Financial Data Analysis

Volatility Models for Return Series

Roman Liesenfeld
(University of Kiel)

October 8, 2008
# Contents

3 Volatility Models for Return Series 3

3.1 ARCH-Models ......................................................... 8
  3.1.1 Specification and Properties ..................................... 8
  3.1.2 Estimation of ARCH Models .................................... 14
3.2 GARCH-Models ......................................................... 18
  3.2.1 Specification and Properties ..................................... 18
  3.2.2 Estimation of GARCH models .................................... 19
3.3 Stochastic Volatility (SV) Model ..................................... 22
  3.3.1 Specification and Properties ..................................... 22
  3.3.2 Generalized Method of Moment (GMM) Estimation .............. 24
  3.3.3 Monte Carlo Markov Chain (MCMC) estimation .................. 29
3 Volatility Models for Return Series

In this chapter, we consider alternative models for the volatility of asset returns, i.e., for the conditional variance/standard deviation of returns.

Important reasons to study the behavior of the volatility are:

1. The volatility is an important factor for option pricing.

   Consider, e.g., the Black–Scholes option pricing formula for a European Call option:

   \[ C_t = P_t \Phi (x_t) - K r^{-\ell} \Phi \left( x_t - \sigma_t \sqrt{\ell} \right), \]

   where

   \[ x_t = \frac{\ln(P_t/Kr^{-\ell})}{\sigma_t \sqrt{\ell}} + \frac{1}{2} \sigma_t \sqrt{\ell} \]

   \[ P_t \] = current price of the underlying stock

   \[ r \] = risk-free interest rate

   \[ K \] = strike price

   \[ \ell \] = time to expiration

   \[ \Phi(\cdot) \] = cdf of a \( N(0,1) \)–distribution

   \[ \sigma_t \] = conditional standard derivation of the return of the underlying stock.

   Hence, the volatility represented by \( \sigma_t^2 \) plays a critical role in option pricing.

2. Volatility is also important for risk management:

   \[ \bullet \text{ Volatility is a measure for risk and} \]

   \[ \bullet \text{ relevant for calculating the value at risk of a financial position.} \]

3. Models for the conditional variance can be used to increase the efficiency of parameter estimation; in particular, in situations where a (F)GLS estimator is used to account for heteroscedasticity due to varying volatility.

Here, we will consider two classes of volatility models:

\[ \bullet \text{ ARCH (autoregressive conditional heteroscedasticity) models} \]

\[ \bullet \text{ SV (stochastic volatility) models.} \]

Lets start with some properties of volatility:

\[ \bullet \text{ The volatility of asset returns is not directly observable. Crude measures are the absolute or squared returns.} \]
• There exist volatility clusters, i.e., volatility tend to be high for certain time periods and low for other periods.

See, e.g., Figure 1 which plots the daily log-returns of IBM stock (1973–1991) and of S&P500 (1980–1986).

• This volatility clustering is typically reflected in a significant positive autocorrelation of squared or absolute returns.

Hence, the return series \( \{ r_t \} \) is either serially uncorrelated or with minor lower order serial correlation, but it is dependent:

See, e.g., Figure 2 which shows the ACF of daily log-returns and squared log-returns of IBM stock (1973–1991) and S&P500 (1980–1986).
Figure 1: Time plot of daily log returns of IBM stock from January 1, 1973 to December 31, 1991 (upper panel) and of S&P500 from January 2, 1980 to December 12, 1986 (lower panel).
Figure 2: Sample ACF of daily log returns of IBM (upper left panel) of S&P500 (upper right panel), and Sample ACF of squared daily log returns of IBM (lower left panel) of S&P500 (lower right panel)
Let's turn to the basic structure of volatility models:

- The key elements of volatility models are assumptions w.r.t. the conditional mean and variance of returns, i.e.,

\[
\mu_t = E[r_t | \mathcal{F}_{t-1}]
\]

\[
\sigma_t^2 = \text{var}[r_t | \mathcal{F}_{t-1}] = E\left[ (r_t - \mu_t)^2 | \mathcal{F}_{t-1} \right],
\]

where

\( \mathcal{F}_{t-1} \) = Information set available in \( t - 1 \), typically past returns.

- The typical assumption for the conditional mean \( \mu_t \) is that of a stationary (low-order) ARMA model such that:

\[
r_t = \mu_t + \varepsilon_t, \quad \text{with} \quad \mu_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i},
\]

\[E[\varepsilon_t | \mathcal{F}_{t-1}] = 0,\]

where \( \varepsilon_t \) is the "shock" or "error term" or mean corrected return. Remember, serial correlation in returns is weak if it is exist at all.

- For the conditional variance we obtain:

\[
\sigma_t^2 = \text{var}[r_t | \mathcal{F}_{t-1}] = \text{var}[\varepsilon_t | \mathcal{F}_{t-1}] = E\left[ \varepsilon_t^2 | \mathcal{F}_{t-1} \right].
\]

Volatility models now specify the manner how \( \sigma_t^2 \) evolves over time.
3.1 ARCH-Models

3.1.1 Specification and Properties

The class of ARCH-models was introduced by Engle (1982)\(^1\) to account for the autoregressive structure of the heteroscedasticity in macroeconomic data. Especially, it has proven to be useful to model the volatility of

- inflation
- interest rates
- stock market returns
- foreign exchange rates

**Specification**

The ARCH(\(m\))-model assumes for the return shocks (mean corrected returns)

\[
\varepsilon_t = \sigma_t u_t \quad \text{with} \quad \sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_m \varepsilon_{t-m}^2,
\]

where

\[
u_t \sim \text{iid}, \quad \text{with} \quad E[u_t] = 0, \quad E[u_t^2] = 1.
\]

and

\[
\alpha_0 > 0, \quad \alpha_i \geq 0.
\]

In practice, \(u_t\) is assumed to follow a \(N(0, 1)\)- or a standardized student-\(t\) distribution.

- Observe that

  - the first two conditional moments of return shocks under a ARCH(\(m\))-model are

  \[
  E[\varepsilon_t | F_{t-1}] = E[\sigma_t u_t | F_{t-1}] = \sigma_t E[u_t | F_{t-1}] = 0
  \]

  \[
  E[\varepsilon_t^2 | F_{t-1}] = E[\sigma_t^2 u_t^2 | F_{t-1}] = \sigma_t^2 E[u_t^2 | F_{t-1}] = \sigma_t^2;
  \]

  - the restrictions \(\alpha_0 > 0, \alpha_i \geq 0\) guarantee that \(\sigma_t^2 > 0\);

  - the specification for \(\sigma_t^2\) implies an autoregressive structure for \(\sigma_t^2\).

  Large past squared shocks \(\{\varepsilon_{t-1}^2\}\) due to large past conditional variances \(\{\sigma_{t-1}^2\}\) imply a large current conditional variance \(\sigma_t^2\).

  Hence, the \(\varepsilon_t\)'s are serially dependent.

Properties of ARCH-Models

To discuss the salient features of ARCH-models it is convenient to focus on the ARCH(1):

\[ \varepsilon_t = \sigma_t u_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2. \]

- The unconditional mean is:
  \[ \mathbb{E} [\varepsilon_t] = \mathbb{E} [\mathbb{E} [\varepsilon_t \mid \mathcal{F}_{t-1}]] = \mathbb{E} [\mathbb{E} [u_t \mid \mathcal{F}_{t-1}]] = 0. \]

- The unconditional variance is:
  \[
  \mathbb{E} [\varepsilon_t^2] = \mathbb{E} [\mathbb{E} [\varepsilon_t^2 \mid \mathcal{F}_{t-1}]] = \mathbb{E} [\mathbb{E} [\sigma_t^2 u_t^2 \mid \mathcal{F}_{t-1}]] \\
  = \mathbb{E} [\sigma_t^2] = \alpha_0 + \alpha_1 \mathbb{E} [\varepsilon_{t-1}^2]
  \]

and under weak stationarity

\[ \mathbb{E} [\varepsilon_t^2] = \frac{\alpha_0}{1 - \alpha_1}, \quad 0 \leq \alpha_1 < 1, \]

where the restriction ensures the positivity of the variance.

- The autocovariance of the shocks is

\[
\text{cov} (\varepsilon_t, \varepsilon_{t-\tau}) = \mathbb{E} [\varepsilon_t \varepsilon_{t-\tau}] = \mathbb{E} [\sigma_t u_t \sigma_{t-\tau} u_{t-\tau}] \\
= \mathbb{E} [\mathbb{E} (\sigma_t u_t \mid \mathcal{F}_{t-\tau-1}) \sigma_{t-\tau} \mathbb{E} (u_{t-\tau} \mid \mathcal{F}_{t-\tau-1})] \\
= 0 \quad \forall \tau \neq 0.
\]

Hence, \( \varepsilon_t \) is serially uncorrelated.

