Modeling The Variance of a Time Series

Peter Bloomfield

Department of Statistics
North Carolina State University

July 31, 2009 / Ben Kedem Symposium
Outline

1 Introduction

2 Time Series Models
   - First Wave
   - Second Wave

3 Stochastic Volatility

4 Stochastic Volatility and GARCH
   - A Simple Tractable Model
   - An Application

5 Summary
Ben has made many contributions to time series methodology.

A common theme is that some unobserved (latent) series controls either:

- the values of the observed data, or
- the distribution the observed data.

In a stochastic volatility model, a latent series controls specifically the variance of the observed data.

We relate stochastic volatility models to other time series models.
Modeling The Variance of a Time Series

Peter Bloomfield

Outline

1. Introduction
2. Time Series Models
   - First Wave
   - Second Wave
3. Stochastic Volatility
4. Stochastic Volatility and GARCH
   - A Simple Tractable Model
   - An Application
5. Summary
Time series modeling is not just about correlation...

http://xkcd.com/552/
The time domain approach to modeling a time series \( \{ Y_t \} \) focuses on the conditional distribution of
\[ Y_t \mid Y_{t-1}, Y_{t-2}, \ldots. \]

One reason for this focus is that the joint distribution of
\[ Y_1, Y_2, \ldots, Y_n \] can be factorized as

\[
f_{1:n}(y_1, y_2, \ldots, y_n) = f_1(y_1) f_{2|1}(y_2 \mid y_1) \cdots f_{n|n-1:1}(y_n \mid y_{n-1}, y_{n-2}, \ldots, y_1).
\]

So the likelihood function is determined by these conditional distributions.
The conditional distribution may be defined by:

- the conditional mean,

\[ \mu_t = \mathbb{E}(Y_t \mid Y_{t-1} = y_{t-1}, Y_{t-2} = y_{t-2}, \ldots) ; \]

- the conditional variance,

\[ h_t = \text{Var}(Y_t \mid Y_{t-1} = y_{t-1}, Y_{t-2} = y_{t-2}, \ldots) ; \]

- the shape of the conditional distribution.
Forecasting

The conditional distribution of $Y_t | Y_{t-1}, Y_{t-2}, \ldots$ also gives the most complete solution to the forecasting problem:

- We observe $Y_{t-1}, Y_{t-2}, \ldots$;
- what statements can we make about $Y_t$?

The conditional mean is our best forecast, and the conditional standard deviation measures how far we believe the actual value might differ from the forecast.

The conditional shape, usually a fixed distribution such as the normal, allows us to make probability statements about the actual value.
Modeling The Variance of a Time Series

Peter Bloomfield

Introduction

Time Series Models

First Wave

Second Wave

Stochastic Volatility

Stochastic Volatility and GARCH

A Simple Tractable Model

An Application

Summary
The first wave of time series methods focused on the conditional mean, $\mu_t$.

- The conditional variance was assumed to be constant.
- The conditional shape was either normal or unspecified.

Need only to specify the form of

$$\mu_t = \mu_t(y_{t-1}, y_{t-2}, \ldots).$$

Time-homogeneous:

$$\mu_t = \mu(y_{t-1}, y_{t-2}, \ldots),$$

depends on $t$ only through $y_{t-1}, y_{t-2}, \ldots$. 
Autoregression

- Simplest form:
  \[ \mu_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p}. \]

- Equivalently, and more familiarly,
  \[ y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t, \]

where \( \epsilon_t = y_t - \mu_t \) satisfies

\[ \mathbb{E}(\epsilon_t \mid y_{t-1}, y_{t-2}, \ldots) = 0, \]
\[ \text{Var}(\epsilon_t \mid y_{t-1}, y_{t-2}, \ldots) = h. \]
Recursion

- Problem: some time series need large $p$.
- Solution: recursion; include also some past values of $\mu_t$:
  \[
  \mu_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \psi_1 \mu_{t-1} + \cdots + \psi_q \mu_{t-q}.
  \]
- Equivalently, and more familiarly,
  \[
  Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}.
  \]
- This is the ARMA (AutoRegressive Moving Average) model of order $(p, q)$. 

