

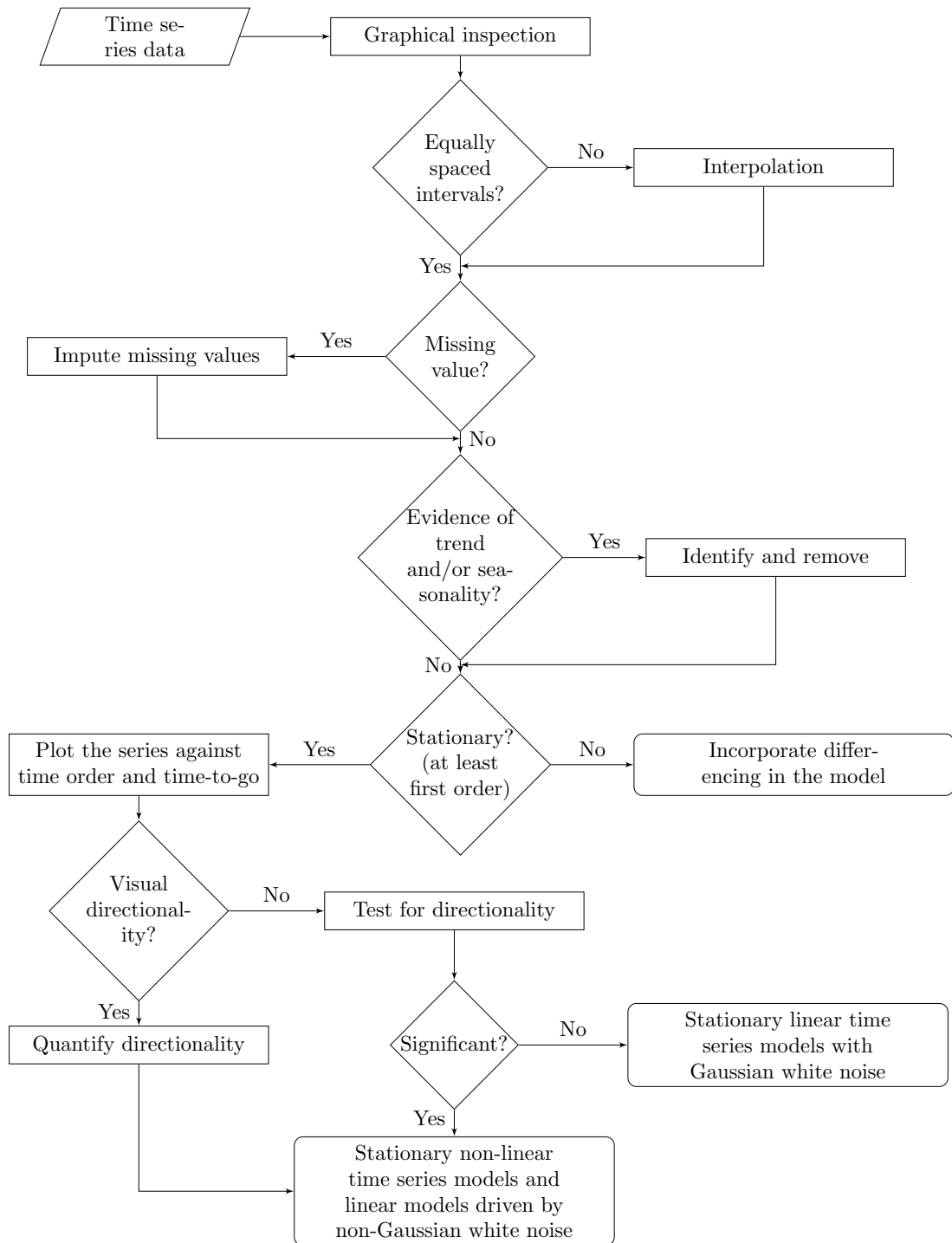
Supplementary Material to “Detecting Directionality in Time Series”

Mahayaudin M. Mansor, David A. Green
and Andrew V. Metcalfe
School of Mathematical Sciences,
University of Adelaide, Adelaide 5005 SA, Australia.

This document provides supporting information to the article. It contains:

1. A flowchart summarizing the general procedure for detecting directionality in time series;
2. R routines to demonstrate reversibility with realizations from: (i) MA(3) with exponential noise, (ii) MA(3) exponential noise with cube errors term, (iii) AR(1) with Gaussian white noise, and directionality with realizations from (iv) AR(1) with uniform noise;
3. R syntax for moving blocks bootstrap and Monte-Carlo simulations, and realizations from AR(1) with Gaussian errors;
4. Application of moving blocks bootstrap to sunspots series (length 315) and EEG (D) series (length 4096);
5. Results from the suite of directionality tests applied to 12 time series, and boxplots of first differences;
6. GARCH model for the residuals of AR models on dtlds tourist arrivals;
7. Poisson regression for the sunspots; and
8. Moving directionality and moving volatility of the Wells Fargo log-returns.

1 Flowchart for detecting directionality in stationary time series



2 R routines to demonstrate reversibility and directionality

A limitation of directionality statistic $\hat{\gamma}_{dab}$ (SKdab) is that it is less sensitive than $\hat{\gamma}_d$ (SKd) because the noise to signal ratio will be doubled by taking absolute values of mean corrected observations. However, it is a useful variant as the following case show. The time series model

$$X_t = \alpha X_{t-1} + \varepsilon_t$$

where ε_t are a sequence of independent uniform random variables is not reversible by Weiss's result (Weiss 1975). However, realizations of length 10^6 with, for example $\alpha = 0.4$, no directionality is apparent in using the statistic $\hat{\gamma}_d$. In contrast $\hat{\gamma}_{dab}$ does detect directionality in the realizations of AR(1) with uniform noise in Section 2.4. Directionality can also be detected indirectly as residuals near uniform rather than Gaussian.