- However, \( \varepsilon_t \) is serially dependent since, e.g., \( \varepsilon_t^2 \) is serially correlated, which will be shown now:

Consider the regression model for \( \varepsilon_t^2 \):

\[
\varepsilon_t^2 = \mathbb{E} [\varepsilon_t^2 \mid \mathcal{F}_{t-1}] + w_t = \sigma_t^2 + w_t \\
= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + w_t,
\]

with

\[
w_t = \varepsilon_t^2 - \sigma_t^2 = \sigma_t^2 u_t^2 - \sigma_t^2 = \sigma_t^2 (u_t^2 - 1).\]
It follows that
\[
\begin{align*}
E[w_t] &= E \left[ \sigma_t^2 \left( E \left[ u_t^2 \mid \mathcal{F}_{t-1} \right] - 1 \right) \right] = 0 \\
E[w_t w_{t-\tau}] &= E \left[ \sigma_t^2 (u_t^2 - 1) \sigma_{t-\tau}^2 (u_{t-\tau}^2 - 1) \right] \\
&= E \left[ E(\sigma_t^2(u_t^2 - 1)|\mathcal{F}_{t-\tau-1}) \sigma_{t-\tau}^2 E(u_{t-\tau}^2 - 1|\mathcal{F}_{t-\tau-1}) \right] = 0
\end{align*}
\]

Therefore, under a stationary ARCH(1), \(w_t\) represents a WN process and \(\varepsilon_t^2\) has a AR(1) representation with an ACF (see, Chapter, 1.2):
\[
\rho_\tau = \alpha_1^\top \quad \tau \geq 0.
\]

- Hence, the ARCH(1) (and higher order ARCH(m)-models) can “explain” the fact that we observe
  - autocorrelation in squared asset returns
  - but typically no or very weak autocorrelation in the returns (see, e.g., Figure 2).

- Figure 3 shows an artificially simulated time series together with the sample ACF for \(\varepsilon_t\) and \(\varepsilon_t^2\) for the following ARCH(1) model
  \[
  \varepsilon_t = \sigma_t u_t, \quad u_t \sim \text{iid } N(0, 1), \quad \sigma_t^2 = 0, 1 + 0, 8 \sigma_{t-1}^2
  \]
Figure 3: Time plot of a simulated Gaussian ARCH(1)-model (upper left panel) and the corresponding sample ACF of the shocks $\varepsilon_t$ (upper right panel) and squared $\varepsilon_t^2$ (lower panel).
- Let's turn to the unconditional fourth moment of the shocks $\varepsilon_t$, which is an important measure for the tail behavior of the return distributions.

  - The conditional fourth moment is:
    \[
    E [\varepsilon_t^4 | F_{t-1}] = E [\sigma_t^4 u_t^4 | F_{t-1}] = E [u_t^4 | F_{t-1}] \cdot \sigma_t^4 = E [u_t^4] \left( \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 \right)^2.
    \]
    (Note, that the existence of $E[\varepsilon_t^4]$ requires the existence of $E[u_t^4]$ which is satisfied, e.g., for Gaussian $u_t$’s with $E[u_t^4] = 3$).

  - Therefore, the unconditional moment is
    \[
    E [\varepsilon_t^4] = E [u_t^4] E \left[ \left( \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 \right)^2 \right] = E [u_t^4] \left( \alpha_0^2 + 2 \frac{\alpha_1 \alpha_0^2}{1-\alpha_1} + \alpha_1^2 E [\varepsilon_{t-1}^4] \right).
    \]
    If $\varepsilon_t$ is fourth order stationary ($E [\varepsilon_t^4] = E [\varepsilon_{t-1}^4]$), then
    \[
    E [\varepsilon_t^4] = \frac{E [u_t^4] \left( \alpha_0^2 + 2 \frac{\alpha_1 \alpha_0^2}{1-\alpha_1} \right)}{1 - E [u_t^4] \alpha_1^2} = \frac{E [u_t^4] (1 + \alpha_1) \alpha_0^2}{1 - \alpha_1 (1 - E [u_t^4] \alpha_1^2)}.
    \]

  - Since, $E[\varepsilon_t^4]$ must be positive, $\alpha_1$ must also satisfy (in addition to $0 \leq \alpha_1 < 1$):
    \[
    1 - \alpha_1^2 E [u_t^4] > 0 \iff \alpha_1^2 < \frac{1}{E [u_t^4]}.
    \]

  - Finally, for the unconditional kurtosis we obtain:
    \[
    K = \frac{E \left[ (\varepsilon_t - E [\varepsilon_t])^4 \right]}{E \left[ (\varepsilon_t - E [\varepsilon_t])^2 \right]^2} = E \left[ \frac{\varepsilon_t^4}{E [\varepsilon_t^4]} \right] = \frac{E [u_t^4] \left( 1 + \alpha_1 \right) \alpha_0^2}{(1 - \alpha_1) \left( 1 - E [u_t^4] \alpha_1^2 \right)} \cdot \frac{(1 - \alpha_1)^2}{\alpha_0^2} = \frac{E [u_t^4] \left( 1 - \alpha_1^2 \right)}{1 - E [u_t^4] \alpha_1^2} > E [u_t^4].
    \]
    Hence, for a Gaussian $u_t$ with $E [u_t^4] = 3$, e.g., we have
    \[
    K > 3,
    \]
    so that a Gaussian ARCH(1) can explain the excess kurtosis typically observed in the empirical distribution of asset returns (see, e.g., Figure 4, Chap. 1.1).
- All these properties of the ARCH(1) also hold for higher order ARCH(m)-models, but the formulas become more complicated.

- In applications, the order of the ARCH(m) must be determined.
  Since under aARCH(m), $\varepsilon_t^2$ has an AR(m) representation, we can use the tools for identifying AR-models to determine $m$, such as the PACF (see, Chapter 1.2.).

<table>
<thead>
<tr>
<th>Weaknesses of ARCH–Models</th>
</tr>
</thead>
</table>

1. It requires rather strong restrictions on the parameters. (For instance, to ensure the positivity of the second and fourth moment $\alpha_1$ in the Gaussian ARCH(1) must satisfy: $0 < \alpha_1^2 < \frac{1}{3}$);

2. ARCH(m) models typically require a large $m$ to account for the observed volatility clustering and hence a rather large number of parameters.

3. ARCH models only provide a mechanical way to describe the volatility behavior. If gives no indication about what causes such behavior to occur. (Note that this is a feature of all univariate time series models).
3.1.2 Estimation of ARCH Models

ARCH models are typically estimated by ML-estimation, which produces asymptotically efficient estimates if the model is correctly specified.

- Consider the following Gaussian-ARCH$(m)$ for the observable returns $\{r_t\}_{t=1}^T$:
  \[ r_t = \varepsilon_t, \quad \varepsilon_t \mid \mathcal{F}_{t-1} \sim N(0, \sigma_t^2) \]
  \[ \sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_m \varepsilon_{t-m}^2. \]

- The likelihood function is given by the product of the conditional densities:
  \[ f(\varepsilon_1, \ldots, \varepsilon_T; \theta) = f(\varepsilon_T \mid \mathcal{F}_{T-1}; \theta) f(\varepsilon_{T-1} \mid \mathcal{F}_{T-2}; \theta) \cdots f(\varepsilon_{m+1} \mid \mathcal{F}_m; \theta) f(\varepsilon_m, \ldots, \varepsilon_1; \theta), \]
  where
  \[ \theta = (\alpha_0, \ldots, \alpha_m)', \quad f(\varepsilon_m, \ldots, \varepsilon_1, \theta) : \text{joint pdf of } (\varepsilon_1, \ldots, \varepsilon_m). \]

Since $f(\varepsilon_m, \ldots, \varepsilon_1; \theta)$ is typically complicated, it is commonly dropped, which can be justified in large samples.

- Therefore, for the Gaussian-ARCH$(m)$, we obtain the following conditional likelihood:
  \[ f(\varepsilon_{m+1}, \ldots, \varepsilon_T \mid \varepsilon_m, \ldots, \varepsilon_1; \theta) = \prod_{t=m+1}^{T} \frac{1}{\sqrt{2\pi \sigma_t^2}} \exp \left\{ -\frac{\varepsilon_t^2}{2\sigma_t^2} \right\}, \]
  where $\{\sigma_t^2\}$ can be calculated recursively for a given value of $\theta$.

- Let
  \[ x_m = (\varepsilon_{m+1}, \ldots, \varepsilon_T)' \]
  then the log-likelihood is:
  \[ \ell(\theta; x_m) = \ln f(x_m \mid \varepsilon_m, \ldots, \varepsilon_1, \theta) \]
  \[ = - \sum_{t=m+1}^{T} \left[ \frac{1}{2} \ln \sigma_t^2 + \frac{1}{2} \ln (2\pi) + \frac{1}{2} \frac{\varepsilon_t^2}{\sigma_t^2} \right], \]
  and the first order conditions for the ML-estimator $\hat{\theta}$ are:
  \[ \frac{\partial \ell(\hat{\theta}; x_m)}{\partial \theta} = 0, \]

\footnote{The assumption $E[r_t \mid \mathcal{F}_{t-1}] = \mu_t = 0$ is imposed for expositional convenience. All considerations also hold for $\mu_t \neq 0$ with $\varepsilon = r_t - \mu_t$.}
with
\[
\frac{\partial \ell(\cdot)}{\partial \theta} = \sum_{t=m+1}^{T} \left( \frac{\varepsilon_t^2 - \sigma_t^2}{2\sigma_t^4} \right) \cdot \begin{pmatrix} 1 \\ \varepsilon_{t-1}^2 \\ \vdots \\ \varepsilon_{t-m}^2 \end{pmatrix}.
\]

- Observe, that the first derivatives are non-linear in \( \theta \). Therefore, the ML-estimates does not have an explicit form and the ML-estimates must be obtained by numerical and iterative optimization procedures.

Such procedures are typically implemented in statistical software packages.

**Model checking**

- The estimated (fitted) model must be examined to check whether the model is adequate or not.

