LIMITED INFORMATION LIKELIHOOD INFERENCE IN STOCHASTIC VOLATILITY JUMP-DIFFUSION MODELS

MICHAEL CREEL AND DENNIS KRISTENSEN

ABSTRACT. We develop a novel method 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 are computationally attractive relative to standard likelihood-based estimators since they rely on low-dimensional auxiliary statistics and so avoid computation of high-dimensional integrals. We also develop a simple filtering algorithm that allows one to track the latent volatility process in real time without any heavy computational burden. Despite their computational simplicity, we find that estimators and filters perform well in practice and leads 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 Stoxx50 equity index. We find evidence of the presence of jumps and in favor of a structural break in parameters. During the recent financial crisis, volatility has a higher mean and variance, and is less persistent than in the before crisis period. Jumps occur slightly less frequently, and are more likely to be negative when they do occur.

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

JEL codes: C13, C14, C15, C33.

Preliminary and Incomplete - Do not circulate

1. INTRODUCTION

Continuous-time models are widely used in finance to model the dynamics of asset prices since this framework allows researchers to employ stochastic calculus when deriving solutions to the models; in particular the use of stochastic calculus provides simple representations for the price of a variety of contingent claims. At the same time, most empirical work is done using discrete-time models since most of the continuous-time models used in theoretical finance cannot be solved solved in closed-form. This problem is further mitigated by the fact that many of the important dynamics factors entering realistic asset pricing models are unobserved and so information about these have to be extracted using advanced 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, the volatility of returns are time-varying, and, second, returns contain jumps; see, for example, Barndorff-Nielsen and Shephard (2006) and Broadie, Chernov and Johannes (2009). Continuous-time stochastic volatility models are widely used in theoretical finance to price 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, these models tend to be estimated in a discrete-time setting due to the aforementioned statistical and computational complexities associated with the continuous-time versions; see, e.g., Dobrev and Szerzen (2010), Eraker, Johannes and Polson (2003), and 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 filter latent states in these. The proposed method utilizes tools developed in the field of Approximate Bayesian Computation (ABC) to obtain computationally simple, yet precise estimators. 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 in, amongst others, Creel and Kristensen (2013) who referred to ABC estimators as indirect likelihood estimators. We show 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 stochastic volatility jump-diffusion models, but the methodology is applicable to many other classes of continuous-time models.

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 some information contained in data is not used in the estimation and so ABC estimators are limited information estimators. ABC can be seen as the Bayesian counterpart to so-called Indirect Inference (II) 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.

In comparison with full likelihood-based estimators, ABC estimators are less efficient since they only utilize limited amount of the information contained in data. This statistical efficiency loss is off-set by computational efficiency gains with ABC being a lot easier and faster to implement relative to full likelihood estimators. The current state of the art for full likelihood-based inference in jump-diffusion models with latent variables is Johannes, Polson and Stroud (2009). Their proposed procedure employs particle filter algorithms to approximate the exact, unknown likelihood. 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. In contrast, our method is very simple to implement, 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.

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.

Another contribution of this paper that it shows how information contained in intra-daily high-frequency data can be used to improve on the estimation and filtering of SV models. In particular, we demonstrate how augmentation of daily returns data by so-called realized volatility (RV) measures lead to more precise estimates of parameters and better forecasts of future volatility. Realized volatility measures are model-free measures of daily integrated volatility. 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, Barndorff-Nielsen and Shephard (2001), Bollerslev and Zhou (2002), Corradi and Distaso (2006), Kanaya and Kristensen (2010) and Todorov (2009). We show that the estimation and filtering of SV models using RV data is straightforward within the framework of ABC, and taking into account first-step estimation errors in the RV measures can easily be handled.

As an empirical application, we use the estimators to learn about the returns dynamics of the Euro Stoxx50 index and show how our filter allows us to track stochastic volatility over time. We find evidence of the presence of jumps and in favor of a structural break in parameters at the beginning of 2008. During the recent financial crisis, volatility has a higher mean and variance, and is less persistent than in the before crisis period. Jumps occur slightly less frequently, and are more likely to be negative when they do occur. Our filtering and smoothing methods identify well trends in volatility and jumps associated with historic events.

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 and 7 contain a simulation study and the empirical application, respectively. We conclude in Section 8.

2. JUMP-DIFFUSIONS AND REALIZED 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,

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

where $\mu_t$ is the time-varying drift, $\sigma_t$ 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 $\lambda_t$. This is a highly general model where the only real restriction is the assumption of finite jump activity.

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. 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,

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

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

$$ QV_{t+1} = \int_0^{t+1} \sigma_t^2 dt + \sum_{i=N_t}^{N_{t+1}} J_i^2. $$

We will refer to the continuous component as the integrated volatility (IV), $IV_{t+1} = \int_t^{t+1} \sigma_t^2 dt$.

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 by the discretely observed prices,

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

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 $QV_{t+1}$,
\[ \sqrt{M}(RV_{t+1} - QV_{t+1}) \to^d N \left( 0, V^{QV}_t \right), \]
where $V^{QV}_t$ is the asymptotic variance. If one is interested in only 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, 2009). Let $\hat{IV}_{t+1}$ be any of these estimators. We can then obtain consistent estimators of both the diffusive and jump component of $QV$ and each of these satisfy a central limit theorem,
\[ \hat{IV}_{t+1} = IV_{t+1} + u_{M,t}^{IV} = \sum_{i=1}^{N_t} j_i^2 + u_{M,t}^{IV}, \]
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 behave as Gaussian white noise sequences.

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 noise measure of the true, efficient price, $y_t = \frac{p_t + i}{M} + \epsilon_{t,i}$, where $\epsilon_{t,i}$ captures the market microstructure noise. In this case, the above measures will be biased such that
\[ \hat{IV}_{t+1} = B_{M,t}^{IV} + IV_{t+1} + u_{M,t}^{IV} = B_{M,t}^{IV} = \sum_{i=1}^{N_t} j_i^2 + u_{M,t}^{IV}, \]
where $B_{M,t}^{IV}$ and $B_{M,t}^{JV}$ capture the biases while $u_{M,t}^{IV}$ and $u_{M,t}^{JV}$ still behave as Gaussian white noise sequences, 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 $B_{M,t}^{JV} \simeq B_{M}^{JV}$ only depend on the intra-daily sampling frequency and so are approximately constant across different days.

