Financial Econometrics and SV models

Santiago Pellegrini* and Alejandro Rodriguez†

December 2007

*Department of Statistics UC3M, e-mail:spellegr@est-econ.uc3m.es
†Department of Statistics UC3M, e-mail:afrodrig@est-econ.uc3m.es

Contents

1 Financial Time Series

Financial time series data may be classified in a compartment different from economic time series. Not only because of their empirical features, but also because their study and analysis is used as an input of a different decision making process. In fact, the financial world concerns about how will be the risk of a given group of financial assets tomorrow and dismisses the information about its mean value. The uncertainty of financial returns plays a central role in many financial models of valuation of financial derivatives, risk management, efficient asset allocation and so on. Thus, if we believe that valuing is the most important task for users of financial time series data, then financial econometricians should put all their effort on analyzing volatility. We just need to take a look over the past 20 years of research work on this area to confirm this pattern.

This course attempts to make a brief survey of a model that is more and more considered to analyze and forecast volatility, called stochastic volatility (SV) model. In particular, we will discuss some of its statistical properties and the some estimation methods. We will intensively use MATLAB for all the computations. Some of the working m-files, together with other material for the course are available here. Let us begin by commenting the empirical properties of financial time series data.

1.1 Empirical Properties of asset returns

In Figure 1, we find the typical features that most financial time series share:

1. Log prices seem to be non-stationary, but without any marked systematic trend... (arguable!)

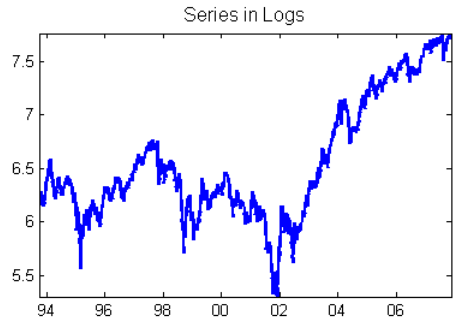
2. Returns (log-difference of the prices) exhibit a constant mean but with periods of clustered volatility (almost axiomatic)

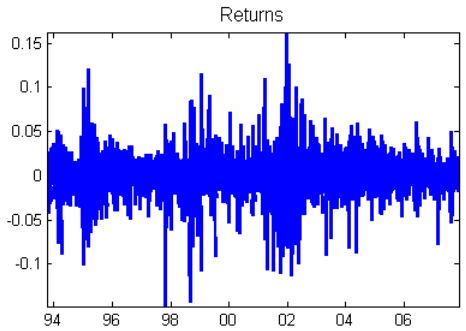
3. Returns are very weakly auto-correlated (emerging countries) or have no correlation at all (highly developed financial markets)

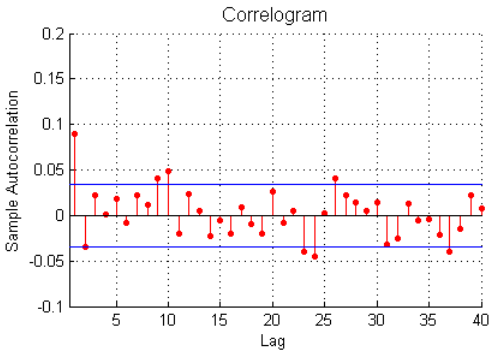
4. Squared returns are highly (positively) correlated, and in general with a high persistence (almost axiomatic)

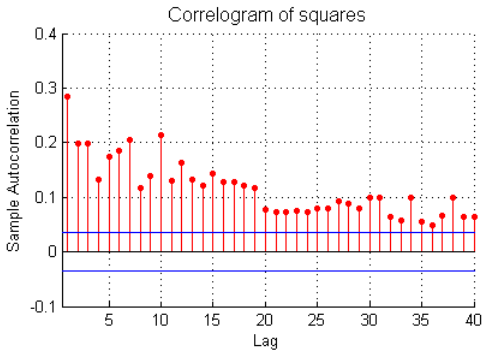
5. Returns are negative skewed (arguable) and exhibit positive excess kurtosis (implying fat tails)

MERVAL INDEX

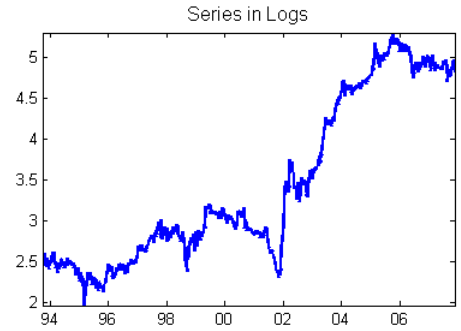


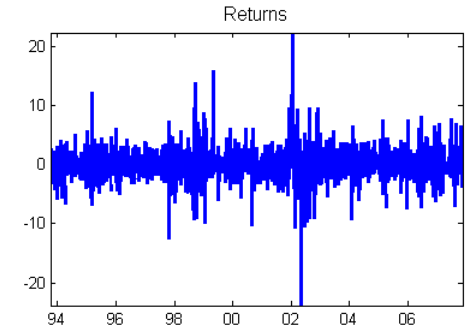


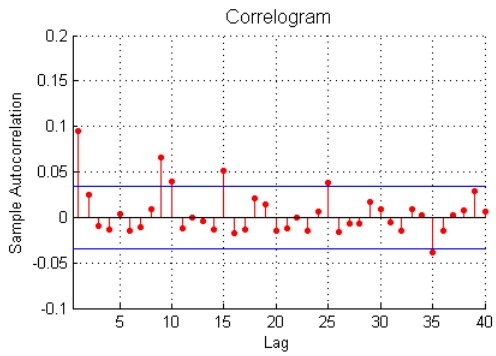


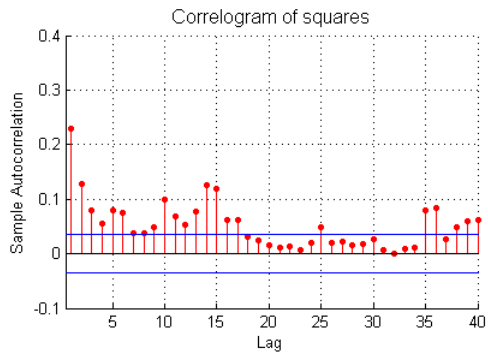


YPF









	MERVAL	YPF
Mean	0.0467 %	0.0711 %
Standard Dev.	2.3076 %	2.1978 %
Skewness	-0.1970	0.2674
Kurtosis	8.1329	16.0699

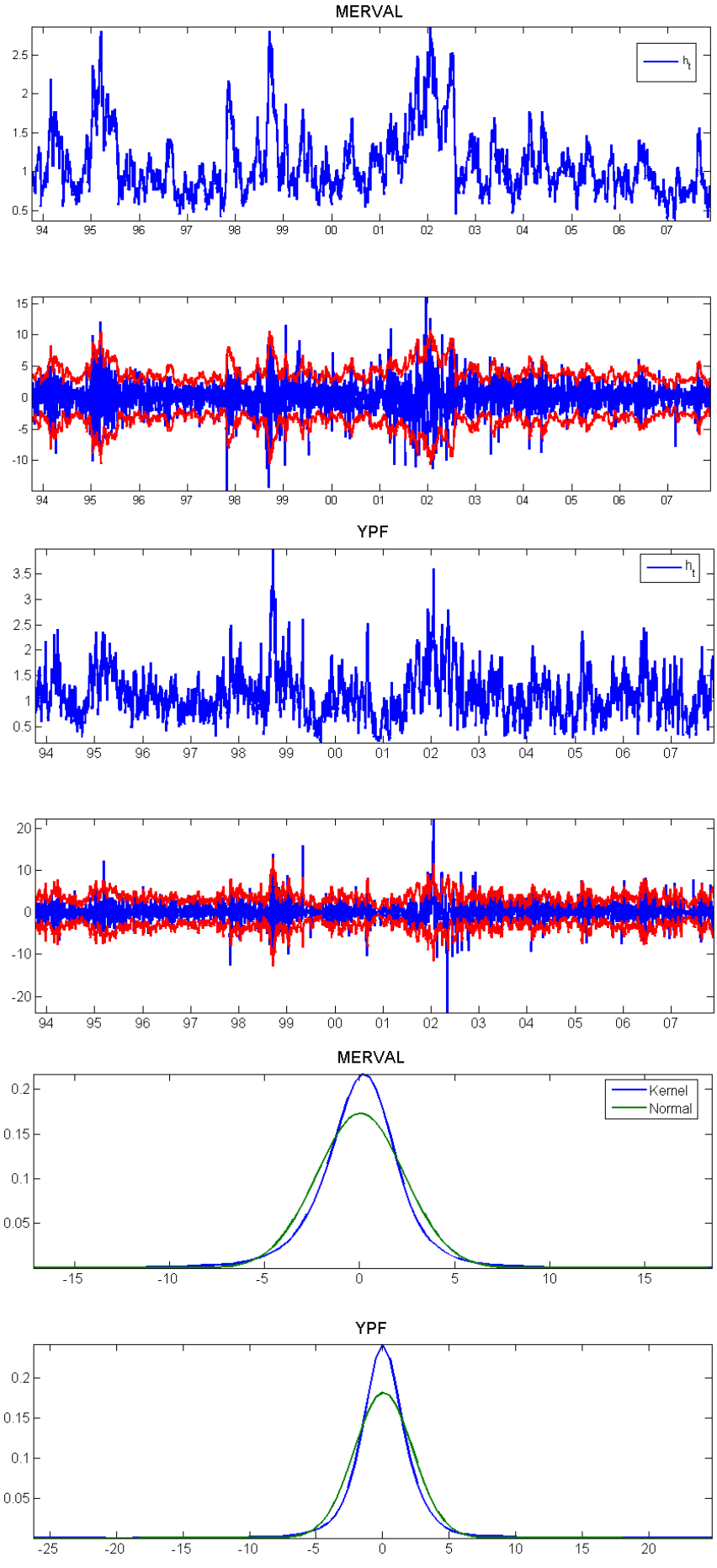


Figure 1: Daily closing price of the MERVAL Index and YPF ORD stocks

A few important conclusions emerge from these features. First, there is (almost) no room for predicting the future mean value of stock returns. Second, there is a LARGE AMOUNT OF INFORMATION COMING FROM LAGGED RETURNS to predict the volatility of stock returns. Last but not least, Normality does not seem an adequate distributional assumption for the log-returns because of the negative skewness and large kurtosis.

1.2 A short comparison between GARCH and SV models

As a "parametrician", we can summarize the stylized facts just commented by modelling the returns, y_t as follows:

$$y_t = \mu_t + \sigma_t \varepsilon_t, \tag{1}$$

where μ_t is often assumed to be either constant or an AR(1) with a parameter close to zero, ε_t an independent random variable with zero mean and unit variance, and σ_t an either deterministic or random process that depends on past values of the returns. When σ_t is expressed as a deterministic function of lagged (squared) returns, we are within the ARCH models (Engle, 1982; Bollerslev, 1986), which have achieved widespread popularity in applied empirical research; see Bollerslev et al. (1992, 1994) for an exhaustive survey of ARCH models. Alternatively, volatility may be modelled as an unobserved component following some latent stochastic process, such as an autoregression. The resulting models are called stochastic volatility (SV) models created by Taylor (1986) and their interest has been increasing during the last years; see Ghysels et al. (1996); Taylor (1994); Shephard (1996); Yu and Meyer (2006); Asai et al. (2006).

