Conditional Variance Dynamics: GARCH & Stochastic Volatility

Md. Nazmul Ahsan [1] Jean-Marie Dufour [2]

[1]Concordia University and CIRANO

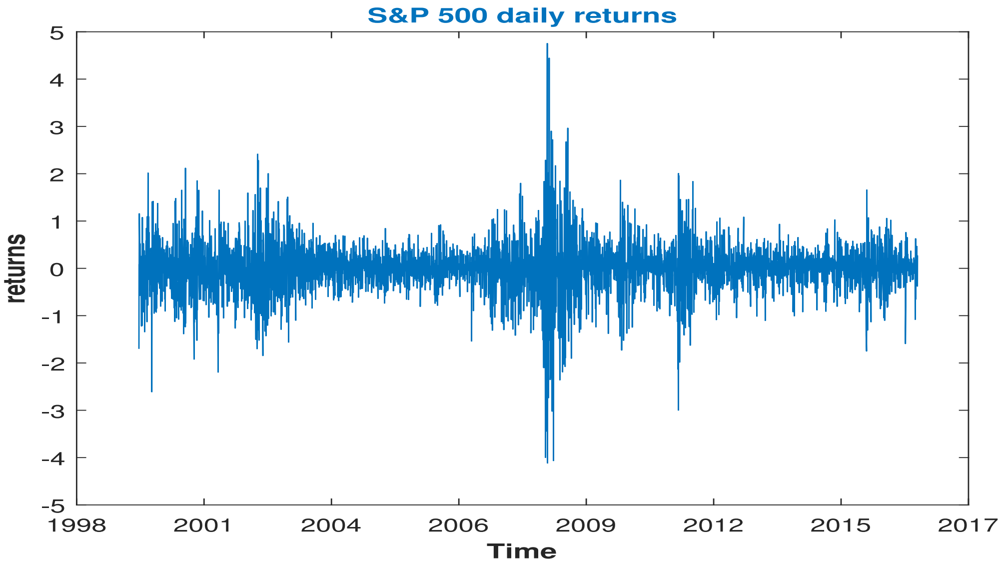
[2]McGill University, CIREQ and CIRANO

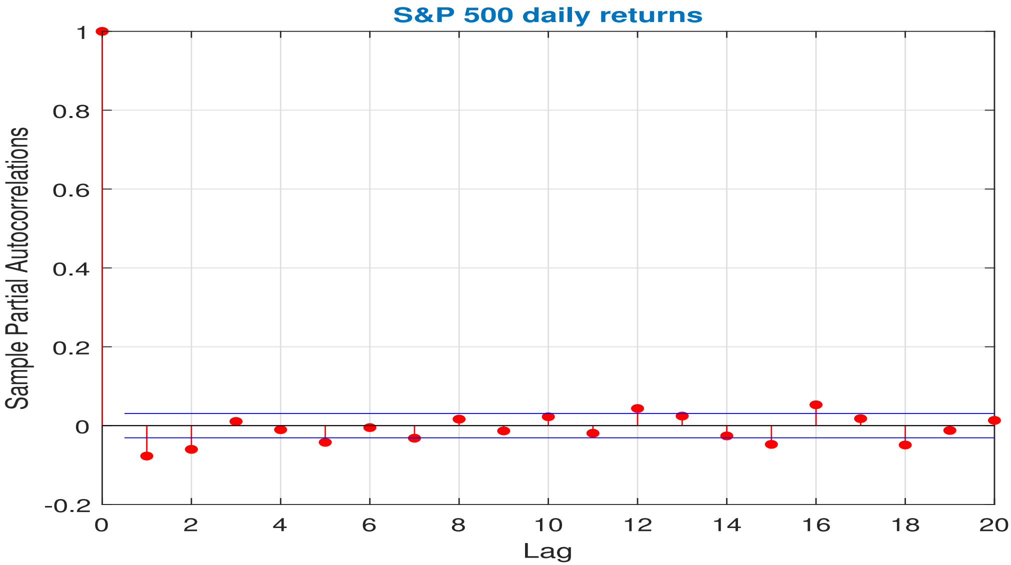
ECON 763: Financial Econometrics

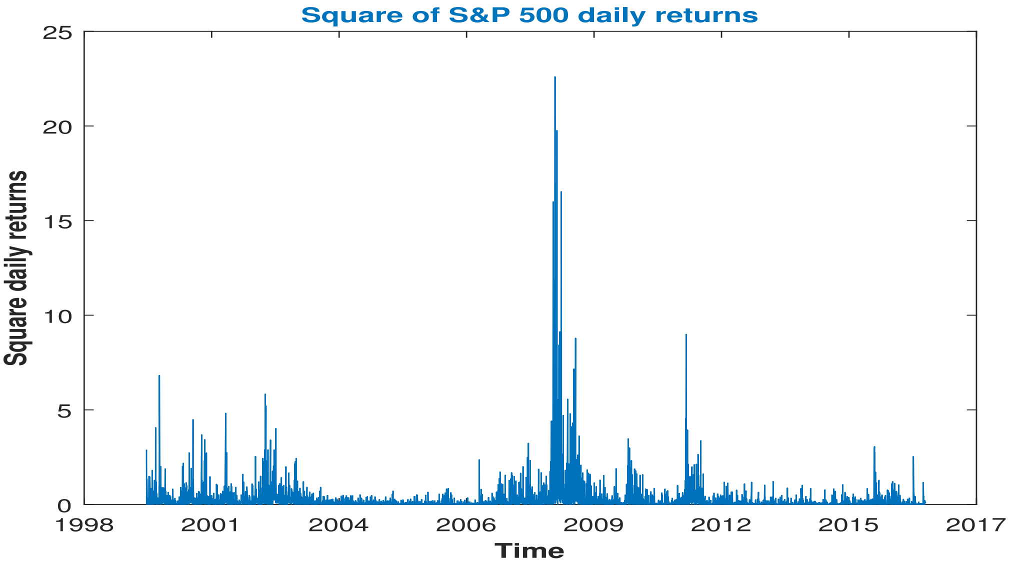
Outline

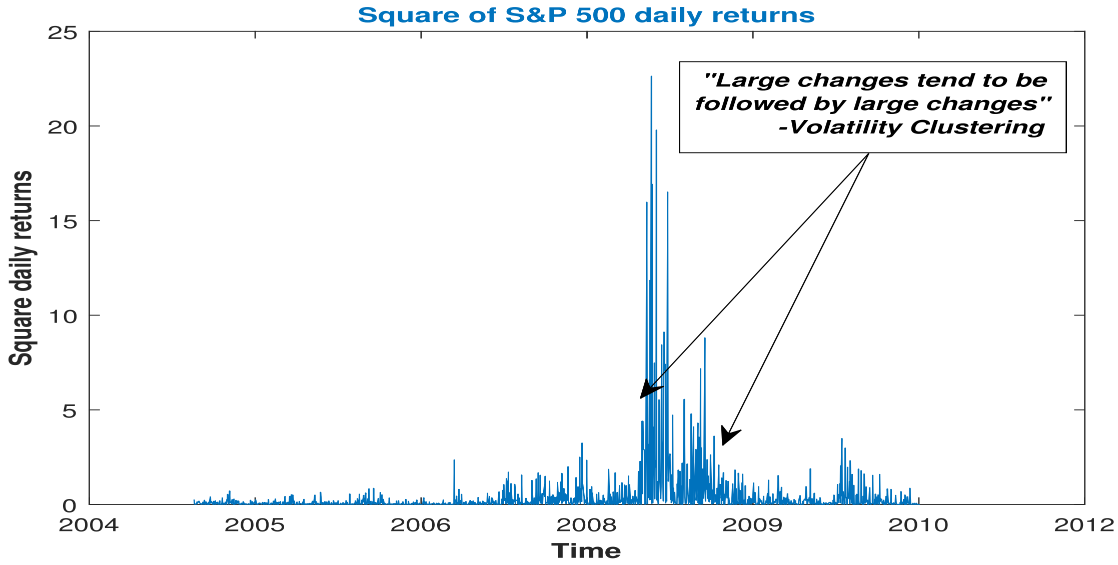
Motivation

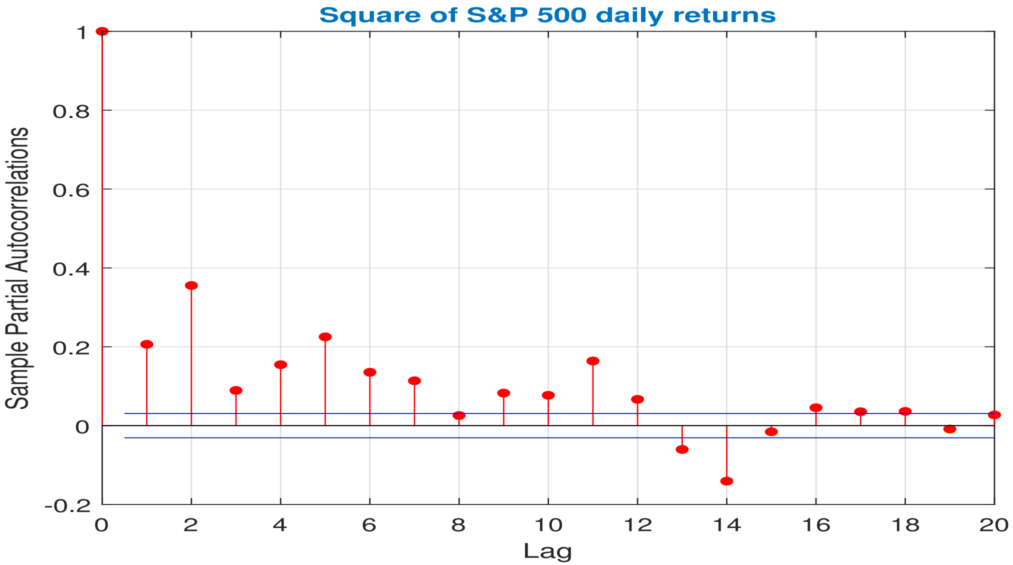
Empirical Stylized Facts




S&P 500 daily returns
Sample Partial Autocorrelations
1
0.8
0.6
0.4
0.2
0
-0.2
0
2
4
6
8
10
12
14
16
18
20
Lag


Square of S&P 500 daily returns
25
20
15
10
5
0
1998
2001
2004
2006
2009
2012
2015
2017
Time
Square daily returns


Square of S&P 500 daily returns
"Large changes tend to be followed by large changes" -Volatility Clustering
Square daily returns
Time
0
5
10
15
20
25
2004
2005
2006
2008
2009
2010
2012


Square of S&P 500 daily returns
Sample Partial Autocorrelations
Lag
1
0.8
0.6
0.4
0.2
0
-0.2
0
2
4
6
8
10
12
14
16
18
20

Motivation

Why do we care about time-varying volatility?

- Dynamic volatility has consequences for many problems of financial decision:

  (1) Risk management; (2) Portfolio allocation; (3) Asset pricing; (4) Hedging and Trading.

  - Example 1 (Portfolio Allocation): Optimal portfolio shares w^* solve:

$$min_w \ w'\Sigma w \qquad s.t. \ \ w'\mu = \mu_p$$

  — Importantly, $w^* = f(\Sigma)$, so if Σ varies, we have $w_t^* = f(\Sigma_t)$.

  - Example 2 (Asset pricing: Derivatives): Black-Scholes formula with constant volatility:

$$P_{Call} = N(d_1)S - N(d_2)Ke^{-r\tau},$$

  where, $d_1 = \frac{\ln(S/K)+(r+\sigma^2/2)\tau}{\sigma\sqrt{\tau}}$ and $d_2 = \frac{\ln(S/K)-(r+\sigma^2/2)\tau}{\sigma\sqrt{\tau}}$.

  — Completely different when σ varies !

- Dynamic macroeconomic models (VAR, DSGE, Time-varying Uncertainty Measures) [see Cogley and Sargent (2005), Primiceri (2005), Benati (2008), Koop, Leon-Gonzalez and Strachan (2009), Koop and Korobilis (2013), Liu and Morley (2014), Jurado, Ludvigson, and Ng (2015), and many recent papers]

Motivation

Conditional volatility models

- Two main classes of models have been proposed for dynamic (random) volatility:

1. GARCH-type models [Engle (1982)] where volatility is modelled as a deterministic process. A GARCH(1,1) model:

$$y_t = \sigma_t z_t, \qquad \sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

where $\alpha_0 > 0$, $\alpha_1 > 0$, $\beta_1 > 0$ for positive variance and z_t's are $i.i.d.$ $N(0,1)$, $y_t = r_t - \mu_r$ is the residual return and σ_t is the volatility at time t.

2. Stochastic volatility (SV) models [Taylor (1982, 1986)] where volatility is modelled as a latent stochastic process. An SV(1) model:

$$y_t = \sigma_t z_t, \qquad \log \sigma_t^2 = \alpha + \phi \log \sigma_{t-1}^2 + v_t,$$

where the vectors $(z_t, v_t)'$ are $i.i.d.$ according to a $N[0, I_2]$ distribution.

Outline

ARCH Models

Simple ARCH Model

- The first and simplest model we will look at is an ARCH model, which stands for Autoregressive Conditional Heteroscedasticity. The AR comes from the fact that these models are autoregressive models in squared returns, which we will demonstrate later in this section.
