# Financial Mathematics

## Stock models: Time series models

A. N. Yannacopoulos

AUEB

Academic year 2023-2024

The binomial model and its limiting form the Black-Scholes model are based on the assumption that returns are independent.

However, real life data often refute this assumption.

This initiates the need to introducing models taking such effects into account: Time series models in which returns at time $t$ are corellated with returns at previous times $t-1, t-2, \cdots, t-k$,

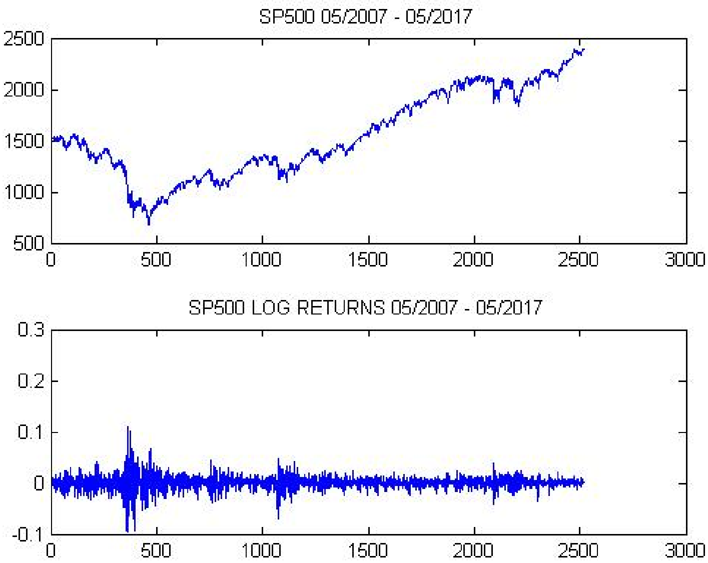
$$y_t = F(y_{t-1}, \cdots, y_{t-k}) + \epsilon$$

An important class of such models are Garch models (which essentially treat the volatility $\sigma^2$ not as a constant but as a stochastic process per se).

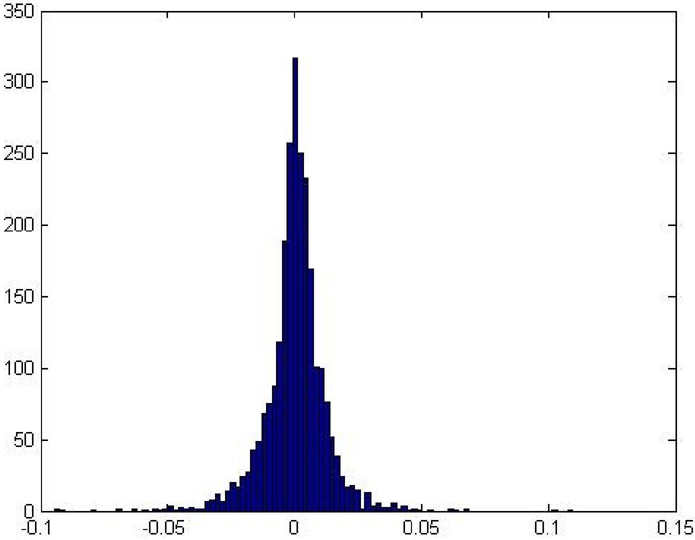
# Stylized facts

Common observations from real life data show that

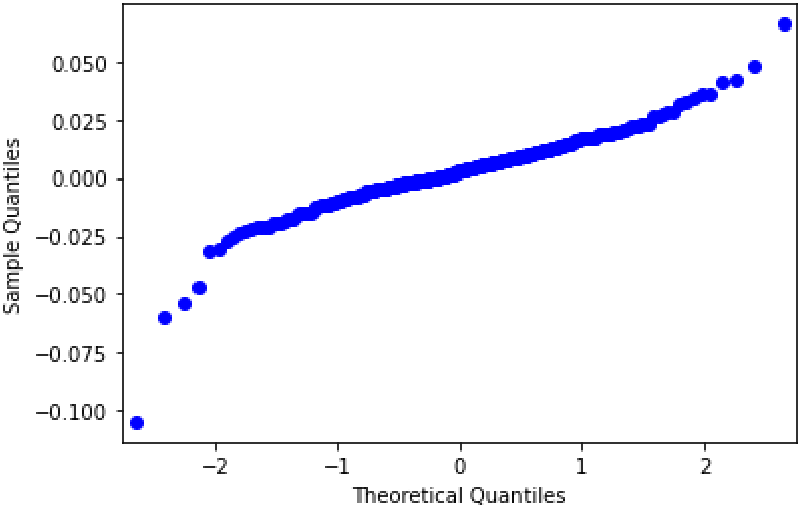
- Returns are correlated
  - Very weak correlations in the actual returns
  - Absolute values or squares of the returns show stronger corellations
- Heavy tails are often observed in the return distributions
- Volatility clustering: Extreme volatility events are often followed by extreme volatility events