Equivalently, and more familiarly,
Modeling The Variance of a Time Series

Peter Bloomfield

Introduction

Time Series Models

First Wave

Second Wave

Stochastic Volatility

Stochastic Volatility and GARCH

A Simple Tractable Model

An Application

Summary
Two Years of S&P 500: Changing Variance
The second wave of time series methods added a focus on the conditional variance, $h_t$.

Now need to specify the form of

$$h_t = h_t(y_{t-1}, y_{t-2}, \ldots).$$

Time-homogeneous:

$$h_t = h(y_{t-1}, y_{t-2}, \ldots),$$

depends on $t$ only through $y_{t-1}, y_{t-2}, \ldots$. 
ARCH

- Simplest form: $h_t$ a linear function of a small number of squared $\epsilon$ s:

$$h_t = \omega + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_q \epsilon_{t-q}^2.$$  

- Engle, ARCH (AutoRegressive Conditional Heteroscedasticity):
  - proposed in 1982;
  - Nobel Prize in Economics, 2003 (shared with the late Sir Clive Granger).
Problem: some time series need large $q$.

Solution: recursion; include also some past values of $h_t$:

$$h_t = \omega + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_q \epsilon_{t-q}^2 + \beta_1 h_{t-1} + \cdots + \beta_p h_{t-p}.$$

Bollerslev, 1987; GARCH (Generalized ARCH; no Nobel yet, nor yet a Knighthood).

Warning! note the reversal of the roles of $p$ and $q$ from the convention of ARMA($p$, $q$).
The simplest GARCH model has \( p = 1, q = 1 \):

\[
h_t = \omega + \alpha \epsilon_{t-1}^2 + \beta h_{t-1}
\]

- If \( \alpha + \beta < 1 \), there exists a stationary process with this structure.
- If \( \alpha + \beta = 1 \), the model is called *integrated*: IGARCH(1, 1).
1. Introduction

2. Time Series Models
   - First Wave
   - Second Wave

3. Stochastic Volatility

4. Stochastic Volatility and GARCH
   - A Simple Tractable Model
   - An Application

5. Summary
In a stochastic volatility model, an unobserved (latent) process \( \{X_t\} \) affects the distribution of the observed process \( \{Y_t\} \), specifically the variance of \( Y_t \).

Introducing a “second source of variability” is appealing from a modeling perspective.
For instance:

- \( \{X_t\} \) satisfies
  \[
  X_t - \mu = \phi (X_{t-1} - \mu) + \xi_t,
  \]
  where \( \{\xi_t\} \) are i.i.d. \( N(0, \sigma^2_\xi) \).

- If \( |\phi| < 1 \), this is a (stationary) autoregression, but if \( \phi = 1 \) it is a (non-stationary) random walk.

- \( Y_t = \sigma_t \eta_t \), where \( \sigma_t^2 = \sigma^2(X_t) \) is a non-negative function such as
  \[
  \sigma^2(X_t) = \exp(X_t)
  \]
  and \( \{\eta_t\} \) are i.i.d. \((0, 1)\)–typically Gaussian, but also \( t \).
Conditional Distributions

- So the conditional distribution of $Y_t$ given $Y_{t-1}, Y_{t-2}, \ldots$ and $X_t, X_{t-1}, \ldots$ is simple:

  $$Y_t | Y_{t-1}, Y_{t-2}, \ldots, X_t, X_{t-1}, \ldots \sim N\left(0, \sigma^2(X_t)\right).$$

- But the conditional distribution of $Y_t$ given only $Y_{t-1}, Y_{t-2}, \ldots$ is not analytically tractable.

- In particular,

  $$h_t(y_{t-1}, y_{t-2}, \ldots) = \text{Var}(Y_t | Y_{t-1} = y_{t-1}, Y_{t-2} = y_{t-2}, \ldots)$$

is not a simple function.
Difficulties

- Analytic difficulties cause problems in:
  - estimation;
  - forecasting.