- The conditional comes from the fact that in these models, next period's volatility is conditional on information this period. Heteroscedasticity means non constant volatility.
- In a standard linear regression where $y_i = \alpha + \beta x_i + \epsilon_i$, when the variance of the residuals, ϵ_i is constant, we call that homoscedastic and use ordinary least squares to estimate α and β. If, on the other hand, the variance of the residuals is not constant, we call that heteroscedastic and we can use weighted least squares to estimate the regression coefficients.
- Let us assume that the return on an asset is

$$r_t = \mu + \sigma_t \epsilon_t$$

where ϵ_t is a sequence of $N(0,\ 1)$ i.i.d. random variables. We will define the residual return at time $t, r_t - \mu$, as

$$u_t = \sigma_t \epsilon_t.$$

ARCH Models

Simple ARCH Model

- In an ARCH(1) model, first developed by Engle (1982),

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2$$

  where $\alpha_0 > 0$ and $\alpha_1 \geq 0$ to ensure positive variance and $\alpha_1 < 1$ for stationarity.
- Under an ARCH(1) model, if the residual return, u_t is large in magnitude, our forecast for next period's conditional volatility, σ_{t+1} will be large.
- We say that in this model, the returns are conditionally normal (conditional on all information up to time $t-1$, the one period returns are normally distributed). We will relax that assumption on conditional normality in a later

ARCH Models

Heavy tails

- We can see right away that a time varying σ_t^2 will lead to fatter tails, relative to a normal distribution, in the unconditional distribution of u_t (see Campbell, Lo, and Mackinlay(1997)).
- The kurtosis of u_t is defined as

$$\mathrm{kurt}(u_t) = \frac{E[u_t^4]}{(E[u_t^2])^2}.$$

- If u_t were normally distributed, it would have a kurtosis of 3. Here,

$$\begin{aligned} \mathrm{kurt}(u_t) &= \frac{E[\sigma_t^4]E[\epsilon_t^4]}{(E[\sigma_t^2])^2(E[\epsilon_t^2])^2} \\ &= \frac{3E[\sigma_t^4]}{(E[\sigma_t^2])^2} \end{aligned}$$

and by Jensen's inequality [for a convex function, $f(x), E[f(x)] > f(E[x])$], $E[\sigma_t^4] > (E[\sigma_t^2])^2$, so

$$\mathrm{kurt}(u_t) > 3.$$

ARCH Models

unconditional variance

- We'll discuss a few properties of an ARCH(1) model in particular. The unconditional variance of u_t is

$$\begin{aligned} \mathrm{Var}(u_t) &= E[u_t^2] - (E[u_t])^2 \\ &= E[u_t^2] \\ &= E[\sigma_t^2 \epsilon_t^2] \\ &= E[\sigma_t^2] \\ &= \alpha_0 + \alpha_1 E[u_{t-1}^2] \end{aligned}$$

and since u_t is a stationary process, the $\mathrm{Var}(u_t) = \mathrm{Var}(u_{t-1}) = E[u_{t-1}^2]$, so

$$\mathrm{Var}(u_t) = \frac{\alpha_0}{1-\alpha_1} .$$

ARCH Models

AR Representation

- An ARCH(1) is like an AR(1) model on squared residuals, u_t^2. To see this, define the conditional forecast error, or the difference between the squared residual return and our conditional expectation of the squared residual return, as

$$v_t \equiv u_t^2 - E[u_t^2 | I_{t-1}] = u_t^2 - \sigma_t^2$$

  where I_{t-1} is the information at time $t-1$.

