or the jump component of the quadratic variation by simple replacing $\log RV_{t,t+h}$ with $\log I\hat{V}_{t,t+h}$ and $\log J\hat{V}_{t,t+h}$,

respectvely. In the case where σ_t^2 is close to being constant, no jumps are present and a large number of intra-daily observations are available, the HAR-RV-J model should provide a reasonably good forecast of the spot volatility. However, when the volatility is time-varying, the sampling error in RV_t is large and/or jumps are present, we in general expect the HAR-RV-J model to do a poor job in forecasting integrated volatility and spot volatility.

6 Implementation in Practice

Next we turn to use of the above described methods, in the context of both Monte Carlo and empirical work with data on the S&P 500 equity index.

6.1 Specific model

The specific model used in the Monte Carlo study and the empirical application is a continuous time stochastic volatility model which potentially contains all of the following elements: non-constant drift, leverage, jumps with dynamic jump intensity, and measurement error. The true log price $p_t = \log(P_t)$ solves the model

$$dp_t = (\mu_0 + \mu_1(h_t - \alpha)/\sigma)\, dt + \sqrt{\exp(h_t)} dW_{1,t} + J_t dN_t. \tag{26}$$

In this equation, h_t is log volatility, J_t is jump size, and N_t is a Poisson process with time-varying jump intensity λ_t. If market microstructure noise is present, p_t is latent and we only observe $\hat{p}_t$ as given in equation (5), where we here assume the measurement errorsϵ_i, $i = 1, ..., n$, are i.i.d. $N\left(0, \sigma_\epsilon^2\right)$. Log-volatility is specified to follow a Vasicek model (and so is a Gaussian process),

$$dh_t = h_t + \kappa(\alpha - h_t)dt + \sigma\left(\rho dW_{1,t} + \sqrt{1-\rho^2} dW_{2,t}\right), \tag{27}$$

where $W_{1,t}$ and $W_{2,t}$ are two independent standard Brownian motions. Finally, the jump intensity process λ_t is modelled as

$$\lambda_t = 1(\lambda_t^* > 0) \text{ where } \lambda_t^* = \lambda_0 + \lambda_0\lambda_1(h_t - \alpha)/\sigma.$$

Thus, the jump intensity is a non-negative censoring of a latent process. This model is similar to the well-known log-Normal volatility model of Wiggins (1987); see also Chesney and Scott (1989), but with jumps similar to, for example, Andersen, Benzoni and Lund (2002). We here model the jump intensity in a different way compared to Andersen et al (2002) though. The parameters are interpreted as follows. Mean drift in log price is given by μ_0, to account for a general trend in prices. Drift is allowed to depend on volatility through μ_1, which is the marginal effect of log volatility as measured in units of standard deviation from its mean, which is α. This parameterization is intended to

ease interpretation of μ_1. It has the additional benefit of reducing the interaction effect of parameters upon statistics used to identify parameters, which helps for their individual identification. The standard deviation of the shock to log volatility is σ, and κ is the mean reversion parameter. The leverage parameter is ρ, which affects the correlation between returns and log volatility. The latent process λ_t^* has mean λ_0, and $\lambda_0\lambda_1$ is the marginal effect of log volatility, again in standard deviations from its mean, on the jump intensity. Thus, when log volatility is one standard deviation above its mean, jump intensity is $1 + \lambda_1$ times the intensity when log volatility is at its mean. Again, this parameterization is intended to allow for ease of interpretation of λ_0 and λ_1. It allows the jump intensity to be zero at times, when log volatility is low, and also to increase when volatility increases. We give some additional comments regarding the parameterization following the presentation of the estimation results. Jump sizes, conditional on the occurrence of a jump, are independent and conditionally normally distributed: $J_t \sim N(\mu_J, \sigma_J^2)$. We collect the 11 parameters of the model in the vector $\theta = (\mu_0, \mu_1, \alpha, \kappa, \sigma, \rho, \lambda_0, \lambda_1, \mu_J, \sigma_J, \sigma_\epsilon)$.

6.2 Data

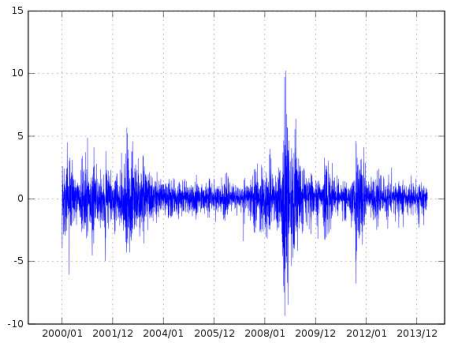
Our data for the S&P 500 index is taken from the Oxford-Man Institute's realized library version 0.2 (see Heber, Lunde, Shephard and Sheppard, 2009) which includes daily measures of returns (r) in percentage terms, two measures of RV ($RV5$ and $RV10$), computed using returns at the 5 minute and 10 minute frequencies, and realized bipower variation (BV) (Barnsdorff-Nielsen and Shepard, 2006), computed using returns at the five minute frequency. The latter is the particular choice for $I\hat{V}$ used in the following. The use of two RV measures are meant to improve on estimation when market microstructure noise as explained earlier. At the time we downloaded the data, stock index returns and the realized measures were available from Jan. 03, 2000 through Jun. 03, 2014. This data source does not include the high frequency intra-day returns that were used to compute RV and BV, only daily measures are available.

The returns series (in percentage terms) is plotted in Figure 1a, and a kernel density plot for returns is given in Figure 1b. The typical volatility clusters and leptokurtosis are apparent, as is the financial crisis of the later part of 2008. Figure 1c plots realized volatility ($RV5$), and Figure 1d plots the measure of jump activity, $RV5_t - BV_t$. We can see that jumps are an important factor in overall volatility. The clusters of jump activity apparent in Figure 1d suggest that a model with dynamic jump intensity may be needed to accurately account for jump activity.

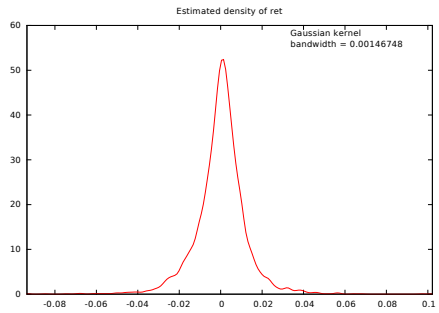
To investigate the effects of the recent financial crisis and to have data for out-of-sample forecasting evaluation, we split the data into 3 portions. The first two portions, 2000-2007 ($n = 1988$ observations) and 2008-2011 (1002 observations) are used to fit the parameters of the models, while the portion Jan. 2012-Apr., 2014 ($n = 581$ observations) is reserved for out-of-sample forecasting purposes.

