Parameter Estimation of Multivariate Factor Stochastic Volatility Models

Jens Christian Wahl

Den Norske Aktuarforening

May 9, 2019

Table of contents

Goal

Use maximum likelihood and Hamiltonian Monte Carlo to estimate parameters in Multivariate Factor Stochastic Volatility (MFSV) models and compare the results with the state of the art method.

Stochastic Volatility and Factor Analysis

Volatility

What is volatility?

Characteristics.

How do we model it?

- ARCH/GARCH models
- Stochastic Volatility

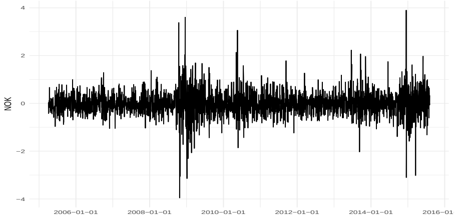


Figure: Log returns of NOK w.r.t EUR

The basic SV model

Let y_t be the log returns of an asset. The basic SV model is then defined as:

$$
\begin{aligned}
y_t &= \sigma_y e^{h_t/2} \epsilon_t, \quad t = 1, \ldots, T, \\
h_{t+1} &= \phi h_t + \sigma_\eta \eta_t, \quad t = 1, \ldots, T - 1,
\end{aligned}
$$

where $\epsilon_t, \eta_t \overset{\text{iid}}{\sim} N(0, 1)$, $-1 < \phi < 1$ and $h_1 \sim N(0, \sigma_\eta^2/(1 - \phi^2))$.

Properties:

- Heavy tailed.
- Autocorrelated squared returns.

Factor Analysis

Capture correlation structure of multivariate data by lower dimensional latent structure.

Let $y = (y_1; : : : ; y_p)$ be a random variable.

Assume y can be represented as a linear function of $q < p$ latent variables (factors).

$$y_1 = {}_{11}f_1 + {}_{12}f_2 + \quad + {}_{1q}f_q + {}_1$$

$$y_2 = {}_{21}f_1 + {}_{22}f_2 + \quad + {}_{2q}f_q + {}_2$$

$$\vdots$$

$$y_p = {}_{p1}f_1 + {}_{p2}f_2 + \quad + {}_{pq}f_q + {}_p:$$

$$y = \; f + \; .$$

Properties

$$\mathrm{Cov}(y) = \quad ' + \quad :$$

There are an infinite number of solutions. Let Q be an orthogonal matrix, such that $QQ' = I$, and $\mathrm{e} = \quad Q$, then

$$\mathrm{e}\,\mathrm{e}' = (\quad Q)(\quad Q)' = \quad (QQ')\quad ' = \quad ' :$$

The Multivariate Factor Stochastic Volatility Model

Model description

Given $y_t = (y_{1t}; : : : ; y_{pt})$, the model can be stated as

$$y_t = \quad f_t + U_t(x_t)^{1=2} \quad {}_t; \qquad f_t = V_t(h_t)^{1=2} \quad {}_t;$$

where

- is a $p \quad q$ loading matrix
- $U_t(x_t) = \text{diag}[\exp(x_{1t}); : : : ; \exp(x_{pt})]$
- $V_t(h_t) = \text{diag}[\exp(h_{1t}); : : : ; \exp(h_{qt})]$
- ${}_t \quad N_q(\mathbf{0}; I)$, and ${}_t \quad N_p(\mathbf{0}; I)$ are independent
- $h_{it} = \quad {}_{h_i} h_{i;t \quad 1} + \quad {}_{h_i} \quad {}_t; \qquad i = 1; : : : ; q$
- $x_{jt} = \quad {}_{x_j} + \quad {}_{x_j}(x_{j;t \quad 1} \quad \quad {}_{x_j}) + \quad {}_{x_j} \quad {}_t; \qquad j = 1; : : : ; p$

Motivation

Simple:

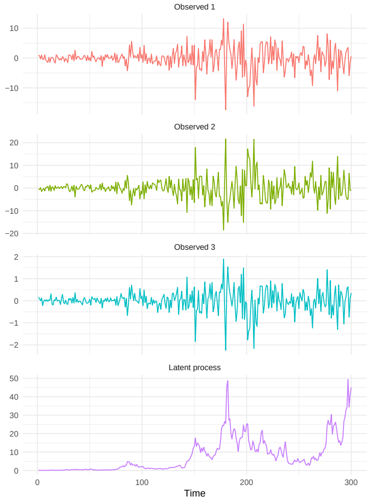
High-dimensional observation space is reduced to a lower-dimensional latent factor space.

Flexible:

Captures volatility clusters.

Robust:

Idiosyncratic variances are time dependent.



Identification issues

Upper triangular of is set to zero to prevent factor rotation.