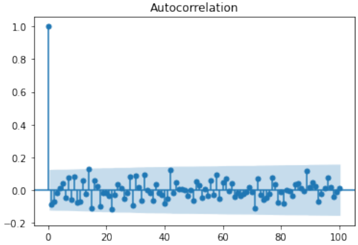

Volatility clustering



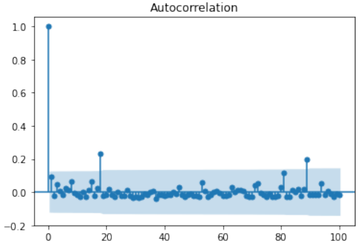

Deviations from normality




Deviations from normality




Autocorrelation function for ln $R_t$




Autocorrelation function for $(\ln R_t)^2$

In the more general case we can consider the returns of a basket of $d$ stocks or indices (probably interrelated) so that the return process

$$\{X(t)\ t \in [0, T]\} = \{(X_1(t), \cdots, X_d)\ :\ t \in [0, T]\}$$

is a vector valued process (where $X_i$ are the returns – or log returns – of stock or index $i$).

Common assumption: The process can be adequately described by the first two moments the mean and the autocovariance function:

$$\mu(t) = \mathbb{E}[X(t)], \quad \mu_i(t) = \mathbb{E}[X_i(t)], \quad i = 1, \cdots, d.$$
$$\Gamma(t,s) = \mathbb{E}[(X(t) - \mu(t))(X(s) - \mu(s))^T],$$
$$\Gamma_{ij}(t,s) = \mathbb{E}[(X_i(t) - \mu_i(t))(X_j(s) - \mu_j(s))], \quad i,j = 1, \cdots, d.$$

- Weak stationarity:

$$\mu(t) = \mu, \quad t \in \mathbb{Z},$$
$$\Gamma(t,s) = \Gamma(t+k, s+k), \quad t,s,k \in \mathbb{Z}.$$

vs

- Stationarity

$$(X(t_1), \cdots X(t_n)) \overset{d}{=} (X(t_1+k), \cdots X(t_n+k)),$$
$$\forall\ t_1, \cdots, t_n, k \in \mathbb{Z},\ n \in \mathbb{N}$$

## Example

Consider the returns $R(t) = \ln\left(\frac{P(t)}{P(t-1)}\right)$ of an index, which are not independent but modelled as

$$R(t) = \rho\, R(t-1) + \epsilon(t), \quad \epsilon(t) \sim_{iid} N(0, \sigma_\epsilon^2).$$

Find the variance of returns for a period of 2 days.

For 2 days

$$\begin{aligned} Var(R(t) + R(t-1)) &= Var(R(t)) + Var(R(t-1)) + 2Cov(R(t-1), R(t)) \\ &= Var(R(t)) + Var(R(t-1)) + 2\rho Var(R(t-1)). \end{aligned}$$

If $Var(R(t)) \simeq Var(R(t-1)) \simeq \sigma^2$ then

$$Var(R(t) + R(t-1)) = 2\sigma^2(1+\rho),$$

Example (Need for models with fat tails)

We often observe the following stylized facts:

- Each year we have 1 or more daily fluctuations that correspond to 4 standard deviations
- Each year we have at least one market that displays daily fluctuations that corresponds to more than 10 standard deviations.

How would you account for these stylized facts?

The empirical probability of observing a 4 standard deviations event is estimated as

$$P_{emp} \simeq \frac{1}{252} = 0.003968$$

If the returns are modelled by the normal distribution then the theoretical probability of such an event would be

$$P = 1 - N(4) = 0.000031671$$

which would correspond to one such event in 125 years! ($t \simeq 1/P$)

Clearly the normal distribution cannot account for such observations.

