ISSN: 3065-0623

Columbia Journal of Entrepreneurship and Management

Research Article

BRIDGING DATA GAPS: A LATENT VARIABLE APPROACH TO MODELING FINANCIAL RANGE DATA

Marco Antonio Rossi

University of Salerno. Address: Dipartimento di Scienze Economiche e Statistiche (DISES), Universita' degli Studi di Salerno, Via Giovanni Paolo II, 84084, Fisciano (SA), Italy

DOI: 10.5281/zenodo.14752880

Abstract

In this paper we introduce a latent variable based model for the dynamics of financial range, the stochastic conditional range (SCR). We propose to estimate its parameters by Kalman filter, indirect inference and simulated maximum likelihood depending on the hypotheses on the distributional form of the innovations. The model is applied to a large subset of the S&P 500 components. A comparison of its fitting and forecasting abilities with the conditional autoregressive range (CARR) model shows that the new approach can provide a competitive alternative.

Keywords: Volatility, range, importance sampling, indirect inference.

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 (2008) on GARCH and stochastic volatility. Most of this literature concentrates on the dynamics of the differences 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 et al. (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 we introduce a latent variable based

ISSN: 3065-0623

Columbia Journal of Entrepreneurship and Management

Research Article

variant of the CARR model: the stochastic conditional range (SCR) model. In this new formulation, the dynamics of the ranges are driven by a latent variable which is supposed to capture the unobserved information flow that reaches the market. The specification of the model is multiplicative, like in the CARR model, but its main difference is that the SCR has two stochastic innovations, one affecting the observed range and the other the latent variable. The model can be seen as characterized by a mixture of distributions, or, in following Cox (1981), as parameter-driven. This specification also shares most of the statistical characteristics of the stochastic conditional duration (SCD) model of Bauwens and Veredas (2004). In section 2, we will present the model and discuss some of its properties.

In section 3 we propose three methods for estimation: maximum likelihood based on Kalman filter or on numerical integration of the latent variable and indirect inference. A comparison on the fitting and predictive capabilities of CARR and SCR models is carried out for a large sample of stocks in section 4. Results show that the SCR provides more reliable estimates of the autocorrelations of the data process, while in terms of forecasting accuracy it is comparable to CARR. Section 5 concludes.

2 The model

Let p_τ the price of a financial asset sampled at frequent (e.g.minutes or seconds) time intervalsτ, and $P_\tau = (p_\tau)$ its logarithm. We define as range the difference $R_t = m(P_t) - min(P_t)$, where $_t$indicates a coarser set of time intervals (e.g. days, weeks) such that $\tau = t-1, t-1+\frac{1}{n}, t-1+\frac{2}{n}, \dots, t,$ (1)

Where$n$ is the number of frequent intervals contained in one of the coarser intervals indexed by *t.*

The conditional autoregressive range CARR(1,1), introduced by Chou (2005), is defined by the following equations:

$$Rt = \Psi t \varepsilon t \quad (2)$$

$$\Psi t = \omega + \alpha Rt{-}1 + \beta \Psi t{-}1 \quad (3)$$

with

$$\omega > 0, \quad \alpha \geq 0, \quad \beta \geq 0,$$

where the baseline range (the error) ε_t has a distribution with density function $p(\varepsilon|I_t)$, which has positive support and unit mean. I_t denotes the information set at time *t-1*, and it includes the past values of R_t and ψ_t. Computing moments and autocorrelation for the CARR(1,1) model is easy and one can obtain the following simple expression:

ISSN: 3065-0623

Columbia Journal of Entrepreneurship and Management

Research Article

$$E(R_t) = \frac{\omega}{1-\alpha-\beta} \quad \text{if } (\alpha+\beta) < 1, \qquad (4)$$

$$E(R_t^2) = E(R_t)^2 \frac{1+\sigma_\varepsilon^2(1-\beta^2-2\alpha\beta)}{1-(\alpha+\beta)^2-\alpha^2\sigma_\varepsilon^2}, \qquad (5)$$

$$\rho_1 = \frac{\alpha(1-\beta^2-\alpha\beta)}{1-\beta^2-2\alpha\beta}, \qquad (6)$$

