Stochastic conditonal range, a latent variable model for financial volatility

Galli, Fausto

University of Salerno

28 February 2014

Stochastic conditonal range, a latent variable model for financial volatility

Fausto Galli*

Abstract

In this paper I introduce a latent variable augmented version of the conditional autoregressive range (CARR) model. The new model, called stochastic conditional-range (SCR) can be estimated by Kalman filter or by efficient importance sampling depending on the hypotheses on the distributional form of the innovations. A predictive accuracy comparison with the CARR model shows that the new approach can provide an interesting alternative.

*DISES and CELPE, University of Salerno. Corresponding author. Address: Dipartimento di Scienze Economiche e Statistiche (DISES) Campus di Fisciano Universita' degli Studi di Salerno Via Ponte Don Melillo 84084 Fisciano (SA), Italy. Email: fgalli@unisa.it, Phone: +39 089 96 8154.

1 Introduction

It is a well known phenomenon that financial time series exhibit volatility clustering. A very large literature on the dynamics of returns has developed since the seminal contributions of Engle (1982), Bollerslev (1986) and Taylor (2007) on GARCH and stochastic volatility. Most of this literature concentrates on the dynamics of closing prices of the reference period as a means of describing the subtle concept of volatility.

Parkinson (1980) suggested that the use of extreme price values can provide a superior estimate of volatility than returns. The potential advantages of using price range as an alternative were also pointed out by Alizadeh, Brandt, and Diebold (2002), who claimed to "show theoretically, numerically, and empirically that range-based volatility proxies are not only highly efficient, but also approximately Gaussian and robust to microstructure noise", while Brandt and Diebold (2006) noticed that range "is a highly efficient volatility proxy, distilling volatility information from the entire intraday price path, in contrast to volatility proxies based on the daily return, such as the daily squared return, which use only the opening and closing prices".

Chou (2005) proposed a dynamic model, the conditional autoregressive range (CARR) for the evolution of high/low range who mimics the structure of the ACD model of Engle and Russell (1998) for inter trade durations. This line of modelling has desirable statistical and empirical properties and the search for its refinements and extensions can draw from the wide body of ACD literature.

In this article I introduce a latent variable augmented version of the CARR model: the stochastic conditional range (SCR) model. The new formulation shares most of the statistical characteristics of the stochastic conditional duration (SCD) model of Bauwens and Veredas (2004). In section 2, I will present the model and discuss some of its properties. In section 3 I will describe how the latent variable present in the the SCR model can filtered or integrated in order to obtain a likelihood that can be easily maximized. A comparison on the predictive accuracy of CARR and SCR models is carried out for a large sample of stocks in section 5. Results show that the SCR can provide a valid alternative to the CARR specification.

2 The model

This section introduces a model for the conditional range with a latent variable. In analogy to the literature on financial durations, where a similar model is colled stochastic conditional duration, SCD, this process will be called stochastic conditional range, SCR.

Let p_τ the price of a financial asset sampled at frequent (e.g. minutes or seconds) time intervals τ, and $P_\tau = ln(p_\tau)$ its logarithm. We define as range the difference $R_t = max(P_t) - min(P_t)$, where t indicates a coarser set of time intervals (e.g. days, weeks) such that

$$\tau = t-1, t-1+\frac{1}{n}, t-1+\frac{2}{n}, \ldots, t, \tag{1}$$

where n is the number of frequent intervals contained in one of the coarser intervals indexed by t.

The stochastic conditional range (SCR) is a process described by the following equations:

$$R_t = e^{\psi_t}\epsilon_t \tag{2}$$

$$\psi_t = \omega + \beta\psi_{t-1} + \sigma u_t \tag{3}$$

where $u_t|I_{t-1}$ has an iid standard normal distribution and $\epsilon_t|I_{t-1}$ is iid and has a distribution with density function $p(\epsilon)$, which has positive support, unit mean and variance σ^2_ϵ. I_t denotes the information set at time $t-1$, and it includes the past values of R_t and ψ_t.

The expected value of the range conditional to the past of the process up to time $t-1$

is

$$E(R_t|I_{t-1}) = e^{\psi_t}$$

and the distribution of R_t results from the mixing of the lognormal distribution of e_t^{ψ} and the distribution of ϵ_t. The condition $|\beta| < 1$ is necessary and sufficient to ensure stationarity and ergodicity for the process ψ_t, and hence for R_t.