We will follow Harvey et al. (1994) to briefly compare both types of models. Let's assume first that $\mu_t = 0$ for all t in (1). Then, in the GARCH(1,1) model, we have that

$$\sigma_t^2 = \gamma + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2, \quad \gamma > 0, \quad \alpha + \beta < 1. \tag{2}$$

Note that by letting the conditional variance be a function of squared previous returns and past variances, it is possible to capture the changes in volatility over time. Moreover, since this model is formulated in terms of the distribution of the one-step ahead prediction error, maximum likelihood estimation is straightforward. This property made GARCH models so popular and widely used in applied research. Of course, the formulation given in (2) may be generalized by adding more lags of both, the squared returns and the past variances. With respect to its dynamics, the conditional volatility in GARCH models may be seen as an ARMA(1,1) to the squared returns, where the AR coefficient is the sum $p = \alpha + \beta$. Therefore, as this sum gets

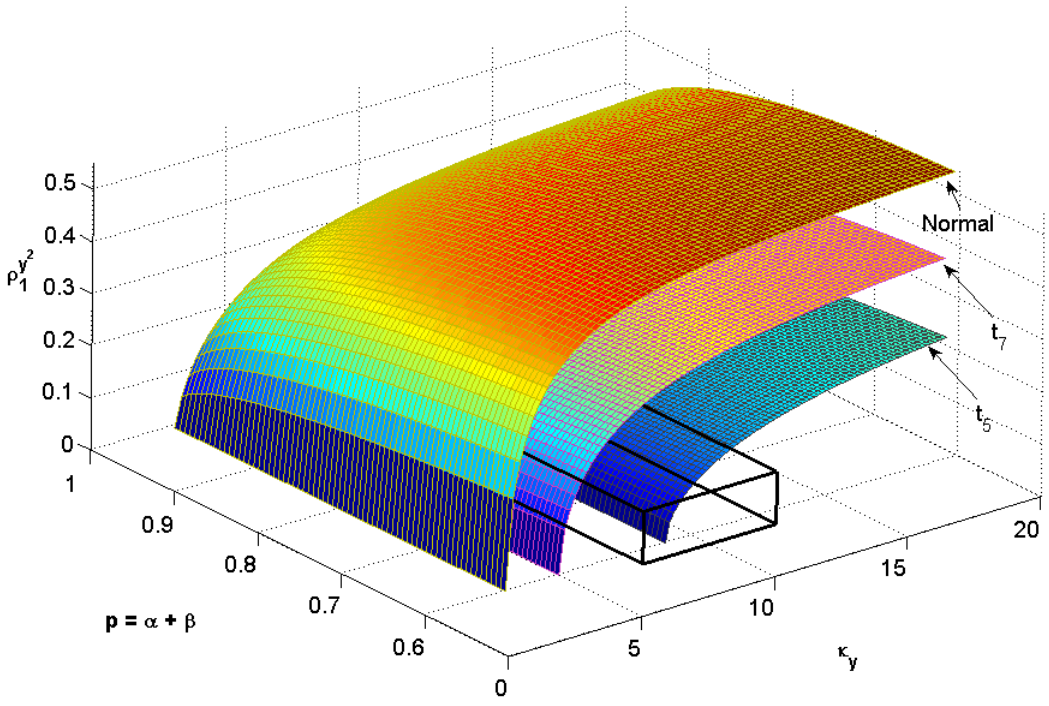


Figure 2: Relationship between autocorrelation, kurtosis and persistence in a GARCH(1,1) model. Source: Carnero et al. (2004)

closer to one, the persistence of the squared observations increases. This means that the autocorrelation function (ACF) of y_t^2 will decay quite slowly, as it is commonly seen in practice. Another feature of the GARCH(1,1) process is the strong relationship between the persistence, p and the kurtosis and lag-one autocorrelation of squares implied by this model. Carnero et al. (2004) study this issue in detail and formulate the lag-one autocorrelation as a function of the persistence and kurtosis to conclude that conditionally Normal GARCH models are not able to simultaneously capture the plausible values of persistence, kurtosis and lag-one autocorrelation of squares present in financial time series data. In other words, GARCH processes may model one or two of the three empirical properties, but fails to capture them all together. Figure 2 extracted from this paper summarizes this issue. It is clear that distributions with fatter tails (like the Student-t) need to be imposed on ε_t to account for the three features.

On the other hand, SV models consider that the dynamics of the logarithm of σ_t^2, denoted h_t, is well described by a latent (unobserved) stochastic process, whose dynamics is normally assumed to be autoregressive. That means

$$h_t = \phi h_{t-1} + \eta_t, \quad 0 \leq \phi < 1 \tag{3}$$

where $\eta_t \sim NID(0, \sigma_\eta^2)$. It is often assumed that η_t and ε_t are mutually independent, although Harvey and Shephard (1996) show that a possible dependence between these two allows the model to pick up the kind of asymmetric behavior that is often seen in stock prices.

Unlike GARCH models, the conditional variance in (3) is driven by an unpredicted shock. Therefore, though more flexible than GARCH, SV models generally lack analytic one-step-ahead forecasts densities $y_t|Y_{t-1}$ and thus also lack of a closed form likelihood. Hence, either an approximation or a numerically intensive method is required to estimate these models. On the other hand, SV models have several advantages that may eventually overcome the estimation issue. **First**, their statistical properties are easier to find and understand. **Second**, they are easier to generalize to the multivariate case. **Third**, they are attractive because they are close to the models used in Financial Theory for asset pricing. **Fourth**, as stated in Carnero et al. (2004), they capture in a more appropriate way the main empirical properties often observed in daily series of financial returns (something we have just started to note above and we will study in depth later). For example, we can show that the fourth moment always exists when h_t is stationary.

So, after this very brief comment of the two models, we should wander why practitioners do not use SV parallel to GARCH models in their everyday work. The reason comes almost entirely from the fact that the exact likelihood function is difficult to evaluate and Maximum Likelihood (ML) estimation of the parameters is not straightforward. The main objective of this course then is to give insights of two alternative ways of estimating the SV parameters, in order to notice that SV models may be really very competitive with respect to GARCH models for empirical applications, specially because (highly) computer intensive methods can be (relatively) easy to handle in modern computers[1]. Before describing the two methods, it is necessary to have a notion of State Space (SS) models because they are the framework usually chosen to define and study unobserved (latent) component models, such as SV models.

2 State Space Models and SV Models

When analyzing time series, sometimes it is useful to decompose the series of interest into trend, seasonal, cyclic and irregular components. These have a direct interpretation but they are unobserved; see Harvey (1989) and Durbin and Koopman (2001) for a comprehensive description of the Unobserved Component (UC) models. We can deal with UC models in a State Space (SS) model framework and using the Kalman filter. The empirical applications of these models are very wide. Next, we mention just a few examples. The evolution of inflation could

[1]For a detailed review of these and other estimation procedures, see Ruiz and Broto (2004).

be represented by a model with long-run level, seasonal and transitory components which may be conditionally heteroscedastic; see, for example, Ball et al. (1990), Evans (1991), Kim (1993) and Broto and Ruiz (2006). Cavaglia (1992) analyzes the dynamic behavior of ex-ante real interest differentials across countries by a linear model in which the ex-post real interest differential is expressed as the ex-ante real interest differentials (underlying unobserved component) plus the cross country differential inflation forecast error. When modelling financial returns, the volatility can also be modelled as an unobserved component as in the Stochastic Volatility (SV) models proposed by Taylor (1986) and Harvey et al. (1994).

2.1 Description and Properties of the SS Models

SS modelling provides a unified methodology for treating a wide range of problem in time series analysis. In this approach it is assumed that the development over time of the system under study is determined by an unobserved series of vectors $\alpha_1, \ldots, \alpha_T$, with which are associated a series of observations $y_1, \ldots, y_T$. The relation between α_t's and the y_t's is specified by the following linear SS model.

$$y_t = \boldsymbol{Z}_t\alpha_t + \boldsymbol{d}_t + \varepsilon_t, \quad t = 1, \ldots, T, \tag{4a}$$

$$\alpha_t = \boldsymbol{T}_t\alpha_{t-1} + \boldsymbol{c}_t + \boldsymbol{R}_t\eta_t, \quad t = 1, \ldots, T. \tag{4b}$$

where y_t is the univariate time series observed at time t, $\boldsymbol{Z}_t$ is a $1 \times m$ vector, $\boldsymbol{d}_t$ is a scalar and ε_t is a serially uncorrelated disturbance with zero mean and variance $\boldsymbol{H}_t$. On the other hand, α_t is the $m \times 1$ vector of unobservable state variables, $\boldsymbol{T}_t$ is an $m \times m$ matrix, $\boldsymbol{c}_t$ is an $m \times 1$ vector, $\boldsymbol{R}_t$ is an $m \times g$ matrix and η_t is a $g \times 1$ vector of serially uncorrelated disturbances with zero mean and covariance matrix $\boldsymbol{Q}_t$. Finally, the disturbances ε_t and η_t are uncorrelated with each other in all time periods. The Hyper-Parameters in the system of matrices $\{\boldsymbol{Z}_t, \boldsymbol{T}_t, \boldsymbol{Q}_t, \boldsymbol{H}_t, \boldsymbol{R}_t, \boldsymbol{c}_t, \boldsymbol{d}_t\}$ can be time-varying, but for our purposes we will assume to be known and *time-invariant*. The specification of the SS system is completed with the initial state vector, α_0, which has distribution with mean $\mathbf{a}_0$ and covariance matrix P_0.

Examples:

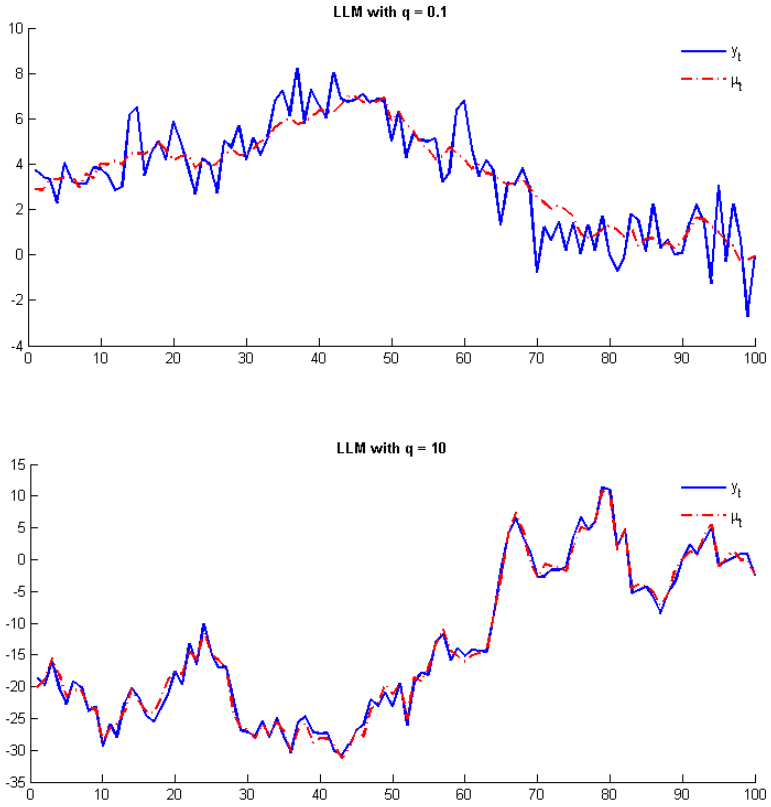
Local Level Model (LLM)

The structural representation of the model is given by

$$y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{IID}(0, \sigma_\varepsilon^2), \tag{5a}$$

$$\mu_t = \mu_{t-1} + \eta_t, \quad \eta_t \sim \mathcal{IID}(0, q\sigma_\varepsilon^2), \tag{5b}$$

where the unobserved state, α_t, is the level of the series, denoted by μ_t, that evolves over time following a random walk. In this model $\boldsymbol{Z} = \boldsymbol{T} = \boldsymbol{R} = 1$, $\boldsymbol{c} = \boldsymbol{d} = 0$, $\boldsymbol{H} = \sigma_\varepsilon^2$ and $\boldsymbol{Q} = q\sigma_\varepsilon^2$, where q is known as the signal to noise ratio.