$$\rho_n = (\alpha+\beta)^{-1} \quad (n > 1). \qquad (7)$$

We introduce the stochastic conditional range (SCR) as the process described as follows:

$$R_t = e^{\psi t}\,\varepsilon_t \qquad (8)$$

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

where $u_t|I_t$ has an *iid* standard normal distribution and $\varepsilon_t|I_t$ has, like in the case of CARR, a distribution defined on the positive axis with unitary mean. The expected value of the range conditional to the past of the process up to time *t-1* is

$$E(R_t|I_t) = e^{\psi t}$$

and the distribution of R_t results from the mixing of the lognormal distribution of $e^{\psi t}$ 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 thes-th autocorrelation of R_t are the following

$$E(R_t) = E(\varepsilon_t)E(e^{\psi_t}) = e^{\frac{\omega}{1-\beta}+\frac{\sigma^2}{1-\beta^2}}, \qquad (10)$$

$$E(R_t^2) = E(R_t)^2\left(\frac{E(\varepsilon_t^2)}{E(\varepsilon_t)^2}e^{\frac{\sigma^2}{1-\beta^2}s-1}\right) + E(R_t)^2, \qquad (11)$$

$$\rho_s = \frac{e^{\frac{\varepsilon\beta}{1-\beta^2}}-1}{E(\varepsilon_t^2)e^{\frac{\sigma_\varepsilon^2}{1-\beta^2}}-1} \qquad (12)$$

for all $s \geq 1$.

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. Weibull distribution is commonly employed in duration analysis and was adopted by Chou (2005) in the CARR model. The justification for the use of the log-normal distribution arises from the result by Alizadeh et al.(2002) on the distribution of daily high and low prices, which

ISSN: 3065-0623

Columbia Journal of Entrepreneurship and Management

Research Article

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.

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 location parameter of the lognormal distribution will be therefore set to $-1/2\sigma_\varepsilon^2$, while the scale parameter of the Weibull will be restricted to be equal to $\Gamma(1+1/\gamma)^{-1}$, where σ_ε^2 and γ are the shape parameters which will be let free to vary.

3 Estimation

In this section we will discuss how the estimation of the SCR model can be performed either by maximum likelihood (ML) or by indirect inference. Concerning ML estimation, we will detail the methods that can be followed in order to deal with the problem of the presence of a latent variable.

3.1 ML with Kalman filter and EIS

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, as in the L-SCR specification, the model can be transformed by taking the logarithms on both sides of equation (8). This yields the following relationships

$$r_t = \ln R_t = \psi_t + \ln\varepsilon_t, \quad (13)$$

$$\psi_t = \omega + \beta\psi_{t-1} + \sigma u_t, \quad (14)$$

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 ($\ln \varepsilon_t$ 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 efficient 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

$$L(\theta, R) = \int g(R|\psi,\theta_1)h(\psi|\theta_2)d\psi = \int \prod_{t=1}^{n} p(R_t|\psi_t,\theta_1)q(\psi_t|\psi_{t-1},\theta_2)d\psi_t \quad (15)$$

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 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.

ISSN: 3065-0623

Columbia Journal of Entrepreneurship and Management

Research Article

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 we 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, we 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 we present a brief summary.

3.2 Indirect inference

An alternative solution for the estimation of the parameters of the SCR models can consist in indirect inference (for a detailed introduction see Gourieroux and Monfort (1996)), a simulation-based method that can be useful in estimating models for which the likelihood function is difficult to evaluate. Indirect inference relies on the possibility of easily simulating data from the model which is object of estimation (the estimand model). Simulations from the estimand are evaluated through a criterion function constructed with an approximate, or auxiliary, model, whose estimation can be performed easily (at least relatively to the estimand model). The auxiliary model does not necessarily provide an accurate description of the true process that generated the data, working more as a window through which to view both the actual data and the ones simulated from the estimand model. The objective of indirect inference is to choose the parameters of the estimand model so that they minimize a distance between the results of the estimation (that can consist in the parameters, the likelihood or the score) of the auxiliary model with the simulated data and the actual ones. A brief summary of this method is presented in the appendix.

