Things we think know

L. C. G. Rogers
Statistical Laboratory, University of Cambridge

January 13, 2019

Abstract

A model is useless if it cannot be taken to data, so it makes sense to begin modelling by looking at the data. Simple exploratory techniques can quickly reveal stylized facts of the data, and may suggest modelling hypotheses, but it is important to be cautious before jumping to conclusions, and to keep an open mind about other possibilities. This paper considers some examples related to asset returns.

1 Models of asset returns.

If S_t denotes the price of an asset at time t, then the celebrated Black-Scholes-Merton model proposes that S evolves as

$$dS_t = S_t(\sigma dW_t + \mu dt), \tag{1}$$

where W is a standard Brownian motion, and $\sigma > 0$, μ are constants. Crucially for derivative pricing, when we work in the pricing measure we have that μ is equal to the riskless rate r, an observation which goes back at least to Merton & Samuelson [4]. The Black-Scholes-Merton model has a number of strengths:

(S1) it is simple and tractable;

(S2) there is only one unknown parameter, σ ;

(S3) it generalizes naturally to many dimensions;

(S4) passing from continuous time to discrete time and back is easy and natural.

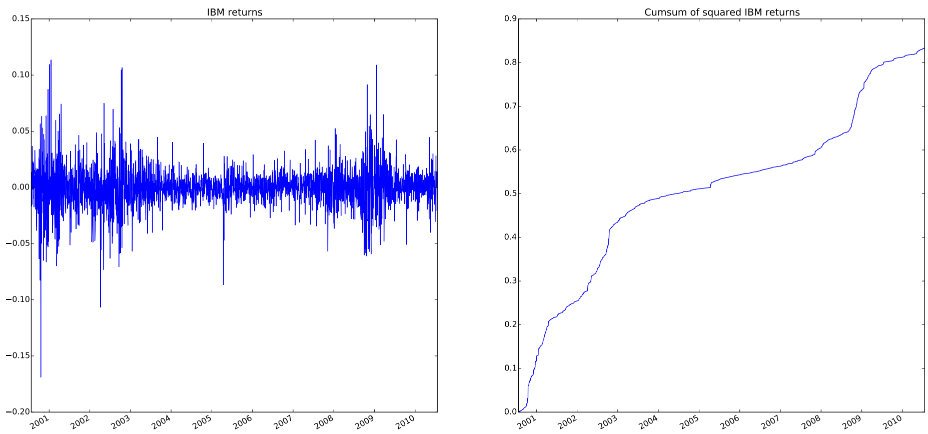
Property **(S1)** allows many closed-form options prices to be obtained, and makes numerical computations easier. Property **(S2)** is the basis of the entire technology of implied volatility. Property **(S3)** is achieved if we let the vector of log-prices of many assets be a Brownian motion with constant drift and covariance. For property **(S4)**, the discrete-time analogue of the Black-Scholes-Merton model is a random walk, and passing back to continuous time is in effect just Donsker's theorem. This is important because any numerical computation has to be done with finite statespace and finite time set. Set against these strengths are some folklore weaknesses, notably:

(W1) volatility is not constant;

(W2) returns[1] are not Gaussian.

We see the first of these from the plot of any asset returns; for example, Figure 1 shows the plot of returns on IBM stock. The right-hand panel is a diagnostic plot of the cumulative sum of squared returns. If returns were IID, then this plot would be a straight line, but it plainly is not. To

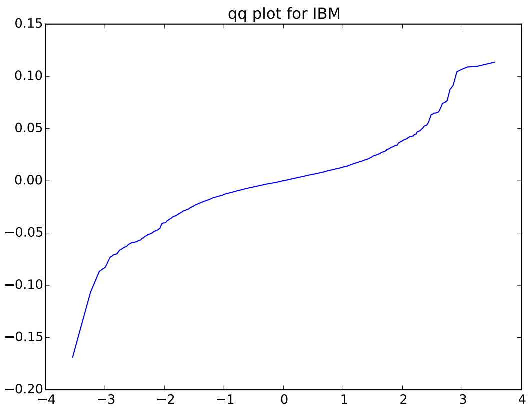
Figure 1: IBM returns.



show that log returns are not Gaussian, we commonly do a *q*-*q* plot, Figure 2. If returns were IID Gaussian, this plot would be a straight line, which visually appears not to be the case. But we need to be cautious - we already *know* from Figure 1 that returns are not IID, so *a fortiori* they are not IID Gaussian - Figure 2 tells us nothing useful. Attempts to model assets as log-Lévy processes relax the Gaussian assumption of the Black-Scholes-Merton model, but retain the obviously wrong assumption of IID returns.