The theoretical first two moments and the s-th autocorrelation of R_t are the following

$$E(R_t) = E(\epsilon_t)E(e^{\psi_t}) = E(\epsilon_t)e^{\frac{\omega}{1-\beta}+\frac{\sigma^2}{1-\beta^2}}, \tag{4}$$

$$var(R_t) = E(R_t)^2 \left(\frac{E(\epsilon_t^2)}{E(\epsilon_t)^2} e^{\frac{\sigma^2}{1-\beta^2}-1} \right), \tag{5}$$

$$\rho_s = \frac{e^{\frac{\sigma^2\beta^s}{1-\beta^2}} - 1}{\frac{E(\epsilon_t^2)}{E(\epsilon_t)^2} e^{\frac{\sigma^2}{1-\beta^2}} - 1} \tag{6}$$

for all $s \geq 1$.[1]

Concerning the distribution of ϵ_t, any law with positive support can be a suitable candidate. In this paper we will use two distributions: the Weibull and the log-normal. The exponential distribution and the Weibull are commonly employed in duration analysis thanks to the flexibility of their hazard function and the direct relationship between the parameters of the density and the shape of the hazard (constant, increasing or decreasing). Because of these features they are popular in the literature on ACD models and were adopted by Chou (2005) in the CARR model. The justification for the use of the log-normal distribution arises from the result by Alizadeh, Brandt, and Diebold (2002) on the distribution of daily high and low prices, which appears to be approximately Gaussian. Depending on the choice of the distribution for ϵ_t, the estimated models will be denoted as W-SCR and L-SCR.

[1] This result is derived by analogy to the corresponding moments computed by Bauwens and Veredas (2004) on SCD models and its proof is available in their paper.

As it was noted above, we restrict the first moment of the baseline range ϵ_t to be equal to one. This is necessary to avoid an identification problem between the expectations of ϵ_t and ψ_t. The parameter of the exponential and the mean parameter of the log-normal distributions will be therefore set to one, while the scale parameter of the Weibull will be restricted to be equal to $\Gamma(1 + 1/\gamma)^{-1}$, where γ is the shape parameter which will be let free to vary.

The original specification of the SCR model can be extended to include exogenous variables, denoted by $x_{t,l}$, in the equation for the logarithm of the conditional mean. If exogenous variables are included, equation (3) becomes

$$\psi_t = \omega + \beta\psi_{t-1} + \sum_{l=1}^{L} \gamma_l x_{t-1,l} + \sigma u_t. \tag{7}$$

When augmented by exogenous variables, the model is denoted by SCRX(L). As choices for variables to be included we will here consider the past values of trading volume, returns and of the range itself.

3 Estimation

In this section I will describe how the estimation of the SCR model can be performed by maximum likelihood (ML).[2] In particular, I will detail the methods that can be followed in order to deal with the problem of the presence of a latent variable. The results presented here refer to the SCR case, but they apply also to the estimation of the SCRX model, as the explanatory variable that are added in this model are observable and do not pose particular problems.

[2]In the literature on SCD models, which share the same functional form with SCR, some alternative approaches are explored. For example Knight and Ning (2008) compare two solutions based on GMM and on empirical characteristic function and Strickland, Forbes, and Martin (2006) follow a Bayesian approach based on MCMC integration of the latent variable.

The distribution of the baseline range ϵ_t plays an important role in deciding how to proceed in the computation of the likelihood function to be maximized.

If ϵ_t is log-normally distributed, that is in the L-SCR specification, the model can be trasformed by taking the logarithms on both sides of equation (2). This yields the following relationships

$$\ln R_t = \psi_t + \ln \epsilon_t, \tag{8}$$

$$\psi_t = \omega + \beta\psi_{t-1} + \sigma u_t, \tag{9}$$

that can be interpreted as the state and transition equations of a linear state-space model. This model can be easily estimated by Kalman filter and the resulting likelihood can be maximized by means of a numerical algorithm.

The reliance of the Kalman filter on the normality of both error components ($\ln \epsilon_t$ and u_t) limits its use to the L-SCR case only. When the distribution of ϵ_t is exponential or Weibull, the Kalman filter will not produce an exact computation of the likelihood anymore. Therefore, it is necessary to resort to the numerical integration of the density of the latent variable to compute an exact likelihood.