2.1 Moving average process of order 3, MA(3), exponential noise

```
set.seed(1)
n=10^6
z=rexp(n)-1
x=rep(0,n);y=rep(0,n)
x[3]=0
for(i in 4:n){
  x[i]=z[i-1]+z[i-2]+z[i-3]
}
y=diff(x)
SKd=(sum((y-mean(y))^3)/n)/sd(y)^3
print(SKd)
[1] -0.0009076836

set.seed(1)
n=10^6
z=rexp(n)-1
x=rep(0,n);y=rep(0,n);xx=rep(0,n)
x[3]=0
for(i in 4:n){
  x[i]=z[i-1]+z[i-2]+z[i-3]
}
xx=abs(x-mean(x))
y=diff(xx)
SKdab=(sum((y-mean(y))^3)/n)/sd(y)^3
print(SKdab)
[1] -0.003919842
```

2.2 MA(3) exponential noise with cube errors term

```
set.seed(1)
n=10^6
z=rexp(n)-1
x=rep(0,n);y=rep(0,n)
x[3]=0
```

```

for(i in 4:n){
x[i]=z[i-1]+2*z[i-2]^3+z[i-3]
}
y=diff(x)
SKd=(sum((y-mean(y))^3)/n)/sd(y)^3
print(SKd)
[1] -0.02307794

set.seed(1)
n=10^6
z=rexp(n)-1
x=rep(0,n);y=rep(0,n);xx=rep(0,n)
x[3]=0
for(i in 4:n){
x[i]=z[i-1]+2*z[i-2]^3+z[i-3]
}
xx=abs(x-mean(x))
y=diff(xx)
SKdab=(sum((y-mean(y))^3)/n)/sd(y)^3
print(SKdab)
[1] -0.02307794

```

2.3 Autoregressive process of order 1, AR(1), with Gaussian white noise

```

set.seed(1)
n=10^6
a=0.4
x=rep(0,n);y=rep(0,n)
for(i in 2:n){
x[i]=a*x[i-1]+rnorm(1)
}
y=diff(x)
SKd=(sum((y-mean(y))^3)/n)/sd(y)^3
print(SKd)
[1] -0.002230051

set.seed(1)
n=10^6
a=0.4
x=rep(0,n);y=rep(0,n);xx=rep(0,n)
for(i in 2:n){
x[i]=a*x[i-1]+rnorm(1)
}
xx=abs(x-mean(x))
y=diff(xx)
SKdab=(sum((y-mean(y))^3)/n)/sd(y)^3
print(SKdab)
[1] -0.001333719

```

2.4 AR(1) with uniform noise

```
set.seed(1)
n=10^6
a=0.4
x=rep(0,n);y=rep(0,n)
for(i in 2:n){
  x[i]=a*x[i-1]+runif(1)
}
y=diff(x)
SKd=(sum((y-mean(y))^3)/n)/sd(y)^3
print(SKd)
[1] 0.0002462343

set.seed(1)
n=10^6
a=0.4
x=rep(0,n);y=rep(0,n);xx=rep(0,n)
for(i in 2:n){
  x[i]=a*x[i-1]+runif(1)
}
xx=abs(x-mean(x))
y=diff(xx)
SKdab=(sum((y-mean(y))^3)/n)/sd(y)^3
print(SKdab)
[1] -0.2427602
```

There is no evidence of directionality from the realizations from models 2.1 MA(3) exponential, 2.2 MA(3) exponential and cubed term, and 2.3 AR(1) Gaussian. This is consistent with the theoretical reversibility. In contrast, the value of $\hat{\gamma}_{dab} = -0.24$ is evidence of directionality in realizations from model 2.4 AR(1) with uniform noise. The negative value of $\hat{\gamma}_{dab}$ corresponds to relatively slow drifts from the mean followed by steeper returns to the mean. The reason is that the uniform distribution of errors does not have tail values and so rapid increases or decreases are reduced in magnitude and tend to be followed by relatively quick returns to the mean. The effect is slight but detectable.

3 R syntax for moving blocks bootstrap and Monte-Carlo simulations, and realizations from AR(1) with Gaussian errors

3.1 R code for moving blocks bootstrap

```
set.seed(1)
data=insert_data_here
N=length(data)
#R is number of replications
#SKd is the skewness of differences
#k is the size of each block
#data.bt is the bootstrap series of length N
R=1000;SKd=rep(0,R);k=50
for(r in 1:R){
```

```

data_bt=rep(0,N)
  for(i in 1:ceiling(N/k)){
    endpoint=sample(k:N,size=1)
    data_bt[(i-1)*k+1:k]= data[endpoint-(k:1)+1]
  }
data_bt=data_bt[1:N]
y=diff(data_bt)
SKd_bt=(sum((y-mean(y))^3)/length(y))/sd(y)^3
SKd[r]= SKd_bt
}
print(mean(SKd))
print(sd(SKd))
hist(SKd,main="SKd of the series, k=50, nrep=1000")

```

3.2 R code for Monte-Carlo simulations on realizations of AR(1) model with Gaussian errors

```

#data = AR(1) + rnorm
#length N=315, then N=4096

set.seed(1)
N=315
x=rep(0,N);x.ar=rep(0,N)
a=0.4
for(j in 2:N){x[j]=a*x[j-1]+rnorm(1)}
data=x
m=ar(data,order.max=1)
R=1000;SKd=rep(0,R)
for(r in 1:R){
  x.ar=rep(0,N)
  for(g in 2:N){x.ar[g]=m$ar[1]*x.ar[g-1]+rnorm(1)*sd(m$res[-1])}
}
y=diff(x.ar)
SKd_ar=(sum((y-mean(y))^3)/length(y))/sd(y)^3
SKd[r]=SKd_ar
}
print(mean(SKd))
print(sd(SKd))

```

3.3 R code for moving blocks bootstrap on realizations of AR(1) model with Gaussian errors

```

#data = AR(1) + rnorm
#length N=315, then N=4096

set.seed(1)
N=315
x=rep(0,N)
a=0.4
for(j in 2:N){x[j]=a*x[j-1]+rnorm(1)}
data=x

```

```

N=length(data)
R=1000;SKd=rep(0,R);k=50
for(r in 1:R){
  data_bt=rep(0,N)
  for(i in 1:ceiling(N/k)){
    endpoint=sample(k:N,size=1)
    data_bt[(i-1)*k+1:k]=data[endpoint-(k:1)+1]
  }
  data_bt=data_bt[1:N]
  y=diff(data_bt)
  SKd_bt=(sum((y-mean(y))^3)/length(y))/sd(y)^3
  SKd[r]=SKd_bt
}
print(mean(SKd))
print(sd(SKd))

```

3.4 Test of code

The script were run for one realization of length 315 and one realization of length 4096 from AR(1) with $\alpha = 0.4$.

```

#method: Monte-Carlo simulations
#when N=315
> print(mean(SKd))
[1] 0.005127861
> print(sd(SKd))
[1] 0.1291475
#when N=4096
> print(mean(SKd))
[1] -0.00174904
> print(sd(SKd))
[1] 0.03767824

#method: moving blocks bootstrap
#when N=315 k=50
> print(mean(SKd))
[1] 0.1799161
> print(sd(SKd))
[1] 0.1356913
#when N=4096, k=1000
> print(mean(SKd))
[1] 0.03968254
> print(sd(SKd))
[1] 0.03691822

```

There is close agreement between Monte-Carlo and moving blocks bootstrap in terms of the standard errors. 0.129 and 0.136, and 0.038 and 0.037 for length 315 and 4096 respectively. The only substantial difference is the mean (0.180) for the bootstrap with length 315, which is consequence of the observed SKd of 0.212 in the realization of the AR(1). Notice that the observed value of 0.183 is not statistically significant as the z -ratio is $0.212/0.136$ which is less than 2.