[1]We shall speak of *returns* instead of the longer but more precise *log returns*.

Figure 2: q-q plot of IBM returns.



2 Alternative asset models.

The Black-Scholes-Merton model is an excellent first choice, but we see that it is not the whole story, so we need to consider other modelling directions. Sadly, there is no consensus about what the next modelling family should be. Among widely-studied choices, we could list:

1. log-Lévy models (which we know do not fit the data);
2. stochastic volatility models such as Heston, Bates, ...
3. local volatility models, popularized by Dupire and others
4. rough volatility models, promoted by Gatheral and co-workers
5. regime switching models, as considered by Elliott and others
6. GARCH models, the favourite of econometricians.

Before investing too much intellectual capital in any particular alternative class of models, it is well worth considering to what extent the strengths **(S3)** and **(S4)** of the Black-Scholes-Merton model survive: if we lose **(S3)** then the alternative class can only deal with one asset at a time, and if we lose **(S4)** then numerics or interpretation will be problematic. The econometricians' favourite, GARCH, fails spectacularly on both counts. But the purpose of this paper is not to survey and argue the strengths and weaknesses of various alternative asset models, rather the purpose is to draw attention to the need for care in exploration of data; and to illustrate this, we now concentrate on the fourth model in the list above, rough volatility models. This topic is relatively recent, but has been received so enthusiastically that it now has its very own website `https://sites.google.com/site/roughvol/home`, and the seminal paper [2] of Gatheral, Jaisson and Rosenbaum boldly declares in its title that 'Volatility is rough'. Let us inspect the evidence for this claim.

3 Rough volatility.

The empirical evidence for rough volatility is based on daily estimates of realized variance made available via the website of the Oxford Man Institute `https://realized.oxford-man.ox.ac.uk`. The reader should visit for more detail on the methodologies used to calculate these estimates from high-frequencey data; suffice it to say that various procedures are used, and produce quite similar estimates. Letting $\hat{\sigma}_t$ denote the estimate on day t of the annualized volatility of the asset, what Gatheral *et al.* calculate in [2] is the quantity

$$m(q,\Delta) \equiv N^{-1} \sum_{t=1}^{N} |\log \hat{\sigma}_{t+\Delta} - \log \hat{\sigma}_t|^q \tag{2}$$

for a range of values $q = 0.5, 1.0, 1.5, 2.0, 2.5, 3.0$ and lags $\Delta \in \{1, \ldots, 50\}$. Their remarkable finding is that apparently