To do this, we start by denoting by R a sequence of n realizations of the range process. R has a conditional density of $g(R|\psi, \theta_1)$, where θ_1 is a parameter vector indexing the distribution and ψ a vector of latent variables of the same dimension of the sample R. The joint density of ψ is $h(\psi|\theta_2)$, with θ_2 a vector of parameters, and the likelihood function for R can be written as

$$L(\theta, R) = \int g(R|\psi, \theta_1)h(\psi|\theta_2)d\psi = \int \prod_{t=1}^{n} p(R_t|\psi_t, \theta_1)q(\psi_t|\psi_{t-1}, \theta_2)d\psi_t \tag{10}$$

the last term of the equation is the result of the sequential decomposition of the integrand in the product of the density of ϵ_t conditional on ψ_t, $p(R_t|\psi_t, \theta_1)$, that in our case will be exponential or Weibull, and the density of ψ_t conditional on its past, $q(\psi_t|\psi_{t-1}, \theta_2)$, which is normal with mean $\omega + \beta\psi_{t-1}$ and variance σ^2.

This high dimensional integral is not analytically solvable and a numerical approach is necessary. There is a very substantial literature on Monte Carlo integration methods, for an interesting survey in the field of stochastic volatility see Broto and Ruiz (2004).

The method I will employ is a refinement of the widespread importance sampling technique, it is called efficient importance sampling (EIS) and was developed by Richard and Zhang (2007). As the authors point out, this method is particularly convenient for an accurate numerical solution of high dimensional "relatively simple" integrals like the ones we need to treat and has already been successfully applied to problems that are similar (see Liesenfeld and Richard (2003) and Bauwens and Hautsch (2006)) or nearly identical (see Bauwens and Galli (2009)) to ours.

For a detailed presentation of the algorithm, I refer the reader to Richard and Zhang (2007). A description of its implementation in the contest of the SCD model, which share the same functional form with the model proposed in this paper is available in Bauwens and Galli (2009). In the appendix, I present a brief summary.

4 Empirical analysis

I carried out the empirical analysis by considering all Standard and Poor's 500 components at the date of February 15, 2014. Data on daily price maxima and minima were downloaded from Yahoo! finance via the `tseries` package in R. The resulting series of ranges were normalized to have a unit mean in order to speed up computation by reducing the search for the intercept in the conditional range function. Out of the original 500 series, 22 of them were composed by less than 1000 observations and were discarded. This choice was somewhat arbitrary, but convergence issues for very limited sample sizes required to set a threshold. Table 1 provides some descriptive statistics of the range series for the remaining 478 stocks. Not all series have a full 10 years length of 2517 observations, but the average sample size after pruning our database of particularly short sets is quite close to the maximum value. It can be noted too that data have a rather low degree of overdispersion (computed as the ratio of sample mean and sample standard deviation), yet maxima tend to be several standard deviations away from the mean. Even visual inspection of some charts revealed that this could be due to an issue of outliers rather than to a particularly fat tail in the baseline distribution. Whether these outliers derive from quirks in recording or from exceptional conditions in the markets is hard to tell. The use of an outlier detection and removal algorithm could be an interesting extension to this analysis and I leave it for further research.

The predictive accuracy of the different models was compared by an insample one-step-ahead analysis. First the full sample is used to estimate the parameters of the models. Then I predicted every observation at time $t = 2, ..., n$ using estimated parameters and observations at time $t-1 = 1, ..., n-1$. An outsample analysis was not performed because splitting the sample in two parts in an already quite short set of data woud either lead to more jittery parameter estimates or to too few forecasts.

The models used in the comparison were a CARR with a Weibull conditional range distribution (W-CARR), an SCR with a lognormal distribution (L-SCR) and an SCR with a Weibull conditional distribution (W-SCR). All models were specified with only one lag of the range (and the conditional range for the CARR model) in the formula for conditional range. The first model was estimated by conditional maximum likelihood. In the second and the third model, likelihood was computed by respectively Kalman filter and EIS. Estimation times runned from less than a second for the CARR model to an average of half a minute the lognormal SCR model and to and average of 5 minutes for the Weibull SCR.

The forecasting accuracy of each estimator for each series has been measured by the mean square (prediction) error, that is the average of the squared difference between predicted and observed values. The significance of the difference between forecast errors of couples of estimators was verified by the Diebold and Mariano (2002) test with a bilateral alternative and a quadratic loss function. Predictions are considered different if the p-value is below 5%.