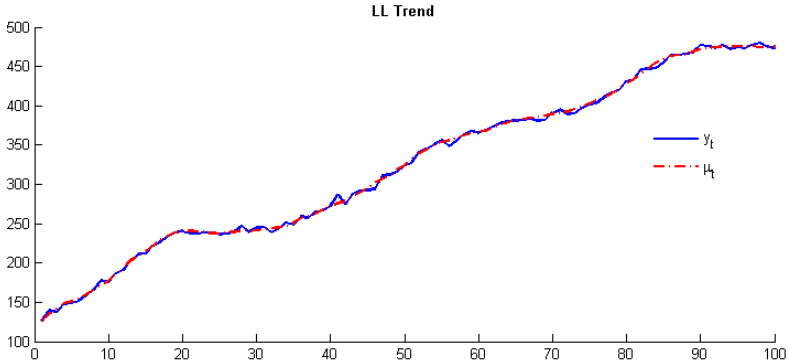
Local Linear Trend (LLT) Model

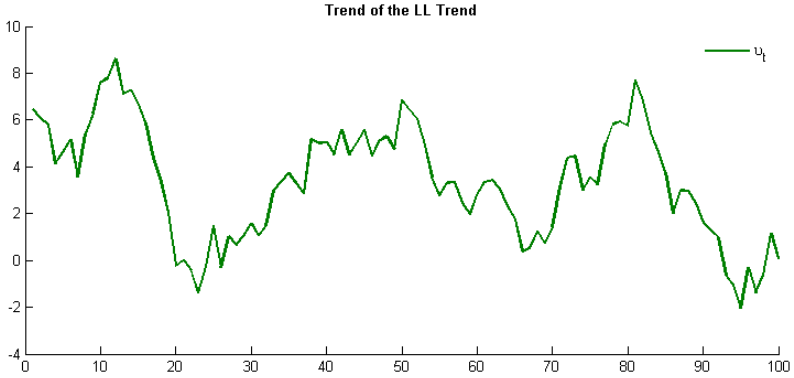
The structural representation of the model is given by

$$y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{IID}(0, \sigma_\varepsilon^2), \tag{6a}$$

$$\mu_t = \mu_{t-1} + \upsilon_t + \eta_t, \quad \eta_t \sim \mathcal{IID}(0, \sigma_\eta^2), \tag{6b}$$

$$\upsilon_t = \upsilon_{t-1} + \xi_t, \quad \xi_t \sim \mathcal{IID}(0, \sigma_\xi^2) \tag{6c}$$





In this model we have:

$$Z = \begin{pmatrix} 1 & 0 \end{pmatrix}, T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, R = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, Q = \begin{bmatrix} \sigma_\eta^2 & 0 \\ 0 & \sigma_\xi^2 \end{bmatrix}$$

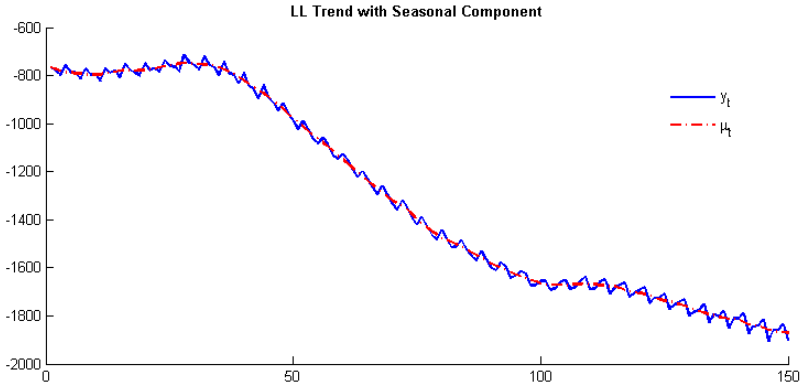
Local Linear Trend (LLT) Model with Seasonal Component

Let s be the number of periods in a year; thus for monthly data $s = 12$, for quarterly data $s = 4$, and so on. To model the seasonal component we use the term γ_t in (7a). The structural representation of the model is given by,

$$\begin{aligned} y_t &= \mu_t + \gamma_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{IID}(0, \sigma_\varepsilon^2), && \text{(7a)} \\ \mu_t &= \mu_{t-1} + \upsilon_t + \eta_t, \qquad \eta_t \sim \mathcal{IID}(0, \sigma_\eta^2), && \text{(7b)} \\ \upsilon_t &= \upsilon_{t-1} + \xi_t, \qquad \xi_t \sim \mathcal{IID}(0, \sigma_\xi^2), && \text{(7c)} \\ \gamma_t &= -\sum_{j=1}^{s-1} \gamma_{t-j} + \omega_t, \qquad \omega_t \sim \mathcal{IID}(0, \sigma_\omega^2) && \text{(7d)} \end{aligned}$$

For the case in which $s = 4$ we have,

$$Z = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \end{pmatrix}, T = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}, R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, Q = \begin{bmatrix} \sigma_\eta^2 & 0 & 0 \\ 0 & \sigma_\xi^2 & 0 \\ 0 & 0 & \sigma_\omega^2 \end{bmatrix}$$



ARIMA(p,q) Models

Autoregressive moving average (ARMA) time series models were introduced by Box and Jenkins in their pathbreaking (1970) book; see Box et al. (1994) for a current version of this book.

$$y_t = \sum_{j=1}^{p} \phi_j y_{t-j} + \sum_{i=1}^{q} \theta_i a_{t-i} + a_t \tag{8}$$

where a_t is a serially uncorrelated disturbance with zero mean and variance σ_a^2. For the particular case in which $p = 1$ and $q = 1$ we have,

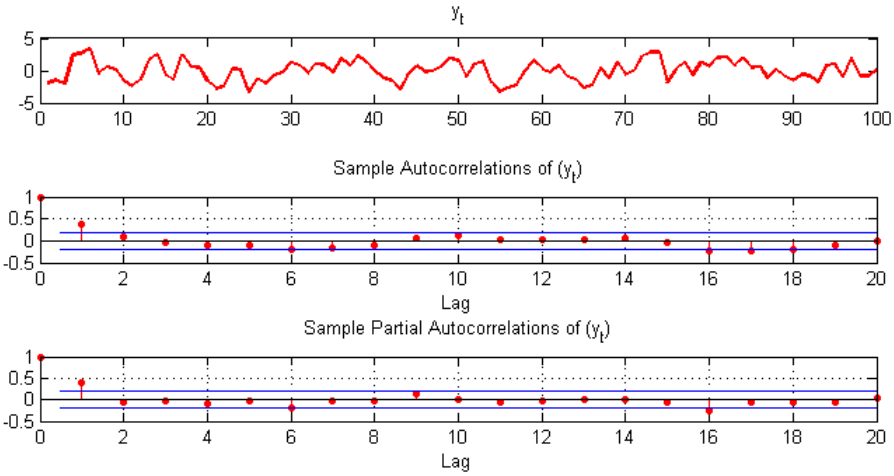
$$y_t = \phi y_{t-1} - \theta a_{t-1} + a_t \tag{9}$$

In general, we can express any ARMA model in a SS framework, see for instance Harvey (1989) and Durbin and Koopman (2001). For this particularized case that expression is given by,

$$\begin{aligned} y_t &= y_t && \text{(10a)} \\ y_t &= \phi y_{t-1} - \theta a_{t-1} + a_t, && \text{(10b)} \\ y_{t-1} &= y_{t-1}, && \text{(10c)} \\ a_t &= a_t && \text{(10d)} \end{aligned}$$

In this model we have:

$$Z = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}, T = \begin{bmatrix} \phi & 0 & -\theta \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, R = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, Q = \sigma_a^2, H = 0$$



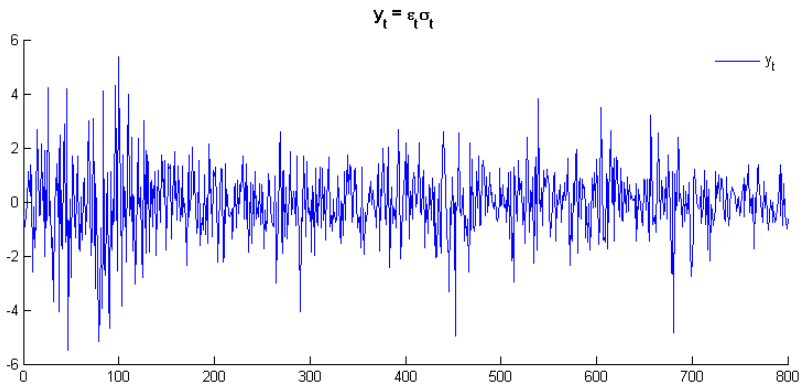
Autoregressive Stochastic Volatility (ARSV) Model

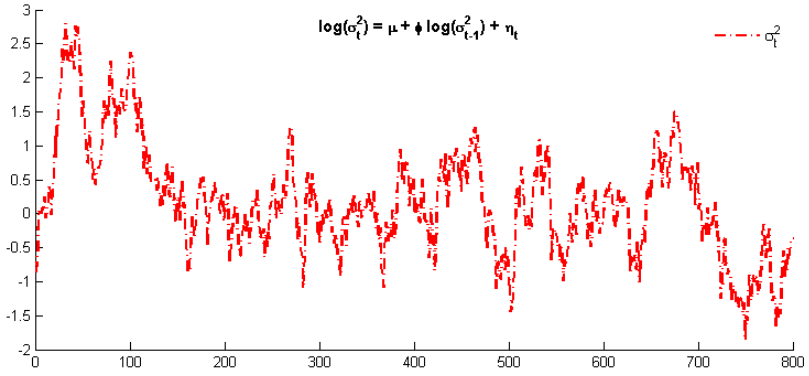
In the simplest case, the conditional log-volatility follows an AR(1) process, given by,

$$y_t = \varepsilon_t \sigma_t, \tag{11a}$$

$$\log\left(\sigma_t^2\right) = \mu + \phi \log\left(\sigma_{t-1}^2\right) + \eta_t, \tag{11b}$$

where ε_t is a strict white noise process with variance 1 and the noise in the conditional log-volatility equation, η_t, is assumed to be a Gaussian white noise process with variance σ_η^2 independent of ε_t. Finally, the parameter μ is related with the marginal variance of y_t.