All the above 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. Moreover, eq. (3) holds with $B_{M,t}^{IV} \simeq B_{M}^{IV}$ and $B_{M,t}^{JV} \simeq B_{M}^{JV}$ under weak assumptions on the measurement errors. Thus, supposing we are not willing to take a stand on intra-daily seasonalities and the particular nature of the market micro-structure noise, these provide useful daily statistics carrying information about the diffusive jump component. On the other hand, these measures are backward-looking and are on their own not informative about future volatility and jump-risks. For that purpose, we have to impose additional structure on the price process. The next section introduces a general class of parametric jump-diffusion models that achieve 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 jump-risks. First, we assume that \( \mu_t = \mu \) which is a reasonable assumption at a daily frequency. Furthermore, suppose that \( \sigma_t = \nu (s_t) \) for some known function \( \nu (s) \), e.g., \( \nu (s) = \exp (s) \), and some underlying state variable \( s_t \) which is a Markov diffusion process solving

\[
d s_t = a (s_t; \theta) \, dt + b (s_t; \theta) \, dW_{2,t},
\]

for some other Brownian motion \( W_{2,t} \) with \( \text{Cov} (dW_{1,t}, dW_{2,t}) = \rho \). Here, the drift, \( a (s_t; \theta) \), and the volatility, \( b (s_t; \theta) \), of \( s_t \) are known up to some parameter \( \theta \), while the correlation coefficient \( \rho \) captures potential leverage effects. The jump size sequence \( J_t \) is assumed to be independent of past price and volatility movements and follow some distribution known up to \( \theta \), \( J_t \sim F (J; \theta) \). Finally, we assume that the Poisson process \( N_t \) has jump intensity \( \lambda_t \) which is allowed to be time-varying as described some parametric model, e.g.,

\[
d \lambda_t = c \left( \lambda_t; \theta \right) dt + d \left( \lambda_t; \theta \right) dW_{3,t},
\]

for a third Brownian motion \( W_{3,t} \). The model is fully parametric with all components specified up to some unknown parameter \( \theta \in \Theta \). In particular, note that \( (p_t, s_t, \lambda_t) \) is a Markov process. 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 \( \theta \) from observations of 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 \( \{ RV_t, IV_t \} \), \( i = 1, ..., n \). In the first case (i), we have only one measurement equation,

\[
p_t \sim f_p (p_t | p_{t-1}, s_{t-1}, \lambda_{t-1}; \theta),
\]

where \( f_p (p_t | p_{t-1}, s_{t-1}, \lambda_{t-1}; \theta) \) is the transition density induced by eq. (1) together with the parametric restrictions above, and two state equations,

\[
s_t \sim f_s (s_t | s_{t-1}; \theta), \quad \lambda_t \sim f_\lambda (\lambda_t | \lambda_{t-1}; \theta),
\]

where \( f_s (s_t | s_{t-1}; \theta) \) and \( f_\lambda (\lambda_t | \lambda_{t-1}; \theta) \) are the transition density induced by eqs. (4) and (5). 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. (2), if noise is not present, or eq. (3).

Whether we are under observation scheme (i) or (ii), the estimation of the SV model is complicated by two facts: First, the transition densities \( f_p, f_s \) and \( f_\lambda \) are, in general, not available on closed form. Second, we do not observe \( s_t \) and \( \lambda_t \) and so face a dynamic
latent variable problem. This means that full-information likelihood inference is computationally very challenging. We instead develop a simpler estimation method in the next section.

4. Limited Information Estimation

We choose to base estimation and inference on a suitably chosen set of sample statistics that should summarize the relevant information in data regarding the jump-diffusion model. For convenience, we let \( y_t \) denote the \( t \)th observation in both cases: (i) \( y_t := p_t \) or (ii) \( y_t = (p_t, RV_t, IV_t) \) for \( i = 1, \ldots, n \). We then choose a set of statistics given data. At the most general level, the set of statistics is a vector mapping taking data into a set, say \( \mathbb{R}^d \) or (ii) \( \mathbb{R}^d_{\geq 1, new \ variables} \), \( Z_n = Z_n (y_1, \ldots, y_n) \in \mathbb{R}^d \). Given the chosen set of sample statistics, we then estimate the parameters through the posterior mean obtained from the corresponding Euler discretization. Choose a set of intradaily time points, \( t_i = i/M \), where \( M \geq 1 \) is the step size, and then compute iteratively, for any given value of

\[
\hat{\theta}_{BIL} = \int_{\Theta} f_n (\theta|Z_n) \, d\theta,
\]

where \( f_n (\theta|Z_n) \) is the posterior distribution,

\[
f_n (\theta|Z_n) = \frac{f_n (Z_n, \theta)}{f_n (Z_n)} = \frac{f_n (Z_n|\theta) \pi (\theta)}{\int_{\Theta} f_n (Z_n|\theta) \pi (\theta) \, d\theta}.
\]

For most models, the indirect likelihood, \( f_n (Z_n|\theta) \), is not available on closed form and so standard simulation-based methods for computing the posterior mean, \( \hat{\theta}_{BIL} \), cannot be employed. We here follow the ABC literature and Creel and Kristensen (2013) and combine simulations and nonparametric regression techniques to obtain an approximate version of \( \hat{\theta}_{BIL} \). First, obtain a swarm of i.i.d. draws \( (\theta^s, Z^s_n) \), \( s = 1, \ldots, S \), from \( f_n (Z_n, \theta) \), and then compute the following kernel regression estimator,

\[
\hat{\theta}_{SBIL} = \bar{E}_S [\theta|Z_n] = \frac{\sum_{s=1}^{S} \theta^s K_h (Z^s_n - Z_n)}{\sum_{s=1}^{S} K_h (Z^s_n - Z_n)},
\]

where \( K_h (z) := K (z/h) / h, K : \mathbb{R}^d \rightarrow \mathbb{R} \) is a so-called kernel function and \( h > 0 \) is a bandwidth. The draws can be obtained by following three steps:

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

Repeating above steps 1-3 \( S \) times, we obtain the desired swarm \( (\theta^s, Z^s_n) \), \( s = 1, \ldots, S \), where \( Z^s_n = Z_n (\theta^s) \). 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 set of intradaily time points, \( t_i = i/M \), where \( M \geq 1 \) is the step size, and then compute iteratively, for any given value of
\[ p_t = p_{t-1} + \mu \frac{1}{M} + \eta (s_{t-1}) \frac{1}{\sqrt{M}} \epsilon_{1,i} + J_i \Delta N_i, \] (8)

\[ s_t = s_{t-1} + a (s_{t-1}) \frac{1}{M} + b (s_{t-1}) \frac{1}{\sqrt{M}} \epsilon_{2,i}, \] (9)

\[ \lambda_t = \lambda_{t-1} + c (\lambda_{t-1}) \frac{1}{M} + d (\lambda_{t-1}) \frac{1}{\sqrt{M}} \epsilon_{3,i}, \] (10)

where \( \epsilon_i = (\epsilon_{1,i}, \epsilon_{2,i}, \epsilon_{3,i}) \) is drawn from a trivariate Normal distribution, potentially correlated if so desired, \( J_i \) is drawn from the jump-size distribution, and \( \Delta N_i \) is drawn from a Poisson distribution with intensity \( \lambda_{t-1} \). We have here suppressed the simulated values’ dependence on \( \theta \). Under sampling scheme (ii), once we have computed approximate trajectories of \( p_t \), we obtain the corresponding realized measures using the formulae from \( QV_t \) and \( \hat{IV}_t \) 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 larger than \( M \) under sampling scheme (ii).