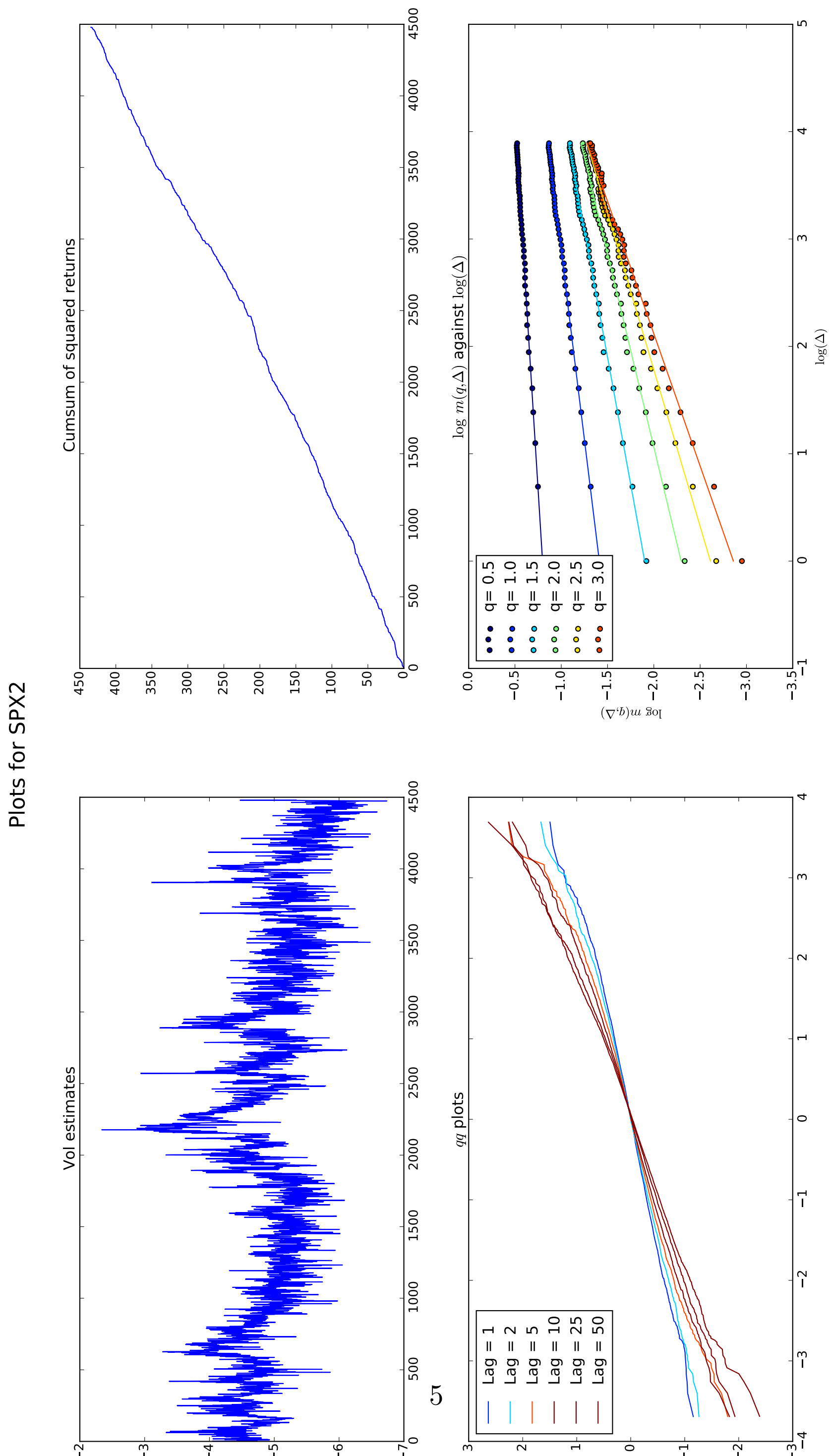
$$m(q,\Delta) \propto \Delta^{\alpha q} \tag{3}$$

for some α which is not the same for all the indices considered in [2], but is generally in the range $[0.07, 0.20]$. This is remarkable, because for any diffusion-based model we would expect that (for short lags at least) $\alpha = 0.5$, from the Brownian scaling. This surprising finding is the basis of the claim that volatility is rough. Moreover, the increments of $\log \hat{\sigma}_t$ appear to have Gaussian distributions. The plots in Figure 3 show these findings. In the top left panel is the daily estimates of volatility, in the top right panel is the cumulative sum of squared changes in volatility, in the bottom left panel are q-q plots of the increments in $\log \hat{\sigma}$ over six different time lags, and in the bottom right panel we have the plots of $\log m(q,\Delta)$ against $\log \Delta$ for six different values of q, with best-fit lines superimposed. The slopes of the best-fit lines are proportional to q, which leads Gatheral *et al.* to propose that $X_t \equiv \log \hat{\sigma}_t$ is a fractional Brownian motion (fBM):

$$X_{t+\Delta} - X_t \sim N(0, v\Delta^{2\alpha}) \tag{4}$$

for some $v > 0$. Fractional Brownian motion has long memory, which is a contentious property; as Gatheral *et al.* state in their paper, 'The evidence for long memory has never been sufficient to satisfy remaining doubters such as Mikosch and Starica.[3]'. One of the main points made in [3] is that when we calculate the ACF of a time series, the interpretation of what

Figure 3: Daily volatility estimates of the S&P500



we see presumes that the series is stationary[2], and departures from this assumption tend to produce the appearance of long-range dependence. In this instance, it may not be an issue; from the top right panel of Figure 3 we see the cumulative sum of squared differences growing linearly, so this at least is consistent with stationarity.