Figure 1: S&P 500 data

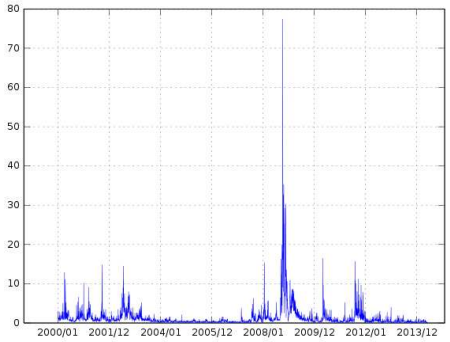
(a) Daily returns, r_t



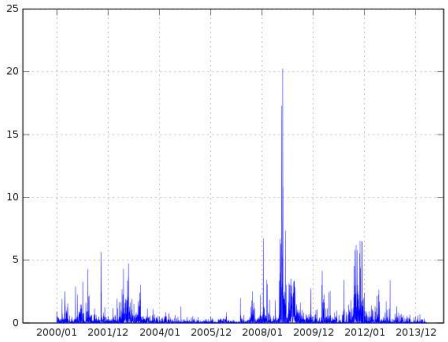
(b) Kernel density, daily returns, r_t



(c) Daily realized volatility, $RV5_t$



(d) Daily Jumps, $RV5_t - BV_t$



6.3 Simulation of the model

The trading period for the S&P 500 index is 6.5 hours per day. To simulate data from the model, we use the Euler approximation method. For the Monte Carlo work, the Euler approximation is done directly using a discrete time interval of 5 minutes, which is the frequency that is used to compute realized measures for the real data. For the estimation work, we simulate using a discrete time interval of 1 minute, recording every fifth point, so that our simulated variables are also computed using 5 minute observations of log price. To allow for the non-trading period each day, we simulate the model over 24 hour days, of which 6.5 hours are the trading period. Daily returns (r) are computed using the closing price at the end of each trading period. The realized measures $RV5_t$, $RV10_t$ and BV_t are computed using the 5 minute observations, only during the trading period. The non-trading period portions of the simulations are discarded, to allow for the overnight effect. We do not use intra-day returns in the estimation, only end of day returns and realized volatility measures. An initial burnin period of 200 days of simulations is discarded from the simulations.

6.4 Implementation of the ABC estimator

This section discusses the specific details of implementation of the ABC estimator as used for the simulation study and the empirical application.

Table 1: Bounds of parameter space

Parameter	True Value	Lower Bound	Upper Bound	Bias	RMSE
μ_0	-0.01	-0.1	0.1	0.010	0.059
μ_1	0.01	-0.1	0.1	-0.010	0.059
α	0.5	-3	3	-0.500	1.803
κ	0.05	0	0.5	0.200	0.247
σ	0.2	0	1	0.300	0.416
ρ	-0.7	-1	0	0.200	0.351
λ_0	0.02	0	0.1	0.005	0.015
λ_1	1	0	3	0.500	1.000
μ_J	-0.005	-0.05	0.05	0.005	0.029
σ_J	1	0	5	1.500	2.082
σ_ϵ	0/0.004	0	0.02		

6.4.1 Parameter space, Monte Carlo design point, and pseudo-prior

The pseudo-prior $\pi(\theta)$ was chosen as a uniform density over the parameter space Θ, which was defined as the Cartesian product of the line segments given by the lower and upper bounds for each parameter as presented in Table 1. The chosen limits are intended to be relatively uninformative for the parameters, given the widely accepted characteristics of data on equity returns and volatility. Below, we present an estimation algorithm that deals effectively with a pseudo prior that is very uninformative. The true parameter values for the Monte Carlo experiment are given in the second column of the Table. These values were chosen to be similar to the estimated values when using the 2000-2007 data (reported below). The last two columns give the bias and the RMSE if the mean of pseudo prior were used as an estimator of the true parameters, so that we can measure the reduction in bias and gains in efficiency due to applying our ABC estimator to the sample data. The true value for σ_ϵ in the last row varies depending on whether or not the DGP contains measurement error.

6.4.2 Selection of statistics

The initial set of auxiliary statistics was determined using the model-based approach described in the previous section, combined with experimentation. A crude indicator of jump activity in period t is a test for $jump_t \equiv RV5_t - BV_t$ being larger than 2.5 standard deviations of the $jump_t$ series. Based upon this, we created a new returns series, $r2$, where $r2_t$ is set to zero if the jump indicator detects a jump. This is intended to give a series that is informative about returns more or less net of jump activity. Another variable, $RVdiff \equiv RV5 - RV10$ was created with the expectation that is would be helpful for identification of σ_ϵ, when the model includes that parameter, following the argument in the final part of Section 2. The initial pool of statistics include means, standard deviations and correlations between r, $RV5$, $RV10$, BV and $r2$, or their logarithms, and descriptive statistics for $RVdiff$. To this we add parameter estimates from several auxiliary models.

These include a simple EGARCH model fit to the $r2$ series, and linear regression models that regress $r2$, BV and $jump_t$ on their own and cross lags. Because a complete verbal description would be tedious, we include the source code of the function that computes the auxiliary statistics the Appendix. This code is commented and is quite self-explanatory. Once this initial pool of statistics is computed, we select the set of statistics to use for estimation using the cross validation method described in Section 4.1.3.

6.4.3 An adaptive algorithm for sampling

The ABC estimator presented in Creel and Kristensen (2013) uses simple sampling from the pseudo prior, and nonparametric estimation. When the prior and the posterior are not similar, direct sampling from the prior can be computationally inefficient, because many draws of the parameter θ^s lead to simulated statistics Z_n^s that are always so far away from the observed statistic, Z_n, so that the associated parameter draw is never retained as one of the neighbors that affects the estimated value. Recognizing this problem, methods of computing estimators using likelihood-free Markov chain Monte Carlo and sequential Monte Carlo have been studied in some detail in the ABC literature (among others, see Marjoram *et al.,* 2003; Sisson, Fan and Tanaka, 2007; Beaumont *et al.* 2009; Del Moral, Doucet and Jasra, 2012). This work treats the prior as a real Bayesian prior that is to be respected, and the goal is to sample from the posterior associated with the prior. For the sequential Monte Carlo methods, this is achieved by properly weighting the particles using importance sampling weights, so as to preserve the original prior. We, on the other hand, treat the pseudo prior as a tool which allows us to compute a point estimator which we interpret from the classical perspective. As such, we omit the weighting that preserves the original prior, because this prior is not especially meaningful to us, except that it provides a means of initiating the simulations. The following algorithm adapts the original pseudo prior iteratively to create a series of pseudo priors such that the final pseudo prior is close to the posterior associated with itself. Then the ABC estimator proposed in Creel and Kristensen (2013) is applied to a final sample drawn from the final pseudo prior. As long as the true parameter value is within the support of the final pseudo prior, the theoretical results of Creel and Kristensen (2013) apply to the estimator that is computed using the final sample of particles. A summary of our algorithm is:

Step 0. generate S_0 particles from pseudo prior and compute the associated statistics Z_n^s. Set the iteration counter i to 0[1].

Iterations:

Step 1. Set iteration counter to $i + 1$

Step 2. select the best α_i% of particles based on proximity of Z_n^s to Z_n

Step 3. from the selected particles, randomly draw, with replacement, S_i particles

Step 4. perturb the particles by adding a mean zero shock to each particle

[1]This initial sample of particles is also used for the cross validation selection of statistics from the large pool of statistics

Step 5. allow recombination of a proportion of the particles by randomly interchanging their elements

Step 6. stop if a stopping criterion is met, otherwise, go to step 1

Step 7. draw a relatively large final set of particles by sampling with replacement from the last iteration particles, perturb each of them, and generate the associated Z_n^s

Step 8. apply the basic ABC method of Creel and Kristensen (2013) to compute the final estimate

Our experience is that algorithm is very effective in restricting attention to the portion of the parameter space that has non-negligible mass. This allows one to specify initially broad bounds on the parameter space, as in Table 1, because the algorithm quickly focuses attention on the region which can generate auxiliary statistics that are close to the observed Z_n. The only drawback to setting a very uninformative pseudo prior is that S_0 in Step 0 may need to be increased to compensate for excessive dispersion of the initial statistics, Z_n^s.