Table 2 displays the main results for the estimation of the three models. It appears to be quite difficult to distinguish the W-CARR and the W-SCR on the basis of sheer MSE, while the L-SCR seems to stand out both in terms of average than in terms of dispersion. When MSE for individual stocks are compared, it turns out that on average SCR forecasts have a lower MSE, but the number of stocks whose MSE is lower for the CARR model are more than half (2/3 in the case of the L-SCR and almost 3/4 in the case of the W-SCR). So the SCR seems to provide "winning" forecasts less often than the CARR, but when it improves, it does it substantially.

When the significance of pairs of forecasts is tested, it turns out that they are only in about one stock of three the CARR and the SCR model forecast in a significantly different way. If finally we restrict our sample to significantly different forecasts only, we see that the gain in terms of MSE is slightly reduced in the case of the W-SCR and reversed (but with a very low mean) for the L-SCR.

I conclude by remarking that statistics on the comparisons between W-SCR and L-SCR, that are nor reported in table 2, display a substantial similarity between the forecasts of the two models (for example, only less than the 9% of the forecasts can be considered different after testing).

5 Conclusion

The stochastic conditional range can be considered a valid alternative to the CARR model when the dynamics of daily price range are to be analyzed. In a forcasting comparison with a large number of stocks, the SCR model improves on the CARR model in terms of expected prediction error, though the percentage of stocks that are better predicted with a CARR specification is higher. The fact that different baseline distribution (log normal and Weibull here) give very similar results should not lead to think that their role is marginal. In a forecasting exercise like the one of this paper, which is a *de facto* first moment analysis, the parameters that are mostly involved are the ones of the conditional range. Other moments of the distribution of the range, such as variance or autocorrelation coefficients, are more sensitive to the shape of baseline and it is likely that a forecasting or prediction comparison of these feature would highlight the importance of the iid distribution chosen for the innovations of the process. We leave this analysis for further research.

References

Alizadeh, S., M. W. Brandt, and F. X. Diebold (2002): "Range-based estimation of stochastic volatility models," *The Journal of Finance*, 57(3), 1047–1091.

Bauwens, L., and F. Galli (2009): "Efficient importance sampling for ML estimation of SCD models," *Computational Statistics and Data Analysis*, 53(6), 1974–1992.

Bauwens, L., and N. Hautsch (2006): "Stochastic conditional intensity processes," *Journal of Financial Econometrics*, 4, 450–493.

Bauwens, L., and D. Veredas (2004): "The stochastic conditional duration model: a latent factor model for the analysis of financial durations," *Journal of Econometrics*, 119(2), 381–412.

Bollerslev, T. (1986): "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

Brandt, M. W., and F. X. Diebold (2006): "A No-Arbitrage Approach to Range-Based Estimation of Return Covariances and Correlations," *Journal of Business*, 79(1), 61–74.

Broto, C., and E. Ruiz (2004): "Estimation methods for stochastic volatility models: a survey," *Journal of Economic Surveys*, 18(5), 613–649.

Chou, R. Y. (2005): "Forecasting Financial Volatilities with Extreme Values: The Conditional Autoregressive Range (CARR) Model," *Journal of Money, Credit, and Banking*, 37(3).

Diebold, F. X., and R. S. Mariano (2002): "Comparing predictive accuracy," *Journal of Business & economic statistics*, 20(1).

Engle, R. (1982): "Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, 50, 987–1007.

Engle, R., and J. R. Russell (1998): "Autoregressive conditional duration: a new approach for irregularly spaced transaction data," *Econometrica*, 66, 1127–1162.

Knight, J., and C. Q. Ning (2008): "Estimation of the stochastic conditional duration model via alternative methods - ECF and GMM," *Econometrics Journal*, 11(3), 593–616.

Liesenfeld, R., and J. F. Richard (2003): "Univariate and multivariate stochastic volatility models: estimation and diagnostics," *Journal of Empirical Finance*, 10(4), 505–531.

Parkinson, M. (1980): "The extreme value method for estimating the variance of the rate of return," *Journal of Business*, 53(1), 61.

RICHARD, J. F., AND W. ZHANG (2007): "Efficient high-dimensional importance sampling," *Journal of Econometrics*, 141(2), 1385–1411.

STRICKLAND, C., C. FORBES, AND G. MARTIN (2006): "Bayesian analysis of the stochastic conditional duration model," *Computational Statistics and Data Analysis*, 50(9), 2247–2267.