- Computationally intensive methods, e.g.:
  - Particle filtering;
  - Numerical quadrature.
Modeling The Variance of a Time Series

Peter Bloomfield

Introduction

Time Series Models
  - First Wave
  - Second Wave

Stochastic Volatility

Stochastic Volatility and GARCH
  - A Simple Tractable Model
  - An Application

Summary
Stochastic Volatility and GARCH

- Stochastic volatility models have the attraction of an explicit model for the volatility, or variance.
- Is analytic difficulty the unavoidable cost of that advantage?
Modeling The Variance of a Time Series

Peter Bloomfield

Outline

1. Introduction

2. Time Series Models
   - First Wave
   - Second Wave

3. Stochastic Volatility

4. Stochastic Volatility and GARCH
   - A Simple Tractable Model
   - An Application

5. Summary
The Latent Process

We construct a latent process by:

\[ X_0 \sim \Gamma \left( \frac{\nu}{2}, \frac{\tau^2}{2} \right), \]

and for \( t > 0 \)

\[ X_t = B_t X_{t-1}, \]

where

\[ \theta B_t \sim \beta \left( \frac{\nu}{2}, \frac{1}{2} \right) \]

and \( \{B_t\} \) are i.i.d. and independent of \( X_0 \).
The observed process is defined for $t \geq 0$ by

$$Y_t = \sigma_t \eta_t$$

where

$$\sigma_t = \frac{1}{\sqrt{X_t}}$$

and $\{\eta_t\}$ are i.i.d. $N(0, 1)$ and independent of $\{X_t\}$.

Equivalently: given $X_u = x_u, 0 \leq u$, and $Y_u = y_u, 0 \leq u < t$,

$$Y_t \sim N(0, \sigma_t^2)$$

with the same definition of $\sigma_t$. 

Since

\[ \text{Var}(Y_0) = \mathbb{E}(X_0^{-1}), \]

we constrain \( \nu > 2 \) to ensure that

\[ \mathbb{E}(X_0^{-1}) < \infty. \]

Requiring

\[ \mathbb{E}(X_t^{-1}) = \mathbb{E}(X_0^{-1}) \]

for all \( t > 0 \) is also convenient, and is met if

\[ \theta = \frac{\nu - 2}{\nu - 1}. \]
Comparison with Earlier Example

This is quite similar to the earlier example, with $\phi = 1$:

- Write $X_t^* = -\log(X_t)$.
- Then
  \[
  X_t^* = X_{t-1}^* + \xi_t^*,
  \]
  where
  \[
  \xi_t^* = -\log(B_t).
  \]
- In terms of $X_t^*$,
  \[
  \sigma_t^2 = \exp(X_t^*).
  \]
A key constraint is that now $\phi = 1$, so $\{X_t^*\}$ is a (non-stationary) random walk, instead of a (stationary) auto-regression.

Also $\{X_t^*\}$ is non-Gaussian, where in the earlier example, the latent process was Gaussian.

Also $\{X_t^*\}$ has a drift, because

$$E(\xi_t^*) \neq 0.$$ 

Of course, we could include a drift in the earlier example.
Matched Simulated Random Walks
Matched Simulated Random Walks
So What?

- So our model is not very different from (a carefully chosen instance of) the earlier example.
- So does it have any advantage?
- Note: the inverse Gamma distribution is the *conjugate prior* for the variance of the Gaussian distribution.
Marginal distribution of $Y_0$:

$$Y_0 \sim \sqrt{h_0} \, t^*(\nu)$$

where

$$h_0 = \frac{\tau^2}{\nu - 2}$$

and $t^*(\nu)$ is the standardized $t$-distribution (i.e., scaled to have variance 1).
Conditional distributions of $X_0$ and $X_1 \mid Y_0$

- **Conjugate prior/posterior property:** conditionally on $Y_0 = y_0$,
  \[ X_0 \sim \Gamma \left( \frac{\nu + 1}{2}, \frac{\tau^2 + y_0^2}{2} \right). \]