If the model is adequate, then the standardized shocks
\[
\frac{\hat{\varepsilon}_t}{\hat{\sigma}_t} = u_t \quad \text{are iid.}
\]

Hence, the standardized residuals (the estimated counterpart of \( u_t \))
\[
\frac{\hat{\varepsilon}_t}{\hat{\sigma}_t} = \hat{u}_t \quad \text{with} \quad \hat{\varepsilon}_t = r_t - \hat{\mu}_t
\]
and
\[
\hat{\sigma}_t^2 = \hat{\alpha}_0 + \hat{\alpha}_1 \hat{\varepsilon}_{t-1}^2 + \cdots + \hat{\alpha}_m \hat{\varepsilon}_{t-m}^2
\]

as well as the squared standardized residuals \( \{\hat{u}_t^2\} \) should be serially uncorrelated.

- To test this implication the Box-Pierce or Ljung-Box statistic for \( \{\hat{u}_t\} \) and \( \{u_t^2\} \) can be used.

**Example**

To illustrate ARCH modelling, I have build an ARCH model for the daily log returns of IBM stock (from January 1, 1973 to December 31, 1991).

- The sample PACF for the squared returns (not presented here) indicates that a ARCH(8) might be appropriate. Hence, I specify the following model:
\[
\begin{align*}
\varepsilon_t &= \mu + \sigma_t u_t, \quad u_t \sim \text{iid } \mathcal{N}(0,1), \\
\sigma_t^2 &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_8 \varepsilon_{t-8}^2
\end{align*}
\]
- I obtained the following fitted model:

\[
\begin{align*}
\hat{r}_t & = 0.034 + \sigma_t u_t \\
\sigma_t^2 & = 0.890 + 0.109 \varepsilon_{t-1} + 0.100 \varepsilon_{t-1}^2 + 0.075 \varepsilon_{t-3} + 0.113 \varepsilon_{t-4}^2 \\
& \quad + 0.052 \varepsilon_{t-5}^2 + 0.064 \varepsilon_{t-6}^2 + 0.033 \varepsilon_{t-7}^2 + 0.065 \varepsilon_{t-8}^2,
\end{align*}
\]

where the \( t \)-statistics for \( H_0 \) that the corresponding coefficient is nil are given in parentheses. \(^3\)

- The following table gives sample autocorrelation of squared returns \( r_t^2 \) and of squared standardized residuals \( \hat{u}_t^2 \) obtained under the fitted ARCH(8) together with the corresponding Ljung–Box statistics:

<table>
<thead>
<tr>
<th>lag ( \ell )</th>
<th>Autocorrelation ( r_t^2 )</th>
<th>Ljung–Box (( p )-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r_t^2 )</td>
<td>( \hat{u}_t^2 )</td>
<td>( r_t^2 )</td>
</tr>
<tr>
<td>1</td>
<td>0.157</td>
<td>-0.004</td>
</tr>
<tr>
<td>2</td>
<td>0.076</td>
<td>-0.002</td>
</tr>
<tr>
<td>3</td>
<td>0.025</td>
<td>-0.006</td>
</tr>
<tr>
<td>4</td>
<td>0.023</td>
<td>-0.010</td>
</tr>
<tr>
<td>5</td>
<td>0.080</td>
<td>-0.008</td>
</tr>
<tr>
<td>10</td>
<td>0.009</td>
<td>-0.004</td>
</tr>
</tbody>
</table>

- Figure 4 shows the estimated conditional standard derivation \( \{\hat{\sigma}_t\} \) obtained under a ARCH(8) model.

---

\(^3\)The \( t \)-statistic are based on the robust Bollerslev–Woolridge standard errors (see, e.g., H94, p.663). These are based on the result, that even if \( u_t \) is actually non-Gaussian, then under certain regularity conditions

\[
\hat{\theta} \overset{d}{\sim} N(\theta, \frac{1}{T} D^{-1} S D),
\]

where

\[
S = \operatorname{plim} \frac{1}{T} \sum_{t=1}^{T} \frac{\partial \ln f(\varepsilon_t | \mathcal{F}_{t-1}, \theta)}{\partial \theta} \frac{\partial \ln f(\varepsilon_t | \mathcal{F}_{t-1}, \theta)}{\partial \theta},
\]

\[
D = \operatorname{plim} \frac{1}{T} \sum_{t=1}^{T} -E \left[ \frac{\partial^2 \ln f(\varepsilon_t | \mathcal{F}_{t-1}, \theta)}{\partial \theta \partial \theta} \mid \mathcal{F}_{t-1} \right].
\]
Figure 4: Time plot of daily log returns of IBM stock from January 1, 1973 to December 31, 1991, and the estimated conditional standard deviation under an ARCH(8)-model.
3.2 GARCH-Models

3.2.1 Specification and Properties

A disadvantage of ARCH($m$) models is that they typically require many parameters to describe the volatility process (see, the example in section 3.1.2).

Bollerslev (1986)\(^4\) proposed a generalization of ARCH models, the generalized ARCH (GARCH) which typically requires less parameters than ARCH Models to describe volatility.

**Specification**

- The most simple GARCH is the GARCH(1,1), which assumes for the return shocks (mean corrected returns)

\[ \varepsilon_t = \sigma_t u_t \quad \text{with} \quad \sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \]

where

\[ u_t \sim \text{iid}, \quad E[ u_t ] = 0, \quad E[ u_t^2 ] = 1 \]

and

\[ \alpha_0 > 0, \quad \alpha_1 \geq 0, \quad \beta_1 \geq 0, \quad (\alpha_1 + \beta_1) < 1. \]

- Observe that

  - the restrictions \( \alpha_0 > 0, \quad \alpha_1 \geq 0, \quad \beta_1 \geq 0 \) ensure the positivity of the variance \( \sigma_t^2 \), while \((\alpha_1 + \beta_1) < 1 \) implies a finite unconditional variance of \( \varepsilon_t \).

  - Furthermore, similar to ARCH models, the GARCH(1,1) can be represented as an ARMA model for \( \varepsilon_t^2 \).

Consider the regression model for \( \varepsilon_t^2 \) (see, section 3.1.1.):

\[ \varepsilon_t^2 = \underbrace{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2} + w_t, \quad \sigma_t^2 = E[\varepsilon_t^2 | F_{t-1}] \]

with

\[ w_t = \varepsilon_t^2 - \sigma_t^2 = \sigma_t^2 (u_t^2 - 1). \]

It follows that

\[ \varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 (\varepsilon_{t-1}^2 - w_{t-1}) + w_t = \alpha_0 + (\alpha_1 + \beta_1) \varepsilon_{t-1}^2 + w_t - \beta_1 w_{t-1}, \]

with

\[ E[ w_t ] = 0, \quad E[ w_t w_{t-\tau} ] = 0 \ \forall \ \tau \neq 0. \]

Hence, \( \varepsilon_t^2 \) has a ARMA(1,1) representation.

Properties of a GARCH(1,1)

- The unconditional mean is: \( E[\varepsilon_t] = 0. \)

- The unconditional variance is obtained from the ARMA representation as follows:
  \[
  E[\varepsilon_t^2] = \alpha_0 + (\alpha_1 + \beta_1)E[\varepsilon_{t-1}^2];
  \]
  under weak stationarity:
  \[
  E[\varepsilon_t^2] = \frac{\alpha_0}{1 - (\alpha_1 + \beta_1)}, \quad (\alpha_1 + \beta_1) < 1,
  \]
  where the restriction ensures the positivity of the variance.

- As in the ARCH-model, \( \varepsilon_t \) is serially uncorrelated, while it can be shown that the ACF of \( \varepsilon_t^2 \) has the form:\(^5\)
  \[
  \rho_\tau = \begin{cases} \frac{\alpha_1(1-\alpha_1-\beta_1^2)}{1-2\alpha_1-\beta_1^4}, & \text{if } \tau = 1 \\ (\alpha_1 + \beta_1)^{\tau-1} \rho_1, & \text{if } \tau > 1 \end{cases},
  \]
  provided that \( \varepsilon_t^2 \) is weakly stationary with \( E[\varepsilon_t^4] < \infty. \)

- The conditional variance of a general GARCH\((m, s)\) model is
  \[
  \sigma_t^2 = \alpha_0 + \sum_{i=1}^{m} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{s} \beta_i \sigma_{t-i}^2,
  \]
  implying the following ARMA representation for \( \varepsilon_t^2 \)
  \[
  \varepsilon_t^2 = \alpha_0 + \sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_i) \varepsilon_{t-i}^2 + w_t - \sum_{i=1}^{s} \beta_i w_{t-i}.
  \]

3.2.2 Estimation of GARCH models

The ML-estimation of a GARCH\((m, s)\)-model is implemented analogously to an ARCH\((m)\) model:

\(^5\)The ACF of \( \varepsilon_t^2 \) under a GARCH\((1,1)\) can be obtained as the ACF of a stationary ARMA\((1,1)\) model.
- Consider the Gaussian–GARCH ($m, s$)

$$
\varepsilon_t \mid \mathcal{F}_{t-1} \sim N(0, \sigma_t^2) \\
\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_m \varepsilon_{t-m}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_s \sigma_{t-s}^2.
$$

- The conditional log-likelihood is

$$
\ell(\theta, x_m) = - \sum_{t=m+1}^{T} \left[ \frac{1}{2} \ln \sigma_t^2 + \frac{1}{2} \ln (2\pi) + \frac{1}{2} \frac{\varepsilon_t^2}{\sigma_t^2} \right],
$$

with

$$
x_m = (\varepsilon_{m+1}, \ldots, \varepsilon_T)' \\
\theta = (\alpha_0, \alpha_1, \ldots, \beta_s)'.
$$

- The starting values for $\sigma_t^2$:

$$
\sigma_j^2, \quad j = 0, -1, -2,
$$

can be set, e.g., equal to the estimate of the unconditional variance:

$$
\frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^2.
$$

**Example**

- I estimated a Gaussian–GARCH(1,1) for the daily log returns of IBM stock and obtained the following fitted model:

$$
r_t = 0.037 + \sigma_t u_t \\
\sigma_t^2 = 0.037 + 0.065 \varepsilon_{t-1}^2 + 0.921 \sigma_{t-1}^2,
$$

where the robust Bollerslev–Wooldridge $t$–statistics are given in parentheses.