6.5 Implementation of Filter and Smoother

The model contains latent variables about which we may wish to learn, in particular, the non-jump component of volatility, h_t, and the jump intensity, λ_t. The limited information approach to filtering is implemented as explained in Section 5 using the long-trajectory simulator and with

$$F_{t-1} = \left(r_{t-1}, ..., r_{t-p}, \log BV_{t-1}, ...\log BV_{t-q}\right) \tag{28}$$

for some lag lengths $p, q \geq 1$.The steps are:

1. Given the estimated parameter vector $\hat{\theta}$, simulate a long series from the model of equations 1-27, retaining end of day returns, realized volatility, the robust measure of volatility, actual log volatility, and total jump size for the day:

$$\left\{\left(\tilde{r}_t, \tilde{BV}_t, \tilde{h}_t, \tilde{\lambda}_t\right)\right\}, \; t = 1, 2, ..., R. \tag{29}$$

2. Use the simulated data together with nonparametric regression to learn about the filtering function M in $g_t = M(F_{t-1}) + e_t$, where $M\left(F_{t-1}\right) = E_\theta\left[g_t | F_{t-1}\right]$, e_t is the smoothing error, and F_{t-1} is chosen as above, according to the variable being filtered.

3. Filter by applying the nonparametric estimate of M obtained in Step 2 to the real data:$\hat{g}_{t|1:t-1} = \hat{M}(F_{t-1})$.

The smoother is implemented in much the same way as limited information filtering, with the only difference being that instead of using only lags of the observable variables, we may also use the contemporaneous values and leads to smooth the latent variables. The method proceeds as outlined in Section 5 with the conditioning variables chosen

depending on which variable we are smoothing:

$$\bar{F}_t = \left(r_t, r_{t\pm1}, ..., r_{t\pm p}, \log BV_t, \log BV_{t\pm1}, ... \log BV_{t\pm q}\right) \tag{30}$$

Given g_t and $\bar{F}_t$, we then proceed as follows:

1. Simulate data from model evaluated at $\theta = \hat{\theta}$, as in equation 29 (or reuse the simulations computed for the filtering)

2. Use the simulated data together with nonparametric regression to learn about the smoothing function M in $g_t = M(\bar{F}_t) + \bar{e}_t$, where $M(\bar{F}_t) = E_\theta[g_t|\bar{F}_t]$, $\bar{e}_t$ is the smoothing error, and $\bar{F}_t$ chosen as above.

3. Smooth by applying the nonparametric estimate of M obtained in Step 2 to the real data:$\hat{g}_{t|1:T} = \hat{M}(F_t)$.

For both the filter and smoother, the number of lags, p and q, and the number of neighbors, k, can be chosen using cross-validation: Minimize out of sample mean squared forecast error, computed using out-of-sample simulated data.

In the case of the linear filter and smoother, one simply replaces the kernel regression step with OLS estimation as described earlier.

7 Monte Carlo and empirical results

This section collects all of the simulation, estimation, and filtering and smoothing results. We consider two versions of the model, with and without measurement error.

7.1 Monte Carlo estimation results

In Table 2 we report the Monte Carlo results for the ABC estimator. The first block of results are for the case where neither the model nor the DGP contains measurement error. These are followed by a second block, where the DGP contains measurement error, which is not accounted for in the model. The third block gives the results for the case of measurement error in the DGP that is accounted for in the model. The true value of the standard deviation of measurement error, when applicable, is set to $\sigma_\epsilon = 0.004$, a value close to the estimated value using the S&P 500 data, reported in the next Section. The Monte Carlo samples were generated using the true values given in Table 1, using a sample size of 2000 observations, which is similar to the size of the pre-crisis (2000-2007) period of our sample for the S&P 500 index, discussed in the next subsection.

Considering a well specified model when the DGP has no measurement error, we see that bias contributes to RMSE in an important way only for λ_1 and μ_J. The standard deviations of the drift parameters are large in comparison to their true values. The parameters that affect the non-jump portion of volatility are all estimated with good precision and little bias. Considering jumps, the base rate of jumps when volatility is at its mean, λ_0,

is estimated with small bias and good precision. The marginal effect of volatility on the jump rate, λ_1, is less well identified than other parameters, it seems. The same is true for the standard deviation of jump size, σ_J, which is has a comparatively large standard deviation. However, overall, we may say that the parameters of the model are estimated quite well.

Next, considering the effect of measurement error that exists in the DGP, but which is ignored by the model, in the second block, it is perhaps surprising to observe that its presence has very little effect on the estimation of the other parameters of the model. The biases and RMSEs of the second block are very similar to those of the first block. For the magnitude of measurement error we consider (which was determined by the estimated value using the real data), and for the frequency of observation that is used to compute the realized measures (5 minutes), there seems to be no need to use a model that accounts for possible existence of measurement error: a model that ignores it gives essentially identical results.

Finally, the third block contains the results for estimation of a model that includes measurement error, when the DGP also does so. Biases and RMSEs are in general slightly greater than for model that ignores measurement error, but we believe that it is fair to say that this model also does a good job of estimating the parameters. One may note that the standard deviation of measurement error, σ_ϵ, is estimated with an upward bias. The true value is 0.004, but the mean of the Monte Carlo replications is 0.007. This suggests that attempting to account for measurement error may make it seem more important than it really is. This, plus the fact that its actual importance is little, makes an even stronger case for simply ignoring its possible presence. Even when it is present, one obtains better estimates of the other parameters by ignoring it, rather than accounting for it. This conclusion is subject to the caveat that for notably higher frequency data, measurement error would be expected to have a more important effect.

7.2 S&P 500 estimation results

Next, Table 3 presents the parameters estimates when the model is fitted to the S&P 500 data set, assuming that there is no measurement error. We give results for the precrisis period 2000-2007 (1988 observations), for the post-crisis period 2008-2011 (1002 observations), and using the entire estimation sample, 2000-2011 (2990 observations). The reported standard error estimates have been computed using 100 bootstrap replications. To assess the reliability of the bootstrap standard errors, one may compare them to the Monte Carlo standard errors of Table 2, which used a sample design close to the parameter estimates for the 2000-2007 period, and which used a very similar sample size (2000 observations). We see that the bootstrap standard errors are very similar to the Monte Carlo standard errors, which leads us to conclude that the bootstrap standard errors are reliable indicators of the precision of the parameter estimates.