Using standard results for kernel regression estimators, we find that, as \( h \to 0 \) and \( Shd \to \infty \) and conditional on \( Z_n \),

\[ \sqrt{Shd} (\hat{\theta}_{SBIL} - \hat{\theta}_{BIL}) \to^d N \left( 0, \frac{\|K\|^2 V(Z_n)}{f(Z_n)} \right), \]

where \( V(Z_n) = \text{Var} [\theta | Z_n] \) and \( f(Z_n) \) is the density of \( F \) see, e.g. Li and Racine (2007). 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. It is important to note that our implementation suffers from a curse-of-dimensionality with the convergence rate, \( \sqrt{Shd} \), slowing down as the dimension \( d \) of the chosen statistic increases.

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 IL 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 choice of statistics.** In the non-model based method, the researcher uses (a relatively large number) 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). 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

\[ r_t = \mu + \sqrt{IV_t} \epsilon_t + \sum_{i=N_{t-1}}^{N_t} J_i, \]

where, \( r_t = p_t - p_{t-1} \) 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. (2). We therefore distinguish between the two sampling cases where realized volatility measures are available or not.

Consider first the case of sampling scheme (i): Without jumps, we only have available daily returns to learn about the volatility dynamics. A natural choice is therefore to base the indirect likelihood inference 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 different 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. 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 \( IV_t \).

4.1.2. Model-based choice of statistics. In the model-specific procedure for choosing the statistics, which is used in the simulation and empirical part (below), 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 Indirect Inference as discussed earlier. 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 statistical point of view. Suppose that such a model has been chosen and can be represented by a (quasi-)likelihood function \( g_n (Y_n | \psi) = g_n (\psi | y_1, ..., y_n) \) given data where \( \psi \) contains the parameters of the auxiliary model. We can then choose \( Z_n \) as the MLE of the auxiliary model,

\[ Z_n = \arg \max_{\psi \in \Psi} g_n (Y_n | \psi), \]
with $Y_n$ being the data generated by the original model. The auxiliary model should be chosen with an eye to 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. In the case of no jumps and under sampling scheme (i), a natural choice is a GARCH-type model; for example,

$$ r_t = \mu + \sigma_t \epsilon_{1,t}, \quad \epsilon_{1,t} \sim N(0, 1) $$

$$ \sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2. $$

If jumps are included in the model, this could be captured by adding a jump component to the first equation,

$$ r_t = \mu + \sigma_t \epsilon_{1,t} + J_t \epsilon_{2,t}, $$

where $J_t$ has the distribution as given by the jump-diffusion model and $\epsilon_{2,t}$ is a Bernoulli variable with jump probability

$$ \lambda_t = a + b r_{t-1}^2 + c \lambda_{t-1}. $$

The above choice of $Z_n$ involves numerical optimization. Recall that the simulated version of BIL takes the form given in eq. (7). For the computation of the indirect likelihood, this requires computation of $Z_n(\theta) = \arg \max_{\psi \in \Psi} g_n(Y_n(\theta) | \psi)$ for each draw $\theta^s$ from the posterior, where

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

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,

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

where $\psi^i$ are i.i.d. draws from the chosen auxiliary prior on $\Psi$. This yields

$$ \hat{\theta}_{SBIL} = \int_{\Theta} \theta \hat{f}_n(\theta | \int_{\Psi} \psi \hat{g}_n(\psi | Y_n) d\psi) d\theta $$

$$ = \frac{\sum_{s=1}^{S} \theta^s K_h \left( \int_{\Psi} \psi \left\{ \hat{g}_n(\psi | Y_n(\theta^s)) - \tilde{g}_n(\psi | Y_n) \right\} d\psi \right)}{\sum_{s=1}^{S} K_h \left( \int_{\Psi} \psi \left\{ \hat{g}_n(\psi | Y_n(\theta^s)) - \tilde{g}_n(\psi | Y_n) \right\} d\psi \right)}. $$

4.2. Bandwidth selection. 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. The most robust bandwidth selection method is so-called cross-validation where $h$ is chosen as

$$ \hat{h} = \arg \min_{h>0} CV(h), $$

where

$$ CV(h) = \sum_{s=1}^{S} \left\| \theta^s - \tilde{E}_{-s} [\theta | Z_n^s] \right\|^2. $$
and \( \hat{E}_{-s}[\theta|Z_n] \) is the leave-one-out estimator,
\[
\hat{E}_{-s}[\theta|Z_n] := \frac{\sum_{i \neq s} \theta^i K_h (Z^i_n - Z^s_n)}{\sum_{i \neq s} K_h (Z^i_n - Z^s_n)}.
\]

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). In practice, to avoid having to evaluate \( \hat{E}_{-s}[\theta|Z_n] \) is large, one can potentially just use a subset of the simulations in the computation of CV \( (h) \).

The above selection method is global in the sense that it minimizes the integrated MSE that averages the estimation error over all potential outcomes of \( Z_n \). Since we are only interested in obtaining a good estimate of \( E[\theta|Z_n] \) as evaluated at the particular observed value of \( Z_n \) that we find from data, there are potential gains from using so-called local bandwidth selection methods since these choose the bandwidth to minimize the pointwise MSE; see, e.g., Fan and Gijbels (1992, 1995), Schucany (1995), Fan et al. (1996), and Prewitt and Lohr (2006).

5. LIMITED INFORMATION FILTERING

Once the model parameters \( \theta \) 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 \( w_t = (p_t, z_t, \lambda_t) \), 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 volatility, we choose \( g(w_{t+m}) = \sigma_{t+m} \). We wish to do so given all information available to us at time \( t \), \( \{y_1, ..., y_t\} \). Suppose we know the data-generating parameter value \( \theta \). We then propose to estimate \( g_{t+m} \) by
\[
g_{t+m|t} := E_\theta [g_{t+m}|F_t],
\]
where \( F_t = F(y_1, ..., y_t) \in \mathbb{R}^d \) is some function of data and \( E_\theta [\bullet|F_t] \) denotes expectations under the model evaluated at \( \theta \). Ideally, one would use all information, \( F_t = (y_1, ..., y_t) \), but this is a high-dimensional computational problem. Instead, we in practice propose to use only a smaller subset, e.g., \( F_t = (y_{1-q}, ..., y_t) \), for some finite, fixed number of lags \( q \geq 1 \) such that the dimension of the forecast information variable remains moderate. This is similar to our limited information estimator, where a moderate sized statistic was used in place of the full data set.

