Bayesian Semiparametric GARCH Models

Xibin (Bill) Zhang; Maxwell L. King

4-5 October 2010

Department of Econometrics and Business Statistics,
Monash University, Australia

MONASH University

Outline

GARCH Models

- The autoregressive conditional heteroscedasticity (ARCH) model of Engle (1982) and the generalised ARCH (GARCH) model of Bollerslev (1986) have proven to be useful in modelling volatilities of asset returns.
- Let $\mathbf{y} = (y_1, \cdots, y_n)'$ denote a vector of n observations of an asset return. The GARCH(1,1) model is often specified as

$$y_t = \sigma_t \, \varepsilon_t,$$
$$\sigma_t^2 = \gamma + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2, \quad (1)$$

  where ε_t, for $t = 1, 2, \cdots, n$, are independent. It is often assumed that $\gamma > 0$, $\alpha \geq 0$, $\beta \geq 0$ and $\alpha + \beta < 1$.
- The assumption of conditional normality of ε_t has contributed to early successes of GARCH models.

Specification of error density

- Weiss (1986) and Bollerslev and Wooldridge (1988) showed that under the assumption of conditional normality, the quasi maximum likelihood estimates (QMLEs) of parameters are consistent when the first two moments of y_t are correctly specified.
- However, enough evidence found by theoretical and empirical studies has shown that it is possible to reject the assumption of conditional normality.
- Such a finding has motivated the investigation of other specifications of the conditional distribution of errors in GARCH models, such as the Student t and other types of heavy-tailed distributions. See for example, Hall and Yao (2003) for detailed discussions

Unknown conditional error density

- Some investigations have been focused on the estimation of the (G)ARCH model without an assumption on the type of conditional density of ε_t.
- Engle and González-Rivera (1991) proposed a semiparametric GARCH model without any assumption on the type of error density. The error density was estimated nonparametrically based on the residuals, which were obtained by applying either the QMLE or OLS to the same parametric model assuming conditional normality.
- The parameters of the semiparametric GARCH model were estimated by maximising the above-derived log-likelihood. This estimation procedure is an one-step semiparametric estimation.

One-step semiparametric estimation procedure

- Their Monte Carlo simulation results showed that this semiparametric approach could improve the efficiency of parameter estimates up to 50% against QMLEs obtained under conditional normality.
- However, the initial estimates of parameters might be inaccurate as a consequence of the assumption of conditional normality in the model used for deriving such estimates. Therefore, the accuracy of resulting kernel estimator of the error density based on residuals might be affected.
- Moreover, this semiparametric estimation procedure is one-step. The estimated parameters will not be used again for improving the kernel density estimator for errors.

On kernel density estimator based on residuals

- A kernel density estimator of either directly observed data or indirect data (such as the calculated residuals) is determined by the choices of kernel and bandwidth.
- In the kernel estimator of the error density based on residuals, once the kernel is chosen, the resulting kernel estimator is completely determined by bandwidth choice, which is crucially important in determining the performance of the density estimator.
- Therefore, the implementation of the above one-step semiparametric estimation method is actually an issue of choosing bandwidth and maximising the approximate likelihood, which was constructed based on the kernel estimator of the error density based on residuals, with respect to parameters.

Our contribution

- We treat the bandwidth involved in the kernel estimator of the error density as an additional parameter in the semiparametric GARCH model.
- Assuming that the error density is approximated by the kernel estimator based on residuals, we are able to derive the approximate likelihood.
- We could maximise the likelihood w.r.t both types of parameters numerically. However, such a direct maximisation may encounter convergence problems, because the likelihood is flat when bandwidth is large.
- Instead, we assume prior densities for the two types of parameters and obtain the posterior of all parameters. The Metropolis-Hastings algorithm is then used to sample both types of parameters.

Our work is related to adaptive estimation

- Linton (1993) proved that under the assumption of a symmetric unknown error density in an ARCH model, the resulting QMLE is asymptotic efficient in the sense of Bickel (1982).
- Drost and Klaassen (1997) argued that adaptive estimation is not possible in GARCH models.
- After a suitable reparameterisation, the estimation problem also exists for a known error density. GARCH parameters are still adaptively estimable.
- Drost and Klaassen (1997) considered the model

$$y_t = h_t^{1/2}(\mu + \sigma\eta_t),$$
$$h_t = 1 + \alpha y_{t-1}^2 + \beta h_{t-1},$$

and presented an one-step estimation procedure.

Other related work

- To deal with the problem of possible misspecification of error density and impose inequality constraints on some parameters of the quasi likelihood, Koop (1992) presented Bayesian semiparametric ARCH models, where the quasi likelihood was set up through a sequence of complicated polynomials.
- His overall finding indicated that the use of Bayesian and semiparametric approaches is feasible and necessary.
- Our Bayesian sampling procedure differs from his in that our likelihood is set up through the leave-one-out kernel estimator of the error density based on residuals. Therefore, either the conditional posteriors or joint posterior can be derived.