Considering the parameters in the order given in the Table, the drift parameters are

Table 2: Monte Carlo results, DGP with and without measurement error (ME), model with and without ME

		DGP no M.E.			DGP with M.E.					
		Model no ME			Model no ME			Model with ME		
Param.	True	Bias	St. Dev.	RMSE	Bias	St. Dev.	RMSE	Bias	St. Dev.	RMSE
μ_0	-0.01	-0.001	0.028	0.028	-0.001	0.028	0.028	-0.001	0.028	0.028
μ_1	0.01	0.005	0.012	0.013	0.004	0.012	0.013	0.004	0.012	0.013
α	0.5	-0.002	0.091	0.090	-0.005	0.090	0.090	-0.041	0.098	0.106
κ	0.05	0.005	0.010	0.011	0.005	0.010	0.012	0.006	0.010	0.012
σ	0.2	0.008	0.015	0.017	0.007	0.015	0.017	0.015	0.015	0.022
ρ	-0.7	0.019	0.063	0.066	0.022	0.065	0.068	0.040	0.057	0.070
λ_0	0.02	-0.001	0.004	0.004	-0.001	0.004	0.004	-0.000	0.004	0.004
λ_1	1	0.313	0.182	0.362	0.3202	0.185	0.370	0.347	0.174	0.388
μ_J	-0.005	0.005	0.006	0.008	0.005	0.006	0.008	0.005	0.007	0.008
σ_J	1	0.033	0.341	0.342	0.037	0.315	0.317	-0.046	0.264	0.268
σ_ϵ	0/0.004	-	-	-	-0.004	0.000	0.004	0.003	0.003	0.004

not significantly different from zero. Mean log volatility, α, shows a very important increase between the pre- and post-crises periods, which is not at all surprising given the increase in volatility that is obvious in Figure 1a. The mean reversion parameter, κ, is quite stable across the two periods, and it indicates that shocks to log volatility are quite persistent, as is well known. The standard deviation of shocks to log volatility, σ, shows a small but probably insignificant increase across the two periods. Leverage, ρ, is quite strong in the 2000-2007 period, and declines somewhat in magnitude in the post-crisis period. The parameter related to the baseline probability of a jump, λ_0, more than doubles in the post-crisis period, compared to its level during the pre-crisis period. These values would give about two expected jumps per year in the pre-crisis period, and a little more than four expected jumps in the post crisis period. On the other hand, the marginal effect of log volatility on the jump rate, λ_1, is significantly different from zero, and is quite stable across the two periods. Our parameterization of the jump rate is somewhat different that that of Andersen, Benzoni and Lund (2002) in that ours depends on the log of spot volatility, while theirs depends on its level, and in that our parameterization allows the jump rate to decline to zero at times, whereas theirs has a positive minimum value for the jump rate. It is interesting to note that we detect a significant role for dynamics in the jump rate, while their estimated effect is not significant. Mean jump size, μ_0, is not significantly different from zero in either period, while the standard deviation of jump size, σ_J, increases somewhat in the post-crisis period. We also present a block of results for a model that imposes parameter constancy over the entire 2000-2011 period, but we believe that the evidence of change in the α and λ_1 parameters is strong enough so that these last results should be discounted. The overall story is one of both the non-jump and the jump components of volatility increasing in the post-crisis period, leverage declining in magnitude, and the remaining parameters being more or less stable.

Table 3: Estimation results, model without measurement error

	2000-2007		2008-2011		2000-2011	
Parameter	Estimate	St. Err.	Estimate	St. Err.	Estimate	St. Err.
μ_0	-0.021	0.022	-0.021	0.039	-0.012	0.026
μ_1	0.003	0.009	-0.013	0.019	-0.008	0.009
α	0.181	0.107	0.785	0.179	0.458	0.098
κ	0.035	0.008	0.036	0.012	0.032	0.006
σ	0.162	0.011	0.204	0.024	0.173	0.010
ρ	-0.851	0.054	-0.713	0.079	-0.842	0.045
λ_0	0.007	0.004	0.016	0.004	0.011	0.005
λ_1	1.235	0.178	1.152	0.167	1.165	0.209
μ_J	0.002	0.006	-0.002	0.006	-0.002	0.007
σ_J	1.201	0.237	1.472	0.367	1.893	0.435

Table 4 gives estimation results for the model that allows for measurement error, again using the three sample periods. The overall results are very similar to those for the model without measurement error: α, σ, λ_0 and σ_J all increase in the post-crisis period, indicating an increase in both the jump and non-jump components of volatility. Leverage declines, and the drift parameters remain insignificantly different from zero. The most notable difference is that λ_1 appears to increase somewhat notably when measurement error is included, compared to the small decline when measurement error is ignored. Of interest is the estimated value of the standard deviation of measurement error, σ_ϵ. The estimated value is 0.003 or 0.004, depending on the period. The bias in estimation of this parameter that was found in the Monte Carlo results (last line of Table 2, third column from the right) is 0.003. It is possible to perform bias reduction by subtracting bootstrap estimates of the bias from the estimated values (see Efron and Tibshirani, 1986, equation 4.4). Because the Monte Carlo design is quite similar to the estimated parameter values, a rough bias correction could be done by subtracting the biases found in Table 2 from the estimated values in Table 4. If we did that for the σ_ϵ parameter, the estimated value would be almost zero. This fact, combined with the previous remarks regarding the possibility of simply ignoring the possible presence of measurement error for data similar to ours, leads us to prefer the results given in Table 3 over those of Table 4. This choice, however, is not of too much importance, as the results are quite similar in the two cases. However, given these considerations, for the filtering and smoothing exercises discussed below, we will use the results for the model without measurement error.

7.3 Monte Carlo results for filtering and smoothing

Next we present filtering and smoothing results for both the non-jump component of volatility, h_t, and the jump intensity, λ_t, using the limited information procedure discussed in Section 5. For filtering, we use our proposed method based on a long simulation of the model at the estimated parameter values, using two lags of r_t and four lags of

Table 4: Estimation results, model with measurement error

	2000-2007		2008-2011		2000-2011	
Parameter	Estimate	St. Err.	Estimate	St. Err.	Estimate	St. Err.
μ_0	-0.022	0.026	-0.009	0.043	-0.014	0.032
μ_1	-0.005	0.010	-0.032	0.018	-0.011	0.010
α	0.146	0.105	0.794	0.197	0.358	0.010
κ	0.036	0.008	0.036	0.010	0.036	0.006
σ	0.165	0.013	0.216	0.022	0.184	0.012
ρ	-0.825	0.057	-0.659	0.076	-0.781	0.048
λ_0	0.012	0.004	0.018	0.004	0.016	0.005
λ_1	0.925	0.195	1.454	0.199	1.141	0.208
μ_J	-0.006	0.006	0.007	0.005	-0.002	0.007
σ_J	0.894	0.234	2.231	0.524	2.256	0.456
σ_ϵ	0.003	0.002	0.004	0.003	0.004	0.002

Table 5: Filtering and smoothing results, simulated data (root mean squared error).

	h	λ
Nonparametric filter	0.259	0.018
Linear filter	0.263	0.025
Nonparametric smoother	0.139	0.010

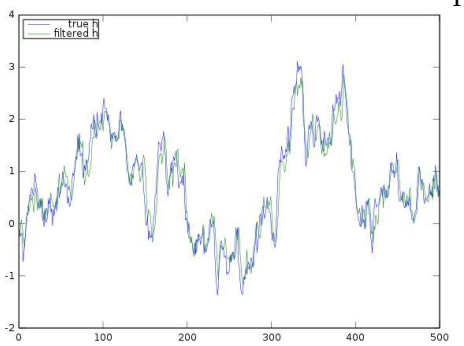
$\log BV_t$ (see equation 28). We give results for both the nonparametric version as well as the linear regression version. As noted above, we use the model without measurement error for filtering and smoothing, using the parameter estimates for the postcrisis period (2008-2011), both for the simulated results, in this section, and for the real data results, in the next section. With these parameter values, we generate a simulation of length 2×10^6 observations. We use the last 2×10^4 observations for one-step ahead forecasting (they are used to obtain F_{t-1} in equation 28). The simulation, excluding this last part, gives the set defined in equation 29. That is, in that equation, $R = 2 \times 10^6 - 2 \times 10^4$. The filtering and smoothing results given in this section do not take into account parameter uncertainty: the parameter values have been taken as given, and so we only explore the accuracy of the limited information filtering and smoothing methods. This is due to the fact that a joint Monte Carlo exploration of parameter estimation and filtering/smoothing accuracy would be computationally very demanding.