The computation of \( g_{t+m|t} \) is done by combining simulations and kernel regression methods, similar to the implementation of the BIL: First, obtain a swarm of i.i.d. draws \( \{g_{t+m|t}^{s}, F_t^{s}\}, s = 1, ..., S \), from the model evaluated at the estimated parameter value, \( \hat{\theta} \) - these draws are obtained in a similar fashion as for the computation of the BIL - and then compute
\[
g_{t+m|t} = \hat{E}_{\hat{\theta}}[g_{t+m}|F_t] = \frac{\sum_{s=1}^{S} g_{t+m|t}^{s} K_h (F_t^{s} - F_t)}{\sum_{s=1}^{S} K_h (F_t^{s} - F_t)}.
\]
Here, the chosen kernel $K$ and bandwidth $h$ will in general differ from the ones used in the computation of the BIL. Under standard regularity conditions for kernel regression estimators,

$$\sqrt{Sh^d} (\hat{g}_{t+m} - \hat{g}_{t+m}^*) \to^d \mathcal{N}\left(0, \|K\|^2 V(F_t|\theta) f(F_t|\theta)\right),$$

where $V(F_t|\theta) = \text{Var}_\theta [g_t + h | F_t]$ and $f(F_t|\theta)$ is the density of $F$, see, e.g. Li and Racine (2007).

The above method can easily be adjusted in order to construct confidence bands on the forecast. This can for example be done by kernel quantile regression,

$$\hat{q}_{t+m} (\tau) = \arg \min_g \sum_{s=1}^S K_h \left(F^s_t - F_t\right) \rho_{\tau} \left(g^s_{t+m} - g\right),$$

where $\rho_{\tau} ()$ is the so-called check-function as known from quantile regression, see Koenker and Bassett (1978). In particular, $[\hat{q}_{t+m} (0.025), \hat{q}_{t+m} (0.975)]$ will provide a consistent 95% forecast interval.

Similarly, 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. For example, $F_t = (y_{t-q}, ..., y_t, ..., y_{t+q})$.

### 6. Features of Simulation and Empirical Studies

Next we turn to use of the above described methods, in the context of Monte Carlo and empirical work with data on the Stoxx50 equity index. Because the Monte Carlo and empirical work share common features, in this section we give a unified presentation.

#### 6.1. Specific models.

We use two models, a simple model without jumps, and an a version with jumps. The first model is a simple continuous time stochastic volatility model with leverage. Log price $p_t = \log (P_t)$, solves the following continuous-time jump-diffusion model,

$$dp_t = \mu dt + \exp \left(\frac{s_t}{2}\right) dW_{1,t}$$

The log volatility process $s_t$ solves

$$ds_t = \kappa (\theta - s_t) dt + \sigma \left(\rho dW_{1,t} + \sqrt{1 - \rho^2} dW_{2,t}\right)$$

where $W_{1,t}$ and $W_{2,t}$ are two independent Weiner processes. The parameters are interpreted as follows: $\mu$ is the mean of returns; $\theta$ is the mean of log volatility; $\kappa$ is the speed of mean reversion of log volatility, such that low values of $\kappa$ imply high persistence of log volatility; $\sigma$ influences the variance of volatility; and $\rho$ is a leverage parameter that affects the correlation between returns and log volatility.

The second model adds jumps that occur with Poisson frequency, and which are conditionally normally distributed. Log price $p_t$, solves the following continuous-time jump-diffusion model,

$$dp_t = \mu dt + \exp \left(\frac{s_t}{2}\right) dW_{1,t} + J_t$$
The jump process \( J_t \) has a Poisson rate \( \lambda \), parameterized so that the number of jumps in a day is distributed Poisson(\( \lambda \)). Jump sizes, conditional on the occurrence of a jump, are independent and conditionally normally distributed: \( J_t | J_t > 0 \sim N(\mu_J, \sigma^2_J) \). The log volatility process \( s_t \) follows the same process as does the model without jumps, given in equation 14.

6.2. Variables. The data available, discussed below, include daily measures of returns \((ret)\), realized volatility \((RV)\) and median realized volatility \((MedRV)\). For each day, indexed by \( t \), the volatility measures are computed based on \( M \) latent high frequency measurements of the intra-day returns, \( ret_{t,m}, m = 1, 2, ..., M \). We do not observe the intra-day returns, only daily returns and volatility measures are observed. The observed variables are computed as:

\[
\begin{align*}
\text{ret}_t &= \log p_t - \log p_{t-1} \\
RV_t &= \sum_{m=1}^{M} ret_{t,m}^2 \\
MedRV_t &= \frac{\pi}{\pi + 6 - 4\sqrt{3}} \left( \frac{M}{M-2} \right) \sum_{m=2}^{M} \text{med} \left( |ret_{t,m-1}|, |ret_{t,m}|, |ret_{t,m+1}| \right)^2
\end{align*}
\]

6.3. Implementation of SBIL. This section discussed the details of implementation of the simulated Bayesian Indirect Likelihood (SBIL) estimator. The experimentation that was done to select statistics, number of neighbors, etc., used the Monte Carlo data, the generation of which is discussed in the next Section. Here, we focus on explaining the modeling choices that are needed to use SBIL, and the rationale for the decisions. In Section 7 we give the results.

6.3.1. Parameter space and pseudo-prior. First, data is simulated from the two models, drawing randomly from a uniform prior over the parameter space defined as the Cartesian product of the line segments given by the lower and upper bounds for each parameter in Table 1. The parameters related to jump are restricted to be 0 when the model is assumed not to present jumps, as indicated in the last 3 rows of the Table. We began with a larger parameter space for the parameters \( \kappa (0, 1) \); \( \theta (-15, 0) \) and \( \rho (-0.99, 0.99) \) but it became apparent from initial Monte Carlo work that the larger region was superfluous, in that points outside the final bounds given in the Table were almost never selected as neighbors to the auxiliary statistics corresponding to the design points. To avoid generating a large number of superfluous simulations, we narrowed the bounds to the given limits.

The chosen limits are intended to be relatively uninformative for the parameters, given the widely accepted characteristics of data on equity returns and volatility. One advantage of a simulation-based method like SBIL is that this sort of experimentation can be done, which can help to reduce the computational burden.

