Financial Time Series

MFM Practitioner Module: Quantitiative Risk Management

John Dodson

November 6, 2019

Invariants

We are generally working with financial timeseries data when calibrating models for the future value of financial variables such as the mark-to-market profit/loss on an asset holding.

- In some cases, such as equity shares, this may mean working with market prices (adjusted for dividends and splits).
- In other cases, such as for bonds or derivatives, it may mean working with derived quantities like yield or implied volatility.

Invariants

If we expect today that the meaning of a financial quantity of interest will remain uniform for the foreseeable future, we term it an invariant quantity. For example, the price or yield on a particular derivative or bond is *not* an invariant because the instrument will expire or mature on a known date.

Invariants

Indexes, Generics, & Synthetics

The challenge of identifying invariants for important classes of financial variables is addressed variously through indexes, synthetics, and generics.

- The S&P 500 equity index purports to represent the performance of typical large-cap U. S. listed equity securities.
- The Fed's CMT indexes purport to represent the performances of typical nominal U. S. Treasury bonds of particular tenors.
- Bloomberg futures generics represent the performances of the 1st, 2nd, etc. contract of a particular futures product.
- The Cboe VIX index purports to represent the performance of a delta-hedged position in one-month S&P 500 index options.

Invariants

Investibility & Relevance of Invariants

If we intend to use an invariant as a proxy for an actual asset, it is important to think carefully about how the performance of the proxy can differ from the performance of the asset.

- The S&P 500 is an investible index whose performance can be replicated by an instantaneously fixed portfolio of equity shares, its performance is influenced by its dynamic composition and the dynamic correlation between constituents, which is obviously not relevant for individual equities.
- Other indexes, such as LIBOR (London interbank offered rate) or OIS (Federal Funds rate overnight index swap), are technically investible, but only by the treasury departments of banks; in particular they are *not* investible to broker-dealers.

Invariants

Seasonality

Some financial timeseries exhibit predictable patterns in time, or seasonality.

- A futures generic must roll whenever new contracts are issued. The actual profit/loss from rolling over a futures position is difficult to predict, and the generic makes no attempt at all.
  - For timeseries analysis purposes, you should omit roll dates from generics for your analysis.
- There may be predictable events, such as earnings announcements or the seasonal consumption patterns of certain commodities, that should be modeled as regimes.
  - This is a specialized topic in econometrics that we will not cover.

Innovations

We generally only care about the most recent level for a risk factor after our timeseries analysis is finished and we are looking at the loss distribution for a particular portfolio. For the analysis, we are more interested in the periodic innovations of the risk factor, such as the log-returns or simple differences.

- You can think of this as the difference operator applied to the index or its (natural) logarithm, $X_t \triangleq \nabla \log S_t$.

Drift

The conditional expected value $\mathrm{E}\left[\nabla \log S_t | \mathcal{F}_{t-1}\right]$ of the log of an index is termed the index drift μ_t.

Volatility

The conditional standard deviation $\sqrt{\mathrm{var}\left[\nabla \log S_t | \mathcal{F}_{t-1}\right]}$ is termed the index volatility σ_t.

Note that the drift and volatility are $\mathcal{F}_{t-1}$-measurable.

Drift with ARMA

White noise is a collection of i.i.d. r.v.'s Z_t with zero mean and finite variance σ_t^2. With $X_t = \mu_t + \varepsilon_t = \mu_t + \sigma_t Z_t$ an innovation of an invariant, we call ε_t the residual.

Autoregressive Moving Average

An ARMA(p, q) process for the drift can be expressed as

$$\mu_t = \phi_0 + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}$$

for parameters $\phi_0, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$.

- AR(1) is a simple model for mean reversion for the innovations around a long-run level $\phi_0/(1-\phi_1)$.

Volatility with GARCH

For financial data there is little to be gained in modeling drifts of timeseries data, because typically $|X_t| \gg \mu_t$.

- Furthermore, if X_t is a log-return, the drift probably ought to include a Jensen term like $-\frac{1}{2}\sigma_t^2$ which certainly does not fit into the ARMA form.

