Sailing in rough waters: examining volatility of fMRI noise

Jenni Leppanen[1], Henry Stone[2], David J. Lythgoe[1], Steven Williams[1], and Blanka Horvath[3]

[1]Department of Neuroimaging, King's College London
[2]Department of Mathematics, Imperial College London
[3]Department of Mathematics, King's College London

August 6, 2020

Corresponding author: Jenni Leppanen, Department of Neuroimaging, Institute of Psychology, Psychiatry and Neuroscience, King's College London, 103 Denmark Hill, London, SE5 8AF, UK.
Email: jenni.leppanen@kcl.ac.uk

Declarations of interest: none

1 Abstract

Background: Functional resonance magnetic imaging (fMRI) noise is usually assumed to have constant volatility. However this assumption has been recently challenged in a few studies examining heteroscedasticity arising from head motion and physiological noise. However, to our knowledge no studies have studied heteroscedasticity in scanner noise. Thus the aim of this study was to estimate the smoothness of fMRI scanner noise using latest methods from the field of financial mathematics.
Methods: A multi-echo fMRI scan was performed on a phantom using two 3 tesla MRI units. The echo times were used as intra-time point data to estimate realised volatility. Smoothness of the realised volatility processes is examined by estimating the Hurst parameter, a parameter $H \in (0, 1)$ governing the roughness (Hölder continuity) of paths in the rough Bergomi model, introduced in [2]. The rough Bergomi model a member of the family of rough stochastic volatility models. A family of models which was recently popularised in mathematical finance by observations indicating that volatility in financial markets is best described by stochastic models where volatility can be modulated by the Hurst parameter H, which usually calibrates to values $H \in (0, 0.5)$ (the rough case), hence inspiring the name of the model family. In this work, calibration of the Hurst parameter H is performed pathwise, using recently developed neural network calibration tools.
Results: In all experiments the volatility calibrates to values well within the rough case $H < 0.5$ and on average fMRI scanner noise was very rough with $H \approx 0.03$. Substantial variability was also observed, which was caused by edge effects, whereby H was larger near the edges of the phantoms.
Discussion: The findings challenge the assumption that fMRI scanner noise has constant volatility (in fact, the lower the value of H, the more pronounced the oscillations of the volatility, and hence the more "severe" is the violation of the constant volatility assumption) and add to the steady accumulation of studies suggesting implementing methods to model heteroscedasticity may improve fMRI data analysis. Additionally, the present findings add to previous work showing that the mean and normality of fMRI noise processes show edge effects, such that signal near the edges of the images is less likely to meet the assumptions of current modelling methods.

Keywords: fMRI, rough noise, volatility, rBergomi, CNN

2 Introduction

A given functional magnetic resonance imaging (fMRI) blood oxygenation level dependent (BOLD) time series can be defined as

$$y_t = \mu_t + \sqrt{v_t}\epsilon_t \tag{1}$$

where the μ_t is the mean, ϵ_t is a one dimensional noise process, and v_t is the volatility of the noise process. Detrending is typically conducted as part of preprocessing to remove signal drift. Thus, a given fMRI time series is often assumed to be a constant process, indicating that v_t in equation (1) could be replaced by constant v. However, recently this assumption has been challenged and there has been increasing interest in exploring time-dependent properties of fMRI noise [18, 33, 36, 43, 55, 59]. It has been shown that factors such as head motion and physiological processes including respiration and pulse can introduce heteroscedasticity to the time series [3, 18, 36, 43]. Heteroscedasticity in turn has been found to complicate the linear modelling, which has led to the introduction of several statistical models to counteract the impact of these artefacts [3, 18, 36]. One limitation of these models is that they cannot explain non-constant volatility arising from unknown or uncontrollable sources, such as scanner noise.

As volatility of a time series cannot be directly observed, a plethora of deterministic and stochastic models have been proposed to estimate it in financial returns data [42, 46, 48]. Over the years, direct comparisons of different volatility models have shown that stochastic models, which assume that logarithm of the volatility process behaves like standard Brownian noise, outperform their deterministic, counterparts providing a better fit to data [5, 8, 24]. This assumption implies in particular that volatility is not constant[1], but exhibits an oscillatory behaviour on any finite time interval. This oscillatory behaviour is governed by a parameter H which in the Brownian case takes the value $H = 0.5$. However, more recently, so-called fractional stochastic volatility models have also been considered (see [1, 2, 22, 23, 25]) where the parameter H is allowed to vary in the range $H \in (0, 1)$. In these models, as mentioned above, the parameter H governs the oscillations of the volatility process, the lower the parameter H, the stronger the oscillations on any finite interval. In particular, the values $H \in (0, 0.5)$ correspond to the *rough* case (i.e. rougher than paths of standard Brownian motion). Indeed, Gatheral et al. [25] found in their 2014 paper that the log-volatility of returns of a given financial time series appears to behave like fractional Brownian motion with the Hurst parameter $H < \frac{1}{2}$, thus introducing the rough fractional stochastic volatility (RFSV) model.

According to the RFSV model, log-volatility is mean reverting and satisfies the fractional Ornstein-Uhlenbeck process [12]

$$\mathrm{d}X_t = \nu \mathrm{d}W_t^H - \alpha(X_t - m)\mathrm{d}t \tag{2}$$

where W_t^H is the fractional Brownian motion with $H < \frac{1}{2}$, and $m \in \mathbb{R}$, $\nu > 0$, and $\alpha > 0$. Volatility of the process is then achieved by

$$\sqrt{v_t} = \exp(X_t) \tag{3}$$

A specific rough stochastic volatility model, introduced as the rough Bergomi (rBergomi) model in [2] by Bayer Friz and Gatheral is described by the system

$$\begin{aligned} \frac{\mathrm{d}S_t}{S_t} &= \sqrt{v_t}\mathrm{d}\left(\rho W_t + \sqrt{1-\rho^2}W_t^{\perp}\right) \\ v_t &= \xi_0(t)\exp\left(\eta\int_0^t (t-s)^{H-1/2}\mathrm{d}W_s - \frac{1}{2}\eta^2 t^{2H}\right), \end{aligned} \tag{4}$$