6.3.2. Selection of statistics. The full set of auxiliary statistics was determined using the model-based approach outlined above, combined with experimentation. Two auxiliary models were used: an EGARCH model fit to standardized and normalized returns, and
an autoregressive model of order 2, fit to the logarithm of the MedRV realized volatility measure. The EGARCH model is (the parameters here are not the same as the parameters of the models of interest, the names are re-used for notational convenience):

\[ \text{ret}_t - \mu = \rho (\text{ret}_{t-1} - \mu) + \epsilon_t \]
\[ \epsilon_t = \exp(\frac{s_t}{2}) u_t \]
\[ s_t = \omega + \alpha |\epsilon_{t-1}| + \gamma \epsilon_{t-1} + \beta s_{t-1} \]

where \( u_t \) is standard Gaussian white noise. This EGARCH model is numerically stable and converges readily using sample-based starting values, so we were able to fit the model to each simulated sample using quasi-Newton methods. There is moderate computational burden due to the use of iterative maximization. It is important to note that the step of generating the simulation draws of the auxiliary statistics can be done prior to the time when estimation or filtering results are needed, so the use of this somewhat costly auxiliary statistic is not an impediment to obtaining the answers needed for financial decisions in a completely timely manner. The full set of auxiliary statistics used is given in Table 2.

It is convenient to reduce the number of auxiliary statistics for estimation of each parameter to the minimal set needed for good identification, so that the dimensionality of the nonparametric fitting step is made as small as possible, preserving good identification and precision of estimation. This allows for obtaining precise results without having to use an extremely large number of simulations. The choice of which statistics to use for estimation of each parameter was guided by economic reasoning, taking into account the model, as well as experimentation. For example, the parameter \( \mu \) in equation 13 is clearly well identified by the sample mean of returns, and statistics such as the estimates of the parameters of the auxiliary AR model for \( \log \text{MedRV} \) are likely to simply exacerbate “curse of dimensionality” problems if used for estimation of \( \mu \). The goal of selection of statistics for each parameter is to keep the dimension as low as possible while ensuring good identification. A formal means of doing this task, possibly by using machine learning methods, remains a topic for future research. The auxiliary statistics selected for estimation of each parameter are detailed in Table 3.

6.3.3. Cross validation to determine number of neighbors. We implement SBIL using \( k \)–nearest neighbors nonparametric regression. In previous work, we have not carefully explored
Table 2. The full set of auxiliary statistics

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EGARCH$\mu$</td>
</tr>
<tr>
<td>2</td>
<td>EGARCH $\rho$</td>
</tr>
<tr>
<td>3</td>
<td>EGARCH $\omega$</td>
</tr>
<tr>
<td>4</td>
<td>EGARCH $\alpha$</td>
</tr>
<tr>
<td>5</td>
<td>EGARCH $\gamma$</td>
</tr>
<tr>
<td>6</td>
<td>EGARCH $\beta$</td>
</tr>
<tr>
<td>7</td>
<td>ret</td>
</tr>
<tr>
<td>8</td>
<td>$\log \text{ret}^2$</td>
</tr>
<tr>
<td>9</td>
<td>skew(ret)</td>
</tr>
<tr>
<td>10</td>
<td>kurtosis(ret)</td>
</tr>
<tr>
<td>11</td>
<td>$\log \text{MedRV}$</td>
</tr>
<tr>
<td>12</td>
<td>$\text{std}(\log \text{MedRV})$</td>
</tr>
<tr>
<td>13</td>
<td>$\text{corr}(\text{ret}, \log \text{MedRV})$</td>
</tr>
<tr>
<td>14</td>
<td>$\log \text{MedRV}$ AR model, constant</td>
</tr>
<tr>
<td>15</td>
<td>$\log \text{MedRV}$ AR model, 1st lag</td>
</tr>
<tr>
<td>16</td>
<td>$\log \text{MedRV}$ AR model, 2nd lag</td>
</tr>
<tr>
<td>17</td>
<td>Autogression, error variance</td>
</tr>
<tr>
<td>18</td>
<td>$\log \text{RV} - \log \text{MedRV}$</td>
</tr>
<tr>
<td>19</td>
<td>$\text{corr}(\log \text{RV}, \log \text{MedRV})$</td>
</tr>
<tr>
<td>20</td>
<td>$\text{std}(\log \text{RV}, \log \text{MedRV})$</td>
</tr>
<tr>
<td>21</td>
<td>$(\log \text{RV} - \log \text{MedRV})^3$</td>
</tr>
</tbody>
</table>

Table 3. Selected auxiliary statistics, by parameter. Enumeration refers to Table 2, second column.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Selected statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>1,7</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>3-6,8,21</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>3-6,8,21</td>
</tr>
<tr>
<td>$\theta$</td>
<td>3, 8, 10-19</td>
</tr>
<tr>
<td>$\rho$</td>
<td>4-6,9,10,13</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>1-6,8-21</td>
</tr>
<tr>
<td>$\mu_J$</td>
<td>1-6,8-21</td>
</tr>
<tr>
<td>$\sigma_J$</td>
<td>1-6,8:21</td>
</tr>
</tbody>
</table>

To choose the number of neighbors, instead, we have used the rule $k = aS^b$ (rounding to an integer) where $a$ is a value such as 1.5, and $b$ has always been set to 0.25. Here, we perform cross validation, as discussed above in Section 4.2, to optimally choose $k$ for each parameter. The exact procedure we used is as follows:

1. Draw $10^6 (\theta^s, Z^s)$ pairs from the pseudo-prior (described below, in Section 7).
2. Randomly select 500 of the pairs as the out-of-sample portion. Note that the true parameter values are different for each pair, having been drawn from the pseudo-prior.
(3) Perform SBIL to estimate the parameters for each of the 500 out-of-sample draws, using the remaining draws as the data for the $k$-nearest neighbors nonparametric regression.

(4) repeat step 3 using different values of $k$

(5) choose the optimal $k$, by selecting the value that gives minimal out-of-sample RMSE, for each of the parameters.

Note that this procedure is not targeted to a specific design point, and it is fully implementable using real data. One could obtain a more accurate approximation to integrated forecast MSE by using a larger simulation set and out-of-sample set, but we find that the optimal $k$ is quite well identified using the chosen values.

This is apparent in Figure 1, which shows the cross validation profiles, for the model with jumps. We see that a one-size-fits-all rule is not appropriate, and that using different numbers of neighbors for estimation of the different parameters can give gains. The parameters $\mu$ and $\rho$ benefit from using a relatively large number of neighbors, while the parameters $\kappa$ and $\theta$ require a small number of neighbors. The other parameters are best estimated using intermediate values. This was done only using the model with jumps. For the model without jumps, which is nested in the model with jumps, we rely on these results.
7. Simulation study

This section reports on implementation of SBIL for estimation of the two models under a known data generating process. We estimate the models using both returns and the high frequency realized measures.

The Monte Carlo results use a sample size of \( n = 3027 \), which is the same as the total in sample portion of the Stoxx50 series, 2000-2011. The true design values for the parameters of the DGPs are given in the second columns of Tables 4 for the no jump DGP and 5 for the jump DGP. These values were chosen based on values found in the literature, and also based upon preliminary exploration using the Stoxx50 data, and are intended to be fairly normal values for financial series such as equity returns, e.g., volatility is quite persistent and leverage is negative. The true probability of a jump is set so that the mean number of jumps is one per 30 days, and the upper bound is set so mean jumps per month are 5. We had less intuition for the parameters of mean jump size, \( \mu_J \) and the standard deviation of jump size, \( \sigma_J \). These were calibrated using exploratory estimation using the Stoxx50 data. The lower and upper bounds of the parameter space are given in Table 1, as was already noted above.

