6. Revision and assignment revisited

MFE Elective, Forecasting & Financial Time Series
Part I: High-Frequency Forecasting

Ryoko Ito[1]

Department of Economics and Nuffield College, Oxford University
Email: ryoko.ito@economics.ox.ac.uk
Website: www.itoryoko.com

Saïd Business School, Oxford, Trinity 2017

Week 1: course aims and problem statement

Course outline: 4 weeks on HF forecasting in finance.
Apply statistical theories and techniques you've learned to a practical problem.
Practice writing a professional report.

Assignment: You are a banker and an MD comes to you...
You only know a volatility model (GARCH) and a bunch of statistical techniques. What can you do?

Week 1: Recap of what we know

GARCH model for financial returns (r_t). Captures

- r_t is mean zero.
- r_t is serially uncorrelated.
- persistence in r_t^2.
- leverage (asymmetric) effect in r_t.

Features:

- GARCH is for two-sided distributions.
- GARCH can lack robustness to large-sized observations.

A robust model is DCS. It can also handle one-sided distributions. (Turns out GARCH is a special case of DCS in any case.)

Week 1: Stylized (peculiar) features of HF financial data

Any good statistical analysis must begin by meta-analyzing the data.
Features of HF (volume) data:

- non-negative distribution
- intra-day patterns (seasonality)
- autocorrelation
- heavy-tailed
- mass of zero-valued observations
- sampling frequency affects all (but the first one) of the above.
- overnight discontinuity

etc.

Week 2: Modeling HF financial data

Any good statistical model must capture stylized features of the data. A good economic model must be intuitive.

Spline-DCS for dynamic scale:

- Handles non-negative distribution. (GB2 is very useful.)
- Exponential link to model dynamic scale (like EGARCH).
- Decomposition of distribution at zero. (Law of total probability, conditional distribution.)
- Decomposition of dynamics into components:
  - Periodic component (spline)
  - Non-periodic component (RW + AR).
- Estimation by ML.
  - Derivatives for analytical standard errors of MLE. (Highlighted the usefulness of AIC and SIC.)
- Parameter restrictions for intuitions and identification.

Asymmetric effect. Announcement effect.

Week 3: Evaluating in-sample estimation results

A good model satisfies what it assumes when it is fitted to the data.
Main assumptions to check:

1. $\varepsilon_{t,\tau} \sim F(\cdot;\theta)$:
   - PIT (distance along the probability axis)
   - QQ-plot (distance along the quantile axis)
   - KS-test (scientific application of PIT)

   The first two can deliver different messages.
2. $\varepsilon_{t,\tau}$ is independent:
   - Ljung-Box test

   Key ideas of statistical hypothesis testing. Power of the test in HF finance: statistical versus economic significance.
3. Dynamic assumptions of the components of $\lambda_{t,\tau}$.
4. Parameter restrictions.

Week 3: Evaluating in-sample estimation results

Gaussian versus non-Gaussian distribution. QMLE versus MLE.

- When we take a departure from Gaussianity, the attention should shift to wanting to know the overall distribution of the data (not just the first two moments).

Week 3: HF forecasting

In-sample fit versus forecasting performance.
(Are we overfitting? Test model stability.)

From ML to density forecasting. We can forecast:

- moment (of higher order, ES)
- quantile (median, VaR)

Easy to do this. We get the picture of the degree of dispersion.

Comparing forecasts. Diebold-Mariano test. MSE versus MAE. Importance of the choice of loss function in HF finance. (What do we really care about?)

Week 4: Competing models

Competing models:

- Average volume at each intra-day bin
- Simple ARMA
- Rolling (moving) average
- Fourier series (Gibbs phenomenon.)
- Spline
- CMEM

All but the last one of these are ad-hoc.
Importance of capturing intra-day diurnal patterns. It dominates intra-day dynamics.

GMM versus ML. Parametric versus non-parametric. Importance of computing time in HF finance.

Week 4: Wrapping up

Key features of HF finance:

- Large data (power of test, non-parametric versus parametric, computing speed, choice of sampling frequency)
- Heavy-tailed data (non-parametric versus parametric, Gaussian versus non-Gaussian, loss function, PIT versus QQ-plot)
- Persistent dynamics with intra-day periodicity (dynamic decomposition, ad-hoc versus Fourier versus spline methods)

We could go far with what we already knew. Perhaps the only new elements were DCS and periodicity.

BUT applying our knowledge hopefully enhanced our understanding of basic statistical theories and methods.

Week 4: Wrapping up

So your MD came to you 4 weeks ago. Note your MD is very busy.

Expectations:

- Short and concise (approx. 10 pages, shorter the better)
- Can put technical expositions in the appendix (here, as many pages as you want)
- Nice story using nice pictures (reinforced by some statistics)
- Neat Matlab codes consistent with your report

Week 4: Writing well

A bible of modern (academic and professional) writing: *Politics and English Language* by Gorge Orwell. (Approx. 20 pages.)

Writing well:

- Clearly state the contribution (key story of your report) in the introduction.
- Shorter the better. Be concise.
- Be explicit about pros AND cons.
- Celebrate creative deep thinking, uniqueness.

Week 4: Writing well

To create a presentable paper, writing goes in 3 phases:

1. Write it up to the highest standard you can achieve in a short period (a dense period of concentrated work)
2. Leave it and sleep on it for a while
3. Come back to it and rewrite until you have a presentable paper

Week 4: Writing well

- Maintain sharp focus on your main story.
- Do NOT write anything irrelevant or your don't really understand, just for the sake of writing or just because it sounds cool. **It's really easy for experts to tell when you're doing this.** (Read Orwell.)
- **You must own every word in your paper.** Thus always ask yourself for **each** sentence you write:
  - does this statement make sense,
  - is it accurate,
  - could it be simpler
  - could it be clearer,
  - is it necessary,
  - are all the words in it necessary?

  (Read Orwell.)