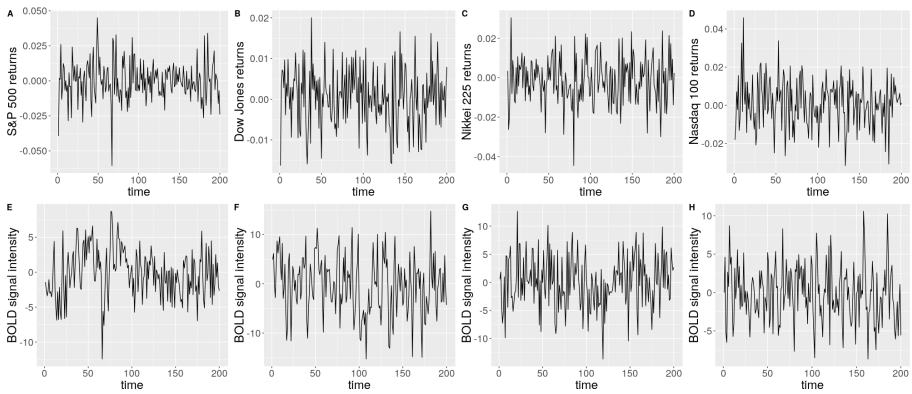
where W_t and $W_t^{\perp}$ represent two independent standard Brownian processes with $\rho \in [-1, 1]$, η is a positive constant that represents the volatility of volatility, and $\xi_0(\cdot)$ describes the initial variance curve, which we assume to be constant. We motivate passing from the original (2) to the driftless rough Bergomi model (4) by the practice of "*detrending*" mentioned above. Volatility processes simulated using the rBergomi model exhibit remarkable sim-

[1] This Brownian assumption implies a Hurst parameter of 0.5 (which is a parameter governing the Hölder regularity of paths) of the stochastic process. It is known, that for standard Brownian motion $H = 0.5$.

ilarity to realised volatility processes. Furthermore, the rBergomi and other rough models introduced since have been found to provide important improvements to forecasting volatility [2, 4, 42, 48].

In addition to improving forecasting accuracy, rough models can be used to assess smoothness of a given process by estimating the H parameter [23, 25, 34, 52]. Estimating the H parameter can provide information about the extent of heteroscedasticity in the series, but requires access to the realised, historical volatility process, which cannot be directly observed. To bypass this difficulty, in finance intra-day data such as 5-minute asset prices returns are used to estimate daily realised volatility [39, 44, 45]. The daily estimates are then combined to form a realised volatility process providing information about daily variances in an asset price over the course of months or years. Considering recent calls to explore the possibility of implementing models from the field of financial mathematics to fMRI [33, 54] and the visual similarities between financial returns data and fMRI BOLD signal (Figure 1), such an approach could be applied to fMRI data as well to examine time-dependent behaviour in volatility of the noise process. Utilising multi-echo acquisition, the data from each echo could be used as intra-time point data to estimate volatility. Thus, in a manner similar to standard combination of data from each echo time, we can produce a realised volatility series. These series could then be used to estimate the smoothness of volatility in fMRI data using models such as the rBergomi.

Figure 1: Financial returns and fMRI BOLD series



A. S&P 500 returns over 200 days during the year 2000; B. Dow Jones returns over 200 days over the year 2004; C. Nasdaq 225 returns over 200 days during the year 2003; D. Nikkei 100 returns over 200 days during the year 2000; E. Demeaned BOLD signal from voxel [28,14,26]; F. Demeaned BOLD signal from voxel [12,27,30]; G. Demeaned BOLD signal from voxel [17,36,24]; H. Demeaned BOLD signal from voxel [37,19,15]. All returns data is from the Oxford-Man database https://realized.oxford-man.ox.ac.uk/. All fMRI data is from participant sub-17821 from dataset ds000258 available at https://openneuro.org/datasets/ds000258.

Estimating rBergomi model parameters, including the H parameter, is computationally expensive and often relies on the use of Monte Carlo based calibration methods [10, 60]. This limits the use of this model in practice despite the benefits it offers [2, 19, 25]. Recently, neural networks have been proposed as an efficient way to solve the calibration problem [27, 28, 47, 49]. Neural networks provide a powerful way of identifying relationships between input parameters and model output and can be particularly useful for models that do not have closed-form solution [27, 28]. Recent work found that neural network calibration framework can be successfully applied to a range of rough stochastic volatility models to aid accurate pricing and hedging [6, 28]

The aim of this paper was to conduct an exploratory empirical study examining the volatility of fMRI noise. We were specifically interested in exploring whether volatility of fMRI noise exhibits time-dependent behaviour that cannot be explained by factors such as head motion and physiological noise. We aimed to collect multi-echo fMRI signal from a phantom to examine thermal noise. Observations collected at each echo time were treated as intra-time point data and were used to estimate realised volatility. The roughness of the realised volatility was assessed by estimating the H parameter, which was accomplished with a neural network approach. As the study was exploratory in nature we did not have prior hypotheses. However, considering the visual similarities between many

financial returns and fMRI BOLD series, we anticipated that the estimated H of the realised volatility processes was in the rough volatility range, $0 < H < \frac{1}{2}$.

3 Material and methods

3.1 FMRI data acquisition

An MRI phantom filled with liquid material was used to acquire multi-echo fMRI signal consisting entirely of thermal noise. The data were acquired with two different 3 Tesla GE Discovery MR750 units using 32-channel receive only head coils (Nova Medical, Wilmington, MA, USA). This was done to ensure the findings were not unique to a specific scanner. The functional multi-echo echo planar imaging (EPI) data consisted of 200 volumes and each volume consisted of 18 slices with the following parameters: 2.5 second repetition time (TR), 80° flip angle, 64 × 64 acquisition matrix, 3mm slice thickness with 4mm slice gap. The fMRI slices were acquired in an ascending order and eight echo times were used: 12 ms, 28 sm, 44 ms, 60 ms, 76 ms, 92 ms, 108 ms, 124 ms. Eight echo times were used as this was the maximum number of echoes that can be acquired with the MR units used.

3.2 FMRI data preprocessing

The phantom data was preprocessed using SPM12 (http://www.fil.ion.ucl.ac.uk/spm). Each echo was preprocessed separately to ensure the echoes could be used as intra-TR data to estimate realised volatility. The following preprocessing steps were taken: slice timing correction was applied first with the middle slice used as a reference slice. The data was then realigned and resliced to correct for head motion and estimate six rigid body transformations. Prior to combining the echoes and estimating realised volatility linear model based de-trending was conducted.