The first two rows of Table 5 give the results. We see that the nonparametric filter gives an overall better forecast than does the linear filter that uses the same conditioning information, but that the difference is not very large. Given that the linear filter is computationally much faster, it may be adequate for many purposes uses. Figures 2a and 2b plot the first 500 of the 20000 forecasts (for visual clarity) of the nonparametric filter results, along with the realized value, for h and λ, respectively. We see that the filter forecasts are quite accurate.

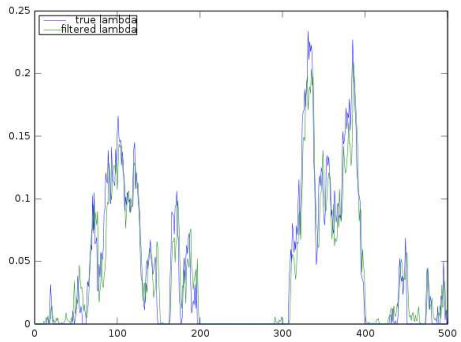
For smoothing, we use the current value and one lag and lead of r_t, and the current

Figure 2: Filtering and smoothing results, simulated data

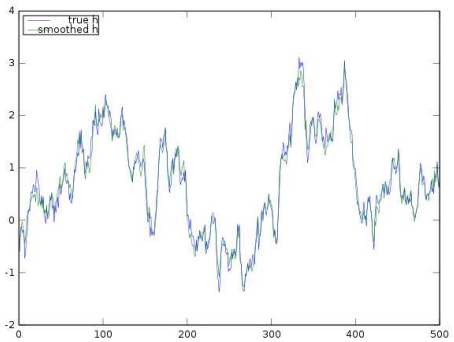
(a) True and filtered log spot volatility, h_t.



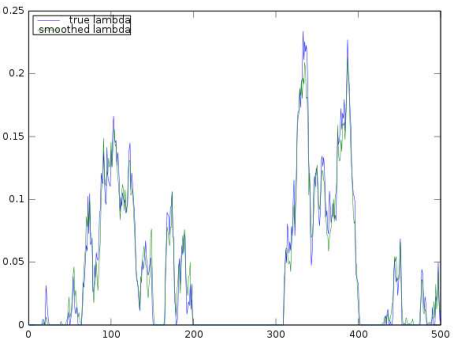
(b) True and filtered jump rate, λ_t.



(c) True and smoothed log spot volatility, h_t.



(d) True and smoothed jump rate, λ_t.



value and two lags and leads of BV_t as the conditioning information (see equation 30). The results are given in the last line of Table 5. The RMSE of the nonparametric smoother is lower than that of the nonparametric filter, as is expected. Figures 2c and 2d plot the first 500 of the 20000 smoothed observations, for h and λ, respectively. We see that the smoothed variables are very accurate.

7.4 Filtering and smoothing of the S&P500 data

Having seen, using simulated data, that the filtering and smoothing methods are reliable, we next turn to filtering and smoothing results for the S&P 500 data. Again, we use the parameter estimates for the model without measurement error, for the 2008-2011 sample period. First, Figure 3a gives the one step ahead forecasts of log spot volatility, h_t, for the out of sample period, 2012-2014. Figure 3b does the same for the jump rate, λ_t. There is a notable overall downward trend in log volatility over the period, and likewise, the jump rate shows a moderation. The peak of both h_t and λ_t occurs on June 25, 2013. Shortly after, on July 26, 2013, Mario Draghi made his famous "whatever it takes" comments. The filtered results show that an extensive period of low volatility and low jump rate followed these remarks, until early 2014, when volatility and the jump rate climbed a bit, again.

We also filter the observable variable log $RV5$, using our procedure and the HAR-RV-J

method, using as conditioning variables those of equation 25. This serves as a check on the reliability of the parametric model that is used to generate the long simulation which underlies the filtering and smoothing results, as the HAR-RV-J method does not rely on the parametric model. Figures 3c and 3d plot the filtering results using the proposed procedure and the HAR-RV-J filter of equation 25, respectively. The results using the two methods are very similar. In both cases, the out of sample root mean forecast error is 0.672. The fact that the two methods give essentially the same results for filtering the observable variable log $RV5$ leads us to conclude that the parametric model that underlies our procedure is adequate for the purpose of filtering. This, plus the reliability that was observed in the previous section using simulated data, lends support to the reliability of the filtering results for the latent variables.

Finally, we present smoothing results for the entire sample, 2000-2014. Figure 3e gives the smoothed log spot volatility, and Figure 3f gives the smoothed jump rate. The high volatility and jump rate of the 2008-2011 period is plain to see. Perhaps most interesting is the observation that both volatility and the jump rate have recently returned to pre-crisis levels of 2004-early 2007. We remark that these smoothing results were obtained using the coefficient estimates from the post-crisis period, using the 2008-2011 data.

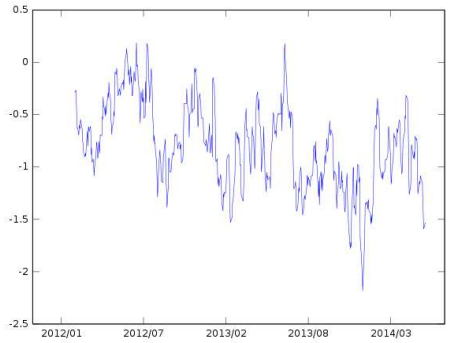
If one wished to compute confidence intervals for the filter and/or smoothing results that take into account parameter uncertainty, a conceptually straightforward procedure would be to repeat the filtering and/or smoothing procedure using each of the bootstrap parameter values that were used to compute the standard error estimates of the estimated parameters, in Section 7.2. This would be somewhat computationally demanding, but simple to implement.

8 Conclusion

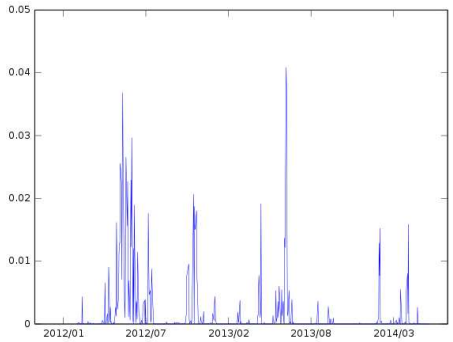
We have here shown how Approximate Bayesian Computation, a particular type of limited information estimation method, can fruitfully be employed in the estimation and analysis of asset pricing models - both for estimation of parameters and filtering of latent states. The method is simpler to implement compared to full likelihood-based estimators, but still delivers precise estimates of parameters and latent state variables. The method is somewhat computationally demanding, but it is very amenable to parallelization using MPI or GPU computing, as in Creel and Zubair (2012), so that it would be possible to obtain estimation and filtering/smoothing results in near real time. We implemented the method using a daily data of returns and realized volatility measures of the S&P 500 index. We found strong evidence of structural breaks in the data around 2007-2008 with the jump and volatility dynamics changing radically around this period. Nevertheless, the smoothing results for the entire sample indicate that volatility and the jump rate have recently returned to the low levels of the 2004-2006 period. Our results indicate that possible measurement error in log price is of a very low level for this data, and it has no significant effect on parameter estimates, at least when realized measures are based on 5