For the indirect inference estimation of the SCR model, we chose two auxiliary models: an AR(10) and an ARMA(1,1). Both models were estimated on the logarithm of the observed and simulated ranges. As a result of the estimation of the two auxiliary models we chose to use their parameters and a simple sum of their squared differences was employed as the distance to minimize to obtain the indirect inference estimator.

ISSN: 3065-0623

Columbia Journal of Entrepreneurship and Management

Research Article

3.3 Estimation of the latent variables

Once estimates the parameters of the models have been obtained, it is possible to compute estimates of the latent variableψ_t. The process described by equations 13 and 14 is in the form of a linear state space model, and this allows to employ Bayesian updating in order to recover estimates for a prediction step, that provides a one-step-ahead prediction of the latent variable ψ_t given the previous observation $r_{(t-1)}$

$$(\psi_t|r_1{:}t-1) = p_\theta(\psi_t|\psi_t-1)p_\theta(\psi_t-1|r_1{:}t-1)d\psi_t-1, \quad (16)$$

and for a filtering (updating) step, which provides an estimate of the value of the latent variable ψ_t given an contemporary observation r_t,

$$p_\theta(\psi_t|r_{1:t}) = \frac{p_\theta(r_t|\psi_t)p_\theta(\psi_t|r_{1:t-1})}{\int p_\theta(y_t|\psi_t)p_\theta(\psi_t|r_{1:t-1})d\psi_t}. \quad (17)$$

When the state space model is Gaussian in both its innovations, the Kalman filter provides simple analytic forms for the predicted and updated values of the latent variable. This is the case only for the L-SCR model. If instead we allow the baseline distribution of the range to follow a different model (like, in our case, a Weibull) the Gaussianity of the process is lost and we had to recur to particle filter, a Monte Carlo method for the numerical evaluation of non-Gaussian state space models (for details see Arulampam et al. (2002).

3.4 Evidence from simulated processes

Table 1 displays the sample means and standard deviations of the estimated parameters of 100 simulations of W-SCR series. The simulated sets have sizes of 1000, 2500 and 10000 observations and the parameters used to generate the data are similar to the average values of the estimates computed later in this paper. The simulated series were estimated by indirect inference with an AR(10) and a ARMA(1,1) auxiliary model and by maximum EIScomputed likelihood. We also report the results of Kalman filter estimation, but only for the three parameters governing the dynamics of the latent variable, the only ones that would be estimated consistently by quasi maximum likelihood.

1000					
		ω	β	σ	γ
DGP		**0.0000**	**0.9750**	**0.1000**	**3.5000**
Indirectinf. AR(10)	Mean	0.0005	0.9767	0.0931	3.4943
	Stddev	*0.0029*	*0.0076*	*0.0117*	*0.1246*

ISSN: 3065-0623

Columbia Journal of Entrepreneurship and Management

Research Article

Indirectinf. ARMA(1,1)	Mean	0.0003	0.9754	0.0918	3.6755
	Stddev	*0.0029*	*0.0074*	*0.0128*	*0.2906*
ML-EIS	Mean	0.0017	0.9763	0.0955	3.4135
	Stddev	*0.0035*	*0.0089*	*0.0094*	*0.1316*
Kalmanfilter	Mean	0.0031	0.9693	0.1945	
	Stddev	*0.0039*	*0.0095*	*0.0122*	
2500					
		ω	β	σ	γ
DGP		**0.0000**	**0.9750**	**0.1000**	**3.5000**
Indirectinf. AR(10)	Mean	0.0002	0.9753	0.1001	3.5508
	Stddev	*0.0017*	*0.0056*	*0.0076*	*0.0803*
Indirectinf. ARMA(1,1)	Mean	0.0001	0.9746	0.0966	3.7535
	Stddev	*0.0017*	*0.0053*	*0.0078*	*0.2125*
ML-EIS	Mean	0.0012	0.9789	0.0915	3.3570
	Stddev	*0.0023*	*0.0066*	*0.0076*	*0.1174*
Kalmanfilter	Mean	0.0033	0.9724	0.1949	
	Stddev	*0.0021*	*0.0059*	*0.0067*	
10000					
		ω	β	σ	γ
DGP		**0.0000**	**0.9750**	**0.1000**	**3.5000**