3.3 T_2^*-weighted realised volatility

As our data from the phantoms contains only noise, we can re-write equation (1) for each echo n at a given time point $t...T$ as

$$y_{n,t} = \sqrt{v_{n,t}}\epsilon_{n,t} \tag{5}$$

$\mu_{n,t} = 0$ as no true brain signal is present.

Observations from each echo time were treated as intra-TR data which were used to estimate realised volatility for each point in the time series. As fMRI signal decays rapidly, the observations from each echo time were weighted to avoid bias [32]. The weighting was based on T_2^* estimates, which were calculated in accordance with methodology used in tedana [17, 30, 31]:

$$S_n = S_0 \exp(-R_2^* \times E_n) \tag{6}$$

where S_n represents the signal intensity at a given echo time n, $R_2^* = \frac{1}{T_2^*}$, E represents the echo time in milliseconds, and S_0 represents the signal intensity at $E = 0$. The value of R_2^* is solved by a log-linear regression.

T_2^*-based weights were then calculated as follows

$$w_n = \frac{E_n \exp(-E/T_2^*)}{\sum_{n=1}^{N} E_n \exp(-E_n/T_2^*)} \tag{7}$$

The weights were used to estimate the mean of the fMRI noise processes at each echo time n at each time point, $t = 1...T$.

$$\bar{x}_t = \frac{\sum_{n=1}^{N} w_n y_{n,t}}{\sum_{n=1}^{N} w_n} \tag{8}$$

Realised volatility at each time point,$t = 1...T$, was estimated by calculating variance between observations at each echo time n.

$$\hat{v}_t = \frac{\sum_{n=1}^{N} w_n (y_{n,t} - \bar{x})^2}{\sum_{n=1}^{N} w_n} \tag{9}$$

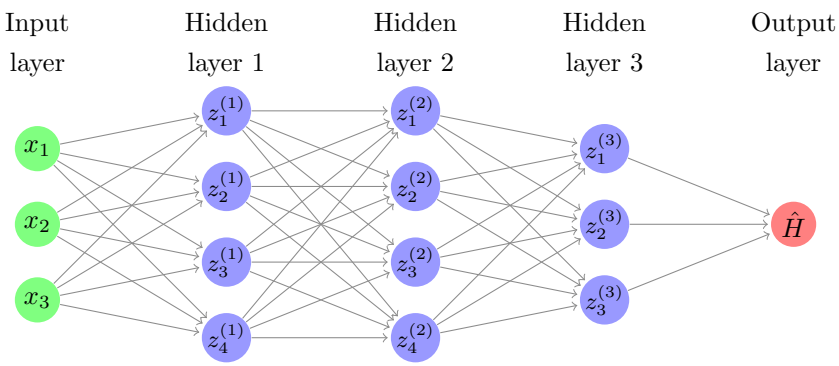
The estimated T_2^*-weighted variance, $\hat{v}_t$, served as a proxy of the unobserved volatility process and was used to investigate the smoothness of the fMRI noise series.

3.4 Estimating smoothness (Hurst parameter) of realised volatility

3.4.1 Neural network architecture

To estimate the smoothness of the fMRI noise volatility series we used a one-dimensional feed-forward convolutional neural network approach introduced by [49]. An introduction to neural networks is given in Appendix A; very simply one can think of a neural network as a composition of affine and non-linear functions that approximates a mapping of inputs to outputs.

The CNN consisted of three kernel layers with kernel size 20. The first convolutional layer had 32 kernels followed by a dropout layer with dropout rate of 0.25, the second had 64 kernels followed by a dropout layer with dropout rate of 0.25, the third had 128 kernels followed by a dropout layer with dropout rate of 0.4, and the fourth dense layer had 128 units followed by a dropout layer with dropout rate of 0.3. Leaky ReLU activation functions followed each layer with $\alpha = 0.1$ and max pooling layers with size 3 were added between each kernel layer.



3.4.2 Neural network training and test

For CNN training and testing, altogether 50,000 sample paths of the normalised rBergomi model log-volatility process, defined as $\tilde{v}_t := \eta \int_0^t (t-s)^{H-1/2} \mathrm{d}W_s$. were simulated, with 200 time points and $H \sim \text{Unif}(0, 1.0)$ and $\eta \sim \text{Unif}(0, 3)$ for each of the 50,000 sample paths. Stone provides a rigorous mathematical justification for this set up [49, Section 3.2, p382].

We took a nested cross-validation approach whereby the simulated sample paths were first divided into training and test datasets with 30% holdout. The training dataset was then further divided into training and validation sets with 20% hold out. Thus, the training dataset consisted of 28,000 training and 7,000 validation sample paths and the final test dataset included 15,000 sample paths.

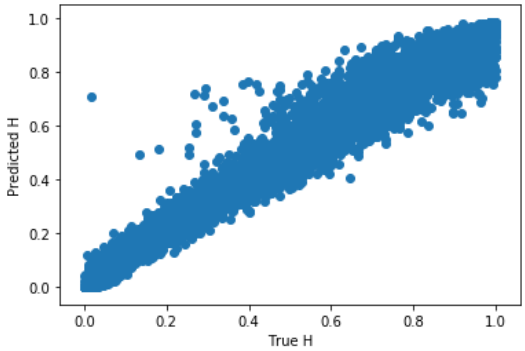
The sample paths were generated using classical methodology which utilises the Cholesky decomposition to achieve exact distribution of the log-volatility paths (add link to code here). The H and η values were sampled from uniform $[0, 0.5]$ and $[0, 3]$ distributions respectively. The sampling was conducted in a manner that ensured that each sample path had a unique H and η enabling better fitting to varying fMRI noise log-volatility processes.

The performance of the trained CNN was assessed using the final test dataset by extracting mean squared error to calculate root mean squared error (RMSE) between the predicted and true H parameters.

$$RMSE = \sqrt{\mathbb{E}((\hat{H} - H)^2)} \tag{10}$$

After training the $RMSE = 0.045$ and the correlation between predicted and true H can be seen in Figure 2.