- Note that v_t is a zero mean, uncorrelated series. The ARCH(1) equation becomes

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \alpha_1 u_{t-1}^2 \\ u_t^2 - v_t &= \alpha_0 + \alpha_1 u_{t-1}^2 \\ u_t^2 &= \alpha_0 + \alpha_1 u_{t-1}^2 + v_t \end{aligned}$$

  which is an AR(1) process on squared residuals.

GARCH Models

Motivation

- In an ARCH(1) model, next period's variance only depends on last period's squared residual so a crisis that caused a large residual would not have the sort of persistence that we observe after actual crises.
- This has led to an extension of the ARCH model to a GARCH, or Generalized ARCH model, first developed by Bollerslev (1986), which is similar in spirit to an ARMA model.
- In a GARCH(1,1) model,

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

where $\alpha_0 > 0, \alpha_1 > 0, \beta_1 > 0$, and $\alpha_1 + \beta_1 < 1$, so that our next period forecast of variance is a blend of our last period forecast and last period's squared return.

GARCH Models

ARMA representation

- We can see that just as an ARCH(1) model is an AR(1) model on squared residuals, an GARCH(1,1) model is an ARMA(1,1) model on squared residuals by making the same substitutions as before, $v_t = u_t^2 - \sigma_t^2$

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\ u_t^2 - v_t &= \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 (u_{t-1}^2 - v_{t-1}) \\ u_t^2 &= \alpha_0 + (\alpha_1 + \beta_1) u_{t-1}^2 + v_t - \beta_1 v_{t-1} \end{aligned}$$

which is an ARMA(1,1) on the squared residuals.

GARCH Models

unconditional variance

- The unconditional variance of u_t is

$$\begin{aligned}
\mathrm{Var}(u_t) &= E[u_t^2] - (E[u_t])^2 \\
&= E[u_t^2] \\
&= E[\sigma_t^2 \epsilon_t^2] \\
&= E[\sigma_t^2] \\
&= \alpha_0 + \alpha_1 E[u_{t-1}^2] + \beta_1 \sigma_{t-1}^2 \\
&= \alpha_0 + (\alpha_1 + \beta_1) E[u_{t-1}^2]
\end{aligned}$$

and since u_t is a stationary process,

$$\mathrm{Var}(u_t) = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}$$

and since $u_t = \sigma_t \epsilon_t$, the unconditional variance of returns, $E[\sigma_t^2] = E[u_t^2]$, is also $\alpha_0/(1 - \alpha_1 - \beta_1)$.

GARCH Models

ARCH (∞) Representation

- Just as an ARMA(1,1) can be written as an AR(∞) , a GARCH(1,1) can be written as an ARCH (∞),

$$
\begin{aligned}
\sigma_t^2 &= \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\
&= \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 (\alpha_0 + \alpha_1 u_{t-2}^2 + \beta_1 \sigma_{t-2}^2) \\
&= \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_0 \beta_1 + \alpha_1 \beta_1 u_{t-2}^2 + \beta_1^2 \sigma_{t-2}^2 \\
&= \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_0 \beta_1 + \alpha_1 \beta_1 u_{t-2}^2 + \beta_1^2 (\alpha_0 + \alpha_1 u_{t-3}^2 + \beta_1 \sigma_{t-3}^2) \\
&\vdots \\
&= \frac{\alpha_0}{1-\beta_1} + \alpha_1 \sum_{i=0}^{\infty} u_{t-1-i}^2 \beta_1^i
\end{aligned}
$$

so that the conditional variance at time t is the weighted sum of past squared residuals and the weights decrease as you go further back in time.

GARCH Models

Weighted combination

- Since the unconditional variance of returns is $E[\sigma^2] = \alpha_0/(1-\alpha_1-\beta_1)$, we can write the GARCH(1,1) equation yet another way

$$\begin{aligned}\sigma_t^2 &= \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\ &= (1-\alpha_1-\beta_1)E[\sigma^2] + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2.\end{aligned}$$

- Written this way, it is easy to see that next period's conditional variance is a weighted combination of the unconditional variance of returns, $E[\sigma^2]$, last period's squared residuals, u_{t-1}^2, and last period's conditional variance, σ_{t-1}^2, with weights $(1-\alpha_1-\beta_1), \alpha_1, \beta_1$ which sum to one.

GARCH Models

Forecasting

- It is often useful not only to forecast next period's variance of returns, but also to make an l-step ahead forecast, especially if our goal is to price an option with l steps to expiration using our volatility model.
- Again starting from the GARCH(1,1) equation for σ_t^2, we can derive our forecast for next period's variance, $\hat{\sigma}_{t+1}^2$

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

$$\begin{aligned}
\hat{\sigma}_{t+1}^2 &= \alpha_0 + \alpha_1 E[u_t^2 | I_{t-1}] + \beta_1 \sigma_t^2 \\
&= \alpha_0 + \alpha_1 \sigma_t^2 + \beta_1 \sigma_t^2 \\
&= \alpha_0 + (\alpha_1 + \beta_1)\sigma_t^2 \\
&= \sigma^2 + (\alpha_1 + \beta_1)(\sigma_t^2 - \sigma^2)
\end{aligned}$$

GARCH Models

Forecasting

- $$\begin{aligned}
\hat{\sigma}^2_{t+2} &= \alpha_0 + \alpha_1 E[u^2_{t+1}|I_{t-1}] + \beta_1 E[\sigma^2_{t+1}|I_{t-1}] \\
&= \alpha_0 + (\alpha_1 + \beta_1)\hat{\sigma}^2_{t+1} \\
&= \sigma^2 + (\alpha_1 + \beta_1)(\hat{\sigma}^2_{t+1} - \sigma^2) \\
&= \sigma^2 + (\alpha_1 + \beta_1)^2(\sigma^2_t - \sigma^2)
\end{aligned}$$

GARCH Models

Forecasting

- $$\begin{aligned}\hat{\sigma}^2_{t+l} &= \alpha_0 + (\alpha_1 + \beta_1)\hat{\sigma}^2_{t+l-1} \\ &= \sigma^2 + (\alpha_1 + \beta_1)(\hat{\sigma}^2_{t+l-1} - \sigma^2) \\ &= \sigma^2 + (\alpha_1 + \beta_1)^l(\sigma^2_t - \sigma^2)\end{aligned}$$

  where we have substituted for the unconditional variance, $\sigma^2 = \alpha_0/(1 - \alpha_1 - \beta_1)$
- From the above equation we can see that $\hat{\sigma}^2_{t+l} \to \sigma^2$ as $l \to \infty$ so as the forecast horizon goes to infinity, the variance forecast approaches the unconditional variance of u_t.
- From the l-step ahead variance forecast, we can see that $(\alpha_1 + \beta_1)$ determines how quickly the variance forecast converges to the unconditional variance.

GARCH Models

Maximum Likelihood Estimation

- In general, to estimate the parameters using maximum likelihood, we form a likelihood function, which is essentially a joint probability density function but instead of thinking of it as a function of the data given the set of parameters, $f(x_1, x_2, \ldots, x_n|\Theta)$.
- We think of the likelihood function as a function of the parameters given the data, $L(\Theta|x_1, x_2, \ldots, x_n)$, and we maximize the likelihood function with respect to the parameters, which is essentially finding the mode of the distribution.