Set $_{h_j} = 0$ to identify the variance of the factor.

Estimation Methods and Implementation

Maximum likelihood with latent variables

If y denotes our observations, $u \in \mathbb{R}^q$ our latent variables and the parameters of interest, the likelihood of is given by

$$L(\ ) = \int_{\mathbb{R}^q} f_Y(y|u) f_u(u)\, du.$$

This integral is often high dimensional and can't be evaluated analytically. We must therefore approximate it.

We apply the Laplace approximation.

Laplace Approximation

The Laplace approximation is given by

$$L(\theta) \approx (2\pi)^{\dim(u)/2} \det(H)^{-1/2} \exp\{-g(\hat{u};\theta)\},$$

where

$$g(u;\theta) = -\log f_y(y|u) f_u(u);$$

$$\hat{u} = \arg\min_u g(u;\theta) \text{ and}$$

$$H = \frac{\partial^2}{\partial u^2} g(u;\theta)\Big|_{u=\hat{u}}$$

Hamiltonian Monte Carlo

- In a Bayesian framework we are interested in finding the density of our parameters given the data, i.e. the posterior $\pi(\theta | y)$.
- In many complex models the posterior is not available in a closed form and we can't sample directly from $\pi(\theta | y)$. We use MCMC methods to approximate $\pi(\theta | y)$.
- Classical MCMC algorithms:
  - Metropolis-Hastings algorithm.
  - Gibbs sampler.
- We use Hamiltonian dynamics to generate efficient proposals in Metropolis-Hastings algorithm.

Deep Interweaving MCMC

Kastner et al. (2017) propose a Gibbs sampler that incorporates the interweaving strategy introduced by Yu and Meng (2011) that combine di erent augmentations schemes by moving back and forth between them when sampling. This leads to moving some of the model parameters back and forth between the observational equatior and state equation.

Implementation

- The R package TMB is used for likelihood estimation. Latent variables are integrated out using the Laplace approximation.
- The probabilistic programming language Stan is used for Hamiltonian Monte Carlo.
- R package factorstochvol for Deep Interweaving MCMC.

Simulation study

An example using MLE

The two dimensional model with one factor:

$$
\begin{aligned}
y_{1t} &= \;{}_1 e^{\frac{1}{2}h_t} \;{}_t + e^{\frac{1}{2}x_{1t}} \;{}_{1t} \\
y_{2t} &= \;{}_2 e^{\frac{1}{2}h_t} \;{}_t + e^{\frac{1}{2}x_{2t}} \;{}_{2t} :
\end{aligned}
$$

Simulate $T = 3000$ observations from the model with parameters $= (0{:}7; 1; 3)$; ${}_h = 0{:}95$; $\log {}_h = \; 1{:}4$; ${}_x = (\; 1{:}3; \; 0{:}8)$; ${}_x = (0{:}95; 0{:}94)$ and $\log {}_x = (\; 1; \; 0{:}8)$.

An example using MLE

Figure: Simulated data

Figure: True and estimated latent processes

An example using MLE

	Estimate	SE	True Value
1	0.764	0.033	0.7
2	1.416	0.062	1.3
h	0.947	0.014	0.95
log h	-1.499	0.144	-1.4
x_1	0.963	0.008	0.95
x_2	0.934	0.013	0.94
log x_1	-1.027	0.082	-1
log x_2	-0.865	0.094	-0.8
1	-1.649	0.212	-1.3
2	-0.662	0.164	-0.8

Analysis of two dimensional model with one factor

We investigate parameter behaviour for two of the methods:

- Maximum Likelihood , where the latent variables are integrated out using the Laplace approximation. Implemented in TMB .
- Deep Interweaving MCMC . Sampling done by the use of the R package factorstochvol .

We generate 1000 datasets, each with $T = 5000$ observations. Priors chosen for DIMCMC are:

- $_i \quad N(0;1)$
- $_i \quad N(0;100)$
- $^2_i \quad G(0{:}5;0{:}5)$
- $(\ _i + 1)=2 \quad B(20;1{:}5)$

Results

	MLE			DIMCMC			
	Estimate	SE	Bias	Estimate	SE	Bias	True value
$_1$	0.693	0.027	-0.007	0.690	0.028	-0.010	0.7
$_2$	1.291	0.051	-0.009	1.281	0.052	-0.019	1.3
$_h$	0.95	0.009	0	0.949	0.009	-0.001	0.95
log $_h$	-1.431	0.091	-0.039	-1.392	0.094	0.008	-1.4
$_{x_1}$	0.953	0.007	0.003	0.948	0.009	-0.002	0.95
$_{x_2}$	0.952	0.008	0.013	0.940	0.013	0.0002	0.94
log $_{x_1}$	-1.068	0.070	0.032	-1.083	0.088	0.017	-1.1
log $_{x_2}$	-0.987	0.079	0.015	-1.005	0.122	-0.005	-1
$_{x_1}$	-1.286	0.126	0.014	-1.313	0.121	-0.013	-1.3
$_{x_2}$	-0.776	0.164	0.024	-0.810	0.162	-0.010	-0.8

Table: Results across the $n = 611$ simulated datasets for which an MLE was obtained.