Figure 2: Scatter plot showing the correlation between predicted and true H parameter



3.5 Evaluation of CNN H estimation

The H parameter provides not only an estimation of roughness of the volatility process, but also provided information about autocorrelation. Therefore, performance of the CNN was first evaluated by examining agreement between the estimated H parameters and the memory in the fMRI noise log-volatility series. Agreement between the CNN H parameter estimates and memory was evaluated by conducting a Spearman correlation test. Memory was estimated by a simple fitting autoregressive fractionally integrated moving average (ARFIMA) $[0, d, 0]$ model to the log-volatility data and calculating the d parameter.

$$\epsilon_t = (1 - B)^d \log(\sigma_t) \tag{11}$$

Where B is the backshift operator and d represents the memory parameter to be calculated. $0 < d < \frac{1}{2}$ indicates the series is a stationary, mean reverting long memory process, while $d < 0$ indicates the series is anti-persistent short memory process. $\frac{1}{2} < d < 1$ indicates the series is a mean reverting, non-stationary long memory process. Although the relationship between smoothness of log-volatility processes and long memory is a complicated one [11,15,25,41], this correlation will give us an indication of the performance of the CNN in estimating H.

4 Results

4.1 The estimated Hurst parameter

The summary statistics of the estimated H of the realised log-volatility series in phantoms 1 and 2 are presented in Table 1. On average the log-volatility series were rough, but there was substantial variability in the H estimates from voxel to voxel in both phantoms. The maximum estimated H remained under 0.5, suggesting that despite the substantial variability fMRI noise was rough across the phantoms.

Table 1: Estimated H exponent in Scanner 1, phantom 1 and Scanner 2, phantom 2

	Phantom 1	Phantom 2
Mean	0.025	0.025
Standard deviation	0.018	0.015
Max	0.476	0.361
Min	0.0004	0.0001

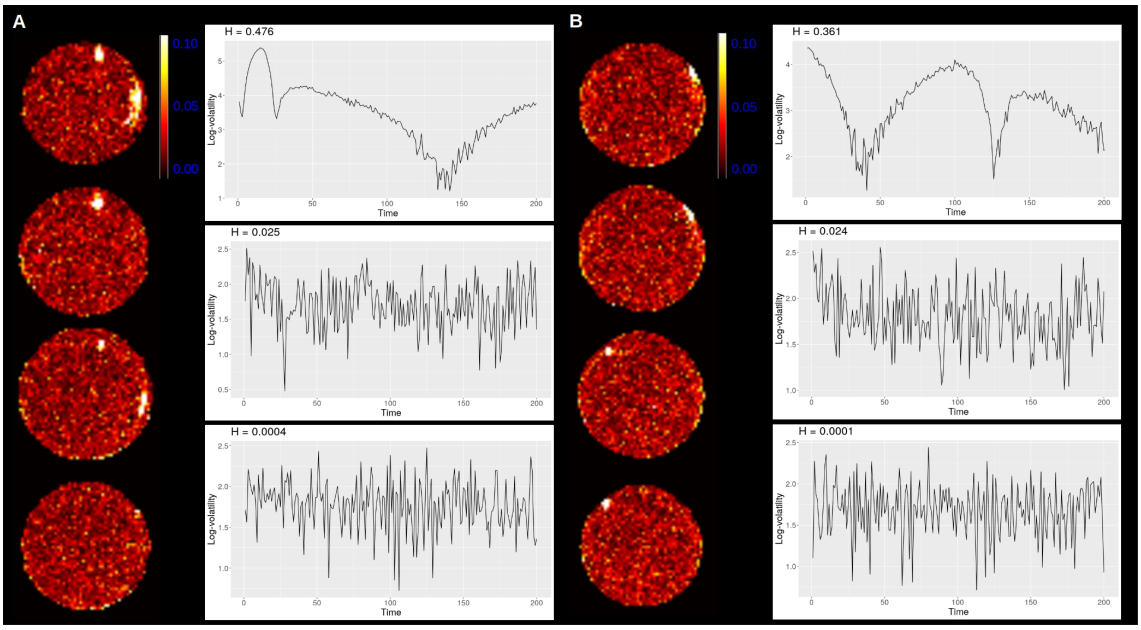
4.2 Spatial pattern in the estimated Hurst parameters

Figure 3 shows how the estimated H varied from region to region across the phantoms. Log-volatility series associate with the maximum, minimum, and H estimates close to the mean are also shown.

Some edge effects were observed, such that the estimated H was generally larger near edges of the phantoms than the middle. In both phantoms the voxels with the maximum H estimates were found near the edge and appeared to form large clusters. In the middle of the phantoms the H estimates were generally very small.

Interestingly, phantom 1 has a small region near the top where the signal intensity was lower than in the nearby voxels, suggesting signal dropout due to a possible air bubble (Supplementary Figure 1). This area consisted of four voxels and one of these voxels had the largest H estimate in phantom 1. This voxel also represents the centre of the cluster near the top of the phantom in Figure 3A. No such signal dropout was seem in phantom 2.

Figure 3: Spatial distribution of estimated H

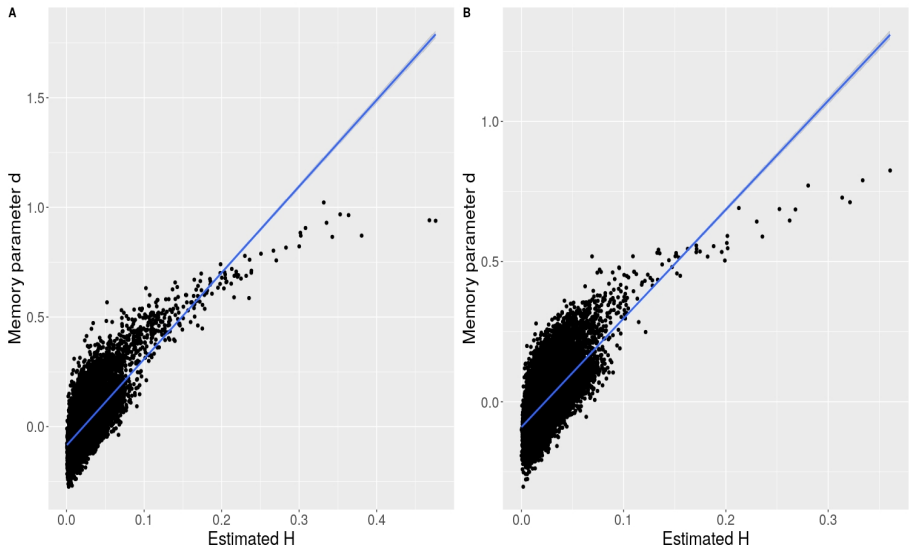


Multi-slice view of scanner 1 phantom 1 (A) and scanner 2 phantom 2 (B).

