Long Memory and Regime Switching: A Simulation Study on the Markov Regime-Switching ARFIMA Model

Yanlin Shi and Kin-Yip Ho

The Australian National University

June 23, 2014

Overview

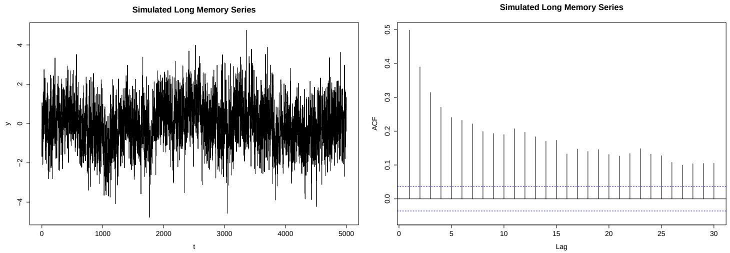
Long Memory

Definition

$$var(S_T) = O(T^{2d+1}) \text{ ,where } S_T = \sum_{t=1}^{T} y_t$$

$\{y_t\}$ is a sequence of time series and T is the number of observations. Then d is the long memory parameter, and $d > 0$ indicates the existence of long memory.

Plots of Simulated Long Memory Series

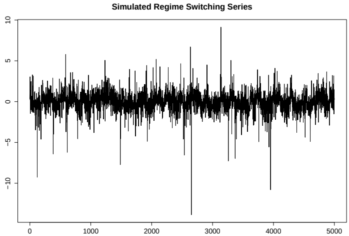


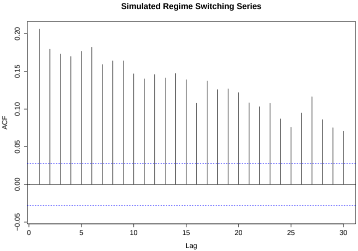
- The long memory series is simulated with $d = 0.35$ and $T = 5000$.
- If long memory exists, sample autocorrelations are significantly different from zero even for large lags.

What is Markov Regime-Swtiching (MRS)?

- In short, y_t has two states $\{1,2\}$, with means μ_1 and μ_2, respectively.
- At time t, $\mathrm{P}(s_{t+1} = k|s_t = j)$, the transition probability of moving from state j to state k ($j, k \in \{1, 2\}$), is a constant defined as p_{jk}.
- The expected time of staying at state j is $1/(1 - p_{jj})$.

Plots of Simulated MRS Series





- The MRS series is simulated with $\mu_1 = -0.5$, $\mu_2 = 0.5$ and $p_{11} = p_{22} = 0.99$.
- ACFs of MRS are also quite significant even at large lags.

Motivation

- An influential study by Diebold and Inoue (2001) provides a theoretical explanation of this phenomenon. They also argue that long memory and regime switching are interchangeable concepts and should not be studied separately.
- Since regime switching is more closely related to the concept ‘business cycle’ (Hamilton, 1989), to distinguish it from long memory is of great financial and economic importance.
- A recent study by Perron and Qu (2010) proposes a test to effectively distinguish the long- and short-memory processes with mean shifts at the first moment of financial series.
- If the effects of regime switching can be appropriately controlled, pure long-memory process should be distinguished from pure regime switching process.

Contribution

- Analyse the quasi maximum likelihood estimation (QMLE) properties of the ARFIMA and MRS models.
- Provide an improved theoretical proof to find the 'real' cause of the confusion.
- Propose a MRS-ARFIMA model which can effectively control for this cause.
- Conduct a series of simulation studies to demonstrate that the proposed MRS-ARFIMA model can distinguish between the pure long-memory and pure regime-switching processes.
- Use empirical results of the London Stock Exchange (FTSE) index to show that our MRS-ARFIMA model can also outperform both the ARFIMA and MRS models.

R/S Statistic

- The R/S statistic (Lo, 1991) is widely used to test the existence of long memory.

R/S statistic

$$Q_T = \frac{1}{\hat{\sigma}_T(q)} \left[\max_{1 \le k \le T} \sum_{t=1}^{k} (y_t - \bar{y}) - \min_{1 \le k \le T} \sum_{t=1}^{k} (y_t - \bar{y}) \right]$$

where $\bar{y} = (1/T) \sum_{t=1}^{T} y_t$, and $\hat{\sigma}_T(q)$ is the square root of the Newey-West estimate of the long run variance with bandwidth q.

V/S Statistic

- The V/S statistic (Giraitis et al., 2003) is modified version of the R/S statistic, which has a better balance of size and power.

V/S statistic

$$M_T = \frac{1}{\hat{s}^2_{T,q} T^2} \left[\sum_{k=1}^{T} \left(\sum_{t=1}^{k} (y_t - \bar{y}) \right)^2 - \frac{1}{T} \left(\sum_{k=1}^{T} \sum_{t=1}^{k} (y_t - \bar{y}) \right)^2 \right]$$

where $\hat{s}^2_{T,q} = \frac{1}{T} \sum_{t=1}^{T} (y_t - \bar{y})^2 + 2 \sum_{t=1}^{q} \omega_t(q) \hat{\gamma}_t$, $\omega_t(q) = 1 - [t/(q+1)]$ and $\hat{\gamma}_t$ is the tth sample covariance (covariance between y_i and y_{i-t}).