2.2 The Kalman filter

Let Y and X be two random vectors, where Y only is observable. We want to know about X using the information of $Y = y$. This can be done using the *conditional probability density*.

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} \tag{12}$$

assuming that $p_Y(y) \neq 0$.

The conditional probability density $p_{X|Y}(x|y)$ with a particular value substituted for y and with x regarded as a variable sums up all the information with knowledge that $Y = y$ conveys about X. Since it is a function rather than a single real number or vector of real numbers, it makes sense to ask if we can obtain a simpler entity. One such estimate would be namely the conditional minimum variance (or minimum Mean Square Error - MSE) estimate.

Let $\hat{x}$ be an estimate of the value taken by X when we know that $Y = y$. Then, $\hat{x}$ is uniquely specified as the *conditional mean of* X *given that* $Y = y$, i.e.,

$$\hat{x} = \mathbb{E}[X|Y = y] = \int_{-\infty}^{\infty} x p_{X|Y}(x|y) dx. \tag{13}$$

Among the different algorithms for estimating the state vector in (4b), the Kalman filter is in the center. It is a recursive procedure for computing the optimal estimator for the state vector at time t, based on the information available at time t. This information consists of the observations up to and including y_t. System of hyper-parameters together with $\mathbf{a}_t$ and P_0 are assumed to be known in all time periods and so do not need to be explicitly included in the information set. The derivation of the Kalman filter rests on the assumption that the disturbances and the initial state vector are normally distributed. A standard result on the multivariate normal distribution allows to calculate recursively the distribution of α_t, conditional on the information set at time t. These conditional distribution are themselves

normal and hence are completely specified by their means and covariance matrices. It is these quantities which Kalman filter computes. When the normality assumption is dropped, there is no longer any guarantee that the Kalman filter will give the conditional mean of the state vector. However, it is still an optimal estimator in the sense that it minimizes the MSE within the class of all linear estimators.

Under the normality distribution, the initial state vector, α_0, has a multivariate normal distribution with mean $\mathbf{a}_0$ and covariance matrix P_0. The distribution of η_t and ε_t also have multivariate normal distribution for $t = 1, \ldots, T$ and are distributed independently of each other and of the α_0.

The state vector at time $t = 1$ is given by,

$$\alpha_1 = \boldsymbol{T}\alpha_0 + \boldsymbol{c} + \boldsymbol{R}\eta_1$$

Thus α_1 is a linear combination of two vectors of random variables, both with multivariate normal distribution, and a vector of constants. Hence it is itself multivariate normal with mean equal to,

$$\mathbf{a}_{1|0} = \boldsymbol{T}\mathbf{a}_0 + \boldsymbol{c},$$

and a covariance matrix given by,

$$P_{1|0} = \boldsymbol{T}P_0\boldsymbol{T} + \boldsymbol{R}\boldsymbol{Q}\boldsymbol{R}.$$

The notation $\mathbf{a}_{1|0}$ indicates the mean of the distribution of α_1 is conditional on the information at time $t = 0$.

In order to obtain the distribution of α_1 conditional on y_1, write,

$$\begin{aligned} \alpha_1 &= \mathbf{a}_{1|0} + (\alpha_1 - \mathbf{a}_{1|0}), && \text{(14a)} \\ y_1 &= \boldsymbol{Z}\mathbf{a}_{1|0} + \boldsymbol{d} + \boldsymbol{Z}(\alpha_1 - \mathbf{a}_{1|0}) + \varepsilon_1, && \text{(14b)} \end{aligned}$$

The equation (14b) is simply a rearrangement of the measurement equation in (4a). From (14) it can be seen that the vector $[\alpha_1' \;\; y_1']'$ has a multivariate normal distribution with mean equal to $[\mathbf{a}_{1|0}' \;\; (\boldsymbol{Z}\mathbf{a}_{1|0} + \boldsymbol{d})']'$ and a covariance matrix given by,

$$\begin{bmatrix} P_{1|0} & P_{1|0}\mathbf{Z}' \\ \mathbf{Z}P_{1|0} & \mathbf{Z}P_{1|0}\mathbf{Z}' + H \end{bmatrix}$$

Applying the multivariate normal properties it turns out that α_1, conditional on a particular

value of y_1, has also a multivariate normal distribution with mean given by,

$$\mathbf{a}_1 = \mathbf{a}_{1|0} + P_{1|0}\boldsymbol{Z}'F_1^{-1}(y_1 - \boldsymbol{Z}\mathbf{a}_{1|0} - \boldsymbol{d}),$$

with covariance matrix given by,

$$P_1 = P_{1|0} - P_{1|0}\boldsymbol{Z}'F_1^{-1}\boldsymbol{Z}P_{1|0},$$

where $F_1 = \boldsymbol{Z}P_{1|0}\boldsymbol{Z}' + H$.

Repeating the steps for $t = 2, \ldots, T$ yields the Kalman filter. For a general time t we get the *prediction equations*,

$$\mathbf{a}_{t|t-1} = \boldsymbol{T}\mathbf{a}_{t-1} + \boldsymbol{c}, \tag{15a}$$

whit covariance matrix given by,

$$P_{t|t-1} = \boldsymbol{T}P_{t-1}\boldsymbol{T}' + \boldsymbol{R}\boldsymbol{Q}\boldsymbol{R}, \text{ for } t = 1, \ldots, T. \tag{15b}$$

Once the new observation, y_t, become available we use the *update equations* for estimate $\mathbf{a}_t$ and P_t,

$$\mathbf{a}_t = \mathbf{a}_{t|t-1} + P_{t|t-1}\boldsymbol{Z}'F_t^{-1}v_t, \tag{16a}$$

where $v_t = (y_t - \boldsymbol{Z}\mathbf{a}_{t|t-1} - \boldsymbol{d})$ is the one-step ahead innovation. Finally, its covariance matrix is given by,

$$P_t = P_{t|t-1} - P_{t|t-1}\boldsymbol{Z}'F_t^{-1}\boldsymbol{Z}P_{t|t-1}, \text{ for } t = 1, \ldots, T. \tag{16b}$$

Let show an example using the LLM as in (5). Assuming a diffuse prior distribution, we obtain $\mathbf{a}_0 = 0$ and $P_0 \longrightarrow \infty$. Using the prediction equations, (15a) and (15b), we get,

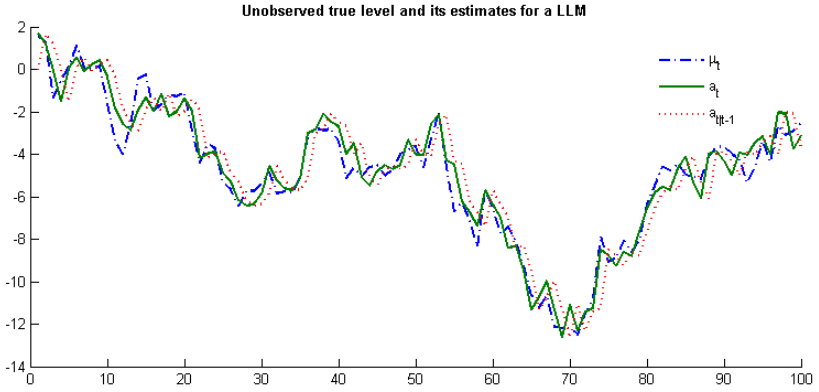
$$\begin{aligned} a_{1|0} &= 0, \\ P_{1|0} &= P_0 + q\sigma_\varepsilon^2 \longrightarrow \infty, \text{ as } k \to \infty \\ F_1 &= P_{1|0} + \sigma_\varepsilon^2 = P_0 + q\sigma_\varepsilon^2 + \sigma_\varepsilon^2 \longrightarrow \infty, \text{ as } k \to \infty. \end{aligned}$$

Once the y_1 become available, using the update equations, (16a) and (16b), we get,

$$a_1 = a_{1|0} + \frac{P_{1|0}}{F_1}(y_1 - a_{1|0}) = \lim_{k\to\infty}\left[a_{1|0} + \frac{P_0 + q\sigma_\varepsilon^2}{P_0 + q\sigma_\varepsilon^2 + \sigma_\varepsilon^2}(y_1 - a_{1|0})\right] = y_1, \tag{17a}$$

$$P_1 = P_{1|0} - P_{1|0}F_1^{-1}P_{1|0} = \lim_{k\to\infty}(P_0 + q\sigma_\varepsilon^2)\left(1 - \frac{P_0 + q\sigma_\varepsilon^2}{P_0 + q\sigma_\varepsilon^2 + \sigma_\varepsilon^2}\right) = \sigma_\varepsilon^2. \tag{17b}$$

Then, using recursively the prediction and the update equations it is possible to obtain one-step ahead estimates of α_t, $a_{t|t-1}$, and update estimates a_t for $t = 2, \ldots, T$.



3 SV Models

ARCH-type models assume that the conditional volatility can be observed s steps ahead. However, a more realistic model for the conditional volatility can be based on modelling it having a predictable component that depends on past information and an unexpected noise. In this case, the conditional volatility is a latent unobserved variable, see Taylor (1986). One interpretation of the latent volatility is that it represents the arrival of new information into the market; see for instance Clark (1973). The main statistical properties of SV models have been reviewed by Taylor (1994), Ghysels et al. (1996) and Shephard (1996).

Usually, ε_t in (11) is specified to have a standard distribution so its variance σ_ε^2 is known. Thus for a normal distribution σ_ε^2 is one while for a Student-t distribution with ν degree of freedom it will be $\nu/(\nu-2)$. Following a convention often adopted in the literature we write $h_t \equiv \log(\sigma_t^2)$,

$$y_t = \sigma \varepsilon_t \exp\left(\frac{1}{2} h_t\right) \tag{18a}$$

where σ is a scale parameter, which removes the need for a constant term in the stationary first-order autoregressive process,

$$h_t = \phi h_{t-1} + \eta_t, \tag{18b}$$

where η_t is assumed to be a gaussian white noise process with zero mean and variance σ_η^2, independent of ε_t. The assumption that the conditional log-volatility has a Normal distribution has several supports in empirical studies for both exchange rate and stock returns; see for example Andersen et al. (2001) and (Andersen et al., 2001, 2003).

3.1 Description and Properties

The following properties of the ARSV model hold even if ε_t and η_t are contemporaneously correlated. First, y_t is a martingale difference. Second, stationarity of h_t implies stationarity of y_t. Third, if η_t is normally distributed, it follows from the properties of the lognormal distribution that $\mathbb{E}[\exp(h_t)] = \exp(\sigma_h^2/2)$, where $\sigma_h^2 = \sigma_\eta^2/(1-\phi^2)$ is the variance of h_t. Hence, if ε_t has a finite variance, the variance of y_t is given by,

$$\mathbb{V}(y_t) = \sigma^2 \sigma_\varepsilon^2 \exp\left(\sigma_h^2/2\right) \tag{19}$$

Similarly if the fourth order moment of ε_t exists, the kurtosis of y_t is $\kappa \exp(\sigma_h^2/2)$, where κ is the kurtosis of ε_t, so y_t exhibits more kurtosis than ε_t. Notice that if κ is finite, the condition for the existence of the kurtosis is the stationarity condition, i.e. $\phi < 1$. Therefore, as far as the model is stationary, the dynamic evolution of the volatility is not further restricted to guarantee the existence of the fourth order moment. However, it is also important to notice that given the kurtosis of returns, the parameter σ_η^2, that allows the volatility to evolve over time, should decrease as the autoregressive parameter, ϕ increases. Therefore, if the kurtosis is constant, the model approaches homoscedasticity as the persistence increases, see Harvey and Streibel (1998).