4.3 The connection between the correlation governed by the Hurst parameter H and autocorrelation

As mentioned earlier, the Hurst parameter does not only govern the roughness of a volatility path, but also the autocorrelation function of the volatility time series. So hence in this section we test for the agreement between the presenence of autocorrelation predicted by the rough volatility model and the autocorrelation predicted by standard ARFIMA models. The relationship between estimated H and ARFIMA$[0, d, 0]$ memory parameter, d, of the log-volatility processes are presented in Figure 4. the correlation was significant and positive in both phantoms: $r = 0.64, p < 0.001$ and $r = 0.65, p < 0.001$ for phantoms 1 and 2, respectively. Additionally, as can be seen in Figure 4, the log-volatility processes with larger estimated H had substantially longer memory than log-volatility processed with very small estimated H. Some log-volatility series with small H estimates still showed long memory, $d > 0$, but some were anti-persistent with $d < 0$. This suggests that the performance of the model was acceptable, providing accurate H estimates.

Figure 4: Relationship between ARFIMA memory parameters d and estimated H



A. Phantom 1; B. Phantom 2. d = ARFIMA$[0, d, 0]$ memory parameter, H = Hurst parameter

5 Discussion

The aim of the present empirical study was to examine the smoothness of fMRI noise volatility. We used multi-echo scans of a phantom from two different MRI scanners to estimate realised volatility. Smoothness of logarithm of the realised volatility processes was estimated using CNN methodology introduced in [49]. The findings indicated that on average the volatility of fMRI noise is very rough with $H \approx 0.03$, but substantial variability was also observed. The variability was caused by the fact that the smoothness of the volatility was not constant across the phantoms, with higher H estimates observed near the edges of the images. Overall, all $H < 0.5$ suggesting that across the phantoms volatility was consistently rough.

The present findings suggest that log-volatility of fMRI noise appears to behave like fractional Brownian motion with $H \approx 0.03$. As anticipated, these findings go some way to mimic the rough volatility pattern observed in high frequency financial data with Hurst estimates varying between 0.02 and 0.14 [23, 25]. Thus, it appears that although fMRI scanner noise on average does not have large fluctuations in volatility over time, the noise processes exhibit some heteroscedasticity that cannot be explained by head motion, physiology, or other known sources of non-constant noise. This finding challenges the assumption that fMRI noise has constant volatility and adds to the steady accumulation of literature exploring heteroscedasticity in fMRI noise [18, 33, 36, 43, 55, 59], further highlighting the importance of taking non-constant noise into consideration during analysis of the time series data

Heteroscedasticity in the fMRI noise process that cannot be easily removed using standard de-noising methods, such as linear regression, may cause problems during data analysis. A previous study examined the impact of heteroscedasticity introduced by simulated head motion spikes on fMRI data analysis [18]. The authors found that a linear modelling approach based on weighted least sum of squares (WLSS) was able to accurately model impulse responses to stimuli if the heteroscedasticity was constant across all voxels [18]. However, when the number of head motion spikes varied from voxel to voxel, the WLSS failed to accurately detect impulse responses. These findings led the authors to propose a heteroscedastic general linear model which incorporates head motion covariates. However, our findings suggest that heteroscedasticity can also be present in the scanner noise, which cannot be easily entered into analysis as a covariate because thermal noise processes cannot be directly observed during brain scanning.

Furthermore, resting state fMRI analyses examining functional connectivity between time series using simple models such as Pearson correlation tests, are also likely to be affected by non-constant noise. Indeed, a few studies have recently proposed using deterministic volatility models, autoregressive conditional heteroscedasticity (ARCH) and generalised ARCH (GARCH), to aid investigation of time-dependent correlations [16, 33, 54]. Models based on ARCH and GARCH have been found to accurately model time-dependent functional connectivity in both simulation and real data experiments, while traditional approaches, including sliding window and exponentially weighted moving average models, have been found to produce more false positive findings [13, 16, 33, 54]. Thus, further investigation into implementing volatility models to fMRI time series analysis could ultimately help improve both resting state and task-based data analysis as the noise in the time series is better understood [37, 43].

Interestingly, present findings also show that the roughness of fMRI noise is not constant across regions in the phantom with the edges showing greater smoothness in the volatility and thus increased heteroscedasticity relative to the centre of the phantom. To an extent these findings mirror those from previous work examining long-range dependence in the mean of fMRI noise [40, 56]. Previous studies have found that the long-range dependence near the edges of the phantom has estimated $H > 0.5$ indicating that the processes are persistence in the mean [40]. Similar edge effects have also been observed in real brain scans [40, 50]. Along the same lines, a previous study investigating normality of fMRI data revealed that signal from the cortex, which lies near the edge of the image, showed greater non-normality than that from the middle of the images [37]. Taken together with the present findings this suggests that fMRI data near the edges of an image appears to be more complex than that near the centre where the dependence in the mean is low and volatility near constant. Such time-dependent behaviour in the noise near the edges complicates data analysis as these effects violate assumptions of most time series modelling

methods and can lead to both spurious regressions and correlations [7, 9, 14, 20, 51, 53, 57, 58]. This is particularly problematic for studies interested in examining brain activity in and around the cortex.

It is also important to note that in the present study, one of the phantoms had a small region of signal dropout, possibly indicating a presence of an air bubble. This region was the centre of one of the clusters where the smoothness of the volatility process was greater than in nearby regions. Previous studies have also found that air bubbles in phantoms can lead to drop in signal intensity, which has been suggested to due to susceptibility artifacts at the air-water boundary [38]. Air bubbles can also introduce phase errors and related magnetic field heterogeneity [29, 35], which could go some way to explain the larger H estimates in one of the clusters in one of the phantoms. Interestingly, such an effect was only found in one of the phantoms, suggesting that all the edge effects could not be explained by air bubbles. Still further investigation of the spatial pattern of volatility in fMRI noise in a gel phantom prepared with warm water, which are less susceptible to air bubbles [26], would be of interest.