The fBM model certainly appears to fit the data well, but there are two main objections to it, which are the same objection, considered operationally and theoretically:

1. the model is highly non-Markovian; in order to predict the future of X, we need to know the *entire* history;
2. what *economic* story could we tell that would result in a model where we need to know the *entire* history in order to predict the future?

In any case, since the value of α varies from one index to another, there is clearly no universal law applicable to all assets. *What could we do that would be better?*

4 A simpler alternative to rough volatility.

When we look again at the top left panel in Figure 3, we see a plot which fluctuates strongly on small time scales, but on longer time scales the level seems to be changing. If the level did not change, we could try to model the data as an Ornstein-Uhlenbeck (OU) process with strong mean reversion and high volatility; as the level appears to be changing, we could try an energetic OU process mean-reverting to a slower one:

$$\begin{aligned} dY_t &= \sigma_Y \, dW'_t - \beta \, Y_t \, dt, && (5) \\ dX_t &= \sigma_X \, dW_t + \lambda \, (Y_t - X_t) \, dt. && (6) \end{aligned}$$

We refer to this model as OU-OU. Figure 4 is the analogue of Figure 3 for data generated by the model (5), (6) with parameter values $\sigma_X^2 = 20$, $\sigma_Y^2 = 0.625$, $\lambda = 210$ and $\beta = 2.5$. The qualitative behaviour is very similar. Moreover, the model is a bivariate diffusion, with linear dynamics and a joint Gaussian distribution, so it really is very nice to work with. It is an example of a multi-scale model of the kind analysed by Papanicolaou, Fouque, Sircar and others; see [1] for a consolidated account of this theme.

So we have two models, the rough volatility model and the OU-OU which both seem to explain the observed data well; is there any way to choose between them? Figure 5 shows the bottom right-hand panel of Figure 4 with a longer range of values for Δ. The linear fits are shown as dashed lines, and the true values (which we know because we know the model which

[2]... just as the q-q plot requires the data to be IID ...