4 Application of moving blocks bootstrap to sunspots series (length 315) and EEG (D) series (length 4096)

4.1 Moving blocks bootstrap on sunspots series

```
#data = sunspots of length 315
#SKd of the observed sunspots
[1] 0.8555702
```

```
#when k=20, nrep=1000
> print(mean(SKD))
[1] 0.5937244
> print(sd(SKD))
[1] 0.3847948
```

```
#when k=50, nrep=1000
> print(mean(SKd))
[1] 0.7069069
> print(sd(SKd))
[1] 0.3316657
```

```
#when k=100, nrep=1000
> print(mean(SKD))
[1] 0.8124426
> print(sd(SKD))
[1] 0.2235839
```

4.2 Moving blocks bootstrap on EEG (D) series

```
#data = EEG (D) of length 4096
#SKd of the observed EEG (D)
[1] 5.404317
```

```
#when k=50, nrep=1000
> print(mean(SKD))
[1] 1.081129
> print(sd(SKD))
[1] 1.928021
```

```
#when k=100, nrep=1000
> print(mean(SKD))
[1] 1.778524
> print(sd(SKD))
[1] 2.296512
```

```
#when k=200, nrep=1000
> print(mean(SKD))
[1] 2.352384
> print(sd(SKD))
[1] 2.253458
```

```
#when k=500, nrep=1000
```



```

> print(mean(SKD))
[1] 3.32431
> print(sd(SKD))
[1] 1.862239

#when k=1000, nrep=1000
> print(mean(SKd))
[1] 3.637276
> print(sd(SKd))
[1] 1.617452

```

The length of the moving block (k) is critical. If the block length is too short, then the directionality is underestimated. However, long block lengths may tend to underestimate the standard error because all bootstrap samples will be rather close to the marginal time series. In practise it will be informative to determine significance using both the Monte-Carlo and moving blocks bootstrap with various block lengths.

```

#plot a histogram
par(mfrow=c(1,2))
>hist(SKd,main="SKd of sunspots, k=50, nrep=1000")
>hist(SKd,main="SKd of EEG (D), k=1000, nrep=1000")

```

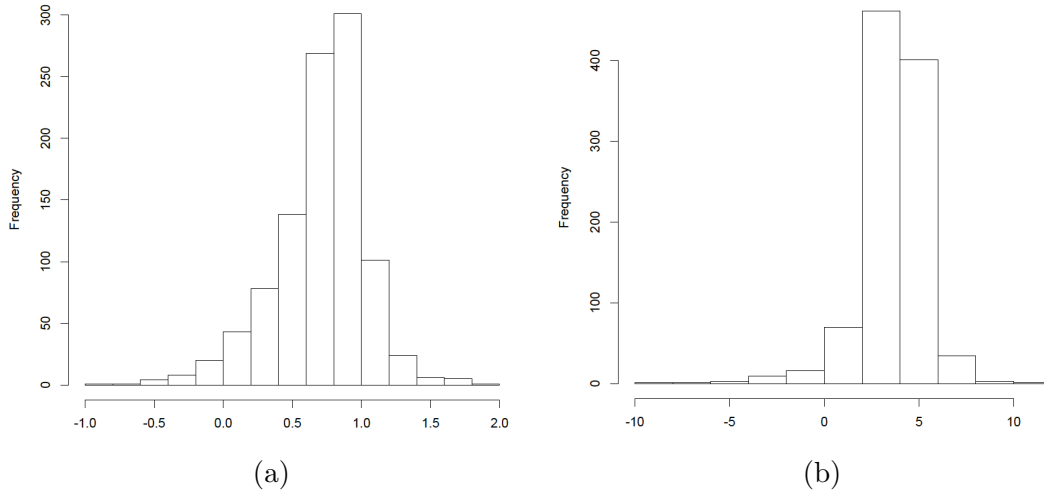


Figure 1: Histogram for the skewness of first differences of the moving blocks bootstrapped of sunspots when $k = 50, nrep = 1000$ (a), and of EEG (D) when $k = 1000, nrep = 1000$ (b).

5 Results from the suite of directionality tests applied to 12 time series, and boxplots of first differences

Table 1: Summary table of test statistics of directionality in the observed time series.