The present study is not without limitations. Realised volatility was estimated after slice timing correction and realignment, but no further preprocessing or de-noising steps were taken prior to estimation. This was done in an attempt to mirror standard multi-echo preprocessing pipelines where the echoes are normally combined prior to further preprocessing steps, such as smoothing, and independent component analysis-base de-noising [21, 31, 32]. This meant that we were unable to examine the impact of de-noising on realised volatility. Additionally, realised volatility was estimated using only eight echo time points as this was the maximum number we were able to collect. In finance, on the other hand, it is common to use high frequency asset price data, with sub-second granularity, to estimate daily volatility. It is difficult to ascertain whether our use of lower frequency data to estimate realised volatility had an impact on the present findings. Finally, the phantom data was collected from two very similar 3 Tesla MR scanners. It is possible that the volatility of fMRI noise from scanners with different field strengths might vary and further investigation of this may be of interest.

6 Conclusions

The aim of the present study was to examine the smoothness of estimated realised volatility of fMRI noise. This was done by conducting two multi-echo scans of a phantom using two different MRI scanner units. Multi-echo data was used to estimate realised volatility by T_2*-weighted variance. Smoothness of the realised volatility data was estimated by following cutting edge methods developed in the field of financial mathematics, namely by training a CNN to predict the Hurst parameter, H. The findings showed that on average fMRI noise is very rough with $H \approx 0.02$. Furthermore, it was shown that roughness of the volatility data varied across the spatially across the phantoms. In both phantom scans the H estimates were larger near the edges, suggesting that volatility was smoother in these regions. Taken together the present findings further challenge the assumption that fMRI scanner noise has constant volatility and highlight the need for further research to investigate how to effectively model the heteroscedasticity during time series analysis.

7 Acknowledgements

JL is supported by Sir Henry Wellcome Postdoctoral Fellowship (213578/Z/18/Z). The funding body did not play an active role in the design of this study, nor in data collection or analysis, nor in writing the manuscript.

References

[1] Elisa Alòs, Jorge A León, and Josep Vives. On the short-time behavior of the implied volatility for jump-diffusion models with stochastic volatility. *Finance and Stochastics*, 11(4):571–589, 2007.

[2] Christian Bayer, Peter Friz, and Jim Gatheral. Pricing under rough volatility. *Quantitative Finance*, 16(6):887–904, 2016.

[3] Erik B. Beall. Adaptive cyclic physiologic noise modeling and correction in functional mri. *Journal of Neuroscience Methods*, 187(2):216 – 228, 2010.

[4] Mikkel Bennedsen, Asger Lunde, and Mikko S. Pakkanen. Decoupling the short- and long-term behavior of stochastic volatility, 2016.

[5] Paul Brockman and Mustafa Chowdhury. Deterministic versus stochastic volatility: implications for option pricing models. *Applied Financial Economics*, 7(5):499–505, 1997.

[6] Hans Buehler, Lukas Gonon, Ben Wood, Josef Teichmann, Baranidharan Mohan, and Jonathan Kochems. Deep hedging: Hedging derivatives under generic market frictions using reinforcement learning-machine learning version. *Available at SSRN*, 2019.

[7] Nunzio Cappuccio and Diego Lubian. Spurious regressions between i(1) processes with long memory errors. *Journal of Time Series Analysis*, 18(4):341–354, 1997.

[8] Joshua C.C. Chan and Angelia L. Grant. Modeling energy price dynamics: Garch versus stochastic volatility. *Energy Economics*, 54:182 – 189, 2016.

[9] Nixon S. Chekenya and Canicio Dzingirai. The impact of the presence of autoregressive conditional heteroscedasticity (arch) effects on spurious regressions. *Scientific African*, 8:e00382, 2020.

[10] Bin Chen, Cornelis W. Oosterlee, and Hans Van Der Weide. Efficient unbiased simulation scheme for the sabr stochastic volatility model. 2011.

[11] Zhiyao Chen, Robert T. Daigler, and Ali M. Parhizgari. Persistence of volatility in futures markets. *Journal of Futures Markets*, 26(6):571–594, 2006.

[12] Patrick Cheridito, Hideyuki Kawaguchi, and Makoto Maejima. Fractional ornstein-uhlenbeck processes. *Electron. J. Probab.*, 8:14 p., 2003.

[13] Ann S. Choe, Mary Beth Nebel, Anita D. Barber, Jessica R. Cohen, Yuting Xu, James J. Pekar, Brian Caffo, and Martin A. Lindquist. Comparing test-retest reliability of dynamic functional connectivity methods. *NeuroImage*, 158:155 – 175, 2017.

[14] Heetaik Chung and Joon Y. Park. Nonstationary nonlinear heteroskedasticity in regression. *Journal of Econometrics*, 137(1):230 – 259, 2007.

[15] Rama Cont. *Volatility Clustering in Financial Markets: Empirical Facts and Agent-Based Models*, pages 289–309. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

[16] A. Duggento, L. Passamonti, M. Guerrisi, and N. Toschi. Simultaneous estimation of the in-mean and in-variance causal connectomes of the human brain. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4371–4374, July 2017.

[17] Elizabeth DuPre, Taylor Salo, Ross Markello, Prantik Kundu, Kirstie Whitaker, and Dan Handwerker. Meica/tedana.

[18] Anders Eklund, Martin A. Lindquist, and Mattias Villani. A bayesian heteroscedastic glm with application to fmri data with motion spikes. *NeuroImage*, 155:354 – 369, 2017.

[19] Omar El Euch and Mathieu Rosenbaum. Perfect hedging in rough heston models. *Ann. Appl. Probab.*, 28(6):3813–3856, 12 2018.

[20] Philip Ernst, Larry Shepp, and Abraham Wyner. Yule's "nonsense correlation" solved!, 2016.

[21] Oscar Esteban, Christopher Markiewicz, Ross W Blair, Craig Moodie, Ayse Ilkay Isik, Asier Erramuzpe Aliaga, James Kent, Mathias Goncalves, Elizabeth DuPre, Madeleine Snyder, Hiroyuki Oya, Satrajit Ghosh, Jessey Wright, Joke Durnez, Russell Poldrack, and Krzysztof Jacek Gorgolewski. fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods*, 2018.

[22] Masaaki Fukasawa. Asymptotic analysis for stochastic volatility: martingale expansion. *Finance and Stochastics*, 15(4):635–654, 2011.

[23] Masaaki Fukasawa, Tetsuya Takabatake, and Rebecca Westphal. Is volatility rough ?, 2019.