7.1. No jump model. Table 4 reports the results for the model without jumps, when the DGP may or may not present jumps. We present results where estimation is based only on returns, and on both returns and realized measures. We also present the performance of the prior mean when used as an estimator, for reference. For most parameters, the prior mean is biased for the true parameter values (compare columns 2 and 3), so reduction in bias when using the SBIL estimator illustrates its good performance in this sense. Columns 5-8 present results for the case that the DGP does not present jumps, so that the model is correctly specified. In columns 5 and 6, we have the mean and RMSE of the SBIL estimator when only returns are used to estimate the parameters, while columns 7 and 8 present the mean and RMSE when both returns and realized measures are used. Comparing columns 6 and 8, we see that RMSE is generally lower, when realized measures are used, with the case of \( \theta \) being exceptionally notable. Focusing on the results that use realized measures, the SBIL estimator has a mean close to the true values (compare columns 2 and 7) and it is considerably less biased than is the prior mean. RMSE of the SBIL estimator is considerably lower than that of the prior mean (compare columns 4 and 8), showing that the estimator is extracting information about all parameters from the sample. When the DGP presents jumps, the no jump model is misspecified. Comparing columns 7 and 8 (model is correct) to columns 9 and 10 (model is misspecified, and is estimated using returns and realized measures), we see that some notable biases appear, for all parameters except \( \mu \). RMSE for all parameters except \( \mu \) is much worse when the model is misspecified than when it is correct. It is clear that if one ignores the presence of jumps, the parameters of the non-jump portion of volatility will not be correctly estimated.
We conclude that: 1) information on realized measures should be used, if available; 2) the no jump model is not robust to the presence of jumps in the DGP and 3) the SBIL estimator give quite accurate results when the model is correctly specified and information of realized measures is available.

7.2. Jump model. Next, we examine the results for the model with jumps, under correct specification (DGP has jumps) and over-parameterization (DGP presents no jumps). The results are in Table 5. When the model is correctly specified (DGP has jumps), we present results for the SBIL estimator for two cases: that where only returns are available, and that where returns are supplemented by realized measures. Columns 7 and 8 give results for estimation using only returns, while columns 9 and 10 give results when returns and realized measures are used. We see that the use of realized measures is very important for estimation of the parameters, with much lower RMSEs when realized information is used. Essentially, SBIL estimation using only returns does not give useful estimates of the parameters, while when realized measures are used to supplement returns, SBIL estimation gives accurate results that have little bias. Low bias is seen by comparing columns 2 and 9, while column 10 shows that RMSE is low. The persistence parameter of the non-jump component of volatility, $\kappa$, and the jump occurrence parameter, $\lambda$, each have a non-negligible bias, suggesting that these two parameters may be difficult to identify separately in a model with jumps. It is possible that a better choice of auxiliary statistics could help to better distinguish the individual effects of these parameters.

When the jump model is estimated, but the DGP presents no jumps, the model is overparameterized, though it does nest the true model, so serious estimation problems should not be encountered. Comparing column 8 of Table 4 and column 6 of 5, we see that this is the case, RMSE increases moderately for all parameters due to the use of an over-parameterized model, with $\theta$ being the most affected parameter. For the estimated parameters of the jump part of volatility, we see that that $\lambda$ is estimated to be 0.015 when it is actually 0 when there are no jumps. Note that the true value of $\lambda$ is on the boundary of the parameter space, and that $\mu_J$ and $\sigma_J$ are not identified when $\lambda = 0$. Because of these issues, the moderate bias of the SBIL estimator of these parameters is not unexpected given that a misspecified model is being used. One could perhaps obtain a better estimate of $\lambda$ by using local linear nonparametric regression instead of nearest neighbors regression, as this other method is know to perform better at the boundaries of the data. We view this as a minor issue, though. Comparing the ninth columns of Tables 4 and

<table>
<thead>
<tr>
<th>Parameters</th>
<th>True values</th>
<th>Ret. only</th>
<th>Ret. and RV</th>
<th>Ret. only</th>
<th>Ret. and RV</th>
<th>Ret. only</th>
<th>Ret. and RV</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.006</td>
<td>0.000</td>
<td>0.001</td>
<td>-0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>0.030</td>
<td>0.101</td>
<td>0.091</td>
<td>0.033</td>
<td>0.008</td>
<td>0.032</td>
<td>0.006</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.200</td>
<td>0.505</td>
<td>0.418</td>
<td>0.209</td>
<td>0.027</td>
<td>0.203</td>
<td>0.012</td>
</tr>
<tr>
<td>$\theta$</td>
<td>-9.000</td>
<td>-10.000</td>
<td>3.055</td>
<td>-10.002</td>
<td>1.467</td>
<td>-9.055</td>
<td>0.144</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-0.600</td>
<td>-0.495</td>
<td>0.304</td>
<td>-0.490</td>
<td>0.129</td>
<td>-0.588</td>
<td>0.069</td>
</tr>
</tbody>
</table>
Table 5. Monte Carlo results, Jump model

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Prior</th>
<th>No jump DGP</th>
<th>Jump DGP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True value</td>
<td>Mean</td>
<td>RMSE</td>
</tr>
<tr>
<td>$\mu$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.006</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>0.030</td>
<td>0.100</td>
<td>0.091</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.200</td>
<td>0.505</td>
<td>0.418</td>
</tr>
<tr>
<td>$\theta$</td>
<td>-9.000</td>
<td>-10.000</td>
<td>3.055</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-0.600</td>
<td>-0.495</td>
<td>0.304</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>$1/30=0.033$</td>
<td>0.083</td>
<td>0.069</td>
</tr>
<tr>
<td>$\mu_J$</td>
<td>0.050</td>
<td>0.000</td>
<td>0.126</td>
</tr>
<tr>
<td>$\sigma_J$</td>
<td>0.050</td>
<td>0.102</td>
<td>0.077</td>
</tr>
</tbody>
</table>

5. we see that there are important differences in the parameters of non-jump volatility across the no-jump and jump models, when the DGP presents jumps. This suggests that a test for presence of jumps may be based upon the differences in the estimates of these parameters, though we leave this is left for future work.

In summary: 1) for estimation of the jump model, use of realized measures is fundamental, and the SBIL estimator does not perform acceptably if only returns are used for estimation; 2) when returns and realized measures are used, the jump model is estimated quite well by SBIL; and 3) the jump model, estimated by SBIL, leads to acceptable results even when jumps are not present.