Method		Based on the original series					Based on the first differences			
Time series		DLQC	MCDB	POT(2)	POT(3)	POT(5)	P_d^+	$\hat{\gamma}_d$	LSK	$\hat{\gamma}_{dab}$
Sunsports $n=315, s=40.24$ AR(9), $\hat{\sigma}_{err}=15.54$		-0.06	0.14	4.71	5.36	4.83	42.5%	0.86	0.15	0.66
			{0.04}	{	$T=81.0$	}	{0.32%}			
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
		[5.57]	[5.43]	[4.43]	[4.97]	[4.38]	[4.25]	[6.64]	[6.15]	[5.84]
ds Rainfall $n=1440, s=69.97$ AR(29), $\hat{\sigma}_{err}=60.09$		-0.01	0.07	0.43	-0.47	0.20	48.4%	0.01	0.02	-0.20
			{0.03}	{	$T=39.9$	}	{0.00%}			
		(0.84)	(0.01)	(0.67)	(0.64)	(0.83)	(0.04)	(0.83)	(0.09)	(0.00)
		[0.19]	[2.91]	[0.40]	[0.46]	[0.19]	[2.00]	[0.20]	[1.71]	[3.67]
Exchange rate $n=6594, s=0.25$ AR(34), $\hat{\sigma}_{err}=0.01$		0.00	0.00	1.44	1.03	0.71	48.4%	0.13	0.03	-0.15
			{0.00}	{	$T=1.53$	}	{1.03%}			
		(0.91)	(0.97)	(0.12)	(0.26)	(0.43)	(0.00)	(0.00)	(0.00)	(0.00)
		[0.01]	[0.00]	[1.57]	[1.08]	[0.77]	[4.14]	[4.18]	[5.50]	[4.56]
SOI $n=1674, s=10.48$ AR(14), $\hat{\sigma}_{err}=7.83$		-0.02	0.01	0.61	0.38	-0.003	49.3%	0.08	0.01	0.15
			{0.02}	{	$T=9.10$	}	{0.42%}			
		(0.19)	(0.93)	(0.53)	(0.70)	(1.00)	(0.32)	(0.16)	(0.34)	(0.01)
		[1.30]	[1.36]	[0.64]	[0.39]	[0.00]	[1.02]	[1.38]	[0.95]	[2.53]
dtds Arrivals $n=451, s=0.055$ AR(14), $\hat{\sigma}_{err}=0.047$		-0.09	0.07	1.17	0.13	-0.07	47.8%	0.19	0.03	0.04
			{0.06}	{	$T=1.04$	}	{0.00%}			
		(0.21)	(0.29)	(0.27)	(0.91)	(0.96)	(0.11)	(0.07)	(0.20)	(0.71)
		[1.24]	[0.49]	[1.15]	[0.12]	[0.06]	[1.57]	[1.76]	[1.33]	[0.36]
ds NGRIP $n=2459, s=1.51$ AR(5), $\hat{\sigma}_{err}=0.62$		-0.02	0.01	2.46	3.38	3.75	47.7%	0.61	0.05	0.54
			{0.01}	{	$T=1.23$	}	{0.00%}			
		(0.00)	(0.13)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
		[11.4]	[1.08]	[2.60]	[3.44]	[3.86]	[3.71]	[12.1]	[5.15]	[9.98]
EEG (A) $n=4096, s=33.9$ AR(35), $\hat{\sigma}_{err}=12.7$		-0.003	0.01	0.199	0.294	1.34	50.1%	0.06	0.01	0.04
			{0.01}	{	$T=31.8$	}	{3.08%}			
		(0.05)	(0.13)	(0.83)	(0.76)	(0.18)	(0.85)	(0.12)	(0.44)	(0.29)
		[1.93]	[1.13]	[0.21]	[0.31]	[1.38]	[0.19]	[1.59]	[0.77]	[1.05]

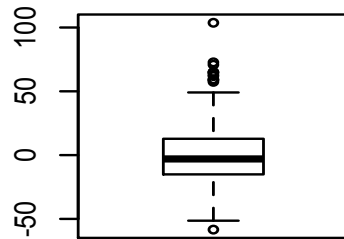
Table 2: Summary table of test statistics of directionality in the detrended and/or deseasonalized time series.

Time series	Method		Based on the original series				Based on the first differences		
	DLQC	MCDB	POT(2)	POT(3)	POT(5)	P_d^+	$\hat{\gamma}_d$	LSK	$\hat{\gamma}_{dab}$
EEG (B) $n=4096, s=46.99$ AR(36), $\hat{\sigma}_{err}=9.38$	0.001	0.001	1.96	1.73	2.14	49.4%	-0.03	0.003	0.07
	{0.005}	{0.005}	{	$T=95.0$	}	{2.47%}			
	(0.49)	(0.93)	(0.05)	(0.09)	(0.03)	(0.23)	(0.54)	(0.73)	(0.07)
	[0.66]	[1.25]	[1.94]	[1.70]	[2.17]	[1.18]	[0.61]	[0.34]	[1.85]
EEG (C) $n=4096, s=0.22$ AR(36), $\hat{\sigma}_{err}=0.03$	0.004	0.004	1.79	0.54	-1.33	52.8%	-1.21	-0.07	1.14
	{0.00}	{0.00}	{	$T=0.19$	}	{4.03%}			
	(0.00)	(0.00)	(0.07)	(0.56)	(0.16)	(0.00)	(0.00)	(0.00)	(0.00)
	[8.60]	[9.38]	[1.81]	[0.56]	[1.42]	[5.25]	[24.0]	[7.88]	[24.4]
EEG (D) $n=4096, s=0.48$	-0.01	0.004	4.02	4.86	5.87	41.9%	5.40	0.20	2.32
	{0.00}	{0.00}	{	$T=5.98$	}	{4.86%}			
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	[15.9]	[45.8]	[4.01]	[4.89]	[6.00]	[16.3]	[113.5]	[25.5]	[50.8]
AR(10), $\hat{\sigma}_{err}=0.06$	[15.5]	[45.7]	[4.16]	[5.13]	[6.05]	[15.4]	[108.1]	[24.3]	[49.7]
ARMA(10,1), $\hat{\sigma}_{err}=0.05$									
EEG (E) $n=4096, s=0.49$ AR(17), $\hat{\sigma}_{err}=0.08$	0.004	0.008	-0.78	-1.49	-1.30	56.6%	-0.75	-0.04	0.97
	{0.00}	{0.00}	{	$T=0.37$	}	{3.74%}			
	(0.00)	(0.00)	(0.44)	(0.122)	(0.18)	(0.00)	(0.00)	(0.00)	(0.00)
	[7.90]	[9.65]	[0.80]	[1.53]	[1.34]	[11.9]	[15.1]	[4.99]	[22.9]
Wells Fargo $n=4290, s=2.463$ AR(35), $\hat{\sigma}_{err}=2.392$	-0.399	0.025	2.038	2.531	2.060	49.2%	0.365	0.008	0.316
	{0.02}	{0.02}	{	$T=2.07$	}	{0.37%}			
	(0.00)	(0.23)	(0.05)	(0.02)	(0.04)	(0.04)	(0.00)	(0.16)	(0.00)
	[12.2]	[0.75]	[2.00]	[2.41]	[2.09]	[1.99]	[12.1]	[1.42]	[10.2]

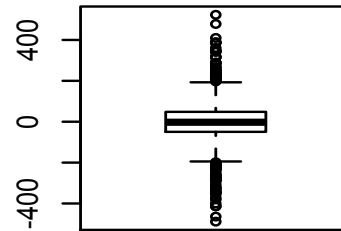
1. In the series description: n is the length of the time series; s is the marginal standard deviation; p in $AR(p)$ is the order of the AR model fitted using AIC; and $\hat{\sigma}_{err}$ is the estimated standard deviation of residuals in the $AR(p)$ model.
2. (number) refers to p -value, [number] refers to the ratio of the test statistic to its standard deviation and {number} in MCDB column is the mean of the test statistic from Monte-Carlo procedure.
3. {number} in P_d^+ refers to proportion of zero in the distribution of differences in original series. T in POT columns refers to threshold value at 80-th percentile in the original series. POT(2), POT(3) and POT(5) are the POT with averaging of 2, 3 and 5 observations before and after the peak.