Generalized Autoregressive Conditional Heteroskedasticity

A GARCH(p, q) process for the conditional variance can be expressed as

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$$

for *non-negative* parameters $\alpha_0, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q$.

Volatility with GARCH

Standardized Residuals

One application of GARCH models is to extract i.i.d. samples from timeseries. We define the standardized residuals as

$$Z_t = \frac{X_t - \mu_t}{\sigma_t}$$

To the extent that the GARCH model is correct, these are strict white noise.

GARCH(1,1)

By far the most common implementation of this model is GARCH(1,1). An important result about this model is that the unconditional variance is

$$\sigma^2 \triangleq \text{var}\left[X_t\right] = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}$$

as long as $\alpha_1 + \beta_1 < 1$.

Volatility with GARCH

In fitting GARCH(1,1) to asset returns, one often sees that $\hat{\alpha}_0$ is close to zero. From the previous slide, we see that $\alpha_0 = 0$ requires that $\alpha_1 = 1 - \beta_1$. So the integrated GARCH(1,1) model has only one parameter.

$$\sigma_t^2 = (1 - \beta_1)\,\varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

Exponentially-weighted moving average

Since (with $\beta_1 = \lambda$) this is equivalent to

$$\sigma_t^2 = \frac{\sum_{i=0}^{\infty} \lambda^i \varepsilon_{t-1-i}^2}{\sum_{i=0}^{\infty} \lambda^i}$$

IGARCH(1,1) is sometimes called the exponentially-weighted moving average (EWMA) model, popularized by RiskMetrics.

- λ can be estimated using the technique discussed below.

Volatility with GARCH

GARCH(1,1) Volatility Process

Under GARCH(1,1), *innovations* of the conditional variance are mean-reverting. You can see this because

$$\nabla\sigma_t^2 = (1-\beta_1-\alpha_1)\left(\sigma^2-\sigma_{t-1}^2\right)+\alpha_1\sigma_{t-1}^2\nabla W_{t-1}$$

where

$$\nabla W_{t-1} = \frac{\varepsilon_{t-1}^2}{\sigma_{t-1}^2}-1$$

is stochastic increment uncorrelated to ε_{t-1} (if the residuals are unskewed).

- The mean-reversion rate is $1-\beta_1-\alpha_1$.
- W_t is not a Brownian motion but rather a more general martingale.

Refresher on Statistical Estimation

In classical statistics, the term sample has two related meanings

- an (unordered) set of N values drawn from the state space of some random variable X, $\{x_1, x_2, \ldots, x_N\}$
- a random variable consisting of N (independent) copies $X_1, \ldots, X_N$ of some random variable $X_i \sim X \; \forall i$.

You can think of the former as a realization of the latter. We can characterize the latter, which we will denote hereafter by $Y^{(N)} \triangleq (X_1, \ldots, X_N)$, as a random variable with

$$f_{Y^{(N)}}(Y) = f_X(X_1) \cdots f_X(X_N)$$

because we have assumed that the draws are independent.

Estimator

An estimator is a function of a sample.

- If the sample is considered to be random, the value of an estimator is a random variable subject to characterization.
- If the estimator is applied to an actual sample, consisting of draws from the sample space, the value is non-random and is called an estimate.

Parameter Estimator

We will be mostly interested in estimating the parameters of a characterization, which we will denote generically by θ. For a univariate normal, for example, $\theta = \left(\mu, \sigma^2\right)'$.

We will denote the parameter estimator by $\hat{\theta}\left(Y^{(N)}\right)$ where $Y^{(N)} = (X_1, \ldots, X_N)$ is the sample represented by N independent copies of the random variable X with a characterization parameterized by θ.

Maximum Likelihood Estimator

Since we have the distribution of the sample, perhaps in terms of sufficient statistics, it is natural to define an estimator for the parameters as the value of the parameters such that the sample observed is “most likely”. That is,

$$\hat{\theta}(y) = \arg\max_{\theta} f_{Y^{(N)}|\theta}(y)$$

where the sample is $y = (x_1, \ldots, x_N)$.