ISSN: 3065-0623

Columbia Journal of Entrepreneurship and Management

Research Article

Indirectinf. AR(10)	Mean	0.0003	0.9756	0.0992	3.5214
	Stddev	*0.0010*	*0.0033*	*0.0039*	*0.0401*
Indirectinf. ARMA(1,1)	Mean	0.0001	0.9758	0.0936	3.7319
	Stddev	*0.0009*	*0.0033*	*0.0052*	*0.1603*
ML-EIS	Mean	-0.0006	0.9761	0.0962	3.4276
	Stddev	*0.0012*	*0.0034*	*0.0057*	*0.1038*
Kalmanfilter	Mean	0.0028	0.9742	0.1957	
	Stddev	*0.0011*	*0.0031*	*0.0035*	

Table 1: Sampling means and standard deviations of 100 estimates of the W-SCR model parameters for simulated series of 1000, 2500 and 10000 observations. The initial parameters of the four estimation methods were chosen to be equal to the simulation parameters plus a zero mean Gaussian error with standard deviation set at 0.05 for ω and β, at 0.01 for σ and at 0.5 for γ.

If any jittered starting value was beyond parameter constraints, a new sample of values was drawn. The parameters ω and β are estimated in a satisfactory way by all models even at the limited sample size of 1000 observations. Sample means and standard deviations are strictly comparable. Estimators seem to converge, in fact as the sample size increases; averages approach the simulated parameters and standard deviation become tighter. σ, the standard deviation of the innovations of the latent variable, seems more problematic to estimate. Indirect inference and ML-EIS seem to underestimate in a similar way, while the Kalman filter grossly overshoots (this last result is consistent with the Monte Carlo results in Bauwens and Galli (2009) for SCD models). Finally concerning γ the shape parameter of the Weibull baseline, it seems that indirect inference with an ARMA(1,1) auxiliary model has a slight loss of efficiency compared with ML-EIS and AR(10)-based indirect inference.

4 Empirical analysis

4.1 The data

We carried out the empirical analysis by considering series of ten years for all Standard and Poor's 500 components ending 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

ISSN: 3065-0623

Columbia Journal of Entrepreneurship and Management

Research Article

mean in order to speed up computation by reducing the search for the intercept in the conditional range function and to have more comparable estimates and forecasts. 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 the numerical algorithms for very limited sample sizes required to set a threshold. Table 2 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 difficult to tell. The use of an outlier detection and removal algorithm, like for instance the one deviced for durations by Chiang and Wang (2012), could be an interesting extension to this analysis and we leave it for further research. Average skewness and kurtosis indicate a strong departure from normality due to the presence of a heavy right tail. Statistics on autocorrelations are reported in the first column of the upper part of table 5 and show the presence of a marked degree of memory. These descriptive statistics are similar to the results obtained by Chou (2005).

	mean	stddeviation	maximum	minimum
observations	2469.176	201.072	2517	1039
means	1	0	1	1
medians	0.808	0.056	0.888	0.571
stddeviations	0.720	0.158	1.489	0.499
minima	0.177	0.046	0.275	<0.001
maxima	9.586	5.294	45.431	4.303
skewnesses	3.745	2.414	25.621	1.758
kurtoses	34.378	91.415	981.904	5.442

Table 2: Descriptive statistics of the 478 stocks used for the empirical analysis.

4.2 Estimation results

The models used in the empirical analysis were a CARR(1,1) with a Weibull conditional range distribution

ISSN: 3065-0623

Columbia Journal of Entrepreneurship and Management

Research Article

(W-CARR), an SCR with a lognormal distribution (L-SCR) and an SCR with a Weibull conditional distribution (WSCR). All models were specified with only one lag of the conditional range (and the range for the CARR model). 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. The W-SCR model were also estimated by indirect inference with an AR(10) and an ARMA(1,1) as auxiliary models. Estimation times runned from less than a second for the CARR model to an average of half a minute for the L-SCR model and the W-SCR with an AR(10) auxiliary model and an average of 3-4 minutes for the W-SCR with ML-EIS and ARMA(1,1) indirect inference.