5.1 Boxplot of the first differences of the observed time series

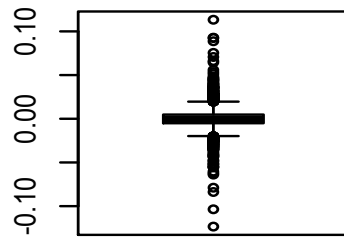
Generally, the boxplots highlight a large proportion of outlying values and the distribution of differences are highly kurtotic. There is clear positive skewness in the case of differences of sunspots, EEG (D) and apparent negative skewness in the differences of EEG (C).



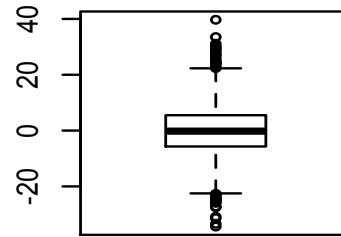
(a) diff(Sunspots)



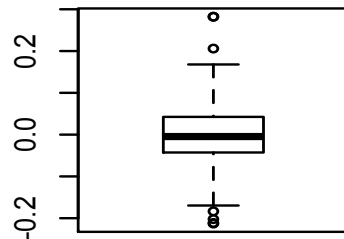
(b) diff(ds Rainfall)



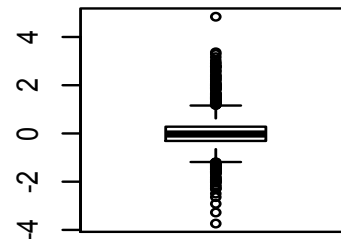
(c) diff(Exchange rate)



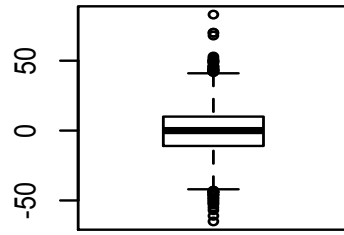
(d) diff(SOI)



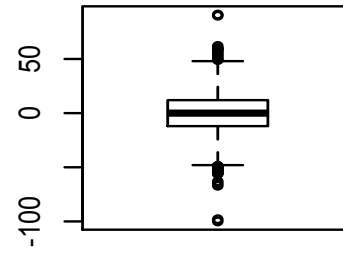
(e) diff(dsdt Arrivals)



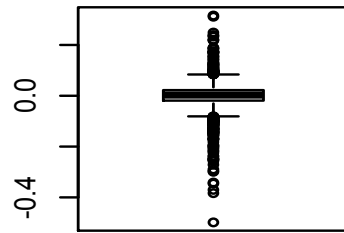
(f) diff(ds NGRIP)



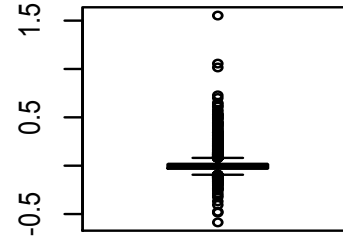
(g) diff(EEG (A))



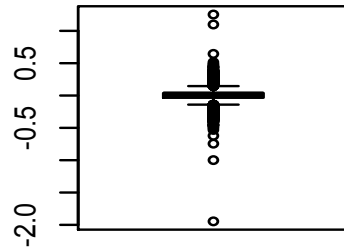
(h) diff(EEG (B))



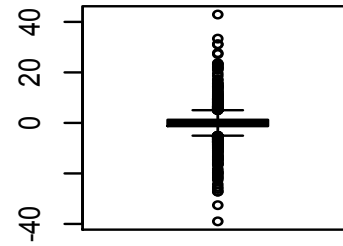
(i) diff(EEG (C))



(j) diff(EEG (D))



(k) diff(EEG (E))



(l) diff(Wells Fargo)

6 GARCH model for the residuals of AR models on dtDs tourist arrivals

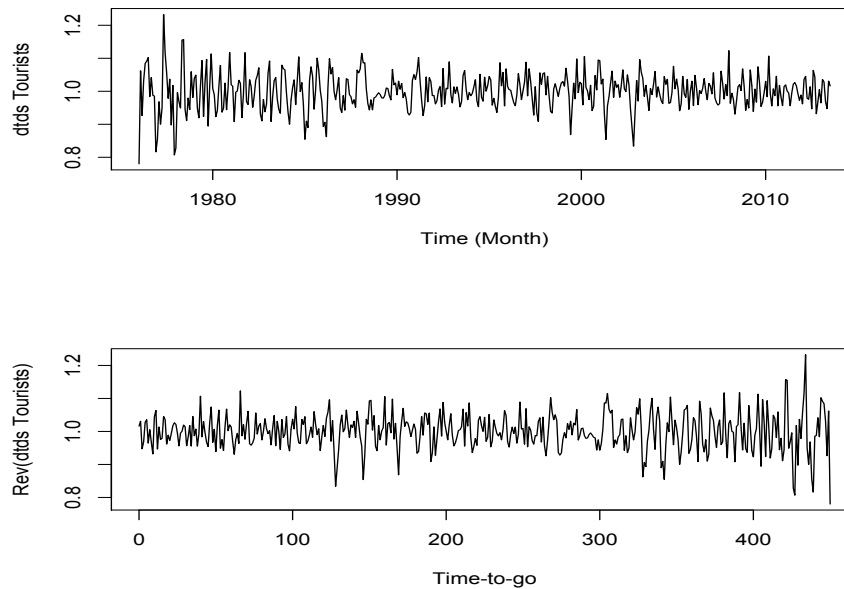


Figure 2: dtas tourist arrivals plotted in time order (above) and reverse time order (below).

