The Journal of Open Source Software

stochvolTMB: An R-package for likelihood estimation of stochastic volatility models

Jens Christian Wahl[1]

1 Norwegian Computing Center

DOI: XX.XXXXX/joss.XXXXX

Software

- Review
- Repository
- Archive

Submitted:
Published:

Summary

Stochastic volatility (SV) models are often used to model financial returns that exhibit time-varying and autocorrelated variance. The first SV model was introduced by Taylor (1982) and models the logarithm of the variance as a latent autoregressive process of order one. Parameter estimation of stochastic volatility models can be challenging and a variety of methods have been proposed, such as simulated likelihood (Liesenfeld and Richard 2006), quasi-maximum likelihood (Harvey, Ruiz, and Shephard 1994) and Markov Chain Monte Carlo methods (MCMC) (Pitt and Shephard 1999; Kastner 2016). `stochvolTMB` estimates the parameters using maximum likelihood, similar to Skaug and Yu (2014). The latent variables are integrated out using Laplace approximation. The models are implemented in `C++` using the `R`-package (R Core Team 2019) `TMB` (Kristensen et al. 2016) for fast and efficient estimation. `TMB` utilizes the `Eigen` library (Guennebaud, Jacob, and others 2010) for numerical linear algebra and `CppAD` (Bell 2005) for automatic differentiation of the negative log-likelihood.

Statement of need

The `stochvolTMB` `R`-package makes it easy for users to do inference, plotting and forecasting of volatility. The `R`-package `stochvol` (Kastner 2016) also performs inference for stochastic volatility models, but differs from `stochvolTMB` since it performs Bayesian inference using MCMC and not maximum likelihood. By using optimization instead of simulations one can obtain substantial speed up depending on the data, model, number of observations and number of MCMC samples.

Implementation

`stochvolTMB` implements stochastic volatility models of the form

$$
\begin{aligned}
y_t &= \sigma_y e^{h_t/2}\epsilon_t, \quad t = 1, \dots, T, \\
h_{t+1} &= \phi h_t + \sigma_h \eta_t, \quad t = 1, \dots, T-1, \\
\eta_t &\overset{\text{iid}}{\sim} \mathcal{N}(0, 1), \\
\epsilon_t &\overset{\text{iid}}{\sim} F, \\
h_1 &\sim \mathcal{N}\left(0, \frac{\sigma_h}{\sqrt{(1-\phi^2)}}\right),
\end{aligned} \tag{1}
$$

where y_t is the observed log return for day t, h_t is the logarithm of the conditional variance of day t and is modelled as an AR(1) process, $\boldsymbol{\theta} = (\phi, \sigma_y, \sigma_h)$ is a vector of the fixed parameters and F denotes the distribution of ϵ_t. Four distributions are implemented for

ϵ_t: (1) The standard normal distribution; (2) The t-distribution with ν degrees of freedom; (3) The skew-normal distribution with skewness parameter α; and (4) The leverage model where (ϵ_t, η_t) are both standard normal with correlation coefficient ρ. The last three distributions add an additional fixed parameter to $\boldsymbol{\theta}$. `stochvolTMB` also supports generic functions such as `plot`, `summary`, `predict` and `AIC`. The plotting is implemented using `ggplot2` (Wickham 2016) and data processing utilizes the `R`-package `data.table` (Dowle and Srinivasan 2019).

The parameter estimation is done in an iterative two-step procedure: (1) Optimize the joint negative log-likelihood with respect to the latent log-volatility $\boldsymbol{h} = (h_1, \ldots, h_T)$ holding $\boldsymbol{\theta}$ fixed, and (2) Optimize the Laplace approximation of the joint negative log-likelihood w.r.t $\boldsymbol{\theta}$ holding $\boldsymbol{h}$ fixed. This procedure is iterated until convergence. Standard deviations for the log-volatility and the fixed parameters are obtained using a generalized delta-method (Kristensen et al. 2016).

Example

As an example we compare the different models on log returns for the S&P index from 2005 to 2018:

```
library(stochvolTMB)
data(spy)
gaussian = estimate_parameters(spy$log_return, model = "gaussian")
t_dist = estimate_parameters(spy$log_return, model = "t")
skew_gaussian = estimate_parameters(spy$log_return, model = "skew_gaussian")
leverage = estimate_parameters(spy$log_return, model = "leverage")
```

To compare competing models we can use model selection tools such as AIC (Akaike (1998)):