GARCH Models

Heavy tails

- If the residual returns were independent of each other, we could write the joint density function as the product of the marginal densities, but in the GARCH model, returns are not, of course, independent.
- However, we can still write the joint probability density function as the product of conditional density functions

$$\begin{aligned} f(r_1, r_2, \ldots, r_T) &= f(r_T | r_1, r_2, \ldots, r_{T-1}) f(r_1, r_2, \ldots, r_{T-1}) \\ &= f(r_T | r_1, r_2, \ldots, r_{T-1}) f(r_{T-1} | r_1, r_2, \ldots, r_{T-2}) f(r_1, r_2, \ldots, r_{T-2}) \\ &\vdots \\ &= f(r_T | r_1, r_2, \ldots, r_{T-1}) f(r_{T-1} | r_1, r_2, \ldots, r_{T-2}) \cdots f(r_1) \end{aligned}$$

GARCH Models

Maximum Likelihood Estimation

- For a GARCH(1,1) model with Normal conditional returns, the likelihood function is

$$
\begin{aligned}
& L(\alpha_0,\ \alpha_1,\ \beta_1,\ \mu | r_1,\ r_2,\ \ldots,\ r_T) \\
= \quad & \frac{1}{\sqrt{2\pi\sigma_T^2}} \exp\left(\frac{-(r_T-\mu)^2}{2\sigma_T^2}\right) \frac{1}{\sqrt{2\pi\sigma_{T-1}^2}} \exp\left(\frac{-(r_{T-1}-\mu)^2}{2\sigma_{T-1}^2}\right) \ldots \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(\frac{-(r_1-\mu)^2}{2\sigma_1^2}\right).
\end{aligned}
$$

GARCH Models

Maximum Likelihood Estimation

- Since the $\ln L$ function is monotonically increasing function of L, we can maximize the log of the likelihood function

$$\ln L(\alpha_0,\ \alpha_1,\ \beta_1,\ \mu | r_1,\ r_2,\ \ldots,\ r_T) \quad = \quad -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{T}\ln\sigma_i^2 - \frac{1}{2}\sum_{i=1}^{T}\left(\frac{(r_i-\mu)^2}{\sigma_i^2}\right)$$

  and for a GARCH(1,1), we can substitute $\sigma_i^2 = \alpha_0 + \alpha_1 a_{i-1}^2 + \beta_1 \sigma_{i-1}^2$ into the above equation, and the likelihood function is only a function of the returns, r_t and the parameters.

- Notice that besides estimating the parameters, $\alpha_0, \alpha_1, \beta_1$, and μ, we must also estimate the initial volatility, σ_1. If the time series is long enough, the estimate for σ_1 will be unimportant.

Outline

Introduction

Why do we care about time-varying volatility?

- Finance: Dynamic volatility has consequences for many problems of financial decision:

  (1) Risk management, (2) Portfolio allocation, (3) Asset pricing, (4) Hedging and Trading

  - Example 1 (Portfolio Allocation): Optimal portfolio shares w^* solve:

$$min_w \ w'\Sigma w \qquad s.t. \ \ w'\mu = \mu_p$$

    — Importantly, $w^* = f(\Sigma)$, so if Σ varies, we have $w_t^* = f(\Sigma_t)$.

  - Example 2 (Asset pricing: Derivatives): Black-Scholes formula with constant volatility:

$$P_{Call} = N(d_1)S - N(d_2)Ke^{-r\tau},$$

    where, $d_1 = \frac{\ln(S/K)+(r+\sigma^2/2)\tau}{\sigma\sqrt{\tau}}$ and $d_2 = \frac{\ln(S/K)-(r+\sigma^2/2)\tau}{\sigma\sqrt{\tau}}$.
    — Completely different when σ varies !

- Macroeconomics: Dynamic volatility is also important for macroeconomic forecasting and measurement of uncertainty, see Cogley and Sargent (2005), Primiceri (2005), Benati (2008), Koop, Leon-Gonzalez and Strachan (2009), Koop and Korobilis (2013), Liu and Morley (2014), Jurado, Ludvigson, and Ng (2015), and many recent papers.

Introduction

What are the alternative classes of models?

- Two main classes of models have been proposed for dynamic (random) volatility:
  1. GARCH-type models [Engle (1982)] where volatility is modelled as a deterministic process. A GARCH(1,1) model:

     $$y_t = \sigma_t z_t, \qquad \sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

     where $\alpha_0 > 0$, $\alpha_1 > 0$, $\beta_1 > 0$ for positive variance and z_t's are $i.i.d.$ $N(0,1)$, $y_t = r_t - \mu_r$ is the residual return (or the error term of any time series regression model) and σ_t is the volatility at time t.
  2. Stochastic volatility (SV) models [Taylor (1982, 1986)] where volatility is modelled as a latent stochastic process. An SV(1) model:

     $$y_t = \sigma_t z_t, \qquad \log \sigma_t^2 = \alpha + \phi \log \sigma_{t-1}^2 + v_t,$$

     where the vectors $(z_t, v_t)^{'}$ are $i.i.d.$ according to a $N[0, I_2]$ distribution.
- Higher-order stochastic volatility [SV(p)] model:

  $$\log \sigma_t^2 = \alpha + \phi_1 \log \sigma_{t-1}^2 + \ldots + \phi_p \log \sigma_{t-p}^2 + v_t.$$

- Ahsan, M. N. and Dufour, J.-M. (2021), Simple estimators and inference for higher-order stochastic volatility models, *Journal of Econometrics* 224(1), 181 – 197.

Introduction

Is there a class of models that has better advantages?

- SV models may be preferable to GARCH-type models for several reasons.
  1. Discrete version of continuous time analogues [Shephard and Andersen (2009), Taylor (1994)].
  2. Empirical evidence suggests that SV models are robust to model misspecification [Carnero, Peña and Ruiz (2004), Chan and Grant (2016)].
  3. Provide more accurate forecasts of volatility [Kim, Shephard and Chib (1998), Yu (2002), Poon and Granger (2003), Koopman, Jungbacker and Hol (2005)].
  4. Statistical properties of SV models are relatively easy to derive [Davis and Mikosch (2009)].
- Empirical popularity of SV models are deterred by two reasons:
  1. No closed form solution for the likelihood function.
  2. Statistical packages for estimating SV models are scarce whereas many statistical packages for GARCH: [EVIEWS, GAUSS, MATLAB, R, S+, SAS, TSP, STATA, PYTHON, OX, etc.]

Introduction

What are the alternative SV estimator?

- Proposed estimation methods of SV models are limited to SV(1):
  - Generalized Method of Moments [Melino and Turnbull (1990), Andersen and Sorensen (1996)];
  - Simulated Maximum likelihood (SML) [Danielsson and Richard (1993), Durham (2006)];
  - Quasi Maximum Likelihood (QML) [Harvey et al. (1994), Ruiz (1994)];
  - Bayesian techniques based on Markov Chain Monte Carlo (MCMC) [Jacquier et al. (1994), Kim et al. (1998), Chib et al. (2002), Flury and Shephard (2011)];
  - Simulated Method of Moments (SMM) [Gallant and Tauchen (1996), Monfardini (1998)];
  - Monte Carlo likelihood (MCL) [Sandmann and Koopman (1998)];
  - Linear-representation based estimation (LR) [Francq and Zakoïan (2006)];
  - Moment based closed-form estimator (DV) [Dufour and Valéry (2006, 2009)];
  - ARMA-SV estimator (ARMA-SV) [Ahsan and Dufour (2019)].

★ These estimation methods are either inefficient and/or very expensive from the computational viewpoint, inflexible across models, not easy to implement in practice, and may not converge. [Broto and Ruiz (2004), Ahsan and Dufour (2019)].

Introduction

Why higher-order stochastic volatility?

- The estimation of SV(p) models is even more challenging than it is for SV(1) models.
- SV(p) models are rarely considered:
  SMM [Gallant et al. (1997)], MCL [Asai (2008)], MCMC [Chan and Grant (2016)].
- Methods proposed for SV(1) models are difficult to extend for an SV(p) models with arbitrary order and may be computationally expensive.
- The motivations for SV(p) models can be described as follows.
  1. The SV(p) model is a natural extension of the SV(1) model, which can only generate geometrically decaying autocovariance function, whereas volatility process generically features persistent memory.
  2. Higher-order autoregressive terms in SV models emerge naturally in some setups [e.g., multi-factor SV models].
  3. SV(p) models provide more flexibility to represent volatility persistence.
  4. Empirical evidence suggests that higher-order SV(p) models may be preferable for in-sample model fitting (in this paper), out-of-sample volatility forecasting and option pricing (in Ahsan and Dufour (2020)).