6.1 Residuals from AR(2) model

```
> #fit AR(2) to the series
> data=ss
> M1=ar(ss, order.max=2)
> M1
Call:
ar(x = ss, order.max = 2)
Coefficients:
1      2
0.2455 -0.0961
Order selected 2  sigma^2 estimated as  0.002858
> M1$order
[1] 2
> sqrt(M1$var.pred)
[1] 0.05346461

> #residuals from AR model
> RES.M1=M1$res[-(1:M1$order)]
> #fit GARCH(1,1) to the residuals
> library(fGarch)
> garch11_RES.M1=garchFit(~ garch(1,1), data = RES.M1, trace = FALSE)

> #the estimated GARCH(1,1) parameters
> garch11_RES.M1@fit$matcoef
      Estimate Std. Error  t value Pr(>|t|)
mu      1.154282e-03 2.151219e-03  0.5365709 0.59156408
omega    7.123456e-05 5.169605e-05  1.3779496 0.16821884
```

```

alpha1 6.466026e-02 2.985989e-02 2.1654556 0.03035281
beta1 9.058234e-01 4.311175e-02 21.0110547 0.00000000

> sum(alpha1_M1,beta1_M1) #a stable GARCH(1,1) if the sum <= 1
[1] 0.9705

> #residuals from M1
> summary(RES.M1)
      Min.    1st Qu.    Median      Mean    3rd Qu.      Max.
-0.207000 -0.030070 -0.002800  0.000228  0.032750  0.235700

> #the residuals via GARCH(1,1) model
> GARCH_RES.M1=RES.M1/sqrt(garch11_RES.M1@h.t)
> summary(GARCH_RES.M1)
      Min.    1st Qu.    Median      Mean    3rd Qu.      Max.
-3.31700 -0.60780 -0.05171  0.01108  0.63530  3.41000

```

6.2 Residuals from AR(14 by AIC) model

```

> #fit AR(aic) to the series
> data=ss
> M2=ar(data)
> M2$order
[1] 14
> sqrt(M2$var.pred)
[1] 0.04712106

> #residuals from AR model
> RES.M2=M2$res[-(1:M2$order)]

> #fit GARCH(1,1) to the residuals
> library(fGarch)
> garch11_RES.M2=garchFit(~ garch(1,1), data = RES.M2, trace = FALSE)
> #the estimated GARCH(1,1) parameters
> garch11_RES.M2@fit$matcoef
      Estimate Std. Error t value Pr(>|t|)
mu      1.815196e-03 2.069610e-03  0.8770713 0.3804479
omega   3.183738e-05 4.002805e-05  0.7953769 0.4263943
alpha1  2.524070e-02 1.803442e-02  1.3995852 0.1616376
beta1   9.561029e-01 3.534603e-02 27.0497953 0.0000000

> sum(alpha1_M2,beta1_M2) #a stable GARCH(1,1) if the sum <= 1
[1] 0.9813

> #residuals from M2
> summary(RES.M2)
      Min.    1st Qu.    Median      Mean    3rd Qu.      Max.
-0.1470000 -0.0280300  0.0004399  0.0006097  0.0300000  0.1437000

> #the residuals via GARCH(1,1) model
> GARCH_RES.M2=RES.M2/sqrt(garch11_RES.M2@h.t)
> summary(GARCH_RES.M2)
      Min.    1st Qu.    Median      Mean    3rd Qu.      Max.

```

```
-3.02100 -0.66630 0.01049 0.02780 0.68550 3.55800
```

```
> #plot qqnorm for the residuals
> par(mfrow=c(2,2))
> qqnorm(RES.M1)
> qqnorm(RES.M2)
> qqnorm(GARCH_RES.M1)
> qqnorm(GARCH_RES.M2)
```

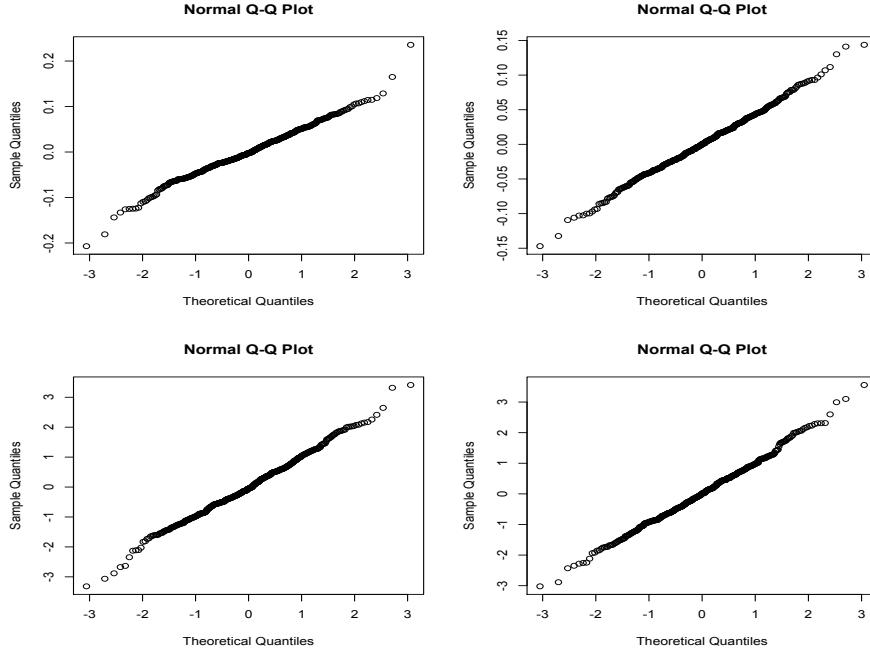


Figure 3: Normal quantile-quantile plot of residuals from AR(2) (above left), AR(14 by AIC) (above right) models on dtDs Arrivals, and the residuals modelled by GARCH(1,1) (below left and right respectively).

7 Poisson regression for the sunspots

We show an auto-correlation function (ACF), spectrum and Fourier line spectrum for the sunspot time series in Figure 4. We modelled the mean annual number of sunspots, X_t , as a sequence of over-dispersed Poisson random variable with a mean at t given by

$$E[X_t] = \beta_0 + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \sum_{j=1}^H \beta_{1j} \cos(2\pi\omega_j t) + \beta_{2j} \sin(2\pi\omega_j t) \quad (1)$$

where $t = 1, 2, \dots, n$ and H is the number of dominant frequencies. The frequencies are ω_j for $j = 1, 2, \dots, H$ and their amplitudes are $\sqrt{\beta_{1j}^2 + \beta_{2j}^2}$. Nine frequencies were identified

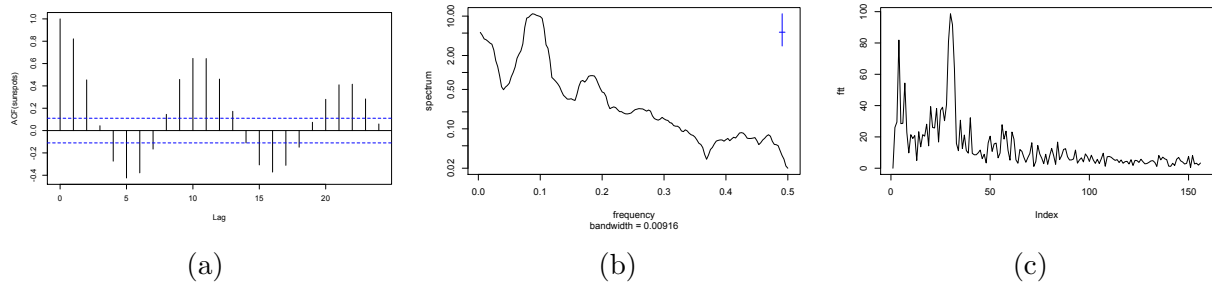


Figure 4: ACF (a), spectrum (b) and the FTT (c) of the sunspots.

as the highest spikes in the Fourier line spectrum, and in order of magnitude these are: 30, 31, 32, 4, 29, 27, 7, 3 and 33.