The ACF of squared observations, derived by Taylor (1986), is given by,

$$\rho_2(\tau) = \frac{\exp(\sigma_h^2 \phi^\tau) - 1}{\kappa \exp(\sigma_h^2) - 1}, \text{ for } \tau \geq 1 \tag{20}$$

If σ_h^2 is small and/or ϕ close to one, Taylor (1986) shows that the ACF in Equation (20) can be approximated by,

$$\rho_2(\tau) \simeq \frac{\exp(\sigma_h^2) - 1}{\kappa \exp(\sigma_h^2) - 1} \phi^\tau \tag{21}$$

The pattern of the approximated autocorrelations in Equation (21) is the same as for the autocorrelations of an ARMA(1,1) model. Consequently, this approximation has often been used to argue that the ACF of squares in GARCH(1,1) and ARSV(1) models are similar. However, in general, the approximated autocorrelations are larger than the true ones and, for small lags, τ, their rate of decay is smaller than the rate of decay of the autocorrelations in Equation (20). As an illustration, Figure 3 plots the true and approximated ACF of squares together with their rates of decay for three alternative ARSV(1) models with parameters ($\phi = 0.95$, $\sigma_\eta^2 = 0.1$), ($\phi = 0.98$, $\sigma_\eta^2 = 0.05$), and ($\phi = 0.99$, $\sigma_\eta^2 = 0.05$), respectively. This figure shows that the use of the approximation could lead to a very distorting picture of the shape of the true ACF. However, since $\alpha + \beta$ and ϕ can be interpreted as measures of

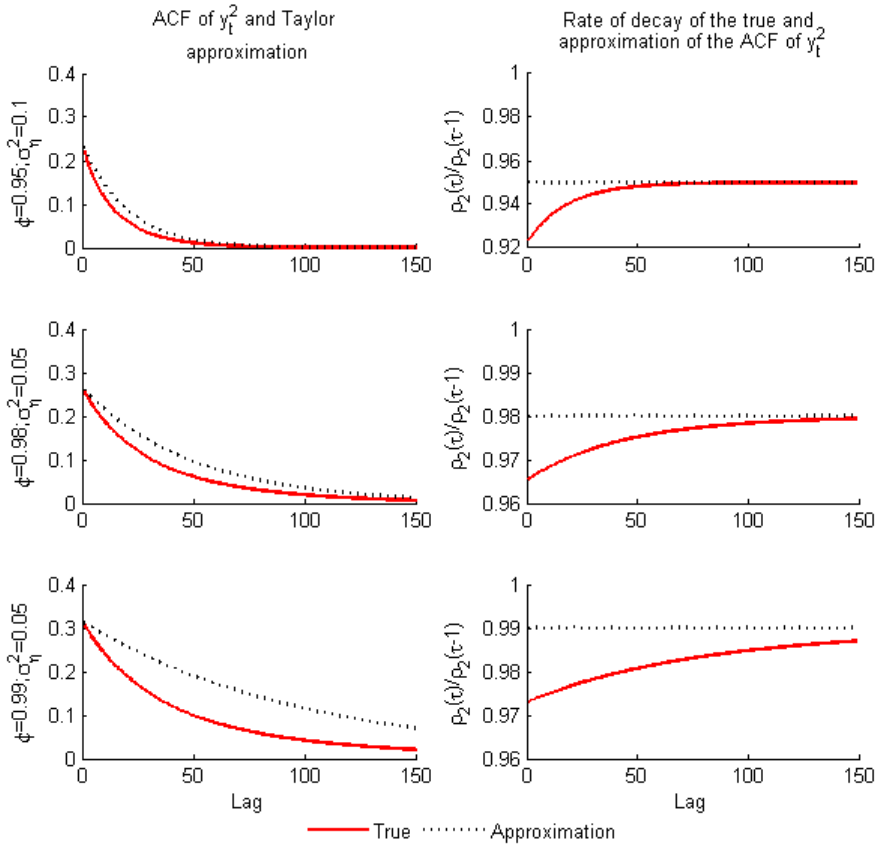


Figure 3: ACF of squared observations and rate of decay of such ACF for symmetric ARSV(1) models with gaussian innovations.

persistence in GARCH(1,1) and ARSV(1) models, respectively, many authors have compared their estimates, see Taylor (1994) and Shephard (1996) for instance. Following them and given that, as Figure 3 illustrates, the rate of decay of the autocorrelations in Equation (20) tends for long lags to ϕ, we also consider this parameter as a measure of the persistence of shocks to volatility in ARSV(1) models, see Kim and Shephard (1998) and Meyer and Yu (2000).

The relationship between kurtosis, persistence of shocks to volatility and first order autocorrelation of squares is different in GARCH and ARSV models. This relationship for a GARCH(1,1) and an ARSV(1) model is given by,

$$\rho_2(1) = \frac{\sqrt{\frac{(\kappa_y-\kappa)(1-p^2)}{(\kappa-1)\kappa_y}}\left[1-p^2+p\sqrt{\frac{(\kappa_y-\kappa)(1-p^2)}{(\kappa-1)\kappa_y}}\right]}{1-p^2+p\frac{(\kappa_y-\kappa)(1-p^2)}{(\kappa-1)\kappa_y}} \tag{22a}$$

where p is the persistence in GARCH(1,1) models, i.e. $\alpha+\beta$, and by

$$\rho_2(1) = \frac{\left(\frac{\kappa_y}{\kappa}\right)^{\phi}-1}{\kappa_y-1}. \tag{22b}$$

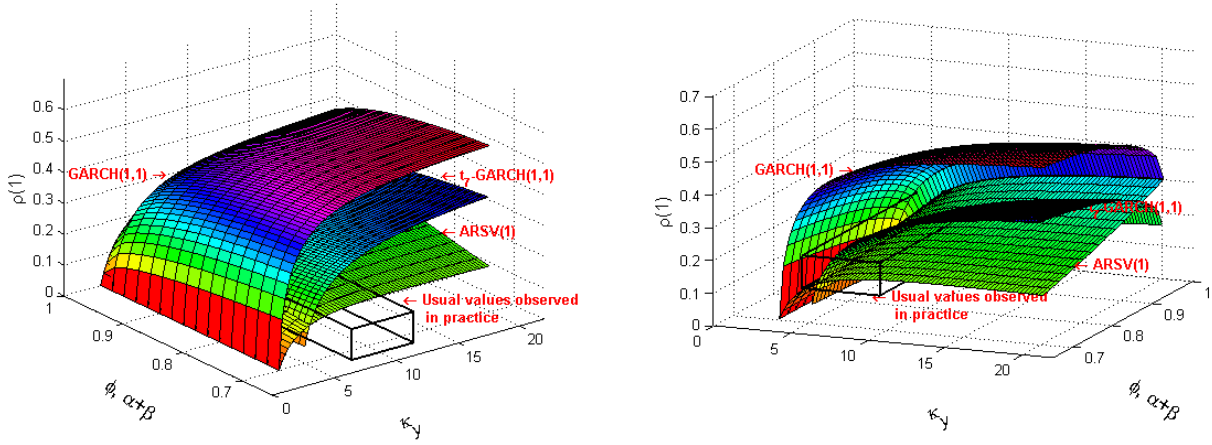


Figure 4: Relationship between kurtosis, first-order autocorrelation of squared observations, and persistence for GARCH(1,1) and ARSV(1) models with gaussian errors. Source: Carnero et al. (2004)

This relationship is plotted in Figure 4 for the Normal ARSV(1) model, together with the corresponding relationship for the Normal GARCH(1,1) and a Student-GARCH(1,1) with 7 degrees of freedom models. Several conclusions can be drawn from this figure. First, it is rather clear that ARSV(1) models are able to generate series with simultaneously higher kurtosis and lower $\rho_2(1)$ than the GARCH(1,1) model for a larger range of values of the persistence. Second, it is possible to observe that, for the same kurtosis and persistence, the autocorrelations of squared observations implied by the Normal ARSV(1) model are smaller than the autocorrelations of the Normal GARCH(1,1) model, except when the volatility approaches the non-stationary region, where both autocorrelations are the same. Furthermore, in ARSV(1) models, for a given kurtosis, the first-order autocorrelation of squares increases with the persistence parameter, ϕ, while in GARCH(1,1) models the first-order autocorrelation decreases with $\alpha + \beta$. Consequently it could be expected that, for a given series of returns with a given kurtosis, if the first-order autocorrelation of squares is small, the persistence estimated in ARSV(1) models is usually lower than in GARCH(1,1) models; see the values of κ_y and $\rho_2(1)$ within the box usually observed in empirical applications. Therefore it is possible to have ARSV(1) models with high kurtosis, low $\rho_2(1)$, and persistence far from the non-stationary region, while in a Normal GARCH(1,1) model, the persistence would be high because it is the only way to achieve both high kurtosis and low $\rho_2(1)$.

Alternatively, as we mentioned before, the GARCH model can generate higher kurtosis without increasing the first-order autocorrelation of squares by having a conditional heavy-tailed distribution. For example, comparing the surfaces of the relationship between κ_y and $\rho_2(1)$ and persistence for the Student- GARCH(1,1) model with 7 degrees of freedom and the

Normal ARSV(1) models, it is possible to observe that these two models are able to generate series with values of the kurtosis and $\rho_2(1)$ within the region defined by the box, where the typical values would be found. Therefore both models may have similar fits when implemented to represent the dynamic evolution of volatility of the same real-time series, see, for instance, Shephard (1996). However, remember that the restrictions on the α parameter of the Student-GARCH model, needed to guarantee that the fourth-order moment is finite, restrict severely the dynamic evolution of the volatility.

3.2 Estimation

Given that the conditional distribution of y_t is, in general, far from the Normality, then the Maximum Likelihood (ML) estimator cannot be obtained by traditional methods. This is the reason why there is a large list of alternative methods to estimate SV models. Next, we describe some of the more promising from an empirical point of view.

3.2.1 Method of Moments

These methods have the difficulty that their efficiency depends on the choice of moments. Furthermore, a particular distribution needs to be assumed for ε_t. Finally, its efficiency is reduced as the process approaches the non-stationarity as it is often the case in the empirical application to financial returns. Melino and Turnbull (1990) proposed to estimate the stationary SV models by Generalized Method of Moments (GMM). However, there is any closed form or criteria to choice the unconditional moments that can be computed, therefore this procedure is relatively inefficient. The Efficient Method of Moments (EMM) seeks efficiency improvements, while maintaining the general flexibility of GMM, by letting the data guide the choice of an auxiliary quasi-likelihood which serves to generate an efficient set of moments; see Andersen, Andersen et al. (1999).

3.2.2 Quasi-Maximum Likelihood (QML) principle

The QML estimator was independently proposed by Nelson (1988) and Harvey et al. (1994) and is based on the Kalman filter.