8. Empirical study

8.1. Data. Based on the simulation results, all estimation is done using both returns and realized measures, because the needed data is available, and use of realized measures has been found to benefit estimation of both the no-jump and jump models.

The data series used are daily returns, realized volatility (RV) and median realized volatility (MedRV) of the Euro Stoxx50 (live) series, taken from the Oxford-Man Institute’s realized library version 0.2 (Heber, Gerd, Asger Lunde, Neil Shephard and Kevin Sheppard, 2009). The data is available from Jan. 03, 2000 through June 28, 2013. This data source does not include the high frequency intra-day returns that were used to compute $RV$ and $MedRV$, only daily measures are available. The realized measures were computed using a frequency of intraday observations of every 5 minutes. Trading days for the Euro Stoxx50 series are 8 hours long, so a 5 minute frequency of observation means that $M = 96$ in the realized volatility definitions in equations 17 and 18.

Plots of the three series are given in Figures 2 and 3, and a kernel estimates of the density of returns is given in Figure 4. The typical features of volatility clusters, leptokurtosis, and some skewness of returns are apparent. There are a number of observations where ordinary realized volatility, $RV$, and the jump-robust measure, $MedRV$, differ substantially, suggesting the presence of jumps in returns.

To investigate the effects of the recent financial crisis and to have date for out-of-sample forecasting, we split the data into 3 portions. The first two portions, 2000-2007 ($n = 2023$ observations) and 2008-2011 (1004 observations) are used to fit the parameters of the
models, while the portion Jan. 2012-June 28, 2013 \((n = 381\) observations) is reserved for out-of-sample forecasting purposes.

8.2. Estimation results. Estimation results using the full in sample period (2000-2011) of the Stoxx50 data are very simple to obtain once the Monte Carlo work has been done, we just replace the 5000 Monte Carlo replications of \(Z_n\) with the actual sample \(Z_n\), and the results are immediate. Then standard errors are computed by parametric bootstrap: the estimated parameter vector is used as the design point to generate a number of bootstrap samples, and SBIL is applied to each of these, exactly as is done in the Monte Carlo work. The across bootstrap sample standard error of the estimates is used as the bootstrap standard error estimate. To investigate model stability across financial crisis that began in late 2008, we also estimate models that allow for a parameter shift at the beginning of 2008,
using the two periods 2001-2007 sample (2023 observations) and using the 2008-2011 period (1004 observations). In Figures 2 and 3, there is a spell of increased volatility in early 2008 which is considerably more active than anything in 2007. For this reason, we choose the break point as the beginning of 2008. One might also consider that October, 2008 could be the break point, but in this paper we only explore the first option.

The estimation results for the full sample and the two sub-samples are given in Table 6. The estimates of the models’ parameters are fairly close to the design values of the Monte Carlo work, so we expect that the lessons learned from analysis of the Monte Carlo results will apply. In particular, when comparing the no jump results to the jump results, we observe the same pattern that occurred in the Monte Carlo work when the no jump model was fit to data that actually contains jumps, and thus is misspecified: the estimated $\sigma$ values are notably higher than the value given by the model with jumps, and the estimated $\kappa$ value is notably lower. Given this, and given our prior belief that the model with jumps is much more likely to be approximately well specified, we focus on the results for this model.

For the whole 2000-2011 sample, and for both of the two sub-samples, we see that there is quite slow mean reversion in log volatility and important negative leverage. Comparing across the two sample periods, 2000-2007 versus 2008-2011, we see some interesting differences: in the second period which highlights the important financial crisis, log volatility becomes less persistent, with the estimated mean reversion parameter $\kappa$ more than doubling, increasing from 0.027 for the full sample to 0.068 for the crisis period. Likewise, the mean ($\theta$) and variance ($\sigma^2$) of the non-jump portion of volatility is higher in the crisis period. Quite interestingly, the parameter related to the probability of jumps
(λ) is lower in the 2008-2011 period, but when jumps do occur, their mean (μ_J) is somewhat lower in the crisis period. To summarize, the post-crisis period seems to present non-jump volatility that is less persistent, with a smaller variance but a higher mean. Leverage seems to have become somewhat less important. Jumps are somewhat less common, and are more likely to be negative when they do occur.

To highlight the differences produced by the financial crisis, Figure 5 plots the bootstrap densities of the parameter estimates. It is quite apparent that at least some of the parameters (most notably, κ, σ, θ, λ, μ_J) changed during the crisis period. For this reason, the model that imposes parameter constancy over the 2000-2011 is likely to suffer from misspecification. We will henceforth focus on the model that uses only the 2008-2011 data.

8.3. Limited information filtering. For the discussion of filtering and smoothing, we adopt the model with jumps. Because of this, we focus on the jump-robust realized volatility measure, MedRV (see equation 18) rather than on the non-robust measure RV (equation 17). A limited information approach to filtering the non-jump portion of volatility, s_t (see equation 14) can be implemented as follows:

1. Given the estimated parameter vector ^θ, simulate a long series for the model \( \{ (ret_r, MedRV_r, s_r, J_r) \} , r = 1, 2, ..., R. \)
2. Use nonparametric regression (we use k nearest neighbors) to fit the nonparametric reduced form forecasting model

\[
\hat{s}_r = \hat{s}_r(ret_{r-1}, ..., ret_{r-p}, MedRV_{r-1}, ... MedRV_{r-q})
\]

where the function \( \hat{s}_r(\cdot) \) represents the nonparametric fitting function. The lags \( p \) and \( q \) can be chosen to minimize out of sample mean squared forecast error, using the simulated data.
Figure 5. Bootstrap densities of parameter estimates, 2000-2007 (plain lines) and 2008-2011 (lines with markers)

(A) $\mu$
(B) $\kappa$
(C) $\sigma$

(D) $\theta$
(E) $\rho$
(F) $\lambda$

(G) $\mu_f$
(H) $\sigma_f$

(3) Filter using the real data by applying the selected nonparametric fitting function to the real data:

$$\hat{s}_t = \hat{s}_t(ret_{t-1}, ..., ret_{t-p}, MedRV_{t-1}, ... MedRV_{t-q})$$

8.3.1. Verification of filtering method using simulated data. We implemented steps 1 and 2 of this procedure by setting the parameter vector $\hat{\theta}$ to the estimated values in column 7 of Table 6. We use a simulation of length $R = 10^6$. The last 20000 elements of the simulated series were used as the out-of-sample portion. It is necessary to use a fairly long out of sample portion in order for out-of-sample RMSE to be stable, because the series we are filtering exhibits stochastic volatility (of course). A long out-of-sample portion ensures that periods of high and low volatility will be encountered with more or less their long-run probabilities. If a short series is used, RMSE will be low if the series happens not to contain a cluster of high volatility, and it will be relatively high if it does. The optimal values of $p$ and $q$ were found to be $p = 1$ and $q = 5$, using $k = 250$ nearest neighbors for the nonparametric fit. The out-of-sample RMSE for forecasting the non-jump component of volatility, $s_r$, is 0.306, and the out-of-sample $R^2$ is 0.81. There is the question of whether or not this represents a good forecast of volatility. A first idea is given in Figure 6, which shows, for clarity, only the first 500 elements of the out-of-sample portion of the
simulation of $h_s$, along with the filtered values. We can see that the filtered series follows the true series fairly closely. Turning to theoretical considerations, the true standard deviation of $s_r$ is given by the parameter $\sigma = 0.250$. This represents the lower bound on out-of-sample RMSE for forecasting $s_r$, if the lags of $s_r$ were observable. Our method, using the observable variable, gives us out-of-sample RMSE 22% larger than the infeasible lower bound. This seems quite reasonable for a limited information method, we believe. We have not yet compared the limited information method to a full information filtering method at this point, but this is an item for future work.