The reason for taking two lags in X_t is that Equation 1 then represents the difference equation of the differential equation form for a linear system with a single mode of vibration with sinusoidal forcing. A Poisson model was chosen to allow for the non-negative variable (modeling the logarithm of the number gave too many unrealistically high values after exponentiation in simulations). The over dispersion factor was 2.3.

8 Moving directionality and moving volatility of the Wells Fargo log-returns

The U.S. banking industry has experienced a series of severe financial and economic crises which have affected the performance of banks (e.g. Dell’Ariccia et al. 2008, Allen & Christa 2013). Therefore, we are interested in comparing directionality during crises with directionality during more stable periods. Allen & Christa (2013) identified two banking crises: credit crunch (1990:Q1 to 1992:Q4), and subprime lending crisis (2007:Q3 to 2009:Q4); three market crises: stock market crash (1987:Q4), Russian debt crisis, and Long-Term Capital Management bailout (1998:Q3 to 1998:Q4); together with dot.com bubble and September 11 attack on the World Trade Center (2000:Q2 to 2002:Q3), in the U.S. as crises periods. Here, we combine the dot.com bubble and the September 11 attack with the 2001-2002 recession (Wikipedia 2016b) as major events that happened in the U.S between 2000 and 2002. We also combine the subprime mortgage crisis with the 2007-2008 GFC, with effects continuing into 2009, the collapse of Lehman Brothers in 2008 and the Great Recession of 2008-2012 (Wikipedia 2016a) as a second period of major shocks in the U.S. during 2007 to 2012. Our definition of stable and unstable periods is summarized in Table 3.

We compare mean, range, skewness, kurtosis, volatility and directionality in the stable and unstable sub-series (Table 4). The volatility is here defined as the standard deviation of the log-returns, and the directionality is measured by skewness of the first differences $\hat{\gamma}_d$. It seems reasonable to expect higher mean values during the stable periods, but the first stable period is quite anomalous. The skewness is somewhat higher during the unstable periods. Directionality is positive and markedly higher during all of the unstable periods than in

Table 3: Stable and unstable periods between May-1999 and Feb-2017.

Period	Status
May to Dec-1999	Stable
Jan-2000 to Dec-2002	Unstable
Jan-2003 to Dec-2006	Stable
Jan-2007 to Dec-2012	Unstable
Jan-2013 to Feb-2017	Stable

any of the stable periods. The range is higher in the unstable periods but the volatility is higher in the first stable period than in the following unstable period. These preliminary results lead us to question whether monitoring directionality, as well as volatility, would give early warning of unstable periods.

Table 4: Summary statistics of the Wells Fargo log-returns by stable and unstable sub-series.

Period	Mean	Range	Skewness	Kurtosis	Volatility	Directionality
Stable	-0.041	14.0	0.15	3.4	2.32	0.11
Unstable	0.028	21.5	0.22	6.1	1.98	0.46
Stable	0.054	7.0	0.13	4.2	0.88	0.04
Unstable	0.007	55.6	0.73	16.0	3.74	0.26
Stable	0.063	12.5	0.15	5.8	1.18	0.05

To investigate this we use a 9-month moving directionality series (MD), 21 days by 9 months is 189 trading days, which is defined in terms of skewness of the first differences $\hat{\gamma}_d$ in the log-returns of Wells Fargo. The formula used to calculate the MD is given in Equation (2),

$$\hat{\gamma}_{d,t} = \frac{\sum_{i=0}^{188} (\Delta x_{t-i} - \overline{\Delta x_t})^3 / m}{\left[\sum_{i=0}^{188} (\Delta x_{t-i} - \overline{\Delta x_t})^2 / m \right]^{3/2}}, \quad (2)$$

where $\Delta x_t = x_t - x_{t-1}$, $t = 189, \dots, 4479$, $m = 189$, $\overline{\Delta x_t} = \sum_{i=0}^{188} \Delta x_{t-i} / m$ and x_t is the full-length time series of log-returns. Similarly for the volatility, given by the marginal standard deviation of log-returns $\{x_t\}$, the 9-month moving volatility series (MV) is defined by Equation (3).

$$\hat{\sigma}_i = \sqrt{\frac{\sum_{i=0}^{188} (x_{t-i} - \bar{x}_t)^2}{m}}, \quad (3)$$

where $\bar{x}_t = \sum_{i=0}^{188} x_{t-i}/m$.

