Discussion of "Should we sample a time series more frequently? Decision support via multirate spectrum estimation" by Guy, Powell, Elliott, and Smith.

Xiao-Li Meng

Department of Statistics, Harvard University
1 Oxford Street, Science Center
Cambridge, MA 02138, USA

A toast and congratulations to the author(s) should be in order whenever a discussant enjoys reading a paper. But often this enjoyment leads to a self-condolence for the discussant: "Well, there goes another paper that I should have written!" The grapes are particularly sour when the discussant has been planting similar varieties in a neighbouring vineyard. In Meng and Xie (2014), which demonstrates that more data *do not* guarantee better results, we studied the impact of the sampling frequency, and the related interplay between data patterns and model assumptions, on estimating the autocorrelation in the simple AR(1) model. The findings were somewhat intriguing, and I have been wondering what they would look like for more complex time series models.

Specifically, suppose in principle we can observe any part of an AR(1) series of indefinite length into the future

$$Y_t = \rho Y_{t-1} + \epsilon_t, \quad \epsilon_t \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2), \quad t = 0, 1, , \dots, \tag{1}$$

but in reality we can only afford taking n observations. What is then the sampling frequency for optimally estimating ρ? Assuming n is sufficiently large to render the adequecy of the Fisher information approximation, we proved that the optimal spacing (between consecutive sampling times) is 1 when $\rho^2 \leq 1/3$, and it is 2 when $1/3 < \rho^2 \leq \sqrt{21/20} - 0.5$, etc. In general, the optimal spacing goes up with ρ^2 at the rate of $[-\log \rho^2]^{-1}$ as $\rho^2 \uparrow 1$, so does the maximal relative gain in efficiency for estimating ρ as compared to using the single spacing.

We also investigated the efficiency gain for estimating ρ from knowing the value of σ^2. When the spacing $s = 1$, the relative gain in Fisher information is bounded above by $1/(n-1)$. However, it can be as high as 50% once $s = 2$ (achieved when $\rho = 0$), and it approaches infinite as $\rho^2 \downarrow 0$ when $s = 3$. From a time-domain perspective, this is due to the fact that once $s > 1$, estimating ρ and estimating σ^2 become tangled because

$$Y_{t+s}|Y_t \sim N\left(\rho^s Y_t,\ k_s(\rho^2)\sigma^2\right), \quad \text{where} \quad k_s(x) = \sum_{j=0}^{s-1} x^j, \tag{2}$$

and hence $\mathrm{Var}(Y_{s+t}|Y_t)$ depends on both ρ and σ^2 when $s > 1$. In other words, the Fisher information matrix for $\{\rho, \sigma^2\}$ is diagonal if and only if $s = 1$.

The authors emphasize the gain of insights from a frequency-domain perspective about missing observations in a time series. Since every time series can be represented equivalently in frequency domain and in time domain, I'd be very interested in gaining additional statistical insights from the authors' perspective about these mathematical findings.

References

Meng, X.-L and Xie, X. (2014) I got more data, my model is more refined, but my estimator is getting worse! Am I just dumb? *Econometric Reviews* **33**, 218-250 .