- Note that the fitted model shows

$$
\hat{\alpha}_1 + \hat{\beta}_1 = 0.986,
$$

which is close to 1, indicating a rather slow decay of the ACF of $\varepsilon_t^2$. This phenomenon is commonly observed in practice and indicates a strong persistence of volatility shocks.
- The following table gives sample autocorrelations of squared standardized residuals $\hat{u}_t^2$ obtained under the fitted GARCH(1,1), together with Ljung–Box statistics:

<table>
<thead>
<tr>
<th>lag $\ell$</th>
<th>Autocorrelation $\hat{u}^2_t$</th>
<th>Ljung–Box ($p$-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.012</td>
<td>0.659 (0.417)</td>
</tr>
<tr>
<td>2</td>
<td>0.005</td>
<td>0.772 (0.680)</td>
</tr>
<tr>
<td>3</td>
<td>-0.003</td>
<td>0.806 (0.848)</td>
</tr>
<tr>
<td>4</td>
<td>0.001</td>
<td>0.809 (0.937)</td>
</tr>
<tr>
<td>5</td>
<td>-0.004</td>
<td>0.902 (0.970)</td>
</tr>
<tr>
<td>10</td>
<td>-0.013</td>
<td>3.419 (0.970)</td>
</tr>
</tbody>
</table>

- I come to some concluding remarks about ARCH and GARCH.

  - The GARCH model has been tremendously successful in empirical work and is regarded as the benchmark model for volatility of asset returns.

  - The literature on (G)ARCH models is enormous and contains a very large number of extensions of the standard (G)ARCH specification (see, e.g. TO2, chapter 3):
    - exponential GARCH (EGARCH)
    - GARCH – in mean
    - integrated GARCH (IGARCH), ...

  - The GARCH model encounters the same weaknesses as the ARCH model:
    - it requires parameter restrictions,
    - it only provides a “statistical description” of volatility.
3.3 Stochastic Volatility (SV) Model

3.3.1 Specification and Properties

The SV model can be obtained from the so called mixture of distribution hypothesis introduced by Clark (1973)\(^6\) and Tauchen and Pitts (1983)\(^7\).

- Let
  \[ \varepsilon_t = \text{daily (mean corrected) log-return}. \]

  Suppose that this is the sum of \(n_t\) intraday log-returns, i.e.:
  \[ \varepsilon_t = \sum_{j=1}^{n_t} \delta_{jt}, \]

  where

  \( \delta_{jt} = \text{\(j\)th intraday log-return (at day \(t\)) assumed to be iid } N(0, \sigma^2) \)

  \( n_t = \text{number of intraday stock price changes triggered by the arrival of new information} \)

  (representing the flow of new information, which is a latent variable).

- To capture the uneven flow of information, \(n_t\) is assumed to be a positive (unobservable) random variable, independent from \(\delta_{jt}\).

  Under these assumptions the conditional distribution of \(\varepsilon_t\) is characterized by

  \[ \operatorname{E} [\varepsilon_t \mid n_t] = 0, \quad \operatorname{var} [\varepsilon_t \mid n_t] = \sigma^2 n_t \]

  and hence,

  \[ \varepsilon_t \mid n_t \sim N \left( 0, \sigma^2 n_t \right). \]

- An alternative representation is

  \[ \varepsilon_t = \sigma_t u_t \quad \text{with} \quad \sigma_t = \sqrt{\sigma^2 n_t}, \quad u_t \sim \text{iid} N(0,1), u_t \Pi n_t; \]

  The standard SV model assumes that the latent log volatility follows a Gaussian AR(1)

  \[ \ln \sigma_t^2 = \gamma + \delta \ln \sigma_{t-1}^2 + \nu v_t, \quad v_t \sim \text{iid} N(0,1). \]

- Observe that this SV model has a similar form as the (G)ARCH models.

---


- However, the main difference between SV and (G)ARCH models is:

  under a (G)ARCH: \( \sigma_t^2 \) is an observable deterministic function of past returns 
  \( \{ \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots \} \) (observation driven model);
  under a SV: \( \sigma_t^2 \) depends on a latent innovation \( v_t \), and hence \( \sigma_t^2 \),
  itself is latent (parameter driven model).

The introduction of the innovation \( v_t \) substantially increases the flexibility of the model in describing volatility process, but it also increases the difficulty in parameter estimation (see below).

Properties of the SV-model

- Defining \( \lambda_t =: \ln \sigma_t^2 \), then the SV model can be written as
  \[
  \varepsilon_t = e^{\lambda_{t}/2} u_t, \quad u_t \sim \text{iidN}(0, 1)
  \]
  \[
  \lambda_t = \gamma + \delta \lambda_{t-1} + \nu v_t, \quad v_t \sim \text{iidN}(0, 1).
  \]

- For \( |\delta| < 1 \), the Gaussian AR(1) process \( \lambda_t \) is stationary with the following unconditional distribution:
  \[
  \lambda_t \sim N(\mu_\lambda, \sigma_\lambda^2), \quad \text{with} \quad \mu_\lambda = \frac{\gamma}{1 - \delta}, \quad \sigma_\lambda^2 = \frac{\nu^2}{1 - \delta^2}.
  \]
  Thus, the volatility \( \sigma_t^2 = e^{\lambda_t} \) and \( \sigma_t = e^{\lambda_t/2} \) are log-normal random variables.

- Using the properties of the log-normal distribution and of the higher-order moments of a standard normal distribution, we obtain for the \( r \)-th even unconditional moment:
  \[
  E[\varepsilon_t^r] = E[e^{\frac{r}{2} \lambda_t} u_t^r] = E[e^{\frac{r}{2} \lambda_t}] E[u_t^r]
  \]
  \[
  = \left[ e^{\frac{r}{2} \mu_\lambda + \frac{1}{2}(\frac{r}{2})^2 \sigma_\lambda^2} \right] \frac{r!}{2^{r/2} (\frac{r}{2})!}, \quad r = 2, 4, \ldots.
  \]

- Note, that all odd unconditional moments are zero, because a Gaussian \( u_t \) implies \( E[u_t^s] = 0, \forall s = 1, 3, \ldots \)

- The unconditional kurtosis is then given by:
  \[
  K = \frac{E[(\varepsilon_t - E[\varepsilon_t])^4]}{E[(\varepsilon_t - E[\varepsilon_t])^2]^2} = \frac{E[\varepsilon_t^4]}{E[\varepsilon_t^2]^2} = 3e^{\sigma_\lambda^2} \geq 3,
  \]
  so that the SV model can explain the excess kurtosis of empirical return distributions.

23
- Under the SV model $\varepsilon_t$ is serially uncorrelated, whereas, e.g., $\varepsilon_t^2$ exhibits serial correlation.

In particular, the ACF of $\varepsilon_t^2$ is obtained from the fact that\(^8\)

$$E\left[\varepsilon_t^2 \varepsilon_{t-\tau}^2\right] = E\left[e^{\lambda_t} e^{\lambda_{t-\tau}}\right] E\left[u_t^2\right] E\left[u_{t-\tau}^2\right] = E\left[e^{\lambda_t + \lambda_{t-\tau}}\right]$$

with

$$(\lambda_t + \lambda_{t-\tau}) \sim N(2\mu_\lambda, 2\sigma_\lambda^2 + 2\delta^2 \sigma_\lambda^2).$$

Hence,

$$E\left[\varepsilon_t^2 \varepsilon_{t-\tau}^2\right] = e^{2\mu_\lambda + \sigma_\lambda^2 (1+\delta^2)}$$

and

$$\text{cov}\left(\varepsilon_t^2, \varepsilon_{t-\tau}^2\right) = e^{2\mu_\lambda + \sigma_\lambda^2 (1+\delta^2)} - e^{2\mu_\lambda + \sigma_\lambda^2} = \left(e^{2\mu_\lambda + \sigma_\lambda^2}\right) \left(\sigma_\lambda^2 \delta^2 - 1\right)$$

and therefore

$$\rho_\tau = \frac{\text{cov}\left(\varepsilon_t^2, \varepsilon_{t-\tau}^2\right)}{E\left[\varepsilon_t^4\right] - E\left[\varepsilon_t^2\right]^2} = \frac{\sigma_\lambda^2 \delta^2 - 1}{3e^{2\mu_\lambda} - 1}, \quad \tau \geq 0.$$

Thus, the SV model, which predicts for $\delta > 0$ a positive serial correlation of the $\varepsilon_t^2$'s behaves similar to the GARCH(1,1) model.

### 3.3.2 Generalized Method of Moment (GMM) Estimation

The main difficulty of using SV models is that, unlike with (G)ARCH models, it is difficult to evaluate the likelihood function. Hence, ML estimation is not straightforward.

**Likelihood function**

- The SV model to be estimated is

  $$\varepsilon_t = e^{\lambda_t/2} u_t, \quad u_t \sim \text{iidN}(0, 1), \quad t : 1 \rightarrow T$$

  $$\lambda_t = \gamma + \delta \lambda_{t-1} + \nu v_t, \quad v_t \sim \text{iidN}(0, 1),$$

where $\lambda_t$ is a latent process with a known starting value $\lambda_0$.