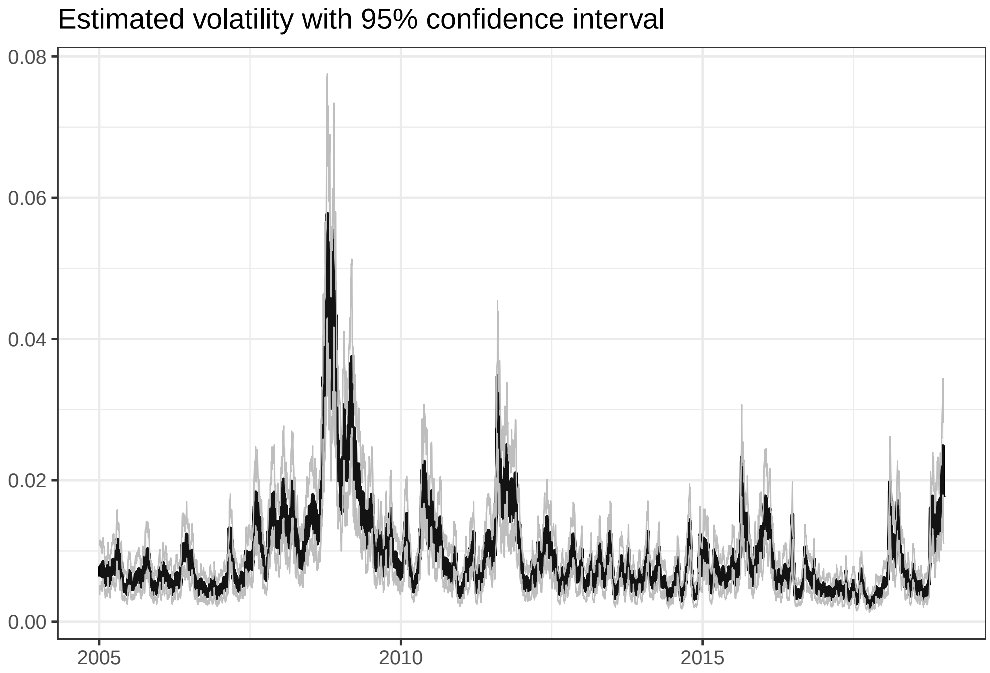
```
AIC(gaussian,
    t_dist,
    skew_gaussian,
    leverage)
```

```
##               df       AIC
## gaussian       3 -23430.57
## t_dist         4 -23451.69
## skew_gaussian  4 -23440.87
## leverage       4 -23608.85
```

The leverage model is preferred in this example. Notice that the Gaussian model performs the worst and shows the importance of having more flexible distributions, even after controlling for the volatility. We can plot the estimated log-volatility with 95% confidence interval

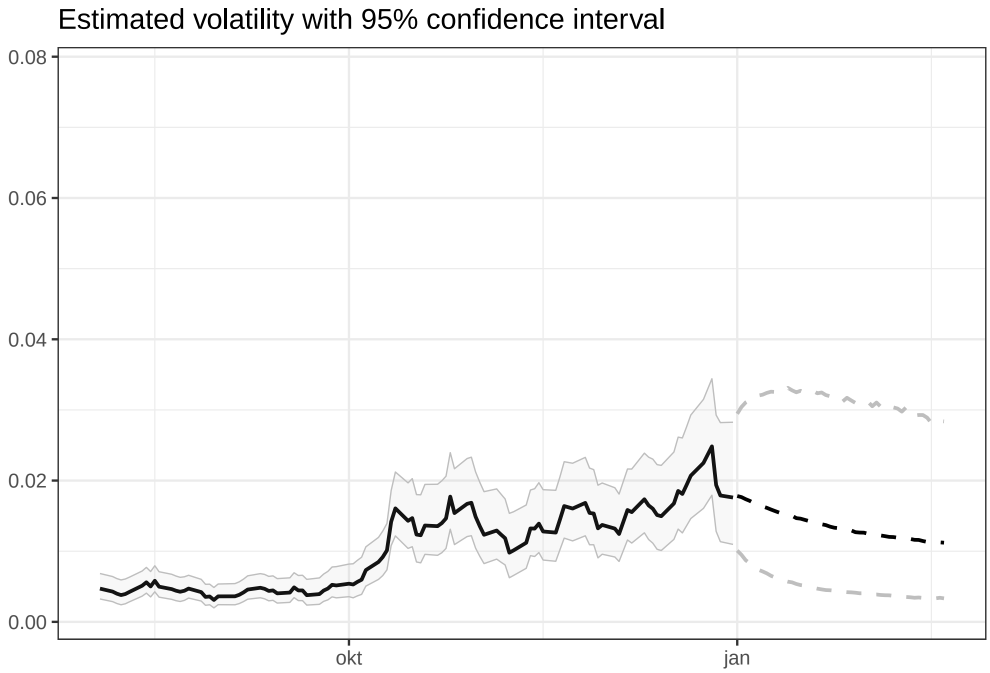
```
plot(leverage, plot_log = FALSE, dates = spy$date)
```