Model (18) is nonlinear, however, transforming y_t by taking logarithms of the squares we

obtain the following system,

$$y_t^* \equiv \log(y_t^2) = \gamma + h_t + \xi_t, \tag{23a}$$

$$h_t = \phi h_{t-1} + \eta_t, \tag{23b}$$

where $\gamma = \log(\sigma^{*2}) + \mathbb{E}[\log(\varepsilon_t^2)]$, $\xi_t = \log(\varepsilon_t^2) - \mathbb{E}[\log(\varepsilon_t^2)]$. ξ_t is a non-Gaussian zero mean white noise process, and its statistical properties depend on the distribution of ε_t. If, for example, $\varepsilon_t \sim N(0,1)$, as it is often assumed in the literature, then, $\mathbb{E}[\log(\varepsilon_t^2)] \approx -1.27$ and $\mathbb{V}[\log(\varepsilon_t^2)] = \pi^2/2$. Moreover, when ε_t is assumed to be Student-t, then $\log(\varepsilon_t^2)$ has also a determined distribution.

System (23) can be thought as a linear SS model. As it has been mentioned above, the distribution of ξ_t is far from the Normality, however, we can still use the Kalman filter for construct the one-step ahead innovations v_t and use them for calculate the likelihood function as ξ_t were Normal. Due to the fact that the true distribution of ξ_t is non-Normal the QML estimator is inefficient, in spite of that, Ruiz (1994) shows that the QML estimator is consistent and asymptotically Normal. The log-likelihood for the ARSV(1) model is given by

$$\log(L) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T} F_t - \frac{1}{2}\sum_{t=1}^{T}\frac{v_t^2}{F_t}, \tag{24}$$

where $F_t = P_{t|t-1} + \sigma_\xi^2$, $v_t = \log(y_t^2) - m_{t|t-1}$ and $m_{t|t-1}$ is the one-step ahead estimates of the log volatility h_t. Table 1 shows the asymptotic standard errors (ASEs) of the GMM and the QML estimator, together with the square root of the relative efficiency of the GMM estimator with respect to the QML estimator. Comparing the ASEs of both estimators, it is possible observe that, in terms of efficiency, the QML estimation method performs better when ϕ is close to one and σ_η^2 is relatively big. On the other hand, when σ_η^2 and ϕ are small, the GMM estimator is more efficient. This result could be expected, because when σ_η^2 and ϕ are small, the variance of the log volatility process, given by $\sigma_\eta^2/(1-\phi^2)$, is very small in relation to the variance of ξ_t. Therefore, the transformation $\log(y_t^2)$ is dominated by ξ_t, and the approximation to normality used by the Kalman filter is very poor. However, in most empirical application with very high frequency financial time series (hourly or daily), it has been observed that the parameter ϕ is very close or exactly one. The estimated values of σ_η^2 are usually between 0.01 and 2.77; see Taylor (1994). With this range of parameter values, there is a little doubt about the better performance of the QML estimator. However, note that there exists an important drawback with the QML estimator. As it has been mentioned above, in most empirical applications with very high frequency financial time series, the parameter ϕ is very close or exactly one. For this range of values of the parameter ϕ the log-likelihood function tends to have a flat shape close

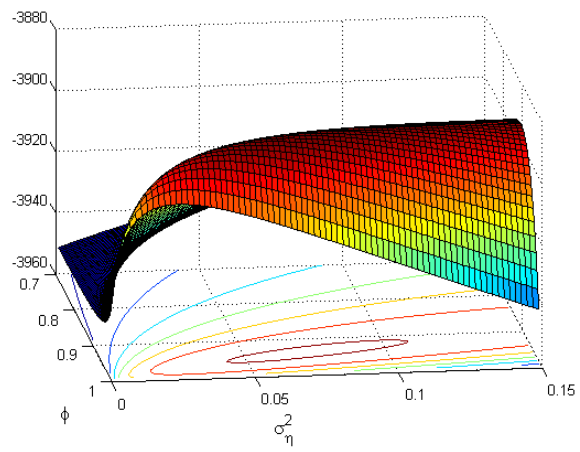
Table 1: Asymptotic standard deviations of the QML and GMM estimator. Source: Ruiz (1994)

	QML	GMM	(Rel. Eff.)$^{1/2}$
$\phi = 0.9$	0.68	12203.65	0.00
$\sigma_\eta^2 = 1$	6.45	28903.37	0.00
$\phi = 0.7$	2.31	29.17	0.08
$\sigma_\eta^2 = 1$	10.74	20.13	0.53
$\phi = 0.9$	1.77	3.77	0.47
$\sigma_\eta^2 = 0.09$	2.24	0.86	2.60
$\phi = 0.7$	15.12	13.94	1.08
$\sigma_\eta^2 = 0.09$	7.62	1.02	7.47
$\phi = 0.95$	0.92	3.00	0.31
$\sigma_\eta^2 = 0.04$	0.95	0.61	1.56
$\phi = 0.97$	0.47	3.54	0.13
$\sigma_\eta^2 = 0.04$	0.69	1.17	0.59
$\phi = 0.99$	0.18	74.36	0.00
$\sigma_\eta^2 = 0.04$	0.47	73.98	0.01

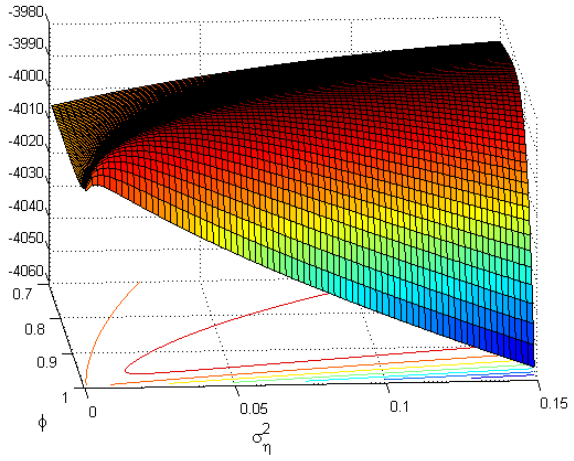
to the optimum point, therefore, sometimes, the algorithm used to optimize the function may not converge to the optimum point. Figure 5 shows the log-likelihood function for different values of ϕ and σ_η^2. It is possible to observe in the figure that, when ϕ is close to one or σ_η^2 is close to zero the shape of the function become more flat close to the maximum. First, when ϕ is close to one the process is close to the non-stationarity then it should be treated as a random walk. Second, when σ_η^2 is close to zero the process is close to the homoscedasticity.

3.2.3 Bayesian Markov Chain Monte Carlo (MCMC) methods

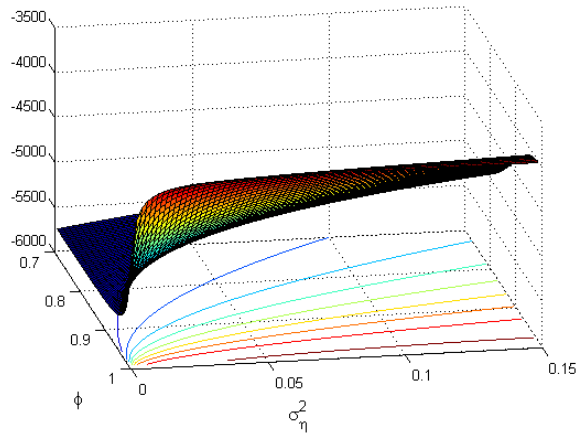
We have already commented that estimation procedures for SV models have grown significantly in the last decades. The increasing use of computer intensive programs based on the development of more powerful computers was one the main factors explaining this growth. Among the *frequentist* methods briefly described above, Bayesian inference has also found its place. In particular Markov Chain Monte Carlo (MCMC) procedures for computing posterior distributions of the parameters have been extensively studied and used in many empirical applications. As Meyer and Yu (2000) point out, the difficulties of multidimensional numerical integration involved in posterior calculations have been overcome with the development of MCMC techniques, as in Jacquier et al. (1994). Together with the development of classical *frequentist* and Bayesian estimation procedures, a vast literature has been devoted to compare the statistical properties of the estimators; see, for instance, Ruiz and Broto (2004). In



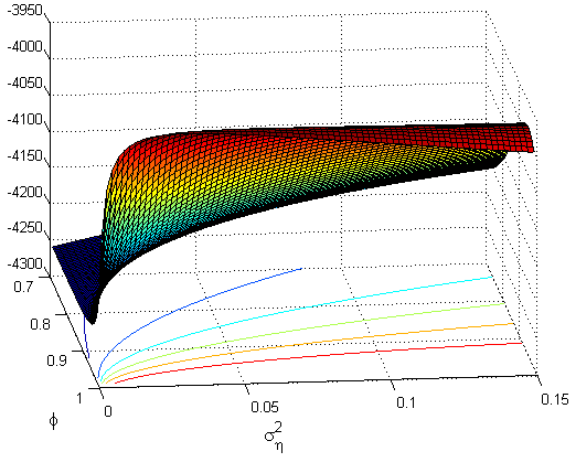
$\phi = 0.9$ and $\sigma_\eta^2 = 0.1$



$\phi = 0.9$ and $\sigma_\eta^2 = 0.04$



$\phi = 0.99$ and $\sigma_\eta^2 = 0.1$



$\phi = 0.99$ and $\sigma_\eta^2 = 0.04$

Figure 5: Log-Likelihood of the Autoregressive Stochastic Volatility Model for different values of parameter ϕ and σ_η^2

general, results are not very much conclusive but in any case, Bayesian-based estimators have good properties, specially in relatively small samples. Moreover, with the availability of special software devoted to calculate numerical integration, such as BUGS (Bayesian Analysis Using Gibbs Sampling)[2], the calculation is no longer a disadvantage of Bayesian estimators.