Contributions

- In this paper, we propose simple estimators for SV(p) models. These include:
  - Simple moment-based estimators — extension of DV and ARMA-SV methods
  - Restricted estimators
  - ARMA-based winsorized estimators — ensures stationarity condition, particularly in a small sample or in the presence of outliers.
- We also propose GMM-type estimators [extension of Andersen-Sorensen (1996)].
- The proposed simple estimators (DV, ARMA-SV) are computationally inexpensive and easy to implement in practice.
- We show that recursive estimation algorithms [Durbin-Levinson-type (DL)] can be applied to SV(p) models.
- The proposed simple estimators are $\sqrt{T}$ consistent and asymptotically normal.
- Because of computational simplicity, we can do Monte Carlo (MC) or Bootstrap type tests using simple estimators, entails that, which may not possible for other simulation based methods.
- In simulation, we compare the performance of our estimators to Bayesian estimators:
  - ★ W-ARMA estimator uniformly outperforms all other estimators in terms of Bias and RMSE — Six times lower RMSE than Bayesian estimator.
  - ★ It is on average 843,220 times faster than the Bayesian estimator for a sample size, T = 2000.

★ For S&P 500 index, both asymptotic and finite sample tests suggest that an SV(2) or SV(3) model could be more appropriate than an SV(1) model.

Outline

SV(p) Models

- Stochastic volatility of order p: The process $\{y_t : t \in \mathbb{N}_0\}$ follows an SV model of the type:

$$y_t = exp(\frac{w_t}{2})\sigma_y z_t, \tag{1}$$

$$w_t = \sum_{j=1}^{p} \phi_j w_{t-j} + \sigma_v v_t, \tag{2}$$

where $\{\phi_j\}_{j=1}^{p}$, σ_y, σ_v are the fixed parameters of the model and $y_t = r_t - \mu_r$ is the residual return at time t

- This is an alternative re-parametrization of SV model with $w_t = \log\sigma_t^2$ and $\sigma_y = exp(\alpha/2(1-\sum_{j=1}^{p}\phi_j))$.
- The vectors $(z_t, v_t)^{'}$, $t \in \mathbb{N}_0$, are $i.i.d.$ according to a $N[0, I_2]$ distribution.
- The process $l_t = (y_t, w_t)^{'}$ is strictly stationary.
- We propose two simple estimators for SV(p) models:
  1. Extension of Dufour and Valéry (EDV) Estimator.
  2. Simple ARMA based (ARMA-SV) Estimator.

  ★ Although the EDV estimator is analytically tractable and computationally simple, it tends to be less precise than ARMA-type estimators.

Simple ARMA-SV Estimator

- It is possible to obtain a more efficient simple estimator by exploiting an ARMA representation of SV model [ARMA-SV].
- Consider the return equation:

$$y_t = exp(\frac{w_t}{2})\sigma_y z_t$$

- Transform y_t by taking logarithms of the squares:

$$\log(y_t^2) = \underbrace{[\log(\sigma_y^2) + E(\log(z_t^2))]}_{\equiv\mu} + w_t + \underbrace{[\log(z_t^2) - E(\log(z_t^2))]}_{\equiv\epsilon_t},$$

$$\log(y_t^2) = \mu + w_t + \epsilon_t. \tag{3}$$

- Since z_t is a Gaussian noise, ϵ_t follows a $\log\chi^2_{(1)}$ distribution with

$$\sigma_\epsilon^2 = \pi^2/2, \quad E(\log(z_t^2)) = -1.27.$$

- Rewrite (3) as

$$y_t^* = w_t + \epsilon_t \tag{4}$$

where $y_t^* = (\log(y_t^2) - \mu)$, and μ is the mean of $\log(y_t^2)$.

Simple ARMA-SV Estimator

State Space Representation of SV model

- Combining (2) and (4), we get a linear state space form:

$$w_t = \sum_{j=1}^{p} \phi_j w_{t-j} + v_t \qquad \text{(State Transition Equation)} \qquad (5)$$

$$y_t^* = w_t + \epsilon_t \qquad \text{(Measurement Equation)} \qquad (6)$$

  where w_t is a logarithm of latent daily volatility, y_t^* is a logarithm of the daily squared residual return corrected by its mean. The v_t's and ϵ_t's are $i.i.d.$ $N(0,\sigma_v^2)$ and $\log \chi^2_{(1)}$ random variables, respectively.

- Several methods have been proposed in the literature that exploits the above state space form of SV model [Nelson (1988), Harvey et al. (1994), Ruiz (1994), Shephard (1994), Breidt and Carriquiry (1996), Harvey and Shephard (1996), Kim et al. (1998), Sandmann and Koopman (1998), Steel (1998), Chib et al. (2002), Knight et al. (2002), Francq and Zakoïan (2006), Omori et al. (2007)].

Simple ARMA-SV Estimator

ARMA Representation of SV Model

- The model defined by (5) and (6) has an ARMA(p,p) representation:

$$y_t^* = \sum_{j=1}^{p} \phi_j y_{t-j}^* + v_t + \epsilon_t - \sum_{j=1}^{p} \phi_j \epsilon_{t-j}, \tag{7}$$

where $v_t + \epsilon_t - \sum_{j=1}^{p} \phi_j \epsilon_{t-j}$ admits an MA(p) process.

- The autocovariances structure of the observed process $\{y_t^*\}$ satisfies the following equations respectively:

$$\gamma_{y^*}(k) = \begin{cases} \phi_1 \gamma_{y^*}(k-1) + \cdots + \phi_p \gamma_{y^*}(k-p) + \sigma_v^2 + \sigma_\epsilon^2; & \text{if } k = 0, \\ \phi_1 \gamma_{y^*}(k-1) + \cdots + \phi_p \gamma_{y^*}(k-p) - \phi_k \sigma_\epsilon^2; & \text{if } 1 \le k \le p, \\ \phi_1 \gamma_{y^*}(k-1) + \cdots + \phi_p \gamma_{y^*}(k-p); & \text{if } k > p, \end{cases} \tag{8}$$

where $\gamma_{y^*}(k) = cov(y_t^*, y_{t-k}^*)$, $y_t^* = (\log(y_t^2) - \mu)$, $\mu = [\log(\sigma_y^2) + E(\log(z_t^2))]$.

Simple ARMA-SV Estimator

- The model defined by (7) has the following analytical closed-form expressions for its parameters, $\theta \equiv (\phi_1, ..., \phi_p, \sigma_y, \sigma_v)'$:

$$\boldsymbol{\phi_p} = \boldsymbol{\Gamma}(p,j)^{-1}\boldsymbol{\gamma}(p,j), \quad j \geq 1, \quad \sigma_y = [\exp(\mu + 1.27)]^{1/2}, \quad \sigma_v = [\gamma_{y^*}(0) - \boldsymbol{\phi}_{\boldsymbol{p}}'\boldsymbol{\gamma}(p) - \pi^2/2]^{1/2}, \tag{9}$$

where $\boldsymbol{\phi_p} := (\phi_1, \ldots, \phi_p)'$, $\boldsymbol{\gamma}(p,j) := [\gamma_{y^*}(p+j), \ldots, \gamma_{y^*}(2p+j-1)]'$, $\boldsymbol{\gamma}(p) := [\gamma_{y^*}(1), \ldots, \gamma_{y^*}(p)]'$ are $p \times 1$ vectors and $\boldsymbol{\Gamma}(p,j)$ is a $p \times p$ matrix

$$\boldsymbol{\Gamma}(p,j) := \begin{bmatrix} \gamma_{y^*}(j+p-1) & \gamma_{y^*}(j+p-2) & \cdots & \gamma_{y^*}(j) \\ \gamma_{y^*}(j+p) & \gamma_{y^*}(j+p-1) & \cdots & \gamma_{y^*}(j+1) \\ \vdots & \vdots & & \vdots \\ \gamma_{y^*}(j+2p-2) & \gamma_{y^*}(j+2p-3) & \cdots & \gamma_{y^*}(j+p-1) \end{bmatrix}.$$

where p is the SV order, $\gamma_{y^*}(k) = \text{cov}(y_t^*, y_{t-k}^*)$, with $y_t^* = [\log(y_t^2) - \mu]$ and $\mu := \mathbb{E}[\log(y_t^2)]$.