[24] René Garcia and Èric Renault. A note on hedging in arch and stochastic volatility option pricing models. *Mathematical Finance*, 8(2):153–161, 1998.

[25] Jim Gatheral, Thibault Jaisson, and Mathieu Rosenbaum. Volatility is rough. *Quantitative Finance*, 18(6):933–949, 2018.

[26] Alexandra Hellerbach, Verena Schuster, Andreas Jansen, and Jens Sommer. Mri phantoms – are there alternatives to agar? *PLOS ONE*, 8(8):1–8, 08 2013.

[27] Andres Hernandez. Model calibration with neural networks. *Available at SSRN 2812140*, 2016.

[28] Blanka Horvath, Aitor Muguruza, and Mehdi Tomas. Deep Learning Volatility. *SSRN Electronic Journal*, feb 2019.

[29] Jun-Ho Kim, Jung-Hyun Kim, So-Hee Lee, Jinhyoung Park, and Seung-Kyun Lee. Fabrication of a spherical inclusion phantom for validation of magnetic resonance-based magnetic susceptibility imaging. *PLOS ONE*, 14(8):1–16, 08 2019.

[30] Prantik Kundu, Noah D. Brenowitz, Valerie Voon, Yulia Worbe, Petra E. Vértes, Souheil J. Inati, Ziad S. Saad, Peter A. Bandettini, and Edward T. Bullmore. Integrated strategy for improving functional connectivity mapping using multiecho fmri. *Proceedings of the National Academy of Sciences*, 110(40):16187–16192, 2013.

[31] Prantik Kundu, Souheil J. Inati, Jennifer W. Evans, Wen-Ming Luh, and Peter A. Bandettini. Differentiating bold and non-bold signals in fmri time series using multi-echo epi. *NeuroImage*, 60(3):1759 – 1770, 2012.

[32] Prantik Kundu, Valerie Voon, Priti Balchandani, Michael V. Lombardo, Benedikt A. Poser, and Peter A. Bandettini. Multi-echo fmri: A review of applications in fmri denoising and analysis of bold signals. *NeuroImage*, 154:59 – 80, 2017. Cleaning up the fMRI time series: Mitigating noise with advanced acquisition and correction strategies.

[33] Martin A. Lindquist, Yuting Xu, Mary Beth Nebel, and Brain S. Caffo. Evaluating dynamic bivariate correlations in resting-state fmri: A comparison study and a new approach. *NeuroImage*, 101:531 – 546, 2014.

[34] Giulia Livieri, Saad Mouti, Andrea Pallavicini, and Mathieu Rosenbaum. Rough volatility: Evidence from option prices. *IISE Transactions*, 50(9):767–776, 2018.

[35] Hanbing Lu, Yousef Mazaheri, Rongyan Zhang, Andrzej Jesmanowicz, and James S. Hyde. Multishot partial-k-space epi for high-resolution fmri demonstrated in a rat whisker barrel stimulation model at 3t. *Magnetic Resonance in Medicine*, 50(6):1215–1222, 2003.

[36] Torben E. Lund, Kristoffer H. Madsen, Karam Sidaros, Wen-Lin Luo, and Thomas E. Nichols. Non-white noise in fmri: Does modelling have an impact? *NeuroImage*, 29(1):54 – 66, 2006.

[37] Wen-Lin Luo and Thomas E. Nichols. Diagnosis and exploration of massively univariate neuroimaging models. *NeuroImage*, 19(3):1014 – 1032, 2003.

[38] Seppo Mattila, Ville Renvall, Jaana Hiltunen, Douglas Kirven, Raimo Sepponen, Riitta Hari, and Antti Tarkiainen. Phantom-based evaluation of geometric distortions in functional magnetic resonance and diffusion tensor imaging. *Magnetic Resonance in Medicine*, 57(4):754–763, 2007.

[39] Michael McAleer and Marcelo C. Medeiros. Realized volatility: A review. *Econometric Reviews*, 27(1-3):10–45, 2008.

[40] Kaare Mikkelsen and Torben Ellegaard Lund. Sampling rate dependence of correlation at long time lags in bold fmri measurements on humans and gel phantoms. *Frontiers in Physiology*, 4:106, 2013.

[41] Thomas Mikosch. Is it really long memory we see in financial returns? Econometrics 0412002, University Library of Munich, Germany, December 2004.

[42] Sovan Mitra. A Review of Volatility and Option Pricing. Papers 0904.1292, arXiv.org, April 2009.

[43] Martin Monti. Statistical analysis of fmri time-series: A critical review of the glm approach. *Frontiers in Human Neuroscience*, 5:28, 2011.

[44] Andrew J. Patton. Data-based ranking of realised volatility estimators. *Journal of Econometrics*, 161(2):284 – 303, 2011.

[45] Andrew J. Patton and Kevin Sheppard. Optimal combinations of realised volatility estimators. *International Journal of Forecasting*, 25(2):218 – 238, 2009. Forecasting Returns and Risk in Financial Markets using Linear and Nonlinear Models.

[46] Ser-Huang Poon and Clive W.J. Granger. Forecasting volatility in financial markets: A review. *Journal of Economic Literature*, 41(2):478–539, June 2003.

[47] Johannes Ruf and Weiguan Wang. Neural networks for option pricing and hedging: a literature review, 2019.

[48] Mawuli Segnon, Thomas Lux, and Rangan Gupta. Modeling and forecasting the volatility of carbon dioxide emission allowance prices: A review and comparison of modern volatility models. *Renewable and Sustainable Energy Reviews*, 69:692 – 704, 2017.

[49] Henry Stone. Calibrating rough volatility models: a convolutional neural network approach. *Quantitative Finance*, 20(3):379–392, 2020.

[50] John Suckling, Anna Barnes, Dominic Job, David Brennan, Katherine Lymer, Paola Dazzan, Tiago Reis Marques, Clare MacKay, Shane McKie, Steve R. Williams, Steven C.R. Williams, Bill Deakin, and Stephen Lawrie. The neuro/psygrid calibration experiment. *Human Brain Mapping*, 33(2):373–386, 2012.

[51] Yixiao Sun. Spurious regressions between stationary generalized long memory processes. *Economics Letters*, 90(3):446 – 454, 2006.

[52] Tetsuya Takaishi. Rough volatility of bitcoin. *Finance Research Letters*, 32:101379, 2020.