Figure 4: Plots for the FIX2000

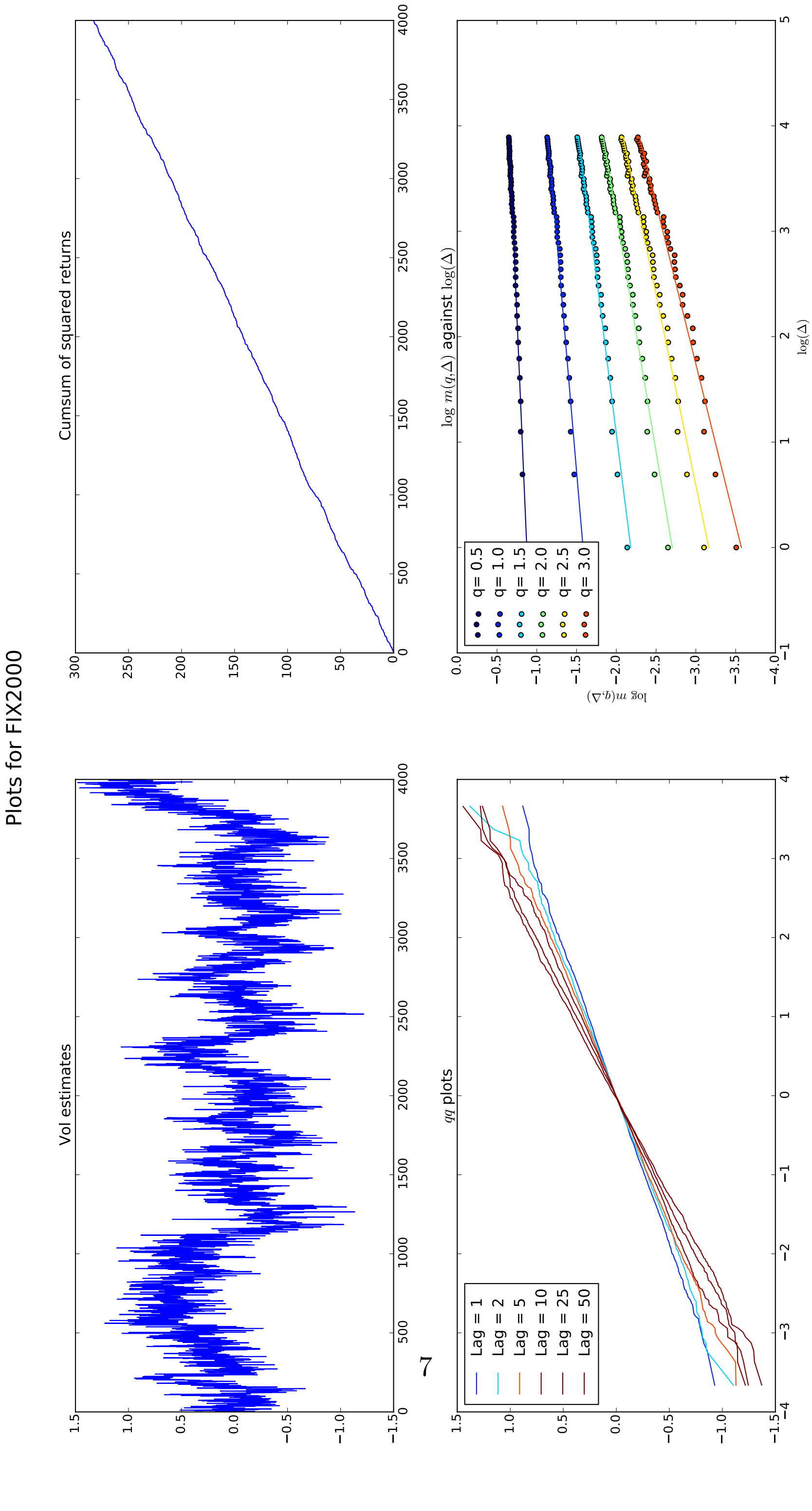
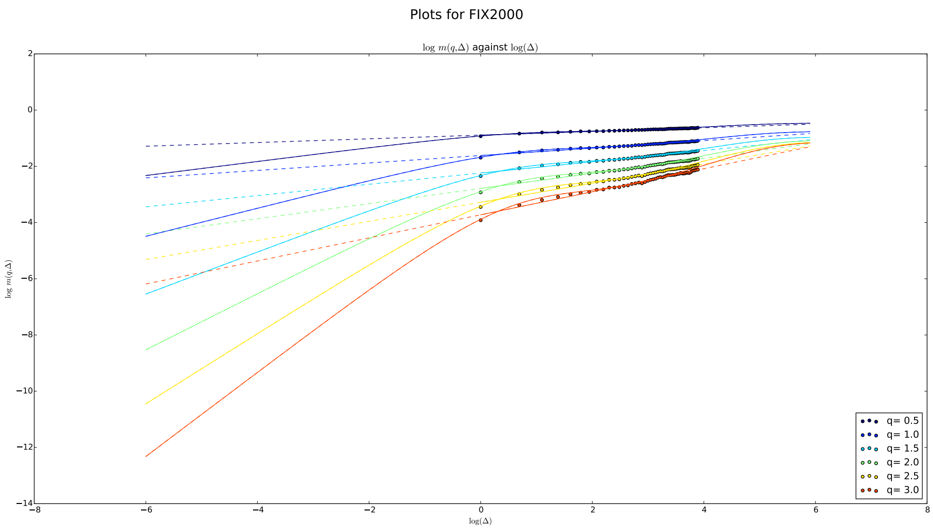