8.3.2. Filtered Stoxx50 data, 2012-mid 2013. Next, we implement step 3 of the filtering method using the real Stoxx50 data, using the real out-of-sample portion of the data, which are the 381 observations corresponding to all of 2012 and 2013, up to June 28. The filtering uses the estimated parameters of the jump model, based upon the 2008-2011 sample period. Filtering could be based upon estimated coefficients that are periodically updated, potentially as every new observation arrives. While this conceptually easy to do, we have not undertaken the exercise. The reader will recall that there were serious doubts regarding the future of the euro during summer of 2012: “Only a few months ago, it was front-page news. Would the euro collapse? Would most of southern Europe go broke, unable to borrow money at any reasonable rate? Would that bring on a new world recession?” (Floyd Norris, Oct. 25, 2012, ‘Euro Survives, but Future Is in Doubt’, New York Times). The ECB’s policy of Outright Monetary Transactions was announced 02 August, 2012, and this program was effective in calming market volatility. These historical notes are reflected in filtered volatility, which is plotted in Figure 7: We observe a rise in volatility during the first part of 2012, followed by a downward trend in the latter
portion of that year, after the announcement of OMT, with a new turn upwards during the first half of 2013. The maximum of filtered volatility occurs on 03 Aug., 2012, and it drops sharply afterwards.

8.4. **Limited information smoothing.** Limited information smoothing can be done 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. While it makes no sense to attempt to filter jumps, which are inherently unpredictable, we can use smoothing to identify the occurrence of jumps in the sample. Thus, smoothing is done for both ordinary volatility, $s$, and for jumps, $J$ (see equations 14 and 15).

The method is:

1. Given the estimated parameter vector $\hat{\theta}$, simulate a long series for the model $\{(ret_r, MedRV_r, s_r, J_r)\}, r = 1, 2, ..., R$. We can use the same long series as was used for filtering.
2. For the latent variable of interest $y \in \{s, J\}$, use nonparametric regression (we use $k$ nearest neighbors) to fit the nonparametric reduced form forecasting model

$$\hat{y}_r = \hat{y}_r(ret_r, ret_{r\pm1}, ..., ret_{r\pm p}, MedRV_r, MedRV_{r\pm1}, ... MedRV_{r\pm q})$$

where the function $\hat{y}_r(\cdot)$ represents the chosen nonparametric fitting function. The 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.
(3) Smooth using the real data by fitting the optimal (in terms of $p, q$ and number of neighbors, $k$) nonparametric forecasting model to the real data:

$$\hat{y}_t = \hat{y}_t(ret_t, ret_{t\pm1}, ..., ret_{t\pm p}, MedRV_t, MedRV_{t\pm1}, ..., MedRV_{t\pm q})$$

8.4.1. **Smoothing simulated data.** Steps 1 and 2 of the smoothing method are used to perform smoothing for the two latent series $s_r$ (non-jump log volatility) and $J_r$ (jumps), using the last 20000 elements of the simulated series as the out-of-sample portion. Relying on the results from filtering, and adjusting for the fact that we use leads as well as lags, we use $p = 1$ and $q = 2$ to illustrate smoothing. These values that were chosen *ad hoc*, without experimentation. Again, we use $k = 250$ neighbors. The out of sample RMSEs are 0.136 for $s$ and 0.003 for $J$, and the last 500 elements of the true and smoothed out-of-sample series are given in Figures 8 and 9. The smoothing results for $s$ are similar to the filtered results, but are naturally more accurate. The smoothed results for jumps show that the method identifies quite well the timing and magnitude of jumps. In the plot of smoothed jumps, a point is plotted when predicted jump size is a small tolerance different from zero, so there are a number of points where the green line (predicted jumps) is slightly different from zero, while the blue line (true jumps) is zero. Nevertheless, it is clear that actual jumps, when they do occur, are identified with good accuracy.

8.4.2. **Smoothing the Stoxx50 data.** Next we apply the smoothing method to the Stoxx50 data, using the out-of-sample period 2012-mid 2013. The plots for smoothed volatility, $s$, and smoothed jumps, $J$, are in Figures 10 and 11, respectively. The smoothed plot for $s$ looks quite similar to the filtered plot in Figure 7, though a bit less choppy. The plot of smoothed jumps shows some clear jump activity. The largest positive jump occurred on 26 July 2012. At this time, there was much speculation that central banks would further
intervene in markets (the ECB’s OMT policy was announced shortly after), and investors were acquiring equities. The largest negative jump occurred 4 Feb. 2013. If one examines the historical chart for the Stoxx50 series, one observed that the identified jumps coincide in date with extreme movements of returns.

One feature that seems apparent in the smoothed jump plot is that there is some clustering of jumps. This apparent feature of the data is not contemplated by the jump model, which assumes that the occurrence of jumps follows a Poisson process. There may be scope for improving the model by allowing for dependence in the jump process.

9. CONCLUSION

Extensions:

- Jumps seem to occur in clusters, so the Poisson assumption is not supported very well. Some form of autoregression in the probability of jumps would probably improve the model. For example, one could assume that the jump intensity evolves according to a CIR diffusion process,
  \[ d\lambda_t = \beta (\alpha - \lambda_t) dt + \sigma \sqrt{\lambda_t} dW_t, \]
  as done in Fang (2000).
- Re-doing the estimation using large number of summary statistics combined with MARS or LASSO to choose which ones of the statistics are informative. This could be done along the lines of XXXX.
- Examining performance of local bandwidth selection procedures (see references in bandwidth selection section).
• formal test of a structural break.
• Using GARCH volatility for filtering and forecasting.
• constructing confidence bands for filtering and forecasting using quantile regression.
REFERENCES


Universitat Autònoma de Barcelona, Barcelona Graduate School of Economics, and MOVE

University College London, Cemmap (Centre for Microdata Methods and Practice, IFS) and CREATES (Center for Research in Econometric Analysis of Time Series, University of Aarhus).