\(^8\)Note that

$$E[\varepsilon_t^2 \varepsilon_{t-\tau}^2] = E\left[\varepsilon_t^2 u_t^2 e^{\lambda_{t-\tau}} u_{t-\tau}^2 | u_{t-\tau}\right] = E\left[\varepsilon_t^2 | u_{t-\tau}\right] \cdot E[u_t^2 | u_{t-\tau}] \cdot E[u_{t-\tau}^2].$$
- Let

\[ E_T = (\varepsilon_1, ..., \varepsilon_T)' : \text{ vector of observable returns} \]
\[ \Lambda_T = (\lambda_1, ..., \lambda_T)' : \text{ vector of latent volatilities} \]
\[ \theta = (\gamma, \delta, \nu)' : \text{ parameters to be estimated.} \]

Then, the likelihood function is obtained by “integrating \( \Lambda_T \) out” of the joint pdf of \( (E_T, \Lambda_T) \):

\[
 f (E_T, \theta) = \int_{\mathbb{R}^T} f (E_T, \Lambda_T; \theta) \, d\Lambda_T \]
\[
 = \int_{\mathbb{R}^T} f (E_T \mid \Lambda_T; \theta) \, f (\Lambda_T; \theta) \, d\Lambda_T,
\]

where the conditional pdf of \( E_T \mid \Lambda_T \) and the marginal pdf of \( \Lambda_T \) are

\[
 f (E_T \mid \Lambda_T; \theta) = \prod_{t=1}^T \frac{1}{\sqrt{2\pi \lambda_t}} e^{-\frac{1}{2} \varepsilon_t^2 e^{-\lambda_t}}
\]
\[
 f (\Lambda_T \mid \theta) = \prod_{t=1}^T \frac{1}{\sqrt{2\pi \nu^2}} e^{-\frac{1}{2} \frac{(\lambda_t - \gamma - \delta \lambda_{t-1})^2}{\nu^2}}.
\]

Hence, the likelihood function is a \( T \)-dimensional integral. An analytical solution to this integration problem is not available. Even an approximation by a quadrature rule is precluded.

- There are different ways to solve this problem to perform estimation; e.g.,
  - application of the (generalized) method of moments;
  - application of efficient importance sampling to evaluate the likelihood by Monte–Carlo integration (ML–EIS);
  - application of a Bayesian procedure based on Monte Carlo Markov Chain (MCMC) to simulate from the posterior distribution of the parameters \( \theta \) and the latent volatilities \( \Lambda_T \).

**Generalized Method of Moments (GMM)**

Using unconditional moments of the returns \( \varepsilon_t \) under the SV model, it is possible to perform GMM estimation of the SV model.
- Examples for unconditional moments which can be used are (see, Chap. 3.3.1):

\[
\begin{align*}
\text{E} [\varepsilon_t^2] &= e^{\mu_\lambda + \sigma_\lambda^2/2} \\
\text{E} [\varepsilon_t^2, \varepsilon_{t-\tau}^2] &= e^{2\mu_\lambda + \sigma_\lambda^2(1+\delta^2)},
\end{align*}
\]

where

\[\mu_\lambda = \gamma/(1-\delta), \quad \sigma_\lambda^2 = \nu^2/(1-\delta^2).\]

- Based on such moments, one can construct a vector of moment restrictions:

\[\text{E} [m(\varepsilon_t, \theta)] = 0,\]

where \(m(\varepsilon_t, \theta)\) is a \(k\)-dimensional moment function, depending on \(\theta\), e.g.,

\[
m(\varepsilon_t, \theta) =
\begin{pmatrix}
|\varepsilon_t| - \sqrt{\frac{\pi}{2}} e^{\mu_\lambda/2+\sigma_\lambda^2/8} \\
e^{\mu_\lambda+\sigma_\lambda^2/2} - e^{\mu_\lambda+\sigma_\lambda^2/2} \\
|\varepsilon_t^3| - 2 \left( \frac{2}{\pi} \right) e^{3\mu_\lambda/2+3\sigma_\lambda^2/8} \\
|\varepsilon_t \varepsilon_{t-1}| - \left( \frac{2}{\pi} \right) e^{\mu_\lambda+(1+\delta^2)\sigma_\lambda^2/4} \\
|\varepsilon_t \varepsilon_{t-3}| - \left( \frac{2}{\pi} \right) e^{\mu_\lambda+(1+\delta^3)\sigma_\lambda^2/4} \\
|\varepsilon_t \varepsilon_{t-5}| - \left( \frac{2}{\pi} \right) e^{\mu_\lambda+(1+\delta^5)\sigma_\lambda^2/4} \\
e^{\mu_\lambda+(1+\delta^2)\sigma_\lambda^2} - e^{2\mu_\lambda+(1+\delta^2)\sigma_\lambda^2} \\
e^{\mu_\lambda+(1+\delta^4)\sigma_\lambda^2} - e^{2\mu_\lambda+(1+\delta^4)\sigma_\lambda^2}
\end{pmatrix}.
\]

- Note, that \(\text{E} [m(\theta, \varepsilon_t)]\) cannot be observed, but (for a given \(\theta\)) it can be consistently estimated by its sample counterpart:

\[m_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} m(\varepsilon_t, \theta).\]

- The GMM estimate for \(\theta\) minimizes a weighted distance between the theoretical moments and their empirical counterparts, i.e.:

\[\hat{\theta} = \arg \min_\theta \quad m_T(\theta)' \cdot W_T \cdot m_T(\theta),\]

where

\[W_T = \text{positive definite } (k \times k) \text{ weighting matrix.}\]

- The weighting matrix \(W_T\) should reflect the relative importance given to matching each of the moments. For \(W = I\), e.g., all moments receive the same weight.
- Lets turn to a review of the properties of the GMM-estimates:

  - Under certain regularity conditions (see, e.g., H94, Chap. 14 and G03, Chap. 18)

    \[
    \begin{align*}
    \text{plim } \hat{\theta} &= \theta \quad \text{(consistency)} \\
    \hat{\theta} &\sim N \left( \theta, \text{asy.var}[\hat{\theta}] \right) \quad \text{(asymptotic normality)}. 
    \end{align*}
    \]

    Note that this holds for any weighting matrix \( W_T \).

  - For \( k > \text{dim } (\theta) \), i.e. situations with more moments than parameters, the system is overidentified, and \( \hat{\theta} \) depends on \( W_T \) (and for different \( W_T \)'s, we obtain different consistent estimates of \( \theta \)).

  - The optimal weighting matrix \( W_T^* \) is that \( W_T \) which maximizes the asymptotic efficiency of the GMM estimator (i.e., minimizes the asymptotic variance of the estimator).

    It can be shown that the optimal weighting matrix is given by:

    \[
    W_T^* = \left\{ \text{asy.var} \left[ \sqrt{T} m_T(\theta) \right] \right\}^{-1} = \left\{ \text{asy.var} \left[ \sqrt{T} \frac{1}{T} \sum_{t=1}^{T} m(\varepsilon_t, \theta) \right] \right\}^{-1}. 
    \]

    - This implies: The larger the variation of a particular moment in \( m \), the smaller its weight in the GMM-objective function.

  - Furthermore, note that, e.g.,

    \[
    \varepsilon_t^2 \quad \text{and} \quad |\varepsilon_t|
    \]

    are typically serially correlated. Hence, the vector process

    \[
    \{m(\varepsilon_t, \theta)\}
    \]

    containing such elements is presumably serially correlated.

    Hence, a consistent estimate of \( W_T^* \) is the Newey-West-estimate:

    \[
    W_T^* = \left\{ \Gamma_0(\hat{\theta}) + \sum_{j=1}^{q} \left( \frac{q-j}{j} \right) \left( \Gamma_j(\hat{\theta}) + \Gamma_j(\hat{\theta})' \right) \right\}^{-1}
    \]

    with

    \[
    \begin{align*}
    \Gamma_j(\theta) &= \frac{1}{T} \sum_{t=0}^{T-j} m(\varepsilon_t, \theta) m(\varepsilon_{t+j}, \theta)' \quad j = 0,1,2,\ldots \\
    q &= \text{bandwidth parameter} \\
    (q-j)/j &= \text{"Bartlett kernel" weighting the autocorrelation at different lags} \\
    \hat{\theta} &= \text{consistent estimate of } \theta.
    \end{align*}
    \]
• Observe that before we can estimate $\theta$, we need an estimate of $W_T^*$, but before we can estimate $W_T^*$, we need an estimate of $\theta$.

In practice, one can use the following iterative procedure:

□ An initial estimate $\hat{\theta}_{(0)}$ is obtained by using $W_T = I$.

□ $\hat{\theta}_{(0)}$ is used to produce an initial estimate of the optimal weighting matrix $W_T^*$ to arrive at a new GMM estimate $\hat{\theta}_{(1)}$.

□ This process is iterated until $\hat{\theta}_{(j)} \simeq \hat{\theta}_{(j+1)}$

• It can be shown that the asymptotic distribution of the GMM estimator based on the optimal weighting matrix is

$$\hat{\theta} \overset{a}{\sim} N \left( \theta, \frac{1}{T} (D_T W_T^{*-1} D'_T)^{-1} \right),$$

where:

$$D_T = \frac{\partial m_T(\theta)}{\partial \theta'}.$$

- **Drawbacks** of the GMM-estimation of SV-models:

  • GMM does not deliver estimated values of the volatilities $\{\sigma_t^2\}$, in contrast, to the ML-estimates of (G)ARCH-models (see chap. 3.2.2).

    Remember that $\sigma_t^2$ is an important variable for the construction of diagnostic tests.