Figure 5: Plot for the FIX2000

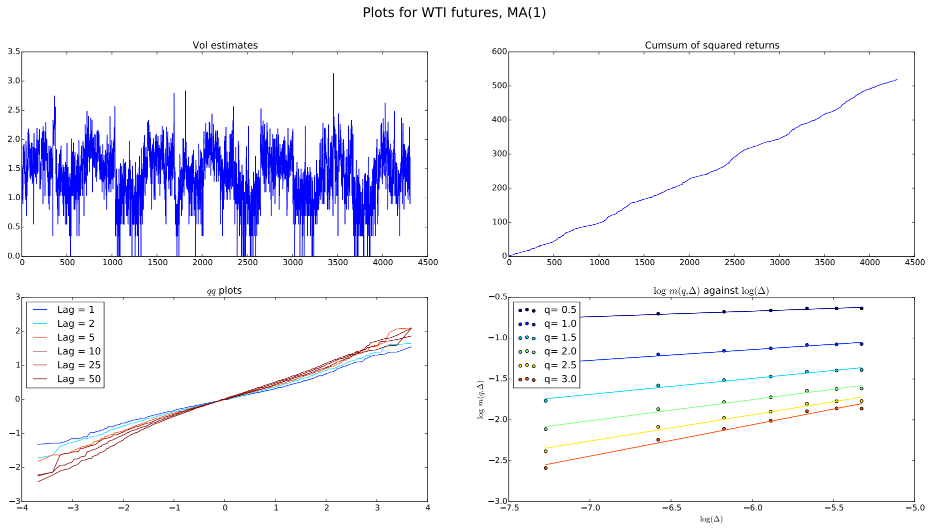


generated the data) are shown as solid curves. There is close agreement *in the original range of* Δ *values*, but as we move to much shorter timescales large differences emerge. So this tells us that if we are to distinguish between the rough volatility story and the OU-OU story, we need to look at these shorter time scales.

5 High-frequency data.

The estimates presented on the Oxford Man website are derived from high frequency data, but that data is not made available. So here we have to be content with something smaller scale, but which is nevertheless indicative. This is based on seven days of WTI futures tick data. Firstly we extract the times at which the mid-price has moved by one tick, which excludes events where the mid-price moves by half a tick, usually caused by the volume at the best bid or best ask momentarily falling to zero. Then we exclude market closed times. Finally we count for each minute the number of one-tick moves which happened in that minute. This gives us for each minute an estimate of the speed of the market; in a diffusion model, the variance σ^2 measures the speed, and in a discrete point process model (which is what we have in high-frequency data) it is the rate of the point process which measures the speed. The number of events in a given interval estimates the rate, so is a suitable proxy for variance in this context. Figure 6 shows what we get;

Figure 6: Plot for WTI high frequency data



once again, the apparent linearity in the bottom right panel is striking.

Zooming in on the data panel, we see Figure 7. The most striking feature is the extreme variability, with the numbers of events in consecutive minutes often differing by an order of magnitude. In view of this, one may wonder whether a model with continuous paths holds at such time scales; a scatter plot reveals Figure 8, and presented in this form it is hard to feel confident that there really is a continuous trajectory here. Maybe what we are seeing is some more regular process plus additive IID noise? To try to understand this possibility, we formed estimates of α for the raw data shown in Figure 7, and then for the MA(2) and MA(3) of these. Intriguingly, for the raw data we get estimates in the range found by Gatheral *et al.*, but for an MA(3) we find the estimates much closer to the limiting value 0.5 we would obtain from a diffusion model!

q	raw	MA(2)	MA(3)
0.5	0.137	0.327	0.482
1.0	0.132	0.319	0.468
1.5	0.130	0.316	0.460
2.0	0.129	0.314	0.455
2.5	0.128	0.313	0.450
3.0	0.128	0.313	0.446