A GARCH model with unknown error density

- Consider the GARCH(1,1) model
$$y_t = \sigma_t \; \varepsilon_t,$$
$$\sigma_t^2 = \gamma + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2, \qquad (2)$$
where conditional on information available at time t, the error $\varepsilon_t \sim f(\varepsilon)$, an unknown density function.
- The constraints on parameters are the same as those in (2). In addition, we restrict γ to be a constant such as 0.05. This is to guarantee the error density is uniquely specified.
- When $f(\varepsilon)$ is known, the likelihood is
$$\ell_0(\mathbf{y}|\gamma, \alpha, \beta) = \prod_{t=1}^{n} \frac{1}{\sigma_t} f\left(\frac{y_t}{\sigma_t}\right).$$

- When $f(\varepsilon)$ is unknown, conditional on parameters, the error density can be approximated by the kernel estimator based on residuals. This estimator is
$$\hat{f}_h(\varepsilon) = \frac{1}{n} \sum_{t=1}^{n} \frac{1}{h} K\left(\frac{\varepsilon - \varepsilon_t}{h}\right).$$
- The leave-one-out estimator of $f(\varepsilon_t)$ is
$$\hat{f}_{h,t}(\varepsilon_t) = \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq t}}^{n} \frac{1}{h} K\left(\frac{\varepsilon_t - \varepsilon_j}{h}\right),$$
which is the kernel estimator of $f(\hat{\varepsilon}_t)$ obtained based on $\{\hat{\varepsilon}_t\}$ without ε_t, for $t = 1, 2, \cdots, n$.
- This is to say that we assume $f(\varepsilon_t) \approx \hat{f}_{h,t}(\varepsilon_t)$.

Likelihood

- Assume that the density of ε_t can be approximated by $\hat{f}_{h,t}(\varepsilon_t)$, for $t = 1, 2, \cdots, n$. Then the likelihood of $\mathbf{y}$ is approximated by

$$\ell(\mathbf{y}|\alpha, \beta, h) = \prod_{t=1}^{n} \frac{1}{\sigma_t} \hat{f}_{h,t}(y_t/\sigma_t). \qquad (3)$$

- Residuals are traditionally used as proxies of errors due to the unavailability of realisations of errors, and therefore, residuals could be used either for testing hypotheses or estimating the error density (Efromovich, 2005).
- In nonparametric regression models, a theory of consistency of the nonparametric estimation of error density based on residuals has been established (Hanson and Johnson, 2002; Cheng, 2004)

Priors

- We follow the prior choices for (α, β) presented by Zhang and King (2008). The prior of each parameter is the uniform density on (0,1).
- An alternative choice of prior for α and β is the beta density. Let $\pi(\alpha, \beta)$ denote the joint prior of (α, β).
- We follow the prior choice of h presented by Zhang, King and Hyndman (2006):

$$\pi(h) \propto \frac{1}{1 + \lambda h^2}.$$

- An alternative choice of priors for bandwidth is

$$\pi(h) \propto \lambda \exp(-\lambda h).$$

In both choices, λ is a hyperparameter.

Posterior of (α, β, h)

- The posterior can be better explained in terms of conditionals.
- Conditional on the bandwidth parameter h, the conditional posterior of (α, β) is

$$\pi(\alpha, \beta | h, \mathbf{y}) \propto \pi(\alpha, \beta) \prod_{t=1}^{n} \frac{1}{\sigma_t} \hat{f}_{h,t}(y_t/\sigma_t). \quad (4)$$

- Conditional on (α, β), the conditional posterior of h is

$$\pi(h | \alpha, \beta, \mathbf{y}) \propto \pi(h) \prod_{t=1}^{n} \frac{1}{\sigma_t} \hat{f}_{h,t}(y_t/\sigma_t). \quad (5)$$

The sampling procedure

- Conditional on h, we calculate $\hat{\varepsilon}_t$, for $t = 1, 2, \cdots, n$, and sample (α, β) from its conditional posterior.
- We use the sampling algorithm presented by Zhang and King (2008) for the t GARCH(1,1) model, in which the t density is replaced by the leave-one-out kernel density estimator of the standardised residuals.
- Conditional on (α, β), we sample h from its conditional posterior.
- we use the sampling algorithm presented by Zhang, King and Hyndman (2006) for kernel density estimation of directly observed data, in which the data are replaced by the standardised residuals
- At both steps, the sampling algorithms are the random-walk Metropolis-Hastings.

Joint posterior of (α, β, h)

- The joint posterior of (α, β, h) is defined as
$$\pi(\alpha, \beta, h|\mathbf{y}) \propto \pi(\alpha, \beta)\pi(h)\ell(\mathbf{y}|\alpha, \beta, h). \quad (6)$$
- We can sample (α, β, h) from its joint posterior using the random-walk Metropolis-Hastings algorithm.

Normalisation of standardised residuals

- The density of residuals can only be estimated up to a scaling constant.
- Linton (1993) suggested two methods to deal with this issue in his adaptive estimation of the ARCH model.
  - $\gamma = 0$; and
  - Unit variance of $\hat{f}_h(\cdot)$.
- We set $\gamma = 0.05$.