Table 3 reports sample means and standard deviations of the estimated parameters. All the estimators of the SCR model yield similar values for ω and for β. The high level of persistence in the data is reflected by the average estimate of β, at a value close to one.

In the CARR case a similar high persistence emerges from the sum of the estimated values of α and β, which is close to one as well.Estimates for σ and γ seem to be sensitive to the method employed and seem to be negatively correlated. Even if the CARR model yields markedly lower estimates for γ than the SCR model, the parameter is always larger than 2 on average. This result is similar to the value obtained in Chou (2005) for the S&P500 index and suggest that an exponential distribution (that could be obtained by setting to one the γ parameter of the Weibull) would not be a suitable model for the baseline range.

		ω	α	β	σ	γ	σ_ε
W-CARR	mean	0.0405	0.2069	0.7509		2.2673	
	stddev	*0.0334*	*0.0642*	*0.0865*		*0.2378*	
L-SCR	mean	-0.0002		0.9782	0.1695		0.3546
	stddev	*0.0014*		*0.0184*	*0.0252*		*0.0238*
W-SCR	mean	-0.0029		0.9784	0.0725	3.9483	
ind. inf. ARMA(1,1)	*stddev*	*0.0028*		*0.0185*	*0.0282*	*0.3637*	
W-SCR	mean	-0.0026		0.9619	0.1033	3.7991	

ISSN: 3065-0623

Columbia Journal of Entrepreneurship and Management

Research Article

ind. inf. AR(10)	*stddev*	*0.0018*		*0.0290*	*0.0308*	*0.3564*	
W-SCR	mean	0.0085		0.9311	0.1495	3.2266	
ML-EIS	*stddev*	*0.0111*		*0.0472*	*0.0386*	*0.2266*	

Table 3: Sample means and standard deviations of estimated parameters.

4.3 Analysis of residuals

For the SCR model, we can define the residual corresponding to the innovation ε_t as

$$\hat{\varepsilon}_t = R_t / e^{\hat{\psi}_t} \qquad (18)$$

where ψ_t are the estimates of the latent variable provided either by the Kalman or the particle filter conditional on the observation of the range at time t (the so called filtered or updated estimates). In the CARR case, the residual is provided by the ratio $R_t/\Psi_t$where Ψ_t is recursively computed by replacing the values of the estimated parameters and of R_{t-1} and Ψ_{t-1} in equation 3. For each stock in the sample we computed the sample correlogram of and checked if the strong autocorrelation present in the data was removed by the estimated dynamic part of the models we used. Results are detailed in table 4. Though none of the models seems to completely explain away the autocorrelation present in the data, residuals display a much limited serial dependence with respect to the samples used for estimation. The first autocorrelation is on average positive and quite high, followed by smaller negative values for the SCR models, while in the CARR case it drops close to zero after the first lag. At higher lags (after 10) SCR residuals' autocorrelations tend to drop to small values while CARR's ones show a tendency to increase.

4.4 Fit of moments and autocorrelations

A comparison of the ability to fit the moments and autocorrelation structure of CARR and SCR models is presented in table 5. Moments and autocorrelations of the data were compared with implicit moments and autocorrelations computed by evaluating for each series formulae 4 to 7 and 10 to 12 with the values of estimated parameters. Except for the W-SCR estimated with AR(10) indirect inference, all models seem to slightly underestimate the average of the data.