Figure 7: Close-up of WTI high frequency data

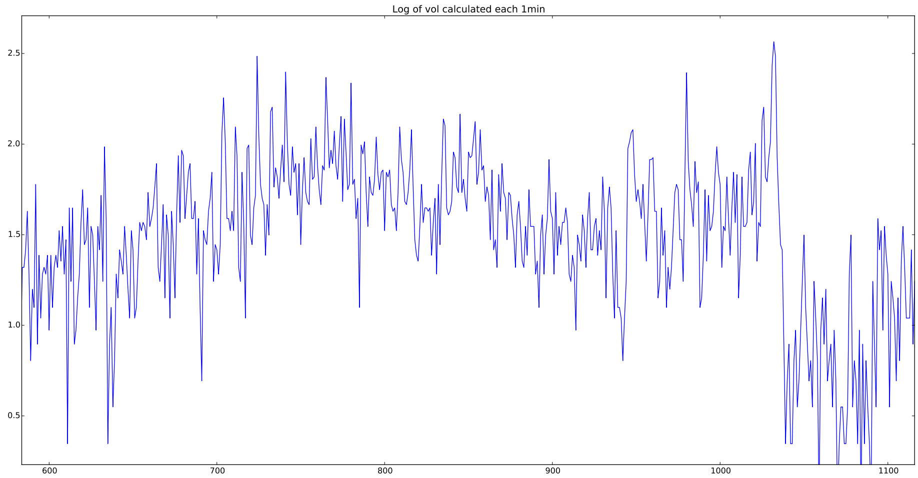
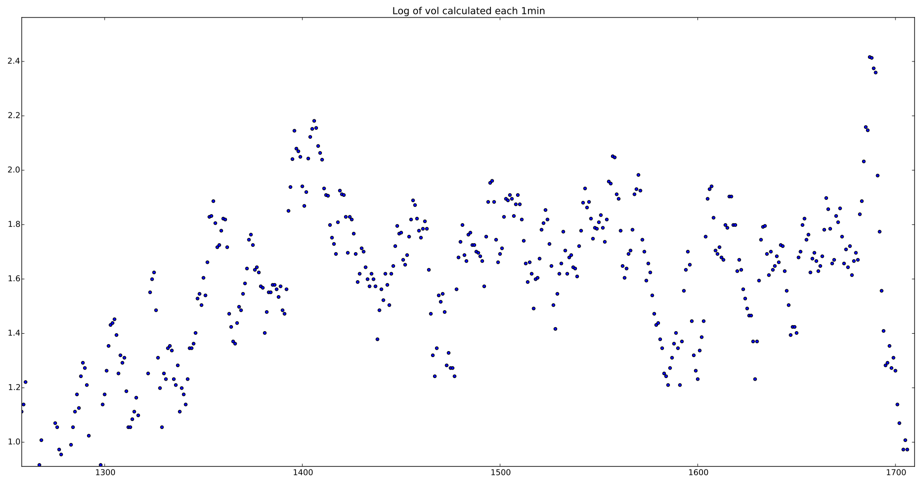


Figure 8: Close-up of WTI high frequency data



Conclusions.

What we have seen in this paper is that the notion that volatility is rough, that is, governed by a fractional Brownian motion, is *not* an incontrovertible established fact; simpler models explain the observations just as well. Neither makes much sense at very high frequency, but at this sort of timescale any estimates will be noisy. However, if we are concerned to model volatility because we want to calculate option prices, then the timescales we care about are days, weeks or months, not minutes, so a model that explains the data well on those timescales is valuable. The OU-OU model proposed here works on those timescales, and moreover is much easier to work with, being a bivariate Gaussian diffusion, amenable to the multiscale option pricing techniques explained in Fouque *et al.* [1].

References

[1] Fouque, J.-P., Papanicolaou, G., Sircar, R., and Sølna, K. *Multiscale stochastic volatility for equity, interest rate, and credit derivatives.* Cambridge University Press, 2011.

[2] Gatheral, J., Jaisson, T., and Rosenbaum, M. Volatility is rough. *Quantitative Finance 18*, 6 (2018), 933–949.

[3] Mikosch, T., and Starica, C. Is it really long memory we see in financial returns. *Extremes and integrated risk management 12* (2000), 149–168.

[4] Samuelson, P., and Merton, R. C. A complete model of warrant pricing that maximizes utility. *IMR; Industrial Management Review 10*, 2 (1969), 17–46.