[53] Wen-Jen Tsay and Ching-Fan Chung. The spurious regression of fractionally integrated processes. *Journal of Econometrics*, 96(1):155 – 182, 2000.

[54] Toktam Valizadeh and Saeid Rezakhah. Flexible cholesky garch model with time dependent coefficients, 2018.

[55] Maziar Yaesoubi, Elena A. Allen, Robyn L. Miller, and Vince D. Calhoun. Dynamic coherence analysis of resting fmri data to jointly capture state-based phase, frequency, and time-domain information. *NeuroImage*, 120:133 – 142, 2015.

[56] Lirong Yan, Yan Zhuo, Yongquan Ye, Sharon X. Xie, Jing An, Geoffrey K. Aguirre, and Jiongjiong Wang. Physiological origin of low-frequency drift in blood oxygen level dependent (bold) functional magnetic resonance imaging (fmri). *Magnetic Resonance in Medicine*, 61(4):819–827, 2009.

[57] Gawon Yoon. Correlation coefficients, heteroskedasticity and contagion of financial crises. *The Manchester School*, 73(1):92–100, 2005.

[58] G Udny Yule. Why do we sometimes get nonsense-correlations between time-series?–a study in sampling and the nature of time-series. *Journal of the royal statistical society*, 89(1):1–63, 1926.

[59] Andrew Zalesky and Michael Breakspear. Towards a statistical test for functional connectivity dynamics. *NeuroImage*, 114:466 – 470, 2015.

[60] Álvaro Leitao, Lech A. Grzelak, and Cornelis W. Oosterlee. On an efficient multiple time step monte carlo simulation of the sabr model. *Quantitative Finance*, 17(10):1549–1565, 2017.

A Artificial Neural Networks

In this appendix we repeat the introduction given by Stone [49, Section 2.1, p382] for the reader's convenience. An artificial neural network is a biologically inspired system of interconnected processing units, where each processing unit is called a layer. Inputs to each layer, apart from the first layer, are outputs from previous layers. A layer is composed of a number of nodes, and each node in a given layer is connected to the nodes in a subsequent layer, thus forming a network; each edge in this network has a weight associated to it. The first processing unit is called the input layer, and the final processing unit is the output layer. The processing unit or units between the input layer and output layer are referred to as hidden layers; typically artificial neural networks have more than one hidden layer.

Convolutional neural networks (CNNs) are a class of artificial neural networks, where the hidden layers can be grouped into different classes according to their purpose; one such class of hidden layer is the eponymous convolutional layer. Below we describe the classes of hidden layers used in our CNN. Of course, this list is not exhaustive, and there exist many classes of hidden layers that we omit for means of brevity. Note also that we describe a CNN in the context of the problem we are trying to solve, where the input data are one dimensional vectors. CNNs can of course also be used on higher dimensional input data, but the fundamental structure and different roles of the hidden layers do not change.

- **Convolutional Layer:** In deep learning, the convolution operation is a method used to assign relative value to entries of input data, in our case one dimensional vectors of time series data, while simultaneously preserving spatial relationships between individual entries of input data. For a given kernel size k and an input vector of length m, the convolution operation takes entries $1, ..., k$ of the input vector and multiplies by the kernel element-wise, whose length is k. The sum of the entries of the resulting vector are then the first entry of the feature map. This operation is iterated $m + 1 - k$ times, thus incorporating every entry in the input data vector into the convolution operation. The output of the convolutional layer is the feature map.

  For example, let $(1, 2, 1, 0, 0, 3)$ be our input vector, and $(1, 0, 1)$ be our kernel; here the kernel size is 3. The first iteration of the convolution operation involves taking the element-wise multiple of $(1, 2, 1)$ and $(1, 0, 1)$: $(1, 0, 1)$ is produced and the sum, equal to 2, is computed. This is the first entry of the feature map. The resulting feature map in this example is then $(2, 2, 1, 3)$.

  Clearly, the centre of each kernel cannot overlap with the first and final entry of the input vector. Zero-padding, sometimes referred to as same-padding, preserves the dimensions of input vectors and allows more layers to be applied in the CNN: zero-padding is simply the extension of the input vector and the setting of the first and final entries as 0, while leaving the other entries unchanged. In our example, the input vector becomes $(0, 1, 2, 1, 0, 0, 3, 0)$ after zero padding.

- **Activation Layer:** The activation layer is a non-linear function σ that is applied to the output of the convolutional layer i.e. the feature map; the purpose of the activation layer is indeed to introduce non-linearity into the CNN. Examples of activation functions include the sigmoid function and the hyperbolic tangent function. In our CNN we use the 'LeakyReLU' activation function, defined as

  $$f_\alpha(x) := \begin{cases} x, & \text{if } x > 0, \\ \alpha x, & \text{otherwise.} \end{cases}$$

  The LeakyReLU activation function allows a small positive gradient when the unit is inactive.

- **Max Pooling Layer:** For a given pooling size p, the max pooling layer returns a vector whose entries are the maximum among the neighbouring p entries in the feature map. For example, for feature map $(1, 3, 8, 2, 1, 0, 0, 4, 6, 1)$ and $p = 3$ the max pooling output is $(8, 8, 8, 8, 8, 2, 4, 6, 6, 6)$.

  Other pooling techniques apply the same idea, but use different functions to evaluate the neighbouring p entries in the feature map. Examples include average pooling, and L2-norm pooling, which in fact uses the Euclidean norm in mathematical nomenclature.

- **Dropout Layer:** Dropout is a well-known technique incorporated into CNNs in order to prevent overfitting. Without the addition of a dropout layer, each node in a given layer is connected to each node in the subsequent layer; dropout temporarily removes nodes from different layers in the network. The removal of nodes is random and determined by the dropout rate d, which gives the proportion of nodes to be temporarily dropped. Note that dropout is only implemented during training; during testing the weights of each node are multiplied by the dropout rate d.

- **Dense Layer:** Also referred to as the fully connected layer, each node in the input layer is connected to each node in the output layer as the name suggests. After being processed by the convolutional, activation, pooling, and dropout layers, the extracted features are then mapped to the final outputs via a subset of the dense layer, an activation function is then applied subsequently. This activation function is chosen specifically for the task that the CNN is required to execute, i.e. binary/multi-class classification, or regression to output a continuous value. The final output from the dense layer has the same number of nodes as the number of classes in the output data.