The following development will be based mainly on Meyer and Yu (2000), who make a detailed description of Bayesian numerical methods for estimating ARSV models. In particular, the use of Markov Chain Monte Carlo (MCMC) methods and the Gibbs sampler. First, let's

[2]BUGS is available free of charge from http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml

redefined the ARSV(1) given in (18) in the following way:

$$y_t|h_t = \varepsilon_t \exp\left(\frac{1}{2}h_t\right), \quad \varepsilon_t \sim N(0,1) \tag{25a}$$

$$h_t|h_{t-1},\omega,\phi,\sigma_\eta^2 = \omega + \phi(h_{t-1}-\omega) + \eta_t, \quad \eta_t \sim N(0,\sigma_\eta^2) \tag{25b}$$

where $h_0 \sim N(\omega, \sigma_\eta^2)$. Note that ω is now the scale parameter. A full Bayesian model for (25) consists of the joint prior distribution of all unobservable parameters, i.e. ω, ϕ, σ_η^2 and h_t, and the observables, y_t. Bayesian inference is then based on the posterior distribution of the unobservables given the data. From now on, the probability density distribution of a random variable x will be denoted by $p(x)$. By successive conditioning, the joint prior density is given by

$$p(\omega,\phi,\sigma_\eta^2,h_0,h_1,\ldots,h_T) = p(\omega,\phi,\sigma_\eta^2)\prod_{t=1}^{T} p(h_t|h_{t-1},\omega,\phi,\sigma_\eta^2). \tag{26}$$

It is usually assumed in the literature that the parameters have independent priors, i.e. $p(\omega,\phi,\sigma_\eta^2)$ is equal to the product of the three marginal densities. For these priors it is commonly assumed that

- $\omega \sim N(0,20)$ (a slightly informative prior)
- $\phi = 2\phi^* - 1, \;\; \phi^* \sim Beta(20, 1.5)$ (which gives a prior mean for ϕ around 0.86)
- $\sigma_\eta^2 \sim Gamma(0.25, 0.25)$ (which gives a prior mean and standard deviation of around 0.05 and 0.13, respectively)

On the other hand, the conditional probability distribution is straightforwardly found to be $p(h_t|h_{t-1},\omega,\phi,\sigma_\eta^2) \sim N(\omega+\phi(h_{t-1}-\omega),\sigma_\eta^2)$ from (25b) and the likelihood is specified according to the observation equation (25a) as

$$p(y_1,y_2,\ldots,y_T|\omega,\phi,\sigma_\eta^2,h_0,h_1,\ldots,h_T) = \prod_{t=1}^{T} p(y_t|h_t). \tag{27}$$

Then, by Bayes' theorem, the joint posterior distribution of the unobservables given the data is proportional to the prior times the likelihood, i.e.:

$$\begin{aligned} p(\omega,\phi,\sigma_\eta^2,h_0,h_1,\ldots,h_T|y_1,y_2,\ldots,y_T) \propto\; & p(\omega)\times p(\phi)\times p(\sigma_\eta^2)\times p(h_0|\omega,\sigma_\eta^2)\times \\ & \prod_{t=1}^{T} p(h_t|h_{t-1},\omega,\phi,\sigma_\eta^2)\times\prod_{t=1}^{T} p(y_t|h_t), \end{aligned} \tag{28}$$

where $\propto$ indicates "proportional to". Unfortunately, although the distributions of the right hand side of (28) are easily derived, it is not sufficient to obtain the full posterior probability. It is also necessary to calculate the normalization constant to make (28) an identity. A general technical difficulty encountered in any application of Bayesian inference is how to find this constant because it often demands the numerical resolution of an integral of high dimension (in this case is T+4: T+1 unobserved log-volatilities, and the three parameters). However, the Gibbs sampler, a special MCMC algorithm, generates a sample of the posterior (28) by iteratively sampling from each of the univariate full conditional distributions. These univariate distributions can be easily constructed from the joint posterior distribution by conditioning properly. Let's define $\theta = (\omega, \phi, \sigma_\eta^2)'$ and $h_{-t} = (h_0, h_1, \ldots, h_{t-1}, h_{t+1}, \ldots, h_T)'$. Then, the Gibbs algorithm can be summarized in the following set of steps:

Step 1 Set $i = 0$ and initialize θ and h_t, $t = 0, 1, \ldots, T$.

Step 2 Sample h_t from $p(h_t|h_{-t}, \theta, y_1, y_2, \ldots, y_T)$ for $t = 1, \ldots, T$.

Step 3 Sample θ from $p(\theta|h_0, h_1, \ldots, h_T, y_1, y_2, \ldots, y_T)$.

Step 4 Update the counter i, go back to Step 2 and continue.

The algorithm above is usually quoted a single-move sampler because the components of h are updated one by one sequentially. Each iteration of the algorithm consists of updating T components of h, and it will run a large number N of iterations before the sampler terminates.

Step 3 follows the standard Bayesian linear regression model theory that starts from a non-informative conjugate prior for θ

$$p(\omega, \phi, \sigma_\eta^2) \quad \propto \quad \sigma_\eta^{-2} \tag{29}$$

and observing that θ and y are conditionally independent given h, the posterior for θ given h is given by

$$p(\theta|h) \quad = \quad N_2\left(\mu, (H'H)^{-1}\sigma_\eta^2\right) \; IG\left((T-2)/2, Ts^2/2\right), \tag{30}$$

where N_k denotes the k-variate Normal distribution, and IG the inverse gamma distribution; $\mu = (H'H)^{-1}H'h_{2,T}$ with $h_{2,T} = (h_2, h_3, \ldots, h_T)'$, H is a $(T-1) \times 2$ matrix whose row t is $(1, h_t)$, $t = 1, 2, \ldots, T-1$, and Ts^2 is the standard sum of squared errors. This comes directly from the fact that we are regressing h_t versus h_{t-1} like in Gaussian AR(1) models.

In Step 2, we need to simulate, for each t, from the single-site conditional distribution

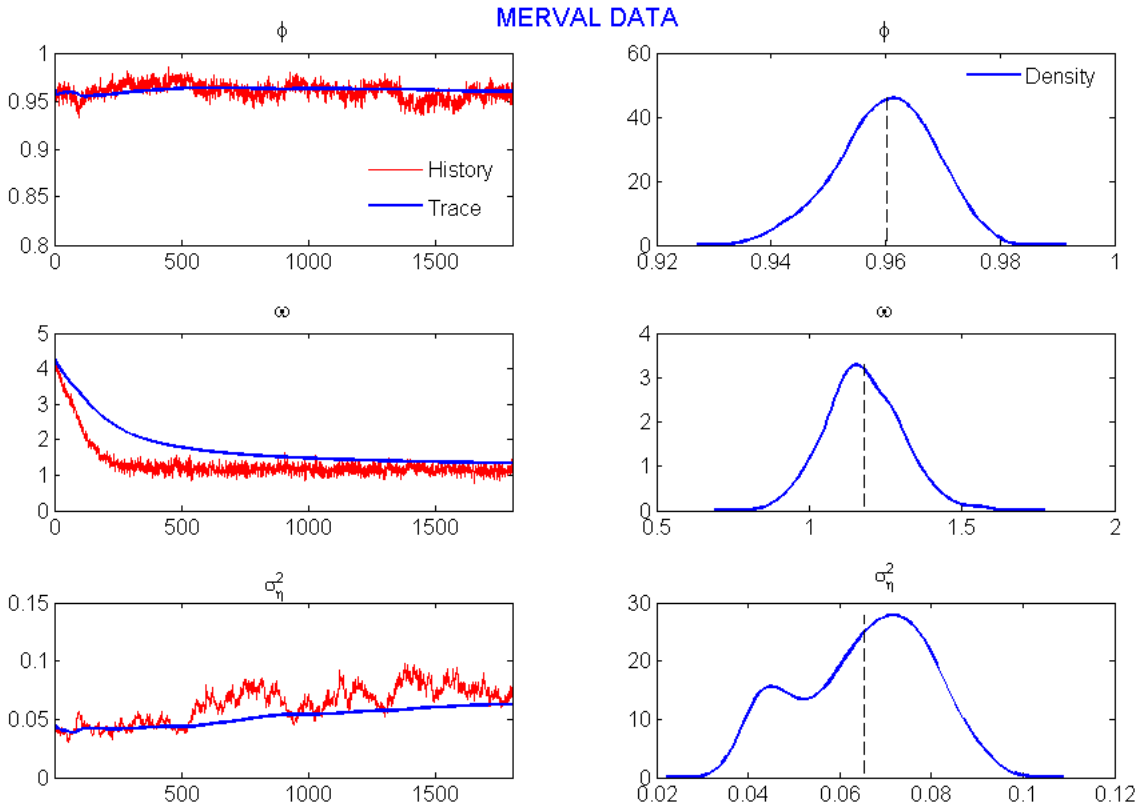


Figure 6: History, trace and kernel estimates of the posterior draws of the SV parameters for the MERVAL returns.

$p(h_t|h_{-t}, \theta, y_1, y_2, \ldots, y_T)$, expressed as

$$p(h_t|h_{-t}, \theta, y_1, y_2, \ldots, y_T) \quad \propto \quad \exp\left\{-\frac{1}{2}\left[h_t + y_t^2/e^{h_t} + (h_t - m_t)^2/b\right]\right\}, \tag{31}$$

where $m_t = [\omega + \phi(h_{t-1} + h_{t+1})]/(1 + \phi^2)$ and $b = \sigma_\eta^2/(1 + \phi^2)$. Applying the Metropolis algorithm locally at each t to (31), Jacquier et al. (1994) propose an inverse gamma proposal density q to approximate the true probability. Other proposal densities such as the Normal distribution have been suggested in the literature; see Kim and Shephard (1998). However, the great deal of this method is how to find the posterior in (31) in an efficient way. Meyer and Yu (2000) state that perhaps a combination of single and block-move sampling strategy may be suitable for ARSV models, specially when the parameter vector is close to the boundary of the parameter space (i.e close to non-stationarity and/or homoscedasticity). BUGS software is able to do this in a relatively easy way. Another drawback of the method described above is the high correlation that appears between consecutive draws in a given chain. This implies that we should either enlarge the chains or pick s-consecutive draws, with s generally between 5 and 20. Figure **??** shows some plots of the posterior draws of ϕ, ω and σ_η^2 for the MERVAL returns.

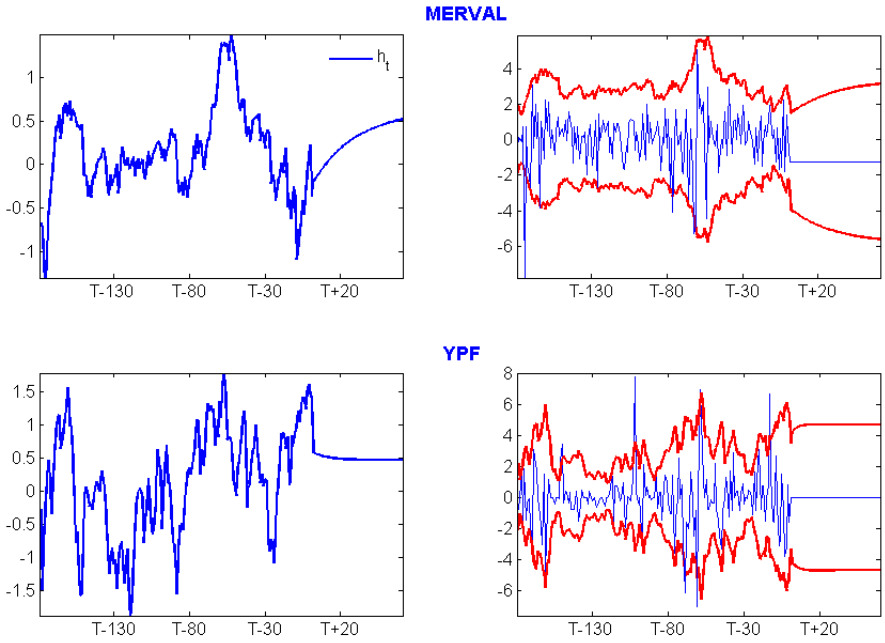


Figure 7: Forecasts of the MERVAL and YPF returns two months ahead

3.3 Prediction

Prediction in simple SV models is almost straightforward with either the quasi maximum likelihood or Bayesian MCMC approach. In the former, once the unobserved volatility, h_t have been estimated the Kalman Filter (or Smoother) provides k-step-ahead forecasts by simply noting that

$$h_{T+k|T} = \phi h_{T+k-1|T}, \quad k = 1, 2, \ldots \tag{32}$$

where $h_{T|T}$ is directly given by the Kalman Filter applied to the whole sample, and the. The k step volatility forecast of y_t for the SV model is calculated by means of the log-normal properties and it is given by

$$E(y^2_{T+1}|T) = \sigma^2 \exp\left(h_{T+1|T} + 0.5\, p_{T+1|T}\right) \tag{33a}$$

$$E(y^2_{T+k}|T) = E(y^2_{T+1}|T) + \sigma^2 \sum_{j=2}^{k} \exp\left[\phi^{j-1} h_{T+1|T} + 0.5\left(\phi^{2(j-1)} p_{T+1|T} + \sum_{i=0}^{k-2} \phi^{2i} \sigma^2_\eta\right)\right], \; k > 1 \tag{33b}$$

where $p_{T+1|T}$ is the variance of $h_{T+1|T}$ and it is also a by-product of the Kalman Filter; see Hol et al. (2004) for further details. As an illustration, Figure 7 shows the 3-months-ahead forecasts of the MERVAL and YPF returns.