- **Beta multiplier property:** conditionally on $Y_0 = y_0$,
  \[ X_1 = B_1 X_0 \sim \Gamma \left[ \frac{\nu}{2}, \theta \left( \frac{\tau^2 + y_0^2}{2} \right) \right]. \]
The conditional distribution of $X_1 | Y_0$ differs from the distribution of $X_0$ only in scale, so conditionally on $Y_0 = y_0$,

$$Y_1 \sim \sqrt{h_1} \, t^*(\nu),$$

where

$$h_1 = \frac{\theta}{\nu - 2} \left( \tau^2 + y_0^2 \right) = \theta h_0 + (1 - \theta) y_0^2.$$

Hmm...so the distribution of $Y_1 | Y_0$ differs from the distribution of $Y_0$ only in scale...I smell a recursion!
The Recursion

- Write \( Y_{t-1} = (Y_{t-1}, Y_{t-2}, \ldots, Y_0) \).
- For \( t > 0 \), conditionally on \( Y_{t-1} = y_{t-1} \),

\[
Y_t \sim \sqrt{h_t} \, t^*(\nu),
\]

where

\[
h_t = \theta h_{t-1} + (1 - \theta) y_{t-1}^2.
\]
The Structure

That is, \( \{ Y_t \} \) is IGARCH(1, 1) with \( t(\nu) \)-distributed innovations.

Constraints:
- \( \omega = 0; \)
- \( \beta = 1 - \alpha = \frac{\nu-2}{\nu-1}. \)

So we can have a stochastic volatility structure, and still have (I)GARCH structure for the observed process \( \{ Y_t \} \).
The Structure

- That is, \( \{Y_t\} \) is IGARCH(1, 1) with \( t(\nu) \)-distributed innovations.

- Constraints:
  - \( \omega = 0; \)
  - \( \beta = 1 - \alpha = \frac{\nu - 2}{\nu - 1}. \)

- So we can have a stochastic volatility structure, and still have (l)GARCH structure for the observed process \( \{Y_t\} \).

Modeling The Variance of a Time Series

Peter Bloomfield

Introduction
Time Series Models
First Wave
Second Wave
Stochastic Volatility
Stochastic Volatility and GARCH
A Simple Tractable Model
An Application

Same Two Years of S&P 500
Two Years of S&P 500

- Data: 500 log-returns for the S&P 500 index, from 05/24/2007 to 05/19/2009.
- Maximum likelihood estimates:
  \[ \hat{\tau}^2 = 4.37 \]
  \[ \hat{\theta} = 0.914 \]
  \[ \Rightarrow \hat{\nu} = 12.6. \]
- With \( \nu \) unconstrained:
  \[ \hat{\tau}^2 = 3.37 \]
  \[ \hat{\theta} = 0.918 \]
  \[ \hat{\nu} = 9.93. \]
Comparison

- Constrained result has less heavy tails and less memory than unconstrained result.
- Likelihood ratio test:

\[-2 \log(\text{likelihood ratio}) = 0.412\]

assuming \(\sim \chi^2(1)\), \(P = 0.521\), so differences are not significant.

- With more data, difference becomes significant.
Outline

1. Introduction

2. Time Series Models
   - First Wave
   - Second Wave

3. Stochastic Volatility

4. Stochastic Volatility and GARCH
   - A Simple Tractable Model
   - An Application

5. Summary
Summary

- Latent processes are useful in time series modeling.
- GARCH and Stochastic Volatility are both valuable tools for modeling time series with changing variance.
- GARCH fits naturally into the time domain approach.
- Stochastic Volatility is appealing but typically intractable.
- Exploiting conjugate distributions may bridge the gap.
Latent processes are useful in time series modeling.

GARCH and Stochastic Volatility are both valuable tools for modeling time series with changing variance.

GARCH fits naturally into the time domain approach.

Stochastic Volatility is appealing but typically intractable.

Exploiting conjugate distributions may bridge the gap.

Thank you!