  • The asymptotic variance and hence the asymptotic efficiency of the GMM estimator depends on the selection of moment restrictions.

    In general, one should use lower-order moments with low variation. However, it remains unclear which and how many moments should be used to estimate the SV-model.

- **Advantages** of the GMM-estimation:

  • It’s easy to implement relative to other approaches, like MCMC.

  • It allows for a specification test based on the test of overidentifying restrictions (Hansen’s $J$-Test).
Example

- I estimated a SV-model for the mean-corrected daily log-returns

\[ \varepsilon_t = r_t - \frac{1}{T} \sum_{t=1}^{T} r_{t-1}. \]

I used the set of 8 moment restrictions given above and obtained the following fitted model

\[ \varepsilon_t = e^{\lambda_t/2} u_t, \quad \lambda_t = 0.017 + 0.960\lambda_{t-1} + 0.161 v_t, \]

where asymptotic standard errors are given in parentheses.\(^9\)

- Note that the fitted model shows

\[ \hat{\delta} = 0.960, \]

which is close to 1, indicating a slow decay of the ACF of \( \varepsilon_t^2 \), similar to the fitted GARCH(1,1)--model.

3.3.3 Monte Carlo Markov Chain (MCMC) estimation

- An alternative to GMM estimation of the SV-model is the MCMC approach with the following attractive features:

  1. It allows to obtain estimated values of the latent volatilities;
  2. it allows to conduct exact finite sample inference
     (remember that GMM inference, e.g., is an asymptotic inference which is valid only for large \( T \)).

- MCMC is a Bayesian estimation approach using simulation techniques based on pseudo random numbers.

- MCMC enables us to make some statistical inference that was not feasible just a few years ago.
  (Remember the integration problem to be solved for the likelihood evaluation of the SV-model; this is a problem which often occurs in the econometrics of non-linear models.)

- The basic element of the MCMC analysis of the SV-model is the joint posterior distribution of \( (\theta, A_T) \):

\(^9\)The estimates are based on the Newey-West estimate of the optimal weighting matrix using a Bartlett kernel with a bandwidth of 5 lags.
\[ f(\theta, \Lambda_T | E_T), \]

where

\[ E_T = \text{vector of observable returns} \]
\[ \Lambda_T = \text{vector of latent volatilities} \]
\[ \theta = \text{parameters}. \]

- Note that the parameters in \( \theta \) are interpreted as random variables and not as “fixed quantities”.

- MCMC uses Markov Chain simulations to produce pseudo random draws from the joint posterior distribution.

- In the following, I review basic concepts which are relevant for MCMC:

  - elements of Bayesian inference
  - Markov processes and Markov chain simulation
  - Gibbs sampling
  - Conjugate prior distributions.

**Elements of Bayesian inference**

- In contrast to the classical inference based on the ML-principle, the Bayesian inference interprets parameters to be estimated as random variables.

- Then, the Bayesian inference combines prior beliefs about the parameters with information about them provided by the data.

In particular, let

\[ \theta : \text{vector of unknown parameters} \]
\[ X : \text{data} \]
\[ f(X | \theta) : \text{likelihood of the data for a given model} \]

Then, the prior belief about \( \theta \) is expressed by specifying for \( \theta \) a
prior distribution $= p(\theta)$.

• Now, a combination of prior beliefs with the data is operationalized by using Bayesian theorem to obtain the conditional pdf of $\theta \mid X$:

$$f(\theta \mid X) = \frac{f(\theta, X)}{f(X)} = \frac{f(X \mid \theta)p(\theta)}{f(X)}$$

with

$$f(X) = \int f(\theta, X)d\theta = \int f(X \mid \theta)p(\theta)d\theta.$$ 

In a Bayesian context

□ the distribution $f(\theta \mid X)$ is the posterior of $\theta$, which is often written as

$$f(\theta \mid X) \propto f(X \mid \theta)p(\theta).$$

□ $f(\theta \mid X)$ can be interpreted as a mixture of prior beliefs and current information in the data.

• A popular “Bayesian estimator” of $\theta$ is the mean of the posterior.$^{10}$

$$E[\theta \mid X] = \int \theta \frac{f(X \mid \theta)p(\theta)}{\int f(X \mid \theta)p(\theta)d\theta} d\theta.$$ 

• The evaluation of such quantities requires integration

  - either exact (when closed form solutions exist),

  - by approximation using a quadrature rule,

  - or by MC methods.

• Well, for the Bayesian analysis of the SV-model, we must integrate using MC methods, since exact integration and quadrature rules are not feasible since

the likelihood $f(X \mid \theta)$ in the posterior has no closed form solution!

$^{10}$Other estimators are the median or the mode of $f(\theta \mid X)$, see, e.g., MGB74, Chap. 7.
MC–Integration and Markov Chain simulation

• Suppose our aim is to evaluate

\[ E[g(\theta)], \quad \theta \sim f(\theta) \]

where \( g(\cdot) \) is some function (e.g., the mean of a posterior).

• Furthermore, suppose that we can generate a simulated random sample from \( f(\theta) \).\(^{11}\) Suppose, e.g., that \( \theta \) is continuous with a cdf \( F(\theta) \). Then, a sample of size \( n \) from \( f(\theta) \) is obtained by

\[ \tilde{u}_j \sim U[0, 1]; \quad i : 1 \to n \quad \text{(available on standard PC’s)} \]

\[ \text{and calculating} \quad \tilde{\theta}_i = F^{-1}(\tilde{u}_i). \]

Then \( \tilde{\theta}_i, (i = 1, ..., n) \) are random draws from \( f(\theta) \).

• Then, an approximation of \( E[g(\theta)] \) is

\[ E[g(\theta)] \simeq \frac{1}{n} \sum_{j=1}^{n} g(\tilde{\theta}_j) = \bar{g}, \]

where

\[ \{\tilde{\theta}_j\}_{j=1}^{n} : \quad \text{simulated random sample from} \quad f(\theta). \]

So the population mean is approximated the sample mean.

• When the \( \tilde{\theta}_j \)'s are independent, then a law of large numbers ensures that

\[ \bar{g} \xrightarrow{a.s.} E[g(\theta)] \quad \text{as} \quad n \to \infty. \]

and

\[ \text{var} \left[ \frac{1}{n} \sum_{j=1}^{n} g(\tilde{\theta}_j) \right] = \frac{1}{n} \text{var}[g(\tilde{\theta}_j)] \to 0 \quad \text{as} \quad n \to \infty. \]

Note that this ensures that the accuracy of the approximation increases with \( n \).

\(^{11}\)How to simulate random variables from a particular distribution is discussed in more details, e.g., in Go8, Appendix E2, or Mo6, Section 6.8.
• **Problem:**
  Independent drawing from $f(\theta)$ is very often not feasible, since $f(\theta)$ has an unknown form (such as the joint posterior for the parameters of the SV-model).

• **Solution:**
  Use Markov–Chain simulation based on a Markov process having $f(\theta)$ as its stationary distribution.

  This is then MCMC.

### Markov–Process

• A stochastic process $\{\theta_j; \theta_j \in \Theta\}$ is a Markov process (Markov chain) if its conditional distribution satisfies:

$$P(\theta_j \in A \mid \theta_{j-1}, \theta_{j-2}, \ldots) = P(\theta_j \in A \mid \theta_{j-1}), \quad A \subset \Theta,$$

so that given $\theta_{j-1}$ the values of $\theta_j$ are independent of $\theta_{j-2}, \theta_{j-3}, \ldots$.

• The function

$$P(\cdot \mid \cdot) = \text{“transition probability function”}.$$

  If $P(\cdot \mid \cdot)$ does not depend on $j$, the Markov chain is (time)-homogeneous.

• One can show that under certain regularity conditions the conditional distribution of $\theta_j \mid \theta_0$ satisfies:\footnote{The regularity conditions are weak, they essentially require that for an arbitrary starting value $\theta_0$, the Markov chain have the chance to visit the full space of $\theta$, see Tierney (1994), “Markov Chain for exploring posterior distributions,” The Annals of Statistics, vol. 21, 1701-1762.}

$$\lim_{j \to \infty} P^{(j)}(\theta_j \in A \mid \theta_0) = \bar{P}(\theta_j \in A),$$

where

$$\bar{P}(-) = \text{unique stationary distribution}$$

$$\theta_0 = \text{starting value of the chain}.$$

So as $j$ increases the impact of $\theta_0$ disappears and the sampled $\{\theta_j\}$ look increasingly like dependent samples from $\bar{P}(\cdot)$.

• Such a Markov chain can be used to construct an MCMC estimate of

$$\mathbb{E}[g(\theta)], \quad \theta \sim f(\theta)$$

as follows:
Suppose we have a Markov Chain with a stationary distribution \( f(\theta) \).

Then, use a simulated dependent sample from this chain to construct the following estimate:

\[
\tilde{g} = \frac{1}{n-m} \sum_{j=m+1}^{n} g(\tilde{\theta}_j),
\]

where

\[
\tilde{\theta}_j \sim P(\cdot | \cdot) \\
\{\tilde{\theta}_0, \ldots, \tilde{\theta}_m\} : \text{ first } m + 1 \text{ draws (burn-in sample)}.
\]

Note that the burn-in sample is discarded. If \( m \) is sufficiently large this ensures that \( \{\tilde{\theta}_{m+1}, \ldots, \tilde{\theta}_n\} \) is a dependent sample from \( f(\theta) \).

Then, according to the ergodic theorem, we have

\[
\tilde{g} \xrightarrow{a.s.} E[g(\theta)] \quad \text{as} \quad n \to \infty.
\]

\( \tilde{g} \) is called Markov Chain simulation or MCMC estimate of \( E[g(\theta)] \).