Important Example

Say $X \sim \mathcal{N}(\mu, \sigma^2)$ and we have a sample $Y^{(N)} = (X_1, \ldots, X_N)$. The density function of the sample is

$$f_{Y^{(N)}}(y) = (2\pi\sigma^2)^{-N/2} e^{\frac{-1}{2\sigma^2}\sum_{i=1}^{N}(x_i - \mu)^2}$$

The MLE is

$$\begin{pmatrix} \hat{\mu} \\ \hat{\sigma^2} \end{pmatrix} = \arg\min_{(\mu,\sigma^2)'} \frac{1}{\sigma^2}\left(\frac{1}{N}\sum_{i=1}^{N} x_i^2 - 2\mu\frac{1}{N}\sum_{i=1}^{N} x_i + \mu^2\right) + \log\sigma^2$$

Maximum Likelihood Estimator

The solution to this (the MLE for a univariate normal) is

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad = \frac{x1}{1'1}$$

$$\hat{\sigma^2} = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \left(\frac{1}{N}\sum_{i=1}^{N} x_i\right)^2 \qquad = \frac{xx'}{1'1} - \frac{1'x'x1}{1'11'1}$$

This result extends to the multivariate case $X \in \mathbb{R}^M$ whereby x has M rows and N columns.

Bias

We can see that the MLE is (slightly) biased.

$$\mathrm{E}\ \hat{\mu} = \mu$$

$$\mathrm{E}\ \hat{\sigma^2} = \frac{N-1}{N}\sigma^2 \quad \text{(prove)}$$

Estimating GARCH

Let us continue to focus on GARCH(1,1). The principal technique for estimating the parameters of a GARCH process is maximum likelihood, but with several caveats:

- We do not know the marginal densities of the residuals and they are not identical
- We do not know know ε_0 or σ_0 (assume $t = 1$ is the first observed innovation)
- While we may assume that they are i.i.d., we may not know the exact density of the standardized residual $f_Z(\cdot)$

We address these through the quasi-MLE, in which we note that the multivariate density is the product of the conditional densities, and we assume that the residuals are normal:

$$\log f_{\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n|\sigma_1}\left(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n\right) = -\frac{1}{2}\sum_{t=1}^{n}\log\left(2\pi\sigma_t^2\right) + \frac{\varepsilon_t^2}{\sigma_t^2}$$

Estimating GARCH

Variance Targeting

Assuming that the unconditional variance of the innovations exists, it is advisable to set the intercept based on the sample variance.

$$\alpha_0 = \hat{\sigma}^2 \left(1 - \alpha_1 - \beta_1\right)$$

Then you are only using the QMLE to estimate α_1 and β_1. N.B.: You should probably put a lower bound on α_1 in this case, otherwise the β_1 could be degenerate.

Initialization

Assuming $t = 1$ is your first innovation, we need a way of determining σ_1^2 in terms of the parameters. That means you need to choose values for ε_0^2 and σ_0^2. One choice is to take both to be σ^2. In combination with variance targeting, this means $\sigma_1^2 = \hat{\sigma}^2$.

Forecasting GARCH

In terms of forecasting, we already have σ^2_{n+1}. Say we are interested in $\mathrm{E}_n\left[\sigma^2_{n+2}\right]$ (the subscript on the expectation represents the sigma algebra), we can write

$$\sigma^2_{n+2} = \sigma^2\left(1-\alpha_1-\beta_1\right) + \sigma^2_{n+1}\left(\alpha_1 Z^2_{n+1} + \beta_1\right)$$

so because $Z_{n+1} \sim \mathrm{SWN}(0,1)$

$$\mathrm{E}_n\left[\sigma^2_{n+2}\right] = \sigma^2_{n+1}\left(\alpha_1+\beta_1\right) + \sigma^2\left(1-\alpha_1-\beta_1\right)$$

Iterating this, we get the general result for integer $m > 0$,

$$\mathrm{E}_n\left[\sigma^2_{n+m}\right] = \sigma^2_{n+1}\left(\alpha_1+\beta_1\right)^{m-1} + \sigma^2\left(1-\left(\alpha_1+\beta_1\right)^{m-1}\right)$$

- The forecasts are a convex combination of the current conditional variance σ^2_{n+1} and the unconditional, or long-run, variance σ^2.