- Setting $j = 1$ in (9) and replacing theoretical moments by their corresponding empirical moments yield the following *simple ARMA-SV* estimator of the SV(p) coefficients:

$$\hat{\boldsymbol{\phi}}_{\boldsymbol{p}} = \hat{\boldsymbol{\Gamma}}(p,1)^{-1}\hat{\boldsymbol{\gamma}}(p,1), \quad \hat{\sigma}_y = [\exp(\hat{\mu} + 1.27)]^{1/2}, \quad \hat{\sigma}_v = [\hat{\gamma}_{y^*}(0) - \hat{\boldsymbol{\phi}}_{\boldsymbol{p}}'\hat{\boldsymbol{\gamma}}(p) - \pi^2/2]^{1/2}. \tag{10}$$

Restricted Estimation

- Proposed simple estimators may yield a solution outside the admissible area, i.e., some of the eigenvalues of the latent volatility process may lie outside the unit circle or equal to unity.
- When this happens, a simple fix is projecting the estimate on the space of acceptable parameter solutions by altering the eigenvalues that lie on or outside the unit circle.
- This can be done in the following two steps:
  1. Given the estimated unstable parameters, we calculate the roots of the characteristic equation and restrict their absolute values to less than unity.
  2. Given these restricted roots, we calculate the constrained parameters which ensure stationarity.

Winsorized ARMA-SV Estimator (W-ARMA-SV)

- We can achieve better stability and efficiency of the ARMA-SV estimator by using "winsorization".
- From (9), it is easy to see that:

$$\boldsymbol{\phi}_{\boldsymbol{p}} = \sum_{j=1}^{\infty} \omega_j \boldsymbol{B}(p,j) = \sum_{j=1}^{\infty} \omega_j \boldsymbol{\Gamma}(p,j)^{-1} \boldsymbol{\gamma}(p,j) \tag{11}$$

for any ω_j sequence with $\sum_{j=1}^{\infty} \omega_j = 1$, where

$$\boldsymbol{B}(p,j) := \boldsymbol{\Gamma}(p,j)^{-1} \boldsymbol{\gamma}(p,j) = [\boldsymbol{B}(p,j)_1, \ldots, \boldsymbol{B}(p,j)_p]'$$

is a $p \times 1$ vector.

- Using (11), we can define a more general class of estimators for $\boldsymbol{\phi}_{\boldsymbol{p}}$ by taking a weighted average of several sample analogs of the $p \times 1$ vector:

$$\tilde{\boldsymbol{\phi}}_{\boldsymbol{p}} := \sum_{j=1}^{J} \omega_j \hat{\boldsymbol{B}}(p,j), \quad \hat{\boldsymbol{B}}(p,j) := \hat{\boldsymbol{\Gamma}}(p,j)^{-1} \hat{\boldsymbol{\gamma}}(p,j) = [\hat{\boldsymbol{B}}(p,j)_1, \ldots, \hat{\boldsymbol{B}}(p,j)_p]', \tag{12}$$

where $\sum_{j=1}^{J} \omega_j = 1$, $1 \le J \le T - p$ and T is the length of time series.

Winsorized ARMA-SV Estimators

Four W-ARMA-SV Estimator

- We consider four winsorized estimators based on the expression given in (12): [see Kristensen and Linton (2006), Hafner and Linton (2017)]
  1. The first winsorized estimate $\hat{\boldsymbol{\phi}}_{\boldsymbol{p}}^{\boldsymbol{m}}$ is the arithmetic mean of sample ratios (equal weights):
  $$\omega_j = 1/J, \quad j = 1, \dots, J. \tag{13}$$
  2. The second estimate $\hat{\boldsymbol{\phi}}_{\boldsymbol{p}}^{\boldsymbol{ld}}$ is a mean of ratios with linearly declining (LD) weights:
  $$\omega_j = (2/J)[1 - (j/(J+1))], \quad j = 1, \dots, J. \tag{14}$$
  3. The third estimate is the median-based estimate which is obtained by taking the median of J estimates of each one of the p components of $\boldsymbol{\phi}_{\boldsymbol{p}}$:
  $$\hat{\boldsymbol{\phi}}_{\boldsymbol{p}}^{\text{med}} = [\hat{\phi}_1^{\text{med}}, \dots, \hat{\phi}_p^{\text{med}}]', \quad \hat{\phi}_i^{\text{med}} = \text{med}\{\hat{\boldsymbol{B}}(p, j)_i : 1 \le j \le J\}, \; i = 1, \dots, p. \tag{15}$$
  4. The fourth estimate is obtained by an OLS regression coefficient (without intercept):
  $$\hat{\boldsymbol{\phi}}_{\boldsymbol{p}}^{\boldsymbol{ols}} = [A(p, J)' A(p, J)]^{-1} A(p, J)' e(p, J) \tag{16}$$
  where $e(p, J)$ is a $(pJ) \times 1$ vector and $A(p, J)$ a $(pJ) \times p$ matrix defined by
  $$e(p, J) = [\hat{\boldsymbol{\gamma}}(p, 1)\omega_1^{1/2}, \dots, \hat{\boldsymbol{\gamma}}(p, J)\omega_J^{1/2}]', \quad A(p, J) = [\hat{\boldsymbol{\Gamma}}(p, 1)\omega_1^{1/2}, \dots, \hat{\boldsymbol{\Gamma}}(p, J)\omega_J^{1/2}]'. \tag{17}$$
- The above estimators are depend on J and for $J = 1$, they are same as the ARMA-SV estimator that given by (10).

Winsorized ARMA-SV Estimators

W-ARMA-SV-OLS: Possible Choices of J

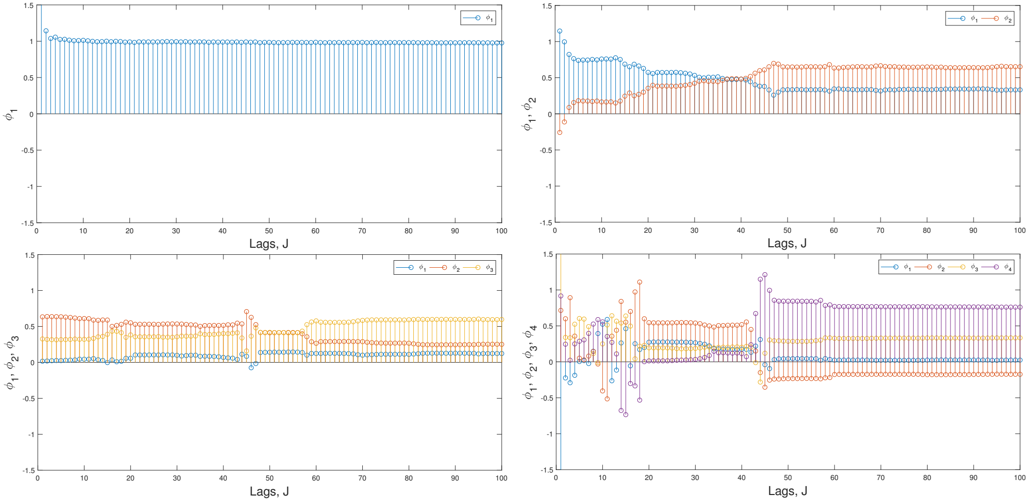


Figure: S&P 500: 1996-2016. W-ARMA-SV-OLS estimators of volatility persistence parameters (ϕ's) as a function of the number of lags (J).

Recursive Estimation Algorithm

- Using simple estimators, one can recursively estimate an $SV(p)$ model from the estimates of an $SV(p-1)$ model.
- By recursion, we can easily estimate an SV model of any order.
- Durbin-Levinson (DL) type algorithm [Levinson (1947), Durbin (1960)].
  - For EDV: Use the DL algorithm that designed for the autoregressive process.