**Gibbs sampling**

- Gibbs sampling is the most popular MCMC method and works as follows:

  Suppose that for \( \theta = (\theta^{(1)}, \theta^{(2)}, \theta^{(3)})' \) the joint distribution \( f(\theta) \) is hard to obtain; but the conditional distributions

  \[
  f \left( \theta^{(1)} \mid \theta^{(2)}, \theta^{(3)} \right), \quad f \left( \theta^{(2)} \mid \theta^{(1)}, \theta^{(3)} \right), \quad f \left( \theta^{(3)} \mid \theta^{(1)}, \theta^{(2)} \right)
  \]

  are known or can, at least, be used to draw samples.

Then, the Gibbs sampler proceeds as follows:

1. Select starting values of \( \theta^{(2)}, \theta^{(3)} : \theta^{(2)}_0, \theta^{(3)}_0 \)
(2) Then, draw \( \tilde{\theta}_1^{(1)} \sim f \left( \theta^{(1)} \mid \theta_0^{(2)}, \theta_0^{(3)} \right) \);
then, draw \( \tilde{\theta}_1^{(2)} \sim f \left( \theta^{(2)} \mid \theta_1^{(1)}, \theta_0^{(3)} \right) \);
then, draw \( \tilde{\theta}_1^{(3)} \sim f \left( \theta^{(3)} \mid \theta_1^{(1)}, \theta_1^{(2)} \right) \).

(3) Next, use the new values \( \tilde{\theta}_1 = \left( \tilde{\theta}_1^{(1)}, \tilde{\theta}_1^{(2)}, \tilde{\theta}_1^{(3)} \right)' \) as starting values to repeat
the iteration in (2) and so on.

☐ These iterations produce a sequence of random draws:

\[
\left( \tilde{\theta}_1, \ldots, \tilde{\theta}_m, \tilde{\theta}_{m+1}, \ldots, \tilde{\theta}_n \right),
\]
where, for a sufficiently large \( m \):

\[
\left( \tilde{\theta}_{m+1}, \ldots, \tilde{\theta}_n \right) \sim f \left( \theta^{(1)}, \theta^{(2)}, \theta^{(3)} \right),
\]
i.e. \( \left( \tilde{\theta}_{m+1}, \ldots, \tilde{\theta}_n \right) \) can be used as a sample from the joint distribution of \( \theta \).

☐ Hence, an MCMC estimate of \( E [\theta^{(\ell)}] \), e.g., is

\[
\tilde{E} \left[ \theta^{(\ell)} \right] = \frac{1}{n-m} \sum_{j=m+1}^{n} \tilde{\theta}_j^{(\ell)}, \quad \ell = 1, 2, 3,
\]
and its variance is

\[
\tilde{\text{var}} \left[ \theta^{(\ell)} \right] = \frac{1}{n-m-1} \sum_{j=m+1}^{n} \left( \tilde{\theta}_j^{(\ell)} - \tilde{E} \left[ \theta^{(\ell)} \right] \right)^2.
\]

☐ The importance of the Gibbs–sampler is that it allows to solve a high-dimensional
problem iteratively by using univariate conditional distributions.

**Conjugate prior distributions**

- As mentioned above, obtaining the posterior distribution for a Bayesian analysis

\[
f (\theta \mid X) = \frac{f (X \mid \theta) p (\theta)}{\int f (X \mid \theta) p (\theta) d\theta}
\]
is typically not simple.

- But there are cases in which the prior and the posterior belongs to same family of
distributions, making closed-form solutions for \( f (\theta \mid X) \) available.

Such priors are called: conjugate priors.
In the following, I review some well-known conjugate priors\(^{33}\):

\[\begin{align*}
\square \text{ Example 1:} \\
\text{Let} \\
X = \{x_i, i : 1 \to n\} : \text{random sample from } N(\mu, \sigma^2) \\
\text{with } \mu : \text{unknown} \\
\sigma^2 : \text{known}.
\end{align*}\]

Consider the prior:
\[p(\mu) \sim N(\mu_0, \sigma_0^2).\]

Then the posterior for \(\mu\) is:
\[f(\mu \mid X) \sim N(\mu_s, \sigma_s^2),\]
with
\[\mu_s = \frac{\sigma^2_n \mu_0 + \sigma_0^2 \bar{x}}{\sigma^2_n + \sigma_0^2}, \quad \text{and} \quad \sigma_s^2 = \frac{\sigma^2_n \sigma_0^2}{\sigma^2_n + \sigma_0^2},\]
where\(^{14}\) \(\bar{x} = \left(\sum_{i=1}^n x_i\right) / n\).

Note that
- the mean \(\mu_s\) is a weighted average of the prior mean \(\mu_0\) and sample mean, \(\bar{x}\), which represents the LSE-estimate of \(\mu\):
  \[
  \hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \text{with variance} \quad \text{var} [\hat{\mu}] = \frac{\sigma^2}{n}.
  \]
- the weights depend on the corresponding variances \(\sigma_0^2\) and \(\sigma^2 / n\).

\[\begin{align*}
\square \text{ Example 2: (multivariate extension of Example 1)}
\end{align*}\]

\(^{33}\)For further examples, see, e.g., TO2, section 10.3.2.

\(^{14}\)This result follows, immediately from the fact, that
\[f(\mu \mid X) \propto f(X \mid \mu) p(\mu) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2} \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{2\sigma_0^2} (\mu - \mu_0)^2},\]
where the r.h.s. of the equation is a density kernel of a \(N(\mu_s, \sigma_s^2)\)-distribution, see e.g., MGB74, p. 347–8.\]
Let

\[ X = \{x_i, i : 1 \to n \} : \text{random sample from a multivariate Gaussian distribution } N(\mu, \Sigma), \]

with \( \mu \) : unknown
\( \Sigma \) : known.

Consider the prior

\[ p(\mu) \sim N(\mu_0, \Sigma_0). \]

Then, the posterior for \( \mu \) is

\[ f(\mu | X) \sim N(\mu_*, \Sigma_*), \]

with

\[ \mu_* = \Sigma_* \left( \Sigma_0^{-1} \mu_0 + \mu \Sigma^{-1} \bar{X} \right) \quad \text{and} \quad \Sigma_* = \left( \Sigma_0^{-1} + n \Sigma^{-1} \right)^{-1}, \]

where \( \bar{X} = (\sum_{i=1}^{n} x_i) / n. \)

\[ \Box \text{Example 3:} \]

Let

\[ X = \{x_i, i : 1 \to n \} : \text{random sample from } N(0, \sigma^2) \]
\( \text{with } \sigma^2 \) : unknown.

Consider the prior:

\[ p(\sigma^2) \sim \frac{1}{\chi^2(\nu)} \text{ (inverted chi-square, with } \nu \text{ d.o.f)}^{15}, \]

i.e.:

\[ \frac{\nu \lambda}{\sigma^2} \sim \chi^2(\nu), \quad \lambda > 0. \]

Then, the posterior of \( \sigma^2 \) is:

\[ f(\sigma^2 | X) \sim \frac{1}{\chi^2(\nu + n)}, \]

i.e.:

\[ 1^{15} \text{The pdf of a random variable } z \text{ following an inverted chi-square distribution with } \nu \text{ d.o.f. is} \]
\[ f(z; \nu) = \frac{2^{-\nu/2}}{\Gamma(\nu/2)} z^{-(\nu + 1)/2} e^{-1/2 z}, \quad z > 0, \]

with \( E[z] = 1/(\nu - 2) \) and \( \text{var}[z] = \frac{2}{(\nu - 2)^2(\nu - 4)}. \)
\[
\frac{\nu \lambda + \sum_{i=1}^{n} x_i^2}{\sigma^2} \sim \chi^2(\nu+n).
\]

**Implementation of MCMC for SV models**

- Let’s turn to the implementation of MCMC to estimate the following SV model for log-returns \( r_t \):

\[
\begin{align*}
    r_t &= \mu + \varepsilon_t, \\
    \varepsilon_t &= \sqrt{\sigma_t^2} u_t, \\
    \ln \sigma_t^2 &= \gamma + \delta \ln \sigma_{t-1}^2 + \eta_t
\end{align*}
\]

where

\[
    u_t \sim \text{iid } N(0,1), \quad \eta_t \sim \text{iidN } (0, \nu^2).
\]

- Let

\[
    \begin{align*}
    R_T &= (r_1, \ldots, r_T)' : \text{observed returns} \\
    \Sigma_T &= (\sigma_1^2, \ldots, \sigma_T^2)' : \text{latent volatilities} \\
    \omega &= (\gamma, \delta, \nu^2)' : \text{volatility parameters}
    \end{align*}
\]

- Now, under a Bayesian framework the parameter vector \((\mu, \omega')\) can be augmented by \(\Sigma_T\)

\[
    \theta = (\mu, \omega', \Sigma_T)' : \text{augmented parameter vector.}
\]

- Hence, a Gibbs sampling approach can be used to estimate \(\theta\), which involves drawing from the following conditional posterior distributions:

\[
    f(\mu \mid R_T, \Sigma_T, \omega), \quad f(\Sigma_T \mid R_T, \mu, \omega), \quad f(\omega \mid R_T, \Sigma_T, \mu).
\]

- This allows us to produce a sample from the joint posterior

\[
    (\mu_j, \omega_j, \Sigma_{T_j}) \sim f(\mu, \omega, \Sigma_T \mid R_T)
\]

with point estimates of parameters:

\[
    \hat{\omega} = \frac{1}{n-m} \sum_{i=m+1}^{n} \omega_i; \quad \hat{\mu} = \frac{1}{n-m} \sum_{i=m+1}^{n} \mu_i.
\]
Let’s turn to the practical implementation based on 3 components using the 3 conditional post-in distributions given above.