Figure 3: Filtering and smoothing results, S&P 500 data

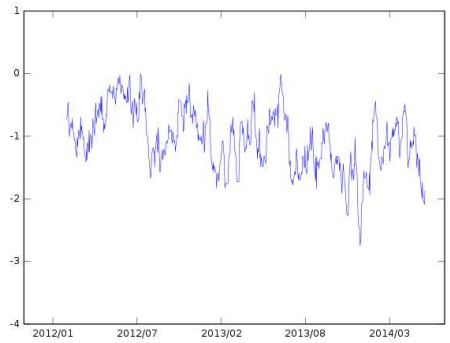
(a) Filtered log spot volatility, h_t, 2012-2014



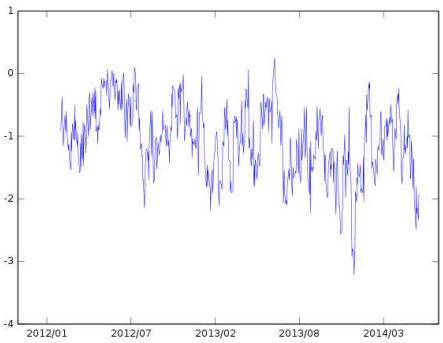
(b) Filtered jump rate, λ_t, 2012-2014



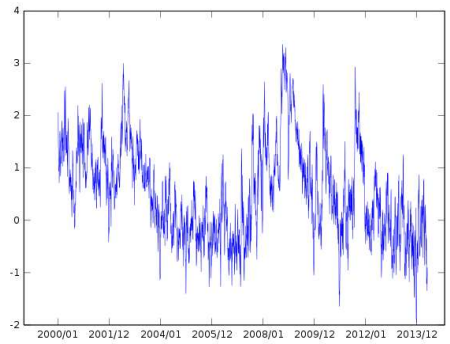
(c) Filtered log $RV5$, 2012-2014



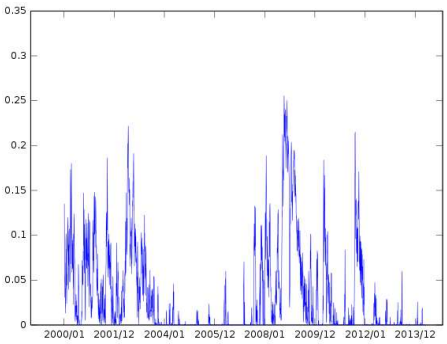
(d) HAR_RV-J filtered log $RV5$, 2012-2014



(e) Smoothed log spot volatility, h_t, 2000-2014



(f) Smoothed jump rate, λ_t, 2000-2014



minute observations of log price.

References

[1] Andersen, T. G., L. Benzoni, and J. Lund (2002). An empirical investigation of continuous-time equity return models. *Journal of Finance* 57, 1239–1284.

[2] Andersen, T. G., T. Bollerslev and F. X. Diebold (2007). Roughing It Up: Including Jump Components in the Measurement, Modeling and Forecasting of Return Volatility. *Review of Economics and Statistics* 89, 701–720.

[3] Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003). Modeling and Forecasting Realized Volatility. *Econometrica* 71, 579–625.

[4] Andersen, T. G., D. P. Dobrev, and E. Schaumburg (2011). Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics* 169, 75–93.

[5] Bansal, R., A.R. Gallant, R. Hussey, and G. Tauchen (1994) Nonparametric estimation of structural models for high-frequency currency market data. *Journal of Econometrics* 66, 251–287.

[6] 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 Series B*, 64, 253–280.

[7] Barndorff-Nielsen, O.E. and N. Shephard (2006). Econometrics of Testing for Jumps in Financial Economics using Bipower Variation. *Journal of Financial Econometrics* 4, 1–30.

[8] Barndorff-Nielsen, O.E., S.E. Graversen, J. Jacod, and N. Shephard (2006). Limit theorems for bipower variation in financial econometrics. *Econometric Theory* 22, 677–719.

[9] Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde, N. Shephard (2008) Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica* 76, 1481–1536.

[10] Beaumont, M.A., J.-M. Cornuet, J.-M. Marin and C.P. Robert (2009) Adaptive approximate Bayesian computation. *Biometrica* 96, 983-990.

[11] Bates, D. (1996). Jumps and stochastic volatility: Exchange rate processes implicit in the PHLX Deutschemark options. *Review of Financial Studies* 9, 69–107.

[12] Bollerslev, T. and H. Zhou (2002). Estimating Stochastic Volatility Diffusion Using Conditional Moments of Integrated Volatility. *Journal of Econometrics* 109, 33–65.

[13] Broadie, M., M. Chernov and M. Johannes (2009). Understanding Index Option Returns. *Review of Financial Studies* 22, 4493–4529.

[14] Calvet, L. and V. Czellar (2012). Accurate Methods for Approximate Bayesian Computation Filtering. Manuscript, Department of Finance, HEC Paris.

[15] Caner, M. and H.H. Zhang (2014) Adaptive Elastic Net for Generalized Methods of Moments. *Journal of Business and Economic Statistics* 32, 30-47.

[16] Carrasco, M., M. Chernov, J.-P. Florens, and E. Ghysels (2007). Efficient estimation of jump diffusions and general dynamic models with a continuum of moment conditions. *Journal of Econometrics* 140, 529–573.

[17] Chan, W. H. and Maheu, J. M. (2002). Conditional jump dynamics in stock market returns. *Journal of Business and Economic Statistics* 20, 377–389

[18] Chernov, M., A.R. Gallant, E. Ghysels, and G. Tauchen (2003) Alternative models for stock price dynamics. *Journal of Econometrics* 116, 225–257.

[19] Chesney, M. and L. O. Scott (1989). Pricing european currency options: A comparison of the modified Black-Scholes model and a random variance model. *Journal of Financial Quantitative Analysis* 24, 267–284.

[20] Corradi, V. and W. Distaso (2006). Semiparametric Comparison of Stochastic Volatility Models via Realized Measures. *Review of Economic Studies* 73, 635–677.

[21] Creel, M. and M. Zubair (2012). High Performance Implementation of an Econometrics and Financial Application on GPUs, *SC Companion: High Performance Computing, Networking Storage and Analysis*, 1147–1153.

[22] Creel, M. and D. Kristensen (2012) Estimation of Dynamic Latent Variable Models Using Simulated Nonparametric Moments. *Econometrics Journal* 15, 490–515.

[23] Creel, M. and D. Kristensen (2013) Indirect Likelihood Inference. UFAE and IAE Working Paper 931.13, Barcelona Graduate School of Economics.

[24] Dean, T., S. Singh, A. Jasra and G.W. Peters (2011) Parameter estimation for Hidden Markov Models with intractable likelihoods. Manuscript, Dep. of Statistics, University College London.

[25] Del Moral, P., A. Doucet and A. Jasra (2012) An adaptive sequential Monte Carlo method for approximate Bayesian computation, *Statistics and Computing*, 22, 1009-1020.

[26] Dobrev, D.P. and P.J. Szerszen (2010) The Information Content of High-Frequency Data for Estimating Equity Return Models and Forecasting Risk. Working paper, Federal Reserve Board.