    Go
  - For ARMA-SV/W-ARMA-SV: Use a *Generalized* DL algorithm proposed by Tsay and Tiao (1984).

    Go

Asymptotic Theory

ARMA based Estimator: Assumptions

We derive the asymptotic properties of $\hat{\theta} \equiv (\hat{\phi}_1, \ldots, \hat{\phi}_p, \hat{\sigma}_y, \hat{\sigma}_v)'$ under the following set of assumptions:

A.1 : The the vectors $(z_t, v_t)'$ are $i.i.d.$ according to a $N[0, I_2]$ distribution.

A.2 : The latent process is strictly stationary with s-th order finite moment, i.e., $E[|w_t|]^s < \infty$.

*** Note that under A.1 and A.2 with $s = 2$, the observed process $\{y_t\}$ is **strictly stationary and ergodic**.

*** Furthermore, these properties are preserved by any continuous transformation of $\{y_t\}$, i.e., $\{f(y_t)\}$ where f is any measurable function.

*** The ARMA based estimator exploits the empirical moments of $y_t^* = \log(y_t^2) - \mu$ and $\log(y_t^2)$. Both y_t^* and $\log(y_t^2)$ are strictly stationary and ergodic.

Asymptotic theory

ARMA-SV Estimator: Consistency, Asymptotic Normality

- Empirical Moments:

R-1. Consistency: Under A.1 and A.2 with $s = 2$, the estimators $\hat{\Gamma}(m) = (\hat{\gamma}_{y^*}(k))_{k=0,\ldots,m}$ and $\hat{\mu}$ in (**??**) satisfy:

$$\hat{\Gamma}(m) \xrightarrow{p} \Gamma(m) \text{ and } \hat{\mu} \xrightarrow{p} \mu.$$

R-2. Asymptotic Distribution: Under A.1, A.2 with $s = 4$, and using R-1, the estimators $\hat{\Gamma}(m) = (\hat{\gamma}_{y^*}(k))_{k=0,\ldots,m}, \hat{\mu}$ satisfy:

$$\sqrt{T}[\hat{\mu} - \mu, \hat{\Gamma}(m) - \Gamma(m)] \xrightarrow{d} N(0, V_M).$$

- Asymptotic Distribution of Simple Estimators: Under A.1, A.2 with $s = 4$, and using R-1 and R-2, the estimator $\hat{\theta} \equiv (\hat{\phi}_1, \ldots, \hat{\phi}_p, \hat{\sigma}_y, \hat{\sigma}_v)'$ given in (**??**)-(**??**) is consistent and asymptotically normal, i.e., $\hat{\theta} \xrightarrow{p} \theta$ and

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} N(0, V).$$

Finite Sample Inference

Monte Carlo Tests

- Simulating an SV model is easy.
- Further, we can simulate the test statistic of SV parameters based on a computationally inexpensive estimator.
- However, if the estimator is computationally expensive then we cannot simulate the test statistic easily. The simulation will run forever.
- Using our proposed computationally simple estimators, one can construct more reliable finite sample inference.
- Monte Carlo test with level α as follows:

  1. Let S_0, be the observed test statistic (based on data).
  2. By Monte Carlo methods, draw N $i.i.d.$ replications of S, i.e., $(S_1,\ldots,S_N)$ under H_0 hypothesis.
  3. Obtain the rank of S_0, $\hat{R}_N[S_0]$ in the series, $S_0, S_1, \ldots, S_N$.
  4. Reject the null hypothesis if $\hat{R}_N[S_0] \geq (N+1)(1-\alpha)+1$, and do not reject, otherwise.

  *** Note that the number of simulations, N, should be chosen such that $\alpha(N+1)$ is an integer, so that the test can control the level exactly.

Outline

Simulation Study

- We consider an SV(2) model with different sets of parameters.
- In our GMM setting, we consider two sets of moments: one set with 24 moments suggested by Andersen and Sørensen (1996) and the other one with 6 lower order moments. The large and small sets are denoted by M_L and M_S and given by

$$M_L = \begin{pmatrix} |y_t|^j - \mu_j(\theta) \quad for \quad j = 1, ..., 4 \\ |y_t||y_{t-j}| - \mu_{1,1}(j|\theta) \quad for \quad j = 1, ..., 10 \\ y_t^2 y_{t-j}^2 - \mu_{2,2}(j|\theta) \quad for \quad j = 1, ..., 10 \end{pmatrix},$$

$$M_S = \begin{pmatrix} |y_t|^j - \mu_j(\theta) \quad for \quad j = 1, ..., 4 \\ y_t^2 y_{t-j}^2 - \mu_{2,2}(j|\theta) \quad for \quad j = 1, 2 \end{pmatrix}.$$

- Different weighting matrix to gain efficiency for the GMM estimators.
- We compare our estimators to two Bayesian estimators:
  — Bayes-1 is based on the precision sampler of Chan and Jeliazkov (2009), where latent volatility states are sampled jointly.
  — Bayes-2 is based on a single-move independent Metropolis-Hastings algorithm within Gibbs sampler, which breaks the joint posterior into univariate conditional posteriors.
  — For both estimators, the posteriors are based on 20000 draws of the sampler, after discarding 5000 draws.
- The simulations use 500 replications .

Simulation Results - SV(2): RMSE

True value	$\mathcal{M}_1 = (\phi_1, \phi_2, \sigma_y, \sigma_v) = (0.30, 0.60, 0.025, 2.5)$							
	$T = 500$				$T = 2000$			
	ϕ_1	ϕ_2	σ_y	σ_v	ϕ_1	ϕ_2	σ_y	σ_v
GMM-6M-E	1.0739	0.9177	0.1060	4.1666	1.0048	0.9817	0.2108	3.8913
GMM-6M-E-R	1.0556	0.9133	0.1060	4.1666	0.9845	0.9761	0.2108	3.8913
GMM-6M-NW	0.8992	0.7567	2.5178	2.0229	0.7536	0.6778	3.3015	1.2647
GMM-6M-NW-R	0.8293	0.7278	2.5178	2.0229	0.7424	0.6737	3.3015	1.2647
GMM-24M-E	0.5506	0.5991	0.1006	4.3062	0.7068	0.7414	0.0697	4.6590
GMM-24M-E-R	0.5247	0.5845	0.1006	4.3062	0.6466	0.7027	0.0697	4.6590
GMM-24M-NW	0.9997	0.8748	3.0388	3.1042	0.9656	0.8586	3.7781	2.0984
GMM-24M-NW-R	0.9744	0.8526	3.0388	3.1042	0.9606	0.8549	3.7781	2.0984
Bayes-1	0.5766	0.5835	1.7308	2.0068	0.2950	0.3267	2.7178	2.1542
Bayes-2	0.1491	0.1437	0.0973	0.6148	0.1680	0.1584	0.0394	0.5928
EDV	0.3801	0.4811	4.4211	1.2310	0.2908	0.3730	1.0332	1.1156
EDV-R	0.4938	0.5656	3.9991	1.4913	0.4455	0.4968	1.5429	1.3793
ARMA-SV	0.1838	0.1806	0.0181	0.1861	0.0861	0.0845	**0.0077**	0.0899
R-ARMA-SV	0.1900	0.1861	**0.0188**	0.1847	0.0861	0.0845	**0.0077**	0.0899
W-ARMA-SV-OLS($J = 10$)	0.1412	0.1377	**0.0188**	**0.1831**	0.0783	0.0763	**0.0077**	**0.0890**
W-ARMA-SV-OLS($J = 100$)	**0.1179**	**0.1150**	**0.0188**	0.1854	**0.0749**	**0.0729**	**0.0077**	0.0893