The Journal of Open Source Software



Future volatility can be simulated from the estimated model. Parameter uncertainty of the fixed effects is by default included and is obtained by simulating parameter values from the asymptotic distribution, i.e. a multivariate normal distribution using the observed Fisher information matrix (inverse Hessian of the negative log-likelihood) as the covariance matrix.

```
set.seed(123)
# plot predicted volatility with 95% confidence interval
plot(leverage, plot_log = FALSE, forecast = 50, dates = spy$date) +
  ggplot2::xlim(c(tail(spy$date, 1) - 150, tail(spy$date, 1) + 50))
```




```
prediction = predict(leverage, steps = 5)
summary(prediction, quantiles = c(0.025, 0.975))
```

```
## $y
##    time quantile_0.025 quantile_0.975          mean
## 1:    1    -0.03663416     0.03752972  1.102221e-04
## 2:    2    -0.03666747     0.03596978 -2.848990e-05
## 3:    3    -0.03602083     0.03644617  8.530205e-05
## 4:    4    -0.03780105     0.03685205 -4.208489e-05
## 5:    5    -0.03651973     0.03629376  7.000557e-05
##
## $h
##    time quantile_0.025 quantile_0.975     mean
## 1:    1    0.404072293       2.487583 1.442077
## 2:    2    0.268454334       2.533374 1.394570
## 3:    3    0.121924793       2.589508 1.347677
## 4:    4   -0.009923701       2.610516 1.304630
## 5:    5   -0.134679921       2.654572 1.259413
##
## $h_exp
##    time quantile_0.025 quantile_0.975       mean
## 1:    1    0.010053253     0.02915122 0.01776023
## 2:    2    0.009335712     0.02994720 0.01749173
## 3:    3    0.008764370     0.03068317 0.01724951
## 4:    4    0.008291696     0.03101963 0.01698012
## 5:    5    0.007747370     0.03151893 0.01668318
```

References

Akaike, Hirotogu. 1998. "Information Theory and an Extension of the Maximum Likelihood Principle." In *Selected Papers of Hirotugu Akaike*, 199–213. Springer.

Bell, Brad. 2005. "CppAD: A Package for C++ Algorithmic Differentiation." http://www.coin-or.org/CppAD.

Dowle, Matt, and Arun Srinivasan. 2019. *data.table: Extension of 'Data.frame'*. https://CRAN.R-project.org/package=data.table.

Guennebaud, Gaël, Benoît Jacob, and others. 2010. "Eigen V3." http://eigen.tuxfamily.org.

Harvey, A., E. Ruiz, and N. Shephard. 1994. "Multivariate Stochastic Variance Models." *The Review of Economic Studies* 61 (2): 247–64. https://doi.org/10.2307/2297980.

Kastner, Gregor. 2016. "Dealing with Stochastic Volatility in Time Series Using the R Package stochvol." *Journal of Statistical Software* 69 (5): 1–30. https://doi.org/10.18637/jss.v069.i05.

Kristensen, Kasper, Anders Nielsen, Casper Berg, Hans Skaug, and Bradley Bell. 2016. "TMB: Automatic Differentiation and Laplace Approximation." *Journal of Statistical Software, Articles* 70 (5): 1–21. https://doi.org/10.18637/jss.v070.i05.

Liesenfeld, Roman, and Jean-François Richard. 2006. "Classical and Bayesian Analysis of Univariate and Multivariate Stochastic Volatility Models." *Econometric Reviews* 25 (2-3): 335–60. https://doi.org/10.1080/07474930600713424.

Pitt, Michael K., and Neil Shephard. 1999. "Time-Varying Covariances: A Factor Stochastic Volatility Approach." In *Bayesian Statistics, 6 (Alcoceber, 1998)*, 547–70. Oxford Univ. Press, New York.

R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. R foundation for statistical computing. https://www.R-project.org/.

Skaug, Hans J., and Jun Yu. 2014. "A Flexible and Automated Likelihood Based Framework for Inference in Stochastic Volatility Models." *Computational Statistics & Data Analysis* 76: 642–54. https://doi.org/https://doi.org/10.1016/j.csda.2013.10.005.

Taylor, S. J. 1982. "Financial Returns Modelled by the Product of Two Stochastic Processes-a Study of the Daily Sugar Prices 1961-75." *Time Series Analysis : Theory and Practice* 1: 203–26. https://ci.nii.ac.jp/naid/10018822959/en/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.