	CARR		L-SCR		W-SCR	W-SCR	W-SCR
					ind.inf.	ind.inf.	ML-EIS

ISSN: 3065-0623

Columbia Journal of Entrepreneurship and Management

Research Article

					ARMA(1,1)		AR(10)			
	mean	*Stdev*	mean	*stdev*	mean	*stdev*	mean	*stdev*	mean	*stdev*
ACF(1)'s	0.067	*0.040*	0.077	*0.026*	0.096	*0.044*	0.054	*0.036*	0.020	*0.041*
ACF(2)'s	0.003	*0.034*	0.000	*0.026*	0.017	*0.028*	-0.015	*0.021*	-0.031	*0.028*
ACF(3)'s	-0.011	*0.036*	-0.022	*0.026*	-0.009	*0.027*	-0.030	*0.022*	-0.033	*0.025*
ACF(4)'s	-0.010	*0.035*	-0.026	*0.024*	-0.016	*0.025*	-0.028	*0.021*	-0.022	*0.025*
ACF(5)'s	0.001	*0.036*	-0.018	*0.026*	-0.012	*0.024*	-0.017	*0.021*	-0.004	*0.024*
ACF(6)'s	0.001	*0.035*	-0.019	*0.023*	-0.014	*0.025*	-0.014	*0.022*	0.000	*0.025*
ACF(7)'s	0.001	*0.033*	-0.019	*0.023*	-0.015	*0.026*	-0.012	*0.023*	0.001	*0.025*
ACF(8)'s	0.003	*0.033*	-0.017	*0.023*	-0.014	*0.025*	-0.009	*0.023*	0.005	*0.024*
ACF(9)'s	0.013	*0.033*	-0.009	*0.022*	-0.008	*0.022*	-0.001	*0.021*	0.014	*0.023*
ACF(10)'s	0.027	*0.031*	0.004	*0.022*	0.006	*0.021*	0.014	*0.021*	0.022	*0.023*
ACF(20)'s	0.026	*0.030*	0.003	*0.021*	0.007	*0.022*	0.014	*0.022*	0.020	*0.020*
ACF(30)'s	0.023	*0.028*	0.002	*0.020*	0.004	*0.021*	0.009	*0.021*	0.013	*0.020*
ACF(40)'s	0.022	*0.027*	0.003	*0.020*	0.005	*0.021*	0.008	*0.022*	0.010	*0.020*
ACF(50)'s	0.015	*0.027*	-0.003	*0.020*	-0.002	*0.019*	0.004	*0.020*	0.006	*0.019*

Table 4: Sample means and standard deviations of the autocorrelations of the residuals ε

The mean square errors of the first moments, computed by taking the average of the squares of differences between the empirical first moment and the implicit one computed on estimated parameters, show that the AR(10)

ISSN: 3065-0623

Columbia Journal of Entrepreneurship and Management

Research Article

W-SCR and the L-SCR seem to evaluate most precisely the mean of the process. Concerning the second moment, once again all models seem to yield estimates that are smaller on average than the sample values computed from the data. Here CARR and again AR(10) W-SCR seem to be the specifications with a lower quadratic loss. Coming to autocorrelations, W-SCD seems to reconstruct the serial dependency of the data with a smaller square loss than LSCD, which tends to overestimate the lower order autocorrelations and underestimate the higher order ones. CARR too has a higher value of MSE at all lags, as it systematically tends to underestimate the serial dependence in the data. It must be noted though that no model seems able to account fully for the apparent long memory in the data and at high order of lags all autocorrelations seem to be underestimated.

4.5 Predictive accuracy

The predictive accuracy of the different models was compared by an in sample one-step-ahead analysis. First the full sample was used to estimate the parameters of the models. Then we predicted every observation at time$t=2,...,n$ using estimated parameters and observations at time $t-1=1,...,n-1$.An out sample 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 forecasting accuracy of each estimator for each series was 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 Bonferroni corrected p-value is below 5%. Table 6 displays the main results for the estimation of the three models. It appears that though MSE's are very similar, CARR tends to predict marginally better than SCR regardless of the method of estimation. The substantial homogeneity in the performance of L-SCR and W-SCR models does not come as a surprise, as the most relevant parameters for forecasting are estimated consistently by Kalman filter. Concerning the slight forecasting edge of the CARR model, this could be due to the presence of two lagged variables in the CARR expression for conditional range. The SCR could be augmented by including the past observed range as a further determinant of the dynamics of its latent component and it would be interesting to evaluate if its forecasting ability improves.

	data		CARR		L-SCR		W-SCR		W-SCR		W-SCR	
							ind.inf.		ind.inf.		ML-EIS	
							ARMA(1,1)		AR(10)			
	mean	*stdev*	mean	*stdev*	mean	*stdev*	mean	*stdev*	mean	*stdev*	mean	*stdev*