Results

	MLE			DIMCMC			
	Estimate	SE	Bias	Estimate	SE	Bias	True value
1	0.717	-	0.017	0.708	0.028	-0.010	0.7
2	1.267	-	-0.033	1.316	0.052	-0.019	1.3
h	0.946	-	-0.004	0.947	0.009	-0.001	0.95
log h	-1.410	-	-0.005	-1.392	0.095	0.008	-1.4
x_1	0.957	-	0.007	0.947	0.009	-0.002	0.95
x_2	0.959	-	0.019	0.936	0.014	0.0002	0.94
log x_1	-1.031	-	0.069	-1.050	0.088	0.017	-1.1
log x_2	-0.973	-	0.027	-0.954	0.122	-0.005	-1
x_1	-1.094	-	0.21	-1.339	0.123	-0.013	-1.3
x_2	-0.97	-	-0.17	-0.832	0.166	-0.010	-0.8

Table: Results across the $n = 389$ simulated datasets for which an MLE was not obtained.

Histogram of estimates

(a) MLE

(b) DIMCMC

Restrictions found through the CGF

By studying the characteristic function $\varphi_Y : \mathbb{R}^n \to \mathbb{C}$ and the cumulative generating function $H_Y(t) = \log \varphi_Y(t)$ we found an expression for the higher order cumulants for the bivariate model:

$$\kappa_{mr} = \frac{\partial^{m+r}}{\partial s_1^m \partial s_2^r} H_{Y_1,Y_2}(s_1,s_2)\Big|_{(s_1,s_2)=(0,0)}$$
$$= \lambda_1^m \lambda_2^r H_{V_h}^{m+r}(0) \quad \text{for} \quad m,r \geq 1,$$

where H_{V_h} is the CGF of the factor process.

Investigate convergence properties when $\lambda_1 = \lambda_2$.

Convergence in 996 out of 1000 datasets.

The Nested Laplace Approximation

The Nested Laplace Approximation

Instead of integrating all latent variables in one go, we can integrate over subsets in a sequential way.

Let $(u_1; u_2)$ be a partitioning of u and assume that u_1 and u_2 are independent.

Apply the Laplace approximation twice, first to u_1 and then u_2:

$$
\begin{aligned}
L(\theta) &= \int f_y(y|u) f_u(u)\, du \\
&= \int f_{u_2}(u_2) \int f_y(y|u_1; u_2) f_{u_1}(u_1)\, du_1 du_2
\end{aligned}
$$

The Nested Laplace Approximation

Applied on two state space models:

- Linear state space model: All methods gave similar results.
- Two dimensional MFSV model: Nested Laplace converged while standard Laplace did not.

Problem: Hessian of outer Laplace approximation

(a) Hessian of outer Laplace (b) Close up of Hessian

Empirical analysis

Empirical analysis

- Dataset containing 26 daily exchange rates w.r.t. EUR on 2650 days. Ranges from April 1, 2005 to August 6, 2015.
- Convergence of the likelihood for the two dimensional model was obtained on 105 out of 325 pair combinations.
- When $_1 = {}_2$ convergence was obtained for 204 datasets.
- Investigate two scenarios:
  - When the likelihood converge.
  - When the likelihood does not converge.
- E ect of di erent priors on .

Likelihood does converge - Australian and Canadia dollar

(a) Log returns

(b) ACF of log returns

(c) ACF of squared log returns

Parameter estimates

	MLE		DIMCMC		HMC	
	Estimate	SE	Estimate	SE	Estimate	SE
1	0.48	0.06	0.55	0.08	0.55	0.06
2	0.36	0.04	0.34	0.05	0.34	0.04
h	0.99	0.004	0.99	0.004	0.99	0.004
h	0.10	0.02	0.09	0.02	0.09	0.016
x_1	0.98	0.008	0.95	0.03	0.96	0.028
x_2	0.99	0.005	0.99	0.005	0.95	0.013
x_1	-2.73	0.41	-5.14	1.09	-5.60	1.99
x_2	-1.68	0.16	-1.57	0.18	-1.52	0.09
x_1	0.23	0.04	0.75	0.21	0.76	0.27
x_2	0.08	0.02	0.08	0.02	0.19	0.03
CPU(s)	38		535		7334	