**Component 1: draws from \( f(\mu \mid R_T, \Sigma_T, \omega) \)**

- First note, that given \( \Sigma_T \), the regression model for \( r_t \) can be rewritten as

\[
\frac{r_t}{\sigma_t} = \mu \frac{1}{\sigma_t} + u_t, \quad u_t \sim \text{iidN}(0, 1) \\
r_t^* = \mu x_t^* + u_t
\]

with a LS estimate of \( \mu \):

\[
\hat{\mu} = \frac{\sum_{t=1}^{T} x_t^* r_t^*}{\sum_{t=1}^{T} (x_t^*)^2}
\]

and

\[
\text{var}[\hat{\mu}] = 1/\sum_{t=1}^{T} (x_t^*)^2.
\]

- Now suppose, a conjugate normal prior for \( \mu \) is used

\[
p(\mu) \sim N(\mu_0; a_0^2).
\]

- Then, the posterior of \( \mu \) is (see Example 1)

\[
f(\mu \mid R_T, \Sigma_T, \omega) \sim N(\mu_*, a_*^2)
\]

with:

\[
\mu_* = \frac{\mu_0/\sum_t (x_t^*)^2 + a_0^2 (\sum_t x_t^* r_t^*)/ \sum_t (x_t^*)^2}{1/\sum_t (x_t^*)^2 + a_0^2} \\
a_*^2 = \frac{a_0^2/\sum_t (x_t^*)^2}{a_0^2 + 1/\sum_t (x_t^*)^2}
\]

which can be used to draw \( \mu_ks \) given \( R_T, \Sigma_T, \omega \).

**Component 2: draws from \( f(\Sigma_T \mid R_T, \mu, \omega) \)**
• First note the problem:
The pdf of $\Sigma_T \mid (\cdot)$ has no closed-form solution. Hence, we can’t directly draw from this posterior.

• A solution is:
Use a ("multimove") sampler to draw $\Sigma_T$ element by element from the conditional posterior:

$$f (\sigma_i^2 \mid R_T, \mu, \omega, \Sigma_{\ell}) , \text{ where } \Sigma_{\ell} : \Sigma_T \text{ without } \sigma_i^2 .$$

• This means: For a vector of starting values $\Sigma_{T,0}$ draw

$$\sigma_{1,1}^2 \text{ from } f (\sigma_1^2 \mid R_T, \mu, \omega, \Sigma_{\ell,0}) ;$$
then, draw

$$\sigma_{2,1}^2 \text{ from } f (\sigma_2^2 \mid R_T, \mu, \sigma_{1,1}^2, \Sigma_{\ell(1,2),0})$$
and so on.

• The components of the posterior are a normal pdf and lognormal pdfs$^{16}$

$$f (\sigma_i^2 \mid R_T, \mu, \omega, \Sigma_{\ell})$$

$$\propto k (\sigma_i^2) = f (\varepsilon_i \mid \sigma_i^2, r_i, \mu) \cdot f (\sigma_i^2 \mid \sigma_{i-1}^2, \omega) f (\sigma_{i+1}^2 \mid \sigma_i^2, \omega)$$

normal lognormal lognormal

Note the Problem:
The normalizing constant needed to transform $k(\cdot)$ into a proper density is unknown
so that we cannot directly draw from $f (\sigma_i^2 \mid \cdot)$.

Solution:

$^{16}$This factorization is obtained by deleting redundant conditioning information and keeping the density kernels for $\sigma_i^2$:

$$f (\sigma_i^2 \mid R_T, \mu, \omega, \Sigma_{\ell})$$

$$= f (\sigma_i^2 \mid \varepsilon_i, \sigma_{i-1}^2, \sigma_{i+1}^2)$$

$$= f (\sigma_i^2, \varepsilon_i, \sigma_{i-1}^2, \sigma_{i+1}^2)$$

$$f (\varepsilon_i \mid \sigma_{i-1}^2) f (\sigma_{i+1}^2 \mid \sigma_i^2, \sigma_{i-1}^2)$$

$$\propto f (\sigma_i^2, \varepsilon_i, \sigma_{i-1}^2, \sigma_{i+1}^2) = f (\varepsilon_i \mid \sigma_i^2, \sigma_{i-1}^2) f (\sigma_i^2, \sigma_{i-1}^2, \sigma_{i+1}^2)$$

$$= f (\varepsilon_i \mid \sigma_i^2) f (\sigma_{i+1}^2 \mid \sigma_i^2) f (\sigma_{i+1}^2 \mid \sigma_i^2)$$

$$\propto f (\varepsilon_i \mid \sigma_i^2) f (\sigma_{i+1}^2 \mid \sigma_i^2) f (\sigma_{i+1}^2 \mid \sigma_i^2) .$$
TO 2 propose to use the so called Griddy–Gibbs algorithm to sample from \( f(\sigma_t^2 | \cdot) \). It works as follows:\(^{17}\)

**Step 1:** Select a grid of points for \( \sigma_t^2 \), say:
\[
\sigma_{11}^2 \leq \sigma_{12}^2 \leq \cdots \leq \sigma_{1m}^2
\]
Then, evaluate the kernel \( k(\cdot) \) of the posterior \( f(\sigma_t^2 | \cdot) \) at these points:
\[
w_j = k(\sigma_{ij}^2), \quad j : 1 \rightarrow m
\]

**Step 2:** The \( w_j \)'s can be used to approximate the cdf of \( f(\sigma_t^2 | \cdot) \) by the following discrete probabilities
\[
F_j = \sum_{r=1}^{j} p(\sigma_{tr}^2), \quad \text{with} \quad p(\sigma_{tr}^2) = w_r / (\sum_{j=1}^{m} w_j)
\]
Now, use the inverse of \( F_j \) to obtain an approximation to the inverse cdf of \( \sigma_t^2 | \cdot \).

**Step 3:** Draw uniforms: \( \hat{u}_i \sim U[0, 1] \) and transform the \( \hat{u}_i \)'s via \( F_j^{-1}(\cdot) \) to obtain draws from \( f(\sigma_t^2 | \cdot) \).
(see, page 112 above).

**Component 3:** Draws from \( f(\omega | R_T, \Sigma_T, \mu) \)

- To draw \( \omega | \cdot \), first partition the prior of \( \omega \):
\[
p(\omega) = p(\alpha)p(\nu^2), \quad \alpha = (\gamma, \delta)',
\]
such that we need the following conditional posteriors:
\[
f(\alpha | \nu^2, R_T, \Sigma_T, \mu) \quad \text{and} \quad f(\nu^2 | \alpha, R_T, \Sigma_T, \mu).
\]

- Suppose we use the following conjugate priors:
\[
p(\alpha) \sim N(\alpha_0, C_0) \quad \text{(multivariate Normal)}
p(\nu^2) \sim \frac{1}{\chi^2(m)} \quad \text{(inverted chi-square)}
\]

- Now, observe that given \((\Sigma_T, R_T, \mu, \nu^2)\), \( \ln \sigma_t^2 \) follows an AR(1):
\[
\ln \sigma_t^2 = \gamma + \delta \ln \sigma_{t-1}^2 + \nu_t
\]

with a LS-estimate:

\[
\hat{\alpha} = \left( \sum_{t=2}^{T} z_t z'_t \right)^{-1} \sum_{t=2}^{T} z_t \ln \sigma_t^2; \quad z_t = (1, \ln \sigma_t^2)'
\]

\[
\text{var} [\hat{\alpha}] = \nu^2 \left( \sum_{t=2}^{T} z_t z'_t \right)^{-1}
\]

- Hence, under the multivariate normal prior, the posterior of \( \alpha \) is (see, Example 2):

\[
f (\alpha \mid \nu^2, R_T, \Sigma_T, \mu) \sim N (\alpha_*, C_*)
\]

with

\[
\alpha_* = C_* \left( \frac{\sum_{t=2}^{T} z_t \ln \sigma_t^2}{\nu^2} + C_0^{-1} \alpha_0 \right)
\]

\[
C_* = \left( \frac{\sum_{t=2}^{T} z_t z'_t}{\nu^2} + C_0^{-1} \right)^{-1}
\]

- Finally, observe that given \((\Sigma_T, R_T, \mu, \alpha)\), we can calculate

\[
\eta_t = \ln \sigma_t^2 - \gamma - \delta \ln \sigma_{t-1}^2, \quad t : 2 \rightarrow T
\]

where \(\eta_t \sim N (0, \nu^2)\).

- Hence, under the inverse chi-square prior the posterior of \(\nu^2\) is (see Example 3)

\[
f (\nu^2 \mid \alpha, \Sigma_T, R_T, \mu) \sim \frac{1}{\chi^2_{(m+T-1)}}
\]

i.e.:

\[
\frac{m \lambda + \sum_{t=2}^{T} \eta_t^2}{\nu^2} \sim \chi^2_{(m+T-1)}
\]

**Example**

(see, TO 2 p. 402–423)

Drawbacks of MCMC estimation of SV-models:
• Compared to GMM, MCMC is complicated to implement

• Estimation results depend on
  – selection of priors
  – length of the burn-in sample.

• It’s unclear how long the burn-in sample must be to ensure that we draw from the stationary distribution (typically larger than 5000).

• Many draws, i.e. cycles through the Gibbs sequence are needed to obtain valid results (often more than 20,000).
  This can be very time consuming.