Example (Continued ...)

Assuming as alternative model the Student distribution with parameter $\nu$ we estimate

$$\begin{aligned} 1 - F_\nu(4) &= 0.028595 \quad \nu = 2, \\ 1 - F_\nu(4) &= 0.014004, \quad \nu = 3, \\ 1 - F_\nu(4) &= 0.0080650, \quad \nu = 4, \\ 1 - F_\nu(4) &= 0.0051617, \quad \nu = 5, \end{aligned}$$

which is closer to the observations.

Hence, models presenting fat tails can describe more closely the observations.

# GARCH processes

In a GARCH model we no longer consider the volatility as a constant, but rather as a stochastic process depending on the previous returns.

That means e.g. that high returns in the past may lead to higher volatility in the future or vice versa.

This is related to the observed phenomenon of volatility clustering

The GARCH model (introduced by Engle and Bollerslev in the 1980's) is now a standard model for stock returns featuring stochastic volatility effects such as volatility clustering.

# Garch(1,1)

Let $X(t) := \ln \frac{S(t)}{S(t-1)}$
According to the GARCH model $X(t)$ follows the law

$$X(t) \mid \mathcal{F}_{t-1} \sim N(\mu(t), \sigma^2(t)),$$
$$\mathcal{F}_{t-1} = \sigma(X(1), \cdots, X(t-1)).$$

where the conditional volatility $\sigma(t)$ depends on past returns and past conditional volatilities as

$$\sigma^2(t) = \alpha_0 + \alpha_1 X^2(t-1) + \beta_1 \sigma^2(t-1).$$

Conditional returns

$$X(t) \mid X(1) \cdots X(t-1) \sim N(\mu(t), \sigma^2(t)),$$

follow the normal distribution.

Unconditional returns will follow a mixture distribution which depending on the specification of $\sigma^2(t)$ may deviate considerably from the normal distribution, hence display fat tails.

Assuming wlog that $\mu(t) = 0$ (else set $\hat{X}(t) = X(t) - \mu(t)$ and rewrite model for that):

- Conditional volatility:

$$\sigma^2(t) = \mathbb{E}[X^2(t) \mid X^2(1), \cdots, X^2(t-1)] =: E_{t-1}[X^2(t)],$$

  i.e. mispecification of $X(t)$ if estimated by $\mathbb{E}[X(t) \mid \mathcal{F}_{t-1}] = 0$ at time $t-1$

vs

- Unconditional volatility

$$Var(X(t)) = \mathbb{E}[X^2(t)],$$

  i.e. mispecification of $X(t)$ is estimated by $\mathbb{E}[X(t)] = 0$ at time 0.

Dynamic form of the model,

$$X(t) = \sigma(t)\epsilon(t), \quad \epsilon(t) \sim N(0,1), \; i.i.d$$
$$\sigma(t)^2 = \omega + \alpha_1 X(t-1)^2 + \beta_1 \sigma(t-1)^2$$

where wlog we consider $\mu(t) = 0$.

Note that

$$\mathbb{E}[X(t) \mid X(u), \; u < t] = 0, \; \forall \, t \in \mathbb{Z},$$

but $Corr(X(t)^2, X(t-h)^2) \neq 0$!

This is compatible with empirical observations such as e.g. volatility clustering

This model presents fat tails, as well as mean reversion effects for the volatility, with the mean level as well as the rate of return to the mean level governed by the parameters $\omega, \alpha_1, \beta_1$.

To make sure that $\sigma^2(t) \geq 0$ we need the assumption

$$\omega \geq 0, \; \alpha_1 \geq 0, \; \beta_1 \geq 0.$$

# Stationarity for the Garch(1,1) model

Stationarity is an important property for a model.