[27] Duffie, D., J. Pan, and K. Singleton (2000). Transform Analysis and Asset Pricing for Affine Jump-Diffusions. *Econometrica* 68, 1343–1376.

[28] Efron, B. (2012). Bayesian Inference and the Parametric Bootstrap. *Annals of Applied Statistics* 6, 1971–1997.

[29] Efron, B. and R. Tibshirani (1986) Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 1, 54-75.

[30] Eraker, B., M. Johannes and N. Polson (2003) The impact of jumps in equity index volatility and returns. *Journal of Finance* 58, 1269–1300.

[31] Fan, J. and I. Gijbels (1992). Variable bandwidth and local linear regression smoothers. *Annals of Statistics* 20, 2008–2036.

[32] Fan, J., I. Gijbels, T. Hu, and L. Huang (1996). A study of variable bandwidth selection for local polynomial regression. *Statistica Sinica* 6, 113–127.

[33] Fang (2000). Option pricing implications of a stochastic jump rate. Working Paper, Department of Economics. University of Virginia.

[34] Fearnhead, P. and D. Prangle (2012). Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation. *Journal of the Royal Statistical Society: Series B* 74, 419–474.

[35] Hansen, B.E. (2006). Interval Forecasts and Parameter Uncertainty. *Journal of Econometrics* 135, 377–398.

[36] Hansen, P.R., Z. Huang and H.H. Shek (2012) Realized GARCH: A joint model for returns and realized measures of volatility. Journal of Applied Econometrics 27, 877–906.

[37] Hansen, P. R., and A. Lunde (2005). A Realized Variance for the Whole Day Based on Intermittent High-Frequency Data. *Journal of Financial Econometrics* 3, 525–554.

[38] Heber, G., A. Lunde, N. Shephard and K. Sheppard (2009) Oxford-Man Institute's Realized Library. Oxford-Man Institute, University of Oxford.

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

[40] Johannes, M., N. Polson and J. Stroud (2009) Optimal Filtering of Jump-Diffusions: Extracting Latent States from Asset Prices. *Review of Financial Studies* 22, 2759–2799.

[41] Kanaya, S. and D. Kristensen (2010) Estimation of Stochastic Volatility Models by Nonparametric Filtering. CREATES Research Paper 2010-67, University of Aarhus.

[42] Kim, S., N. Shephard and S. Chib (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies* 65, 361–393.

[43] Koenker, R. and G. Bassett, Jr. (1978). Regression Quantiles. *Econometrica* 46, 33–50.

[44] Koopman, S.J. and M. Scharth (2013). The Analysis of Stochastic Volatility in the Presence of Daily Realized Measures. *Journal of Financial Econometrics* 11, 76–115.

[45] Kristensen, D. (2009). Uniform Convergence Rates of Kernel Estimators with Heterogeneous, Dependent Data. *Econometric Theory* 25, 1433–1445.

[46] Li, Q. and J. Racine (2007) *Nonparametric Econometrics: Theory and Practice*. Princeton: Princeton University Press.

[47] Mancini, C. (2009) Non-parametric Threshold Estimation for Models with Stochastic Diffusion Coefficient and Jumps. *Scandinavian Journal of Statistics* 36, 270–296.

[48] Marjoram, P., J. Molitor, V. Plagnol and S. Tavaré (2003) *Proceedings of the National Academy of Sciences*, 100, 15324-15328.

[49] Monfardini, C. (1998). Estimating Stochastic Volatility Models Through Indirect Inference. *Econometrics Journal* 1, 113–128.

[50] Podolskij, M. and M. Vetter (2009). Bipower-type estimation in a noisy diffusion setting. *Stochastic Processes and their Applications 119*, 2803–2831

[51] Prewitt, K. and S. Lohr (2006). Bandwidth selection in local polynomial regression using eigenvalues. *Journal of the Royal Statistical Society - Series B* 68, 135–154.

[52] Schucany, W. (1995). Adaptive bandwidth choice for kernel regression. *Journal of the American Statistical Association* 90, 535–540.

[53] Sisson, S.A., Y. Fan and M. Tanaka (2007) Sequential monte carlo without likelihoods. *Proceedings of the National Academy of Science*s, 104, 1760-1765.

[54] Takahashi, M., Y. Omori and T. Watanabe (2009) Estimating stochastic volatility models using daily returns and realized volatility simultaneously. *Computational Statistics and Data Analysis* 53 (2009) 2404–2426

[55] Todorov, V. (2009). Estimation of Continuous-time Stochastic Volatility Models with Jumps using High-Frequency Data. *Journal of Econometrics* 148, 131–148.

[56] Wiggins, J. (1987). Option values under stochastic volatility: Theory and empirical estimates. *Journal of Financial Economics* 19, 351–372.

[57] Zhang, P. (1991). Variable Selection in Nonparametric Regression with Continuous Covariates. *Annals of Statistics* 19, 1681–2284.

[58] Zhang, L. (2006). Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli* 12, 1019-1043.