ISSN: 3065-0623

Columbia Journal of Entrepreneurship and Management

Research Article

1^{st}mom	1.000	*0.000*	0.954	*0..060*	0.933	*0.111*	0.953	*0.111*	1.005	*0.066*	0.944	*0.121*
2^{nd}mom	1.543	*0.278*	1.313	*0.360*	1.024	*0.086*	1.090	*0.204*	1.296	*0.103*	1.127	*0.191*
ACF(1)'s	0.601	*0.102*	0.502	*0.156*	0.697	*0.091*	0.595	*0.087*	0.621	*0.064*	0.531	*0.059*
ACF(2)'s	0.550	*0.110*	0.483	*0.159*	0.682	*0.092*	0.581	*0.087*	0.596	*0.067*	0.473	*0.065*
ACF(3)'s	0.525	*0.111*	0.465	*0.163*	0.668	*0.095*	0.569	*0.089*	0.573	*0.074*	0.462	*0.072*
ACF(4)'s	0.513	*0.110*	0.447	*0.167*	0.654	*0.099*	0.556	*0.092*	0.551	*0.080*	0.472	*0.087*
ACF(5)'s	0.503	*0.111*	0.431	*0.171*	0.640	*0.102*	0.544	*0.093*	0.530	*0.086*	0.484	*0.098*
ACF(6)'s	0.494	*0.111*	0.416	*0.175*	0.627	*0.104*	0.533	*0.095*	0.511	*0.090*	0.493	*0.107*
ACF(7)'s	0.482	*0.113*	0.402	*0.179*	0.615	*0.107*	0.522	*0.097*	0.492	*0.094*	0.496	*0.114*
ACF(8)'s	0.475	*0.112*	0.389	*0.182*	0.602	*0.109*	0.511	*0.098*	0.474	*0.098*	0.506	*0.120*
ACF(9)'s	0.476	*0.113*	0.376	*0.185*	0.590	*0.111*	0.500	*0.099*	0.457	*0.101*	0.530	*0.124*
ACF(10)'s	0.468	*0.110*	0.364	*0.188*	0.578	*0.113*	0.490	*0.101*	0.441	*0.104*	0.536	*0.127*
ACF(20)'s	0.414	*0.107*	0.271	*0.207*	0.476	*0.128*	0.400	*0.108*	0.314	*0.116*	0.491	*0.128*
ACF(30)'s	0.380	*0.105*	0.212	*0.215*	0.394	*0.134*	0.330	*0.110*	0.229	*0.111*	0.416	*0.122*
ACF(40)'s	0.345	*0.108*	0.172	*0.218*	0.329	*0.134*	0.273	*0.108*	0.171	*0.101*	0.357	*0.129*
ACF(50)'s	0.307	*0.104*	0.143	*0.218*	0.276	*0.131*	0.228	*0.103*	0.130	*0.090*	0.243	*0.117*

ISSN: 3065-0623

Columbia Journal of Entrepreneurship and Management

Research Article

		MSE	MSE	MSE	MSE	MSE
1 stmo m		0.006	0.002	0.007	0.000	0.009
2ndmom		0.159	0.395	0.315	0.116	0.301
ACF(1)'s		0.031	0.013	0.005	0.005	0.010
ACF(2)'s		0.026	0.021	0.006	0.007	0.010
ACF(3)'s		0.025	0.024	0.007	0.007	0.007
ACF(4)'s		0.026	0.023	0.007	0.006	0.004
ACF(5)'s		0.028	0.023	0.007	0.005	0.003
ACF(6)'s		0.029	0.021	0.006	0.005	0.003
ACF(7)'s		0.029	0.021	0.006	0.005	0.003
ACF(8)'s		0.032	0.020	0.006	0.005	0.004
ACF(9)'s		0.036	0.017	0.005	0.006	0.006
ACF(10)'s		0.037	0.016	0.005	0.006	0.008
ACF(20)'s		0.052	0.008	0.004	0.017	0.013
ACF(30)'s		0.065	0.005	0.007	0.030	0.017
ACF(40)'s		0.068	0.004	0.009	0.038	0.019
ACF(50)'s		0.068	0.005	0.010	0.038	0.021

Table 5: Upper table: sample means and standard deviations of sample moments and autocorrelations of the 478 S&P 500 stocks with more than 1000 observations and theoretical moments and autocorrelations computed from estimated parameters. Lower table: averages of the squared differences between implicit theoretical moments and sample moments computed for each stock.