The Garch(1,1) model is stationary if $\alpha_1 + \beta_1 < 1$

**Theorem**

*If $\alpha_1 + \beta_1 < 1$ then the stochastic process*

$$
\begin{aligned}
X(t) &= \sqrt{H(t)}\epsilon(t), \\
H(t) &= \omega \left\{ 1 + \sum_{i=1}^{\infty} A(t-1) \cdots A(t-i) \right\} \\
A(k) &= \alpha_1 \epsilon^2(k) + \beta_1 \geq 0
\end{aligned}
$$

*is the unique stationary solution to the Garch(1,1) model.*

Proof: Express the model as

$$\begin{aligned}\sigma^2(t) &= \omega + \alpha_1 X(t-1)^2 + \beta_1 \sigma^2(t-1) \\ &= \omega + \alpha_1 \sigma(t-1)^2 \epsilon(t-1)^2 + \beta_1 \sigma^2(t-1) \\ &= \omega + (\alpha_1 \epsilon(t-1)^2 + \beta_1)\sigma^2(t-1) \\ &= \omega + A(t-1)\sigma^2(t-1),\end{aligned}$$

where

$$A(t-1) := \alpha_1 \epsilon(t-1)^2 + \beta_1 \geq 0$$

and show by induction that for any $N > 0$,

$$\sigma^2(t) = \omega[1 + \sum_{i=1}^{N} A(t-1)\cdots A(t-i)] + A(t-1)\cdots A(t-N-1)$$

Define the process

$$H(t,N) := \omega(1 + \sum_{i=1}^{N} A(t-1)\cdots A(t-i))$$
$$= \omega(1 + A(t-1) + A(t-1)A(t-2) + \cdots + A(t-1)\cdots A(t-N)),$$

and observe that $(H(t,N))_N$ is a monotone sequence of random variables hence (if bounded) converges a.s. to some limit $H(t)$,

$$H(t) := \lim_{N\to\infty} H(t,N), \text{ a.s}$$

From the definition of $H(t,N)$

$$H(t,N) = \omega + A(t-1)H(t-1,N-1), \ \forall N, t$$

hence taking the limit as $N \to \infty$

$$H(t) = \omega + A(t-1)H(t-1), \ \forall t$$

The condition for $H(t)$ finite is

$$-\infty \leq \gamma := \mathbb{E}[\ln(\alpha_1 \epsilon^2(t-1) + \beta_1)] < 0$$

From the equation for $H(t)$ we see that for every $M$ it holds that

$$H(t) = \omega[1 + A(t-1) + A(t-1)A(t-2) + \cdots + A(t-1)A(t-2)\cdots A(t-M)] \\ + A(t-1)A(t-2)\cdots A(t-M-1)H(t-M-1)$$

Whether the limit is finite or not depends on the convergence of the series

$$\sum_{M=1}^{\infty} b_M, \ \ b_M = A(t-1)\cdots A(t-M)$$

By the Cauchy criterion this series converges if $\lambda = \limsup b_M^{1/M} < 1$.

But

$$b_M^{1/M} = \exp\left(\frac{1}{M}\sum_{i=1}^{M} \ln A(t-i)\right) \to e^{\gamma} =: \lambda, \text{ a.s}$$

from the law of large numbers.

By Jensen's inequality we have

$$\mathbb{E}[\ln A(t)] \leq \ln \mathbb{E}[A(t)] = \ln(\alpha_1 + \beta_1) < 0 \text{ αν } \alpha_1 + \beta_1 < 1.$$

This concludes the proof. QED

# Weak stationarity

Strong stationarity may be too strong a condition. Often weaker versions of stationarity are sufficient.

We will ask for covariance stationarity instead,

$$C(t_1, t_2) := \mathbb{E}[X(t_1)X(t_2)] = C(t_1 - t_2), \ \forall \ t_1, t_2 > 0.$$

(where we assume for simplicity that $\mathbb{E}[X(t)] = 0$).

For a covariance stationary process

$$\mathbb{E}[X^2(t)] = \mathbb{E}[X^2(0)]^2, \ \forall \, t > 0.$$

By the GARCH(1,1) model $X(t) = \sigma(t)\epsilon(t)$, with $\sigma(t)$ fully known if we know $X(t-1), \sigma(t-1)$.