APPENDIX

Below is the listing of the GNU Octave function that computes the entire pool of auxiliary statistics. Not all statistics are used, the selection method chooses which statistics to use from the set generated by this function.

```
function Z = aux_stat(data)
 bad_data = false;
 % check for bad inputs
 if (sum(any(isnan(data)))||sum(any(isinf(data)))||sum(any(std(data)
    ==0)))
   Z = -1000*ones(60,1);
 else
   rets = data(:,1);
   RV5 = data(:,2);
   RV10 = data(:,3);
   BV = data(:,4);
   MedRV = data(:,5);
   jumps1 = RV5 - BV;
   jumps2 = RV5 - MedRV;
   RV5= RV5.*(RV5>0);
   RV10= RV10.*(RV10>0);
   BV= BV.*(BV>0);
   MedRV = MedRV.*(MedRV>0);
   jumps1 = jumps1.*(jumps1>0);
   jumps2 = jumps2.*(jumps2>0);
```

```
20
21    % bound data
22    test = rets > 1000;
23    rets = test*1000+ (1-test).*rets;
24    test = rets < -1000;
25    rets = test*(-1000)+ (1-test).*rets;
26    test = RV5 > 10000;
27    RV5 = test*10000+ (1-test).*RV5;
28    test = RV10 > 10000;
29    RV10 = test*10000+ (1-test).*RV10;
30    test = BV > 10000;
31    BV = test*10000+ (1-test).*BV;
32    test = MedRV > 10000;
33    MedRV = test*10000+ (1-test).*MedRV;
34    test = jumps1 > 10000;
35    jumps1 = test*10000+ (1-test).*jumps1;
36    test = jumps2 > 10000;
37    jumps2 = test*10000+ (1-test).*jumps2;
38
39    % select
40    RVs = RV5;
41    rob = BV;
42    jumps = jumps1;
43    RVdiff = log(RV5)-log(RV10); % magnitude of measurement error
44
45    % detect jumps
46    test1 = abs(jumps);
47    s = std(test1);
48    test1 = test1 < 2.5*s;
49    test2 = abs(log(jumps));
50    s = std(test2);
51    test2 = test2 < 2.5*s;
52    test = test1 | test2;
53    rets2 = rets.*test; % if jump, sets to zero
54    RVdiff = RVdiff.*test; % if jump, sets to zero
55    test = 1-test;
56    Z = mean(test); % good for lam0
57    m = moving(lag(test,1),10); % to detect clustering
58    temp = [test, m];
59    temp = temp(11:end,:);
60    y = temp(:,1);
61    x = [ones(size(y)), temp(:,2)];
62    z = ols(y,x);
63    Z = [Z; z];
64
65    % use logs
```

```
66    rob = log(rob);
67    RVs = log(RVs);
68
69    % detrend rets and regress on lag (captures drift: mu0 and mu1)
70    t = 1:rows(rets2);
71    t = t';
72    temp = [rets2 t lags(rets2,1)];
73    temp = temp(2:end,:);
74    y = temp(:,1);
75    x = [ones(rows(y),1) temp(:,2:end)];
76    z = ols(y,x);
77    rets2 = y-x*z;
78    rets2 = [0; rets2]; % keep conformable
79    Z = [Z; z];
80
81    % basic stats
82    data = [rets rets2 RVs rob jumps];
83    z = dstats(data, 0, true);
84    z = z(:,1:4);
85    z = z(:);
86    z = z([1,3:end],:); % drop mean of rets2, which is above
87    Z = [Z; z]; % means, sd, skew + kurt
88
89    % correlations
90    c = corr(data);
91    c = triu(c,1);
92    c = vec(c);
93    c = c(c !=0);
94    Z = [Z; c];
95
96    % demean, so constants not needed (they are in dstats, above)
97    data = st_norm(data);
98
99    if (sum(any(isnan(data)))||sum(any(isinf(data)))||sum(any(std(data)
          ==0)))
100     bad_data = true;
101   else
102     rets = data(:,1);
103     rets2 = data(:,2);
104     RVs = data(:,3);
105     rob = data(:,4);
106     jumps = data(:,5);
107
108     % jump robust: rate of decline of coefs should be good for kappa
109     temp = [rob lags(rob, 4) lags(jumps,1)];
110     temp = temp(5:end,:);
```

```
111     y = temp(:,1);
112     x = [sum(temp(:,2:3),2)/2 sum(temp(:,4:end-1),2)/2 temp(:,end)] ;
113     [z, junk, junk, ess, rsq] = mc_ols(y,x,"", true);
114     z = [z; z(3,:)/z(2,:)];
115     sig = sqrt(ess/(rows(x)-columns(x)));
116     Z = [Z; z; sig; rsq];
117
118     % jumps
119     temp = [jumps lags(rob, 10)];
120     temp = temp(11:end,:);
121     y = temp(:,1);
122     x = [sum(temp(:,2:end),2)/10];
123     [z, junk, junk, ess, rsq] = mc_ols(y,x,"", true);
124     sig = sqrt(ess/(rows(x)-columns(x)));
125     Z = [Z; z; sig; rsq];
126
127     % for kappa and sig
128     temp = [rob lags(rob,1) lag(rets2,1)];
129     temp = temp(2:end,:);
130     y = temp(:,1);
131     x = temp(:,2:end);
132     x = [x x.*x];
133     [z, junk, e_rob, ess, rsq] = mc_ols(y,x,"", true);
134     sig = sqrt(ess/(rows(x)-columns(x)));
135     Z = [Z; z; sig; rsq];
136
137     % egarch RETS
138     y = st_norm(rets2);
139     params =[0; 0.2; 0.9];
140     % use SA for estimation of egarch
141     ub = [0.2; 0.8; 1];
142     lb = [-0.2; 0; 0];
143     nt = 3;
144     neps = 3;
145     ns = 1;
146     rt = 0.5;
147     maxevals = 10000;
148     functol = 1e-5;
149     paramtol = 1e-3;
150     verbosity = 0;
151     minarg = 1;
152     control = { lb, ub, nt, ns, rt, maxevals, neps, functol, paramtol,
            verbosity, 1};
153     [z junk converge] = mle_estimate(params, y, 'egarch_small', '',
           control, 0);
154     %% finish with BFGS
```

```
155     %control = {100,2,0,1,0,1e-8,1e-6,1e-6};
156     %[z junk converge] = mle_estimate(params, y, 'egarch', '', control
            , 0);
157     if (converge == 2) converge = 1; endif
158     if (!(converge==1))
159       bad_data = true;
160     endif
161     Z = [Z; z];
162
163     % rets
164     temp = [rets2 lag(rets2,1) lag(rob,1)] ;
165     temp = temp(2:end,:); % keep obs lined up
166     y = temp(:,1);
167     x = temp(:,2:end);
168     x = [x x.*x];
169     [z, junk, e_ret, ess, rsq] = mc_ols(y,x,"", true);
170     sig = sqrt(ess/(rows(x)-columns(x)));
171     Z = [Z; z; sig; rsq];
172     z = corr(e_ret, e_rob); % capture rho
173     Z = [Z; z];
174    endif
175    if (sum(any(isnan(Z)))||sum(any(isinf(Z))))
176      Z = -1000*ones(60,1);
177    endif
178    if bad_data
179      "egarch no converge"
180      Z = -1000*ones(60,1);
181    endif
182   endif
183 endfunction
```

Research Papers 2013

2014-14: Mikko S. Pakkanen and Anthony Réveillac: Functional limit theorems for generalized variations of the fractional Brownian sheet

2014-15: Federico Carlini and Katarzyna Łasak: On an Estimation Method for an Alternative Fractionally Cointegrated Model

2014-16: Mogens Bladt, Samuel Finch and Michael Sørensen: Simulation of multivariate diffusion bridges

2014-17: Markku Lanne and Henri Nyberg: Generalized Forecast Error Variance Decomposition for Linear and Nonlinear Multivariate Models

2014-18: Dragan Tevdovski: Extreme negative coexceedances in South Eastern European stock markets

2014-19: Niels Haldrup and Robinson Kruse: Discriminating between fractional integration and spurious long memory

2014-20: Martyna Marczak and Tommaso Proietti: Outlier Detection in Structural Time Series Models: the Indicator Saturation Approach

2014-21: Mikkel Bennedsen, Asger Lunde and Mikko S. Pakkanen: Discretization of Lévy semistationary processes with application to estimation

2014-22: Giuseppe Cavaliere, Morten Ørregaard Nielsen and A.M. Robert Taylor: Bootstrap Score Tests for Fractional Integration in Heteroskedastic ARFIMA Models, with an Application to Price Dynamics in Commodity Spot and Futures Markets

2014-23: Maggie E. C. Jones, Morten Ørregaard Nielsen and Michael Ksawery Popiel: A fractionally cointegrated VAR analysis of economic voting and political support

2014-24: Sepideh Dolatabadim, Morten Ørregaard Nielsen and Ke Xu: A fractionally cointegrated VAR analysis of price discovery in commodity futures markets

2014-25: Matias D. Cattaneo and Michael Jansson: Bootstrapping Kernel-Based Semiparametric Estimators

2014-26: Markku Lanne, Jani Luoto and Henri Nyberg: Is the Quantity Theory of Money Useful in Forecasting U.S. Inflation?

2014-27: Massimiliano Caporin, Eduardo Rossi and Paolo Santucci de Magistris: Volatility jumps and their economic determinants

2014-28: Tom Engsted: Fama on bubbles

2014-29: Massimiliano Caporin, Eduardo Rossi and Paolo Santucci de Magistris: Chasing volatility - A persistent multiplicative error model with jumps

2014-30: Michael Creel and Dennis Kristensen: ABC of SV: Limited Information Likelihood Inference in Stochastic Volatility Jump-Diffusion Models