Under the Bayesian MCMC framework, forecasting is straightforward once we have obtained the posteriors of each parameter and the volatilities. One strategy is simply to obtain the mean of each posteriors and work as in the QML case. However, it seems more interesting to use all the information to find the distribution of the s-steps-ahead forecast of both, h and y. Following Jacquier et al. (1994), it is necessary first to compute the predictive density of a vector of future volatilities given the sample data, $p(h_f|y)$, where $h_f = (h_{T+1}, \ldots, h_{T+k})'$. To achieve it, the authors compute the joint posterior distribution of the h vector including both the sample and future $h's$ and simply marginalize on the future values. Finally, after obtaining $h_f|y$, they find the values of $y_f|h_f, \theta$ easily obtained from $i.i.d.$ normal draws. With these values, it is possible to obtain not only the point and interval predictions, but also the entire prediction density for each k.

4 Extensions

There is a broad variety of different SV models in the literature that go beyond the simple ARSV(1) considered in this course. These models have also been constructed, tested and applied in many financial time series. We will only review how these models deal with asymmetry (remember, one of the empirical stylized facts we have seen at first) and the multivariate case. For a deep and exhaustive survey of asymmetric SV models, see Yu (2005), and for multivariate SV models, see Yu and Meyer (2006); Asai et al. (2006).

4.1 Asymmetry

The usual claim in finance is that when there is bad news, which decreases the price and hence increases the debt-to-equity ratio (i.e. financial leverage), it makes the firm riskier and tends to increase future expected volatility. As a result, the leverage effect must correspond to a negative relationship between volatility and price/return. The empirical evidence of this leverage effect regards as the asymmetric response in the volatility when there is bad news (stronger response) or good news (weaker response). In our examples, we see the MERVAL having a negative skewness. Harvey and Shephard (1996) derive a model that captures this leverage effect within the SV framework. They show that by allowing the noises ε_t and η_t in (18) to be mutually correlated. This simple change to the basic SV model proves to be useful for capturing the leverage effect in asset returns. In order to estimate this model, the authors use the QML approach of an SS model that changes a bit with respect to the one considered previously.

For other alternative ways of modelling the leverage effect in SV models, see Yu (2005).

4.2 Multivariate SV Models

As Asai et al. (2006) point out, the Euler scheme of time diffusion models for a vector of m continuous time log-prices, S_t, leads to the discrete time version of the Multivariate stochastic volatility model

$$y_t = H_t^{1/2}\varepsilon_t, \quad \varepsilon_t \sim N(0, I_m) \tag{34a}$$

$$f(\text{vech}(H_t))) = a(\text{vech}(H_{t-1})) + f(\text{vech}(H_{t-1})) + b(\text{vech}(H_{t-1}))\eta_{t-1}, \quad \eta_t \sim N(0, \Sigma_\eta) \tag{34b}$$

where $y_t = S_t - S_{t-1}$, $vech(.)$ is the operator that stacks each column of the lower triangular matrix, and f, a, b are all known functions. Without further restrictions, H_t may not be positive definite, and thus the validity of a well-defined econometric model is not guaranteed. The paper of Harvey et al. (1994) is seminal in this area because the authors specify a multivariate version of the SV model by imposing (34) to be of the form

$$y_t = H_t^{1/2}\varepsilon_t, \tag{35a}$$

$$H_t^{1/2} = \text{diag}\{\exp(h_{1t}/2), \ldots, \exp(h_{mt}/2)\} = \text{diag}\{\exp(h_t/2)\}$$

$$h_{t+1} = \mu + \phi \circ h_t + \eta_t, \tag{35b}$$

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} P_\varepsilon & O \\ O & \Sigma_\eta \end{pmatrix} \right],$$

where $h_t = (h_{1t}, \ldots, h_{mt})'$ is an $m \times 1$ vector of unobserved log-volatility, μ and ϕ are $m \times 1$ parameter vectors, the operator $\circ$ denotes the Hadamard (or element-by-element) product, Σ_η is a positive definite covariance matrix, and $P_\varepsilon = \{\rho_{ij}\}$ is the correlation matrix, that is, P_ε is a positive definite matrix with $\rho_{ii} = 1$ and $|\rho_{ij}| < 1$ for any $i \neq j$, $i, j = 1, \ldots, m$. It can be seen that the positive definiteness of H_t and parsimony are achieved by restricting H_t to be a diagonal matrix, with the diagonal element following the exponential function of a Gaussian vector AR(1) process. The number of parameters in the basic model is $2m + m^2$. In Asai et al. (2006), a great number of other more complex multivariate SV models are reviewed. For example, it is possible to account for leverage effects, factor models, or time-varying correlation models. Of course, the complexity increases significantly while dealing with large m, and so the estimation procedures to make inference from them. In any case the QML or Bayesian MCMC approaches are still valid as estimation procedures; see Yu and Meyer (2006) for a review of them, focusing mainly in Bayesian inference.

As an illustration, Table 2 shows the estimation results of a bivariate model of the two returns considered in this course, using the Bayesian MCMC of BUGS.

	Coefficient	Std. Error	z-Statistic	p-value.
ϕ(M)	0.9302	0.0111	83.8881	0.0000
σ_η^2(M)	0.1627	0.0147	11.1058	0.0000
ϕ(Y)	0.9689	0.006677	145.11	0.0000
σ_η^2(Y)	0.0476	0.0054	8.8895	0.0000
ρ_ε(M,Y)	0.8354	0.1067	7.8291	0.0000
ρ_η(M,Y)	0.5272	0.1234	4.2726	0.0000
Log likelihood	-15229.9	Akaike info criterion		9.2060
Parameters	6	Schwarz criterion		9.2170
Diffuse priors	0	Hannan-Quinn criter.		9.2099

Table 2: Estimations of the Bivariate SV model of Harvey et al. (1994) fitted to the Merval and YPF returns.

References

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association 96*, 42–55.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001). The distribution of realized stock return volatility. *Journal of Financial Economics 61*, 43–76.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003). Modelling and forecasting realized volatility. *Econometrica 71*, 579–625.

Andersen, T. G., H.-J. Chungb, and B. E. Sorensen (1999). Efficient method of moments estimation of a stochastic volatility model: A monte carlo study. *Journal of Econometrics 91*, 61–87.

Asai, M., M. McAleer, and J. Yu (2006). Multivariate stochastic volatility: A review. *Econometric Reviews 25*, 145–175.

Ball, L., S. G. Cecchetti, and R. J. Gordon (1990). Inflation and uncertainty at short and long horizons. *Broking Papers on Economic Activity 1990*, 215–254.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics 51*, 307–327.

Bollerslev, T., R. Y. Chou, and K. F. Kroner (1992). ARCH modelling in finance: A selective review of the theory and empirical evidence. *Journal of Econometrics 52*, 5–59.

Bollerslev, T., R. F. Engle, and D. B. Nelson (1994). ARCH models. *The Handbook of Econometrics 2*, 2959–3038.

Box, G. E., G. M. Jenkins, and G. C. Reinsel (1994). *Time Series Analysis: forecasting and control*. Englewood Cliffs.

Broto, C. and E. Ruiz (2006). Using auxiliary residuals to detect conditional heteroscedasticity in inflaction. *Manuscript, Universidad Carlos III de Madrid*.

Carnero, M. A., D. Peña, and E. Ruiz (2004). Persistence and kurtosis in GARCH and stochastic volatility models. *Journal of Financial Econometrics 2*, 319–342.

Cavaglia, S. (1992). The persistence of real interest differentials: A Kalman filtering approach. *Journal of Monetary Economics 29*, 429–443.

Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica 41*, 135–155.

Durbin, J. and S. J. Koopman (2001). *Time Series Analysis by State Space Methods*. New York: Oxford University Press.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. *Econometrica 50*, 987–1008.

Evans, M. (1991). Discovering the link between inflation rates and inflation uncertainty. *Journal of Money, Credit and Banking 23*, 169–184.

Ghysels, E., A. C. Harvey, and E. Renault (1996). Stochastic volatility. In G. S. Maddala and C. R. Rao (Eds.), *Handbook of Statistics*, Volume 14, Amsterdam. North-Holland.

Harvey, A. and M. Streibel (1998). Testing for a slowly changing level with special reference to stochastic volatility. *Journal of Econometrics 87*, 167–189.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

Harvey, A. C., E. Ruiz, and N. G. Shephard (1994). Multivariate stochastic variance models. *Review of Economics Studies 61*, 247–264.

Harvey, A. C. and N. Shephard (1996). Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business and Economic Statistics 14*, 429434.

Hol, E., S. J. Koopman, and B. Jungbacker (2004, August). Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. Computing in Economics and Finance 2004 342, Society for Computational Economics. available at http://ideas.repec.org/p/sce/scecf4/342.html.

Jacquier, E., N. G. Polson, and P. E. Rossi (1994). Bayesian analysis of stochastic volatility models. *Journal of Business and Economics Statistics 12*, 371–389.

Kim, C. J. (1993). Unobserved-component time series models with markov-switching heterskedasticity: Changes in regimen and the link between inflation rates and inflation uncertainty. *Journal of Business and Economic Statistics 11*, 341–349.

Kim, S. and N. Shephard (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *The Review of Economic Studies 65*, 361–393.

Melino, A. and S. M. Turnbull (1990). Pricing foreign currency options with stochastic volatility. *Journal of Econometrics 45*, 239–265.

Meyer, R. and J. Yu (2000). BUGS for a bayesian analysis of stochastic volatility models. *The Econometrics Journal 3*, 198–215.

Nelson, D. B. (1988). The time series behavior of stock market volatility and returns. *Unpiblished Ph D Dissertation Massachusetts Institute of Technology, Cambridge, MA*.

Ruiz, E. (1994). Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics 63*, 289–306.

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

Shephard, N. G. (1996). Statistical aspects of ARCH and Stochastic Volatility. In D. R. Cox, D. V. Hinkley, and O. E. Barndoff-Nielsen (Eds.), *Time Series Models in Econometrics, Finance and Other Fields*, London. Chapman & Hall.

Taylor, S. J. (1986). *Modelling Financial Time Series*. New York: John Wiley & Sons.

Taylor, S. J. (1994). Modelling stochastic volatility: A review and comparative study. *Mathematical Finance 4*, 183–204.

Yu, J. (2005, August). On leverage in a stochastic volatility model. *Journal of Econometrics 127*(2), 165–178. available at http://ideas.repec.org/a/eee/econom/v127y2005i2p165-178.html.

Yu, J. and R. Meyer (2006). Multivariate stochastic volatility models: Bayesian estimation and model comparison. *Econometric Reviews 25*, 361–384.