TAYLOR, S. J. (2007): "Modelling financial time series," .

A Appendix: brief description of the EIS numerical integration method

An importance sampling estimate for the integral

$$G(y) = \int_{\Lambda} g(y, \lambda) p(\lambda) d\lambda, \tag{11}$$

where g is an integrable function with respect to a density $p(\lambda)$ with support Λ and the vector y denotes an observed data vector (which in our context corresponds to the observed ranges) is provided by

$$\bar{G}_{S,m}(y, a) = \frac{1}{S} \sum_{i=1}^{S} g(y, \tilde{\lambda}_i) \frac{p(\tilde{\lambda}_i)}{m(\tilde{\lambda}_i, a)}, \tag{12}$$

where the $\tilde{\lambda}_i$'s now denote draws from the IS density m.

The aim of *efficient* importance sampling (EIS) is to minimize the MC variance of the estimator in (12) by selecting optimally the parameters a of the importance function density m given a functional form for m (here, the Gaussian density).

A convenient choice for the auxiliary sampler $m(\psi_t, a_t)$ is a parametric extension of the natural sampler $q(\psi_t|\psi_{t-1}, \theta_2)$, in order to obtain a good approximation of the integrand without too heavy a cost in terms of analytical complexity. Following Liesenfeld and Richard (2003), we use by the following specification:

$$m(\psi_t, a_t) = \frac{q(\psi_t|\psi_{t-1}, \theta_2)e^{a_{1,t}\psi_t + a_{2,t}\psi_t^2}}{\int q(\psi_t|\psi_{t-1}, \theta_2)e^{a_{1,t}\psi_t + a_{2,t}\psi_t^2} d\psi_t}, \quad (13)$$

where $a_t = (a_{1,t}\ a_{2,t})$. This specification is rather straightforward and has the advantage that the auxiliary sampler $m(\psi_t, a_t)$ remains Gaussian.

The parameters a_t can be chose such that they minimize the sampling variance

$$\hat{a}_t(\theta) = \arg\min_{a_t} \sum_{j=1}^{S} \{\ln\left[f(R_t, \tilde{\psi}_t^{(j)}|\tilde{\psi}_{t-1}^{(j)}, R_{t-1}, \theta)\chi(\tilde{\psi}_t^{(j)}, \hat{a}_{t+1})\right] - c_t - \ln(k(\tilde{\psi}_t^{(j)}, a_t))\}^2, \quad (14)$$

where c_t is constant that must be estimated along with a_t. If the auxiliary sampler $m(\psi_t, a_t)$ belongs to the exponential family of distributions, the problem becomes linear in a_t, and this greatly improves the speed of the algorithm, as a least squares formula can be employed instead of an iterative routine. EIS-ML estimates are finally obtained by maximizing $\tilde{L}(\theta; R, a)$ with respect to θ. The number of draws used (S in equation 12) can be quite low and for all estimations in this article is equal to 50.

	mean	std deviation
observations	2469.075	201.271
means	1	0
std deviations	0.721	0.159
minima	0.178	0.045
maxima	9.591	5.298

Table 1: Descriptive statistics of the 478 stocks used for the predictive accuracy analysis.

	mean	sd	max	min	
MSE W-CARR	0.278	0.124	0.155	1.161	
MSE L-SCR	0.276	0.114	0.163	1.058	
MSE W-SCR	0.278	0.122	0.164	1.030	
stocks with smaller forecast MSE with L-SCR than with W-CARR					33.8%
stocks with smaller forecast MSE with W-SCR than with W-CARR					27.7%
% MSE reduction when forecasting with L-SCR wrt W-CARR	0.002	0.049	-0.216	0.276	
% MSE reduction when forecasting with W-SCR wrt W-CARR	0.003	0.063	-0.292	0.347	
significantly different L-SCR and W-CARR forecasts					31.9%
significantly different W-SCR and W-CARR forecasts					34.0%
% MSE reduction when forecasting with L-SCR wrt W-CARR (significantly different forecasts only)	-0.001	0.052	-0.216	0.250	
% MSE reduction when forecasting with W-SCR wrt W-CARR (significantly different forecasts only)	0.002	0.071	-0.292	0.347	

Table 2: MSE comparison and Diebold and Mariano (2002) results for 478 stocks of the Standard and Poor's 500 used for the predictive accuracy analysis.