Estimate of latent processes

Likelihood does not converge - Croatian kuna and Philippines peso

(a) Log returns

(b) ACF of log returns

(c) ACF of squared log returns

Parameter estimates when +(1)=2 B (20; 1:5)

	MLE		DIMCMC		HMC	
	Estimate	SE	Estimate	SE	Estimate	SE
1	-0.003	-	-0.0007	0.0015	0.063	0.015
2	0.135	-	0.255	0.094	0.005	0.008
h	0.100	-	0.544	0.145	0.982	0.006
h	1.751	-	0.973	0.303	0.305	0.045
x_1	0.929	-	0.913	0.991	0.517	0.088
x_2	0.971	-	0.991	0.0005	0.937	0.016
x_1	-4.78	-	-4.75	0.116	-6.31	0.287
x_2	-1.287	-	-1.83	0.518	-1.171	0.083
x_1	0.431	-	0.483	0.052	1.374	0.166
x_2	0.167	-	0.109	0.039	0.220	0.031
CPU(s)	125		557		5743	

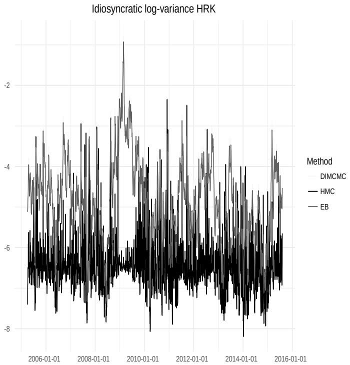
Parameter estimates when +(1) = 2 B (10; 3)

	MLE		DIMCMC		HMC	
	Estimate	SE	Estimate	SE	Estimate	SE
1	-0.003	-	-0.0008	0.0016	0.052	0.006
2	0.135	-	0.326	0.070	0.001	0.009
h	0.100	-	0.273	0.152	0.556	0.004
h	1.751	-	0.867	0.228	1.193	0.151
x_1	0.929	-	0.905	0.019	0.974	0.009
x_2	0.971	-	0.987	0.006	0.929	0.018
x_1	-4.78	-	-4.87	0.1122	-5.99	0.430
x_2	-1.287	-	-2.238	0.6065	-1.187	0.078
x_1	0.431	-	0.501	0548	0.391	0.058
x_2	0.167	-	0.171	0.062	0.235	0.035
CPU(s)	125		557		4845	

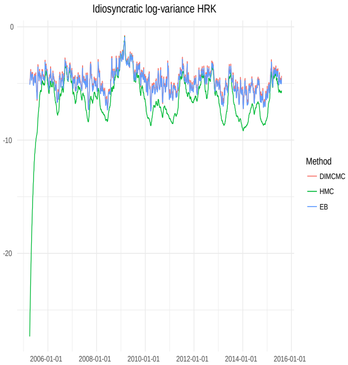
Estimated log variance of factor

(a) (+ 1) =2 B (20; 1:5) (b) (+ 1) =2 B (10; 3)

Estimated log variance HRK

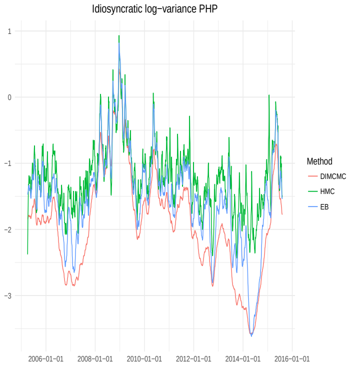


(a) (+ 1)=2 $B(20; 1{:}5)$

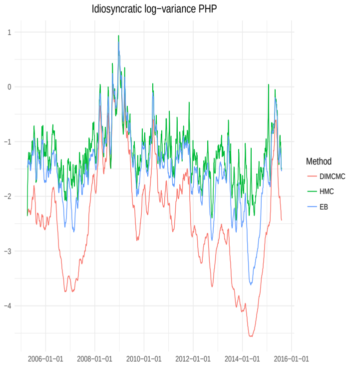


(b) (+ 1)=2 $B(10; 3)$

Estimated log variance PHP



(a) (+ 1)=2 $B(20; 1{:}5)$



(b) (+ 1)=2 $B(10; 3)$

Conclusion and future work

Conclusion

- Investigated two new estimation methods for MFSV model: ML and HMC.
- Convergence of likelihood was unstable and very data dependent.
- Restricting the loading vector by letting $_1 = {}_2$ made convergence more stable.
- The nested Laplace approximation was accurate, but computationally slow.
- All methods gave similar results on real-world data, but HMC was sensitive to the choice of priors.

Future work

Investigate simulating from the posterior using Riemann Manifold Hamiltonian Monte Carlo (RMHMC) to adjust for big local curvature in the parameter space.