Hence,

$$X(t) \mid X(t-1), \sigma(t-1) \sim N(0, \sigma^2(t)),$$
$$Var(X(t) \mid X(t-1), \sigma(t-1)) = \sigma^2(t).$$

Applying the law of total variance

$$Var(Y) = \mathbb{E}[Var(Y \mid W)] + Var(\mathbb{E}[Y \mid W]),$$

for $Y = X(t)$ and $W = (X(t-1), \epsilon(t-1))$ we have that

$$Var(X(t)) = \mathbb{E}[\sigma^2(t)]$$

From the GARCH(1,1) model taking expectations

$$\mathbb{E}[\sigma^2(t)] = \omega + \alpha_1 Var(X(t)) + \beta_1 \mathbb{E}[\sigma^2(t-1)].$$

By covariance stationarity $Var(X(t)) = \bar{\sigma}^2$, constant, hence $\mathbb{E}[\sigma^2(t)] = \bar{\sigma}^2$, so the above yields

$$\bar{\sigma}^2 = \frac{\omega}{1 - \alpha_1 - \beta_1}.$$

Hence $\alpha_1, \beta_1$ must satisfy

$$\alpha_1 + \beta_1 < 1,$$

which can be proved that it is a necessary and sufficient condition for autocovariance stationarity.

In conclusion for the GARCH(1,1) model for $\alpha_1 + \beta_1 < 1$ the unconditional variance is constant and equal to

$$\bar{\sigma}^2 = \frac{\omega}{1 - \alpha_1 - \beta_1}.$$

# GARCH(1,1) model and predictions

If the returns are provided by the GARCH(1,1) model can we provide an estimate for $\mathbb{E}_{t-1}[X^2(t+n)]$?

This corresponds to the prediction error for the return $X(t+n)$ given the information about the market until time $t-1$.

Note that

$$\mathbb{E}_{t-1}[X^2(t+n)] \neq \sigma^2(t+n) := \mathbb{E}_{t+n-1}[X^2(t+n)]!$$

# Calculating $\mathbb{E}_{t-1}[X^2(t+1)]$

Start from the definition

$$\sigma^2(t) := \mathbb{E}_{t-1}[X^2(t)]$$

that yields $\sigma^2(t) \in m - \mathcal{F}_{t-1}$.

By the properties of conditional expectation

$$\mathbb{E}_{t-1}[X^2(t+1)] = \mathbb{E}_{t-1}[\underbrace{\mathbb{E}_t[X^2(t+1)]}_{\sigma^2(t+1)}] = \mathbb{E}_{t-1}[\sigma^2(t+1)]$$

$$\underbrace{==}_{Garch} \mathbb{E}_{t-1}[\omega + a_1 X^2(t) + b_1\sigma^2(t)] = \omega + a_1 \underbrace{\mathbb{E}_{t-1}[X^2(t)]}_{\sigma^2(t)} + b_1\mathbb{E}_{t-1}[\underbrace{\sigma^2(t)}_{\in m-\mathcal{F}_{t-1}}]$$

$$= \omega + (a_1+b_1)\sigma^2(t) = \bar{\sigma}^2(1-a_1-b_1) + (a_1+b_1)\sigma^2(t)$$

$$= \bar{\sigma}^2 + (a_1+b_1)(\sigma^2(t) - \bar{\sigma}^2).$$

Defining

$$\begin{aligned} \gamma &:= a_1 + b_1, \\ \bar{\sigma}^2 &:= \frac{\omega}{1-\gamma}, \end{aligned}$$ asymptotic volatility

we conclude that for the conditional volatility $\mathbb{E}_{t-1}[X^2(t+1)]$ it holds that

$$\mathbb{E}_{t-1}[X^2(t+1)] = \bar{\sigma}^2 + \gamma(\sigma^2(t) - \bar{\sigma}^2) \tag{1}$$

where $\sigma^2(t) = \mathbb{E}_{t-1}[X^2(t)]$.

# Calculating $\mathbb{E}_{t-1}[X^2(t+2)]$

By the properties of conditional expectations

$$\mathbb{E}_{t-1}[X^2(t+2)] = \mathbb{E}_{t-1}[\underbrace{\mathbb{E}_{t+1}[X^2(t+2)]}_{\sigma^2(t+2)}] = \mathbb{E}_{t-1}[\sigma^2(t+2)]$$

$$\underbrace{==}_{Garch} \mathbb{E}_{t-1}[\omega + a_1 X^2(t+1) + b_1 \sigma^2(t+1)]$$

$$= \omega + a_1 \underbrace{\mathbb{E}_{t-1}[X^2(t+1)]}_{\text{known by (1)}} + \mathbb{E}_{t-1}[\underbrace{\sigma^2(t+1)}_{\sigma^2(t+1):=\mathbb{E}_t[X^2(t+1)]}]$$