- We estimated the GARCH(1,1) model of daily returns for 8 indices: S&P500, Nasdaq100, DJIA, NYSE, Nikkei225, FTSE100, DAX and Australian All Ordinaty (AOI).
- We considered two specifications of the error density: unknown density and the Student t density.
- We applied Bayesian sampling procedures under both specifications.
- We presented not only the parameter estimates, but also the marginal likelihood values.
- We found that our Bayesian semiparametric model is favored against the t GARCH for S&P500, Nasdaq100 and Nikkei225.
- Our Bayesian semiparametric model is not worse than the t GARCH model for NYSE and AOI.

	Semiparametric GARCH		t GARCH	
	Mean	SIF	Mean	SIF
S&P500				
σ_0^2	0.6671	8.17	0.4843	8.41
γ	—	—	0.0071	80.00
α	0.0736	35.90	0.0646	61.61
β	0.9038	27.77	0.9075	67.70
h	0.2541	26.98	—	—
ν	—	—	6.7497	22.17
ln margin. likelihood	-1851.50		-1854.07	
DJIA				
σ_0^2	0.8633	7.57	0.5455	7.05
γ	—	—	0.0061	83.04
α	0.1120	28.28	0.0656	53.51
β	0.8740	24.30	0.9073	57.81
h	0.3459	7.38	—	—
ν	—	—	6.9095	25.95
ln margin. likelihood	-1768.07		-1763.01	

	Semiparametric GARCH		t GARCH	
	Mean	SIF	Mean	SIF
Nasdaq100				
σ_0^2	1.6638	5.64	1.5415	7.21
γ	—	—	0.0244	87.06
α	0.1072	13.44	0.0660	64.34
β	0.8781	16.68	0.9041	82.40
h	0.2468	5.08	—	—
ν	—	—	6.6941	19.23
ln margin. likelihood	-2054.9		-2062.77	
NYSE				
σ_0^2	0.8764	5.12	0.5965	8.12
γ	—	—	0.0084	82.36
α	0.0920	26.32	0.0697	52.63
β	0.8930	24.44	0.9041	62.25
h	0.3170	14.91	—	—
ν	—	—	7.4469	14.07
ln margin. likelihood	-1931.43		-1932.18	

	Semiparametric GARCH		t GARCH	
	Mean	SIF	Mean	SIF
FTSE				
σ_0^2	0.8512	5.37	0.6097	7.46
γ	—	—	0.0192	82.99
α	0.1488	23.01	0.0968	73.88
β	0.8331	24.35	0.8745	88.72
h	0.3388	9.88	—	—
ν	—	—	11.3587	19.03
ln margin. likelihood	-1904.36		-1898.18	
DAX				
σ_0^2	1.1682	7.11	1.3091	8.58
γ	—	—	0.0247	91.18
α	0.1480	25.44	0.0797	61.45
β	0.8345	23.41	0.8866	87.43
h	0.3949	12.16	—	—
ν	—	—	9.5068	13.40
ln margin. likelihood	-2032.78		-2021.53	

	Semiparametric GARCH		t GARCH	
	Mean	SIF	Mean	SIF
Nikkei225				
σ_0^2	1.8959	6.42	6.1623	5.46
γ	—	—	0.0374	78.05
α	0.1227	11.24	0.0907	72.51
β	0.8438	9.42	0.8811	89.08
h	0.2324	9.70	—	—
ν	—	—	13.9171	17.92
ln margin. likelihood	-2108.19		-2117.43	
AOI				
σ_0^2	0.8436	5.90	0.6058	7.38
γ	—	—	0.0298	70.82
α	0.1647	24.90	0.0981	67.81
β	0.8109	22.32	0.8669	80.69
h	0.2965	15.78	—	—
ν	—	—	13.8270	12.25
ln margin. likelihood	-1865.99		-1866.33	

Localised bandwidths

- Sometimes there are spurious bumps in the tails of the kernel density estimator because only a small number of observations contribute to the resulting kernel estimator (Koekemoer and Swanepoel, 2008).
- We have worked out a sampling algorithm when localised bandwidths are used in the kernel estimator of error density.

Value-at-Risk (VaR)

- We are interested in the left-tail behavior of the density of y_t conditional on σ_t.
- Our sampling algorithm can produce an estimate of the error density, and therefore, we are able to approximate the conditional VaR.

- This paper aims to estimate parameters of a GARCH model with an unknown error density through Bayesian sampling.
- The bandwidth involed in the kernel estimator of error density is treated as a parameter.
- We have derived the posterior of both types of parameters and built up a sampling procedure. The proposed sampling algorithm has achieved very good mixing performance.
- We applied the sampling algorithm to semiparametric GARCH(1,1) models of daily returns for 8 stock indices.
- The Bayesian semiparametric model is clearly favored against the t GARCH for S&P500, Nasdaq100 and Nikkei225.