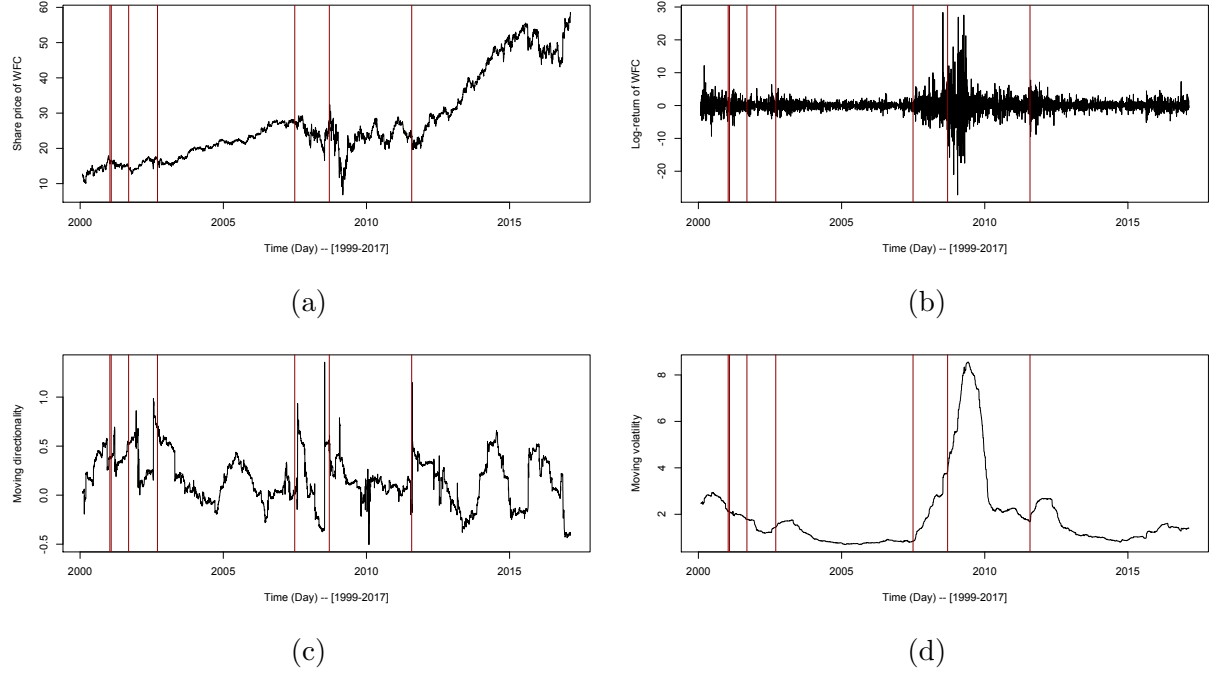


Figure 5: Daily share price (a), daily log-return (b), 9-month moving directionality (c), and 9-month moving volatility (d) of Wells Fargo.

The plots of the MD, $\{\hat{\gamma}_{d,t}\}$, and the MV, $\{\hat{\sigma}_i\}$, for the Wells Fargo together with its share prices and log-returns are illustrated in Figure 5. We use vertical lines in the plots to show the dates when the crises are reported in news. For example, 15-Jan-2001 for the 2001 dotcom crash, 11-Sep-2001 for the 2001 Sept 11 attack on the World Trade Centre, the subprime mortgage crisis was first widely reported on 1-Jul-2007 and the collapse of Lehman Brothers was reported on 15-Sep-2008. No exact dates are found to indicate the recessions, but the U.S. unemployment rate rose to 4.2% on 1-Feb-2001, the National Association of Securities Dealers Automated Quotations (NASDAQ) reached a 6-year low on 24-Sept-2002 and U.S. stock market fell dramatically on 1-Aug-2011 for the 2011 recession. These vertical lines are particularly helpful in this case as the share prices move in short to long-term downward trends after the shocks in Figure 5(a). In particular, major declines in share prices after the 1-Jul-2007 and the 15-Sep-2008 during the crisis period. Similarly, more variability in the log-returns during the 2007-2012 unstable period in Figure 5(b).

The MD appears to be less smooth than the MV, and the MD soars on two occasions during crises periods in Figure 5(c). In particular, the fourth vertical line of 24-Sept-2002 and the second last vertical line of 14-Sept-2008. In contrast, MV appears to be less influenced by crises. We have seen similar effects in the moving directionality at 3-month, 6-month and 12-month.

We illustrate the relationship between directionality and volatility for the Wells Fargo

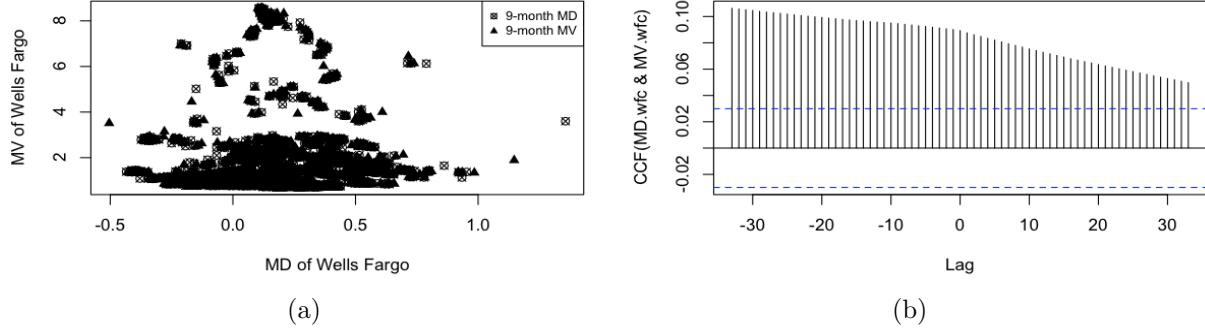


Figure 6: Scatter plot (left) and CCF (right) of the 9-month moving directionality and the 9-month moving volatility of Wells Fargo.

daily log-return series using scatter plot of the $\{\hat{\gamma}_{d,t}\}$ versus the $\{\hat{\sigma}_i\}$ in Figure 6(a). There is no obvious linear pattern from the plots, and the correlation coefficient at a lag of 0 is 0.09. This low correlation coefficient indicates that directionality could provide additional information to volatility. This might be expected as volatility is a second order characteristic and directionality is third order. Moreover, in the case of Wells Fargo, the cross-correlation function (CCF) in Figure 6(b) shows directionality leads the occurrence of volatility over this period.

These features of directionality suggest that it may have potential as an early warning signal of financial crises and consequent falling share prices. We have carried out preliminary work on an investment strategy of buying when the MD is high and selling when it is low for a portfolio of U.S. bank shares. Over the period of May 3, 1999 to February 17, 2017 the return was 2.15% per annum, which was higher than achieved with MV. This work is continuing.

References

- Allen, N. B. & Christa, H. S. B. (2013), ‘How does capital affect bank performance during financial crises?’, *Journal of Financial Economics* **109**(1), 146–176.
- Dell’Ariccia, G., Detragiache, E. & Rajan, R. (2008), ‘The real effect of banking crises’, *Journal of Financial Intermediation* **17**(1), 89 – 112. Financial Contracting and Financial System Architecture.
- Weiss, G. (1975), ‘Time-reversibility of linear stochastic processes’, *Journal of Applied Probability* **12**(4), 831–836.
- Wikipedia (2016a), ‘Financial crisis of 2007–2008’, <https://en.wikipedia.org> (last accessed 11 November 2016).
- Wikipedia (2016b), ‘List of economic crises’, <https://en.wikipedia.org> (last accessed 5 October 2016).