Simulation Results - SV(2): RMSE

True value	$\mathcal{M}_2 = (\phi_1, \phi_2, \sigma_y, \sigma_v) = (0.95, -0.85, 0.025, 2.5)$							
	$T = 500$				$T = 2000$			
	ϕ_1	ϕ_2	σ_y	σ_v	ϕ_1	ϕ_2	σ_y	σ_v
GMM-6M-E	0.8756	0.5273	0.1158	2.0796	0.8268	0.6287	0.1595	2.0601
GMM-6M-E-R	0.8628	0.4996	0.1158	2.0796	0.8000	0.5608	0.1595	2.0601
GMM-6M-NW	1.1262	0.8471	2.2744	0.8664	1.0421	0.7413	2.2072	0.6990
GMM-6M-NW-R	1.1237	0.8436	2.2744	0.8664	1.0421	0.7413	2.2072	0.6990
GMM-24M-E	0.6349	1.3978	0.5772	6.8450	0.3983	1.5975	0.6545	6.6632
GMM-24M-E-R	0.7613	1.1746	0.5772	6.8450	0.6201	1.3403	0.6545	6.6632
GMM-24M-NW	1.2920	0.8144	3.6846	1.5207	1.1021	0.7327	3.7249	1.1020
GMM-24M-NW-R	1.2900	0.8101	3.6846	1.5207	1.1021	0.7327	3.7249	1.1020
Bayes-1	0.2017	0.7453	3.7717	1.9174	0.3480	1.0969	8.7618	2.1449
Bayes-2	0.1824	0.2981	0.0176	0.4300	0.1408	0.3096	0.0072	0.4754
EDV	1.0498	0.0847	1.0031	1.8172	0.9293	0.1124	1.1515	1.8843
EDV-R	1.2065	0.3540	4.8938	2.4661	1.0175	0.2597	6.3289	2.4868
ARMA-SV	0.0336	0.0395	**0.0020**	**0.1868**	**0.0164**	0.0186	**0.0010**	**0.0930**
R-ARMA-SV	0.0336	0.0395	**0.0020**	**0.1868**	**0.0164**	0.0186	**0.0010**	**0.0930**
W-ARMA-SV-OLS($J = 10$)	0.0331	**0.0302**	**0.0020**	0.1902	0.0168	**0.0155**	**0.0010**	0.0980
W-ARMA-SV-OLS($J = 100$)	**0.0330**	0.0306	**0.0020**	0.1969	0.0170	0.0157	**0.0010**	0.0993

Simulation Results - SV(2)

Number of inadmissible values / non-convergence and relative computing time

	Panel A: Inadmissible values / non-convergence (out of 500)				Panel B: Relative computing time with respect to the ARMA-SV estimator	
	$T = 500$		$T = 2000$			
Estimators	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_1$	$\mathcal{M}_2$	$T = 500$	$T = 2000$
GMM-6M-E	23	6	25	20	734.81	717.67
GMM-6M-NW	60	3	13	0	1019.49	1467.50
GMM-24M-E	13	477	39	498	1752.19	1785.62
GMM-24M-NW	63	5	16	0	3091.74	4059.37
Bayes-1	33	7	9	3	55750.12	127080.14
Bayes-2	0	0	0	0	440922.67	1146786.30
EDV	98	497	88	498	0.99	0.99
ARMA-SV	12	0	0	0	1.00	1.00
W-ARMA-SV-OLS($J = 10$)	0	0	0	0	1.38	1.36
W-ARMA-SV-OLS($J = 100$)	0	0	0	0	8.42	7.46

Outline

Empirical Results

Evidence of Higher-order Stochastic Volatility: W-ARMA-SV-OLS Estimates

S&P 500 index, 1928 - 2016, number of observations: 23372				
	$p = 1$		$p = 2$	
	Coefficient	Std. Error	Coefficient	Std. Error
ϕ_1	0.9938***	(0.0357)	0.6887***	(0.0719)
ϕ_2			0.2863***	(0.0734)
σ_y	0.3356***	(0.0167)	0.3356***	(0.0167)
σ_v	0.6533***	(0.0623)	0.6166***	(0.3204)
	$p = 3$		$p = 4$	
	Coefficient	Std. Error	Coefficient	Std. Error
ϕ_1	0.5477***	(0.1204)	0.3633***	(0.2183)
ϕ_2	-0.4264***	(0.0936)	-0.0251	(0.2142)
ϕ_3	0.6211***	(0.0122)	0.6305***	(0.0174)
ϕ_4			0.0005	(0.0169)
σ_y	0.3356***	(0.0167)	0.3356***	(0.0142)
σ_v	0.6211**	(0.3993)	0.6133	(0.5456)

***, **, and * indicate significance at 1%, 5%, and 10% levels

Empirical Results

Finite Sample Inference: Monte Carlo Tests

- The p-values tabulated in the earlier tables are based on the asymptotic approximation.
- Simulation studies reveal that asymptotic p-values can be markedly different and may be quite unreliable in finite samples:
  - Ahsan (2020) in the context of stochastic volatility model
  - Schwert (2002) in the context of ARMA models
  - Park and Mitchell (1980), Miyazaki and Griffiths (1984), Dejong, Nankervis, Savin and Whiteman (1992) in the context of AR(1) models
- To tackle this problem, we propose Local Monte Carlo tests / parametric Bootstrap in order to make a reliable inference.

Empirical Results

Finite Sample Inference: Local Monte Carlo / Parametric Bootstrap Tests

$S\&P$ 500 index, 1928 - 2016, number of observations: 23372

	$p = 1$					
	Coefficient	S_0	Asy. tests	LMC tests		
				$N = 19$	$N = 99$	$N = 999$
ϕ_1	0.9938	27.84	0.00	0.05	0.01	0.001
σ_y	0.3356	20.06	0.00	0.05	0.01	0.001
σ_v	0.6533	10.48	0.00	0.05	0.01	0.001
Time (in seconds)			0.69	1.5	4.5	38.2
	$p = 2$					
	Coefficient	S_0	Asy. tests	LMC tests		
				$N = 19$	$N = 99$	$N = 999$
ϕ_1	0.6887	9.58	0.00	0.05	0.01	0.001
ϕ_2	0.2863	3.90	0.00	0.05	0.01	0.001
σ_y	0.3356	20.06	0.00	0.05	0.01	0.001
σ_v	0.6166	1.92	0.03	0.10	0.06	0.075
Time (in seconds)			0.70	2.6	10.3	95.1

Empirical Results

Finite Sample Inference: Local Monte Carlo / Parametric Bootstrap Tests

$S\&P$ 500 index, 1928 - 2016, number of observations: 23372						
	$p=3$					
	Coefficient	S_0	Asy. tests	LMC tests		
				$N=19$	$N=99$	$N=999$
ϕ_1	0.5477	4.55	0.00	0.05	0.01	0.001
ϕ_2	-0.4264	-4.55	0.00	0.05	0.01	0.001
ϕ_3	0.8489	69.67	0.00	0.05	0.01	0.001
σ_y	0.3356	20.06	0.00	0.05	0.01	0.001
σ_v	0.6211	1.56	0.06	0.10	0.09	0.082
Time (in seconds)			0.79	12.2	60.2	622.1

Empirical Results

Finite Sample Inference: Local Monte Carlo / Parametric Bootstrap Tests

$S\&P$ 500 index, 1928 - 2016, number of observations: 23372						
	$p=4$					
	Coefficient	S_0	Asy. tests	LMC tests		
				$N=19$	$N=99$	$N=999$
ϕ_1	0.3633	1.69	0.05	0.05	0.01	0.001
ϕ_2	-0.0251	-0.12	0.45	0.85	0.88	0.865
ϕ_3	0.6305	37.68	0.00	0.05	0.01	0.001
ϕ_4	0.0005	0.03	0.49	0.70	0.65	0.623
σ_y	0.3356	20.06	0.00	0.05	0.01	0.001
σ_v	0.6133	0.67	0.25	0.20	0.15	0.185
Time (in seconds)			0.97	20.7	105.0	1237.2

Outline

Concluding Remarks

- The simple W-ARMA-SV estimator uniformly outperforms other estimators in terms of statistical efficiency and time.
- Our results cast doubt on the advice that one should use a large number of moments, consistent with Buse (1992), Bekker (1994), Chao and Swanson (2007).
  —Using too many moments can be very costly from an efficiency viewpoint.
- Empirical results can be summarized as follows:
  1. Daily returns could be better modeled as an SV(p) model.
  2. It is easy to construct simulation-based inference for SV(p) models using simple estimators.
  3. SV(p) models are superior in forecasting daily volatility.