When the significance of pairs of forecasts is tested, it turns out that only about in one stock in fifteen the CARR and the W-SCR model forecast in a significantly different way. The proportion reduces of a half when the LSCR is concerned. If finally we restrict our sample to significantly different forecasts only, we see that the gain of
CARR in terms of MSE is slightly reduced in the case of the W-SCR while it remains substantially the same for the LSCR. We conclude by remarking that statistics on the comparisons between W-SCR and L-SCR, that are

ISSN: 3065-0623

Columbia Journal of Entrepreneurship and Management

Research Article

not reported in table 6, display a substantial similarity between the forecasts of the two models (for example, only less than the 1% of the forecasts can be considered different after testing).

	mean	sd	min	max	
MSE W-CARR	0.2738	0.1288	0.1564	1.3703	
MSE L-SCR	0.2759	0.1136	0.1632	1.0583	
MSE W-SCR ind.inf. AR(10)	0.2758	0.1155	0.1597	1.0745	
MSE W-SCR ind.inf. ARMA(1,1)	0.2763	0.1189	0.1601	1.1181	
MSE W-SCR ML-EIS	0.2756	0.1171	0.1613	1.0690	
significantly≠					
L-SCR and W-CARR					3.8%
significantly≠					
W-SCR AR(10) and W-CARR					7.9%
significantly≠					
W-SCR ARMA(1,1) and W-CARR					7.9%
significantly≠					
W-SCR ML-EIS and W-CARR					7.8%

Table 6: MSE comparison and Diebold and Mariano (2002) results for in sample one-step-ahead forecasts.

5 Conclusion

The SCR is a simple model for the dynamics of financial range. Its estimation is feasible and can be achieved with several techniques, a few of them have been proposed here. In an empirical analysis on a large subset of the stocks composing S&P 500, SCR seemed to improve on the CARR model in reconstructing the autocorrelation structure of the data and was only slightly less efficient in forecasting. Extensions of the models are possible and could be explored in future research. Concerning the latent variable, a long memory version, a version with a more complex process than an AR(1) and a departure from normality of the innovation in the latent variable process could be of interest. Concerning the baseline range distribution, a mixture distribution could be useful in accounting for the consequences of heterogeneity in the information in the market.

ISSN: 3065-0623

Columbia Journal of Entrepreneurship and Management

Research Article

References

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

Bauwens, L., & Galli, F. (2009). Efficient importance sampling for ml estimation of scd models. Computational Statistics and Data Analysis, 53(6):1974–1992.

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

Bauwens, L.,& Veredas, D. (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.,& Diebold, F. X. (2006). A no-arbitrage approach to range-based estimation of return covariances and correlations. Journal of Business, 79(1):61–74.

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

Chiang, M. H., & and Wang, L. M. (2012).Additive outlier detection and estimation for the logarithmic autoregressive conditional duration model.Communications in Statistics-Simulation and Computation, 41(3):287–301.

Chou, R. Y. (2005). Forecasting financial volatilities with extreme values: The conditional autoregressive range (carr) model. Journal of Money, Credit, and Banking, 37(3).

Cox, D. R.(1981). Statistical analysis of time series: Some recent developments [with discussion and reply]. Scandinavian Journal of Statistics, pages 93–115.

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

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

Engle, R. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of united kingdom inflation. Econometrica, 50:987–1007.

Gourieroux, C., Monfort, A., &Renault, E.. Indirect inference. (1993).Journal of applied econometrics, 8(S1):S85– S118.

Gourieroux, C.,& Monfort, A. (1996).Simulation-based Econometric Methods. Oxford University Press.

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

Liesenfeld, R.&Richard, J. F. (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., &Zhang, W. (2007). Efficient high-dimensional importance sampling. Journal of Econometrics, 141(2):1385–1411.

Arulampalam, M. S., Maskell, S., Gordon, N., & Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. Signal Processing, IEEE Transactions on, 50(2):174–188.

Strickland, C., Forbes, C., & Martin, G. (2006). Bayesian analysis of the stochastic conditional duration model. Computational Statistics and Data Analysis, 50(9):2247–2267. Taylor, S. (2008). Modelling financial time series. World scientific.