$$= \omega + a_1 \mathbb{E}_{t-1}[X^2(t+1)] + b_1 \mathbb{E}_{t-1}[\mathbb{E}_t[X^2(t+1)]]$$

$$= \omega + a_1 \mathbb{E}_{t-1}[X^2(t+1)] + b_1 \mathbb{E}_{t-1}[X^2(t+1)] = \omega + \gamma \mathbb{E}_{t-1}[X^2(t+1)]$$

$$\underbrace{==}_{(1)} \omega + \gamma(\bar{\sigma}^2 + \gamma(\sigma^2(t) - \bar{\sigma}^2)) = \bar{\sigma}^2(1-\gamma) + \gamma(\bar{\sigma}^2 + \gamma(\sigma^2(t) - \bar{\sigma}^2))$$

$$= \bar{\sigma}^2 + \gamma^2(\sigma^2(t) - \bar{\sigma}^2).$$

We then have,

$$\mathbb{E}_{t-1}[X^2(t+1)] = \bar{\sigma}^2 + \gamma(\sigma^2(t) - \bar{\sigma}^2)$$
$$\mathbb{E}_{t-1}[X^2(t+2)] = \bar{\sigma}^2 + \gamma^2(\sigma^2(t) - \bar{\sigma}^2)$$

where $\sigma^2(t) = \mathbb{E}_{t-1}[X^2(t)]$ and

$$\begin{aligned} \gamma &:= a_1 + b_1, \\ \bar{\sigma}^2 &:= \frac{\omega}{1-\gamma}, \quad \text{asymptotic volatility} \end{aligned}$$

With an induction step we can show that

$$\mathbb{E}_{t-1}[X^2(t+n)] = \bar{\sigma}^2 + \gamma^n(\sigma^2(t) - \bar{\sigma}^2)$$

# In conclusion ...

We showed inductively that

$$\mathbb{E}_{t-1}[X^2(t+n)] = \bar{\sigma}^2 + \gamma^n(\sigma^2(t) - \bar{\sigma}^2), \ \forall\, t, n$$

where

$$\gamma = a_1 + b_1, \ \bar{\sigma}^2 = \frac{\omega}{1-\gamma}, \ \sigma^2(t) = \mathbb{E}_{t-1}[X^2(t)]$$

This is the best prediction we may have concerning the volatility of the market $n$ times ahead, given the information of the market up to time $t-1$.

Since $\gamma < 1$ then as $n$ grows this prediction tends to the asymptotic value $\bar{\sigma}^2$, i.e.

$$\mathbb{E}_{t-1}[X^2(t+n)] \to \bar{\sigma}^2, \ \text{ as } n \to \infty.$$

with the convergence speed depending on $\gamma$.

# Calibrating the GARCH(1,1) model

The model can be calibrated to market returns data $\{X_1, \cdots, X_n\}$ using maximum likelihood.

According to the model $\sigma(t)$ is fully determined by $X_{t-1}, \cdots, X_1, X_0, \sigma(0)$ (by simply iterating the GARCH equation) and since

$$f_{X_t \mid X_{t-1}, \cdots, X_0, \sigma_0}(x_t \mid x_{t-1}, \cdots, x_0, \sigma_0) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{x_t^2}{2\sigma^2(t)}\right)$$

by the independence of $\epsilon(t)$ we have

$$f_{X_1, \cdots, X_n \mid X_0, \sigma_0}(x_1, \cdots, x_n \mid x_0, \sigma_0) = \prod_{i=1}^{n} f_{X_t \mid X_{t-1}, \cdots, X_0, \sigma_0}(x_t \mid x_{t-1}, \cdots, x_0, \sigma_0)$$

Hence, the likelihood of the sample is

$$L(\omega, \alpha_1, \beta_1; X) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2(t)}} \exp\left(-\frac{X_i^2}{2\sigma(t)^2}\right)$$

with $\sigma^2(t)$ provided by the iteration

$$\begin{aligned} \sigma^2(t) &= \omega + \alpha_1 X_{t-1}^2 + \beta_1 \sigma^2(t-1), \\ \sigma^2(0) &= \sigma_0^2. \end{aligned}$$

Maximizing the likelihood with respect to $\omega, \alpha_1, \beta_1$, possibly setting $\sigma_0$ equal to

$$\sigma_0^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2.$$

provides the best estimate to the parameters of the model, which can then be used for prediction.