ARFIMA Model

- Autoregressive Fractionally Integrated Moving Average (ARFIMA) model is widely used in estimating the long memory parameter of financial series at the first moment.

ARFIMA(p, d, q) Model

$$\varphi(L)(1-L)^d(y_t-\mu)=\theta(L)\varepsilon_t$$

where $\varphi(L)=1-\sum_{i=1}^{p}\varphi_i L^i$ and $\theta(L)=1-\sum_{j=1}^{q}\theta_j L^j$

L denotes the lag operator, d is the long memory parameter, and μ is the expectation of y_t.

ARFIMA Model

- Hosking (1981) defines the term $(1-L)^d$ as follows.

Fractional Differencing Operator

$$(1-L)^d = \sum_{k=0}^{\infty} \delta_k(d) L^k, \ \delta_k(d) = \frac{k-1-d}{k}\delta_{k-1}(d), \ \delta_0(d) = 1$$

Generally, k is set to 1000 to proxy the infinity process.

Properties of ARFIMA Model

- The roots of $\varphi(L)$ and $\theta(L)$ need to be outside the unit circle.
- ε_t is assumed to be an independently and identically (iid) Normally distributed innovation with zero mean and variance σ_ε^2.
- $|d| > 1/2$, y_t is non-stationary; when $0 < d < 1/2$, $\{y_t\}$ is stationary and has long memory; and when $-1/2 < d < 0$, $\{y_t\}$ is stationary and has short memory.
- When $-1/2 < d < 1/2$, Sowell (1992) describes how to compute the exact maximum likelihood estimate (MLE) of d.

QMLE of ARFIMA Model

- Significant evidence suggests that financial series is rarely Normal but typically leptokurtic and exhibits heavy-tail behaviour (Bollerslev, 1987; Susmel and Engle, 1994).
- Student-t distribution is a widely used alternative, which can accommodate the excess kurtosis of the innovations (Bollerslev, 1987).
- QMLE based on Normal distribution is asymptotically consistent, but may not be efficient.

Simulation Study: QMLE of ARFIMA Model

- 500 replicates of ARFIMA(0,d,0) model with $\mu = 0$, $\sigma_\varepsilon^2 = 1$ and $v = 3$.

Table 1 : ARFIMA(0,d,0) Model with Normal Distribution

d	T	$Bias_d$	$RMSE_d$	SE_d	$Bias_{\sigma_\varepsilon^2}$	$RMSE_{\sigma_\varepsilon^2}$	$SE_{\sigma_\varepsilon^2}$
0.15	3000	-0.0016	0.0149	0.0148	-0.0202	0.1819	0.1810
	4000	-0.0009	0.0125	0.0124	-0.0079	0.1887	0.1887
	5000	-0.0005	0.0107	0.0107	0.0147	0.3136	0.3135
0.25	3000	-0.0015	0.0144	0.0144	-0.0017	0.2310	0.2313
	4000	-0.0005	0.0116	0.0116	-0.0071	0.1829	0.1829
	5000	-0.0008	0.0113	0.0113	-0.0165	0.2001	0.1996
0.35	3000	-0.0003	0.0141	0.0141	0.0171	0.4355	0.4356
	4000	-0.0006	0.0126	0.0126	-0.0142	0.1617	0.1612
	5000	-0.0010	0.0106	0.0106	-0.0123	0.1493	0.1489
0.45	3000	0.0011	0.0149	0.0149	-0.0017	0.4433	0.4437
	4000	0.0008	0.0126	0.0126	-0.0028	0.2146	0.2148
	5000	0.0007	0.0109	0.0109	-0.0204	0.1791	0.1781

Simulation Study: QMLE of ARFIMA Model

Table 2 : ARFIMA(0,d,0) Model with Student-t Distribution

d	T	$Bias_d$	$RMSE_d$	SE_d	$Bias_{\sigma_\varepsilon^2}$	$RMSE_{\sigma_\varepsilon^2}$	$SE_{\sigma_\varepsilon^2}$
0.15	3000	-0.0008	0.0102	0.0102	0.0047	0.1037	0.1037
	4000	0.0004	0.0091	0.0091	0.0112	0.0851	0.0844
	5000	-0.0006	0.0076	0.0076	0.0078	0.0788	0.0785
0.25	3000	-0.0006	0.0103	0.0103	0.0070	0.1072	0.1071
	4000	-0.0005	0.0087	0.0087	0.0053	0.0891	0.0891
	5000	-0.0005	0.0081	0.0080	0.0036	0.0782	0.0782
0.35	3000	0.0005	0.0100	0.0100	0.0134	0.1086	0.1079
	4000	0.0000	0.0091	0.0091	0.0050	0.0903	0.0902
	5000	-0.0004	0.0084	0.0084	0.0022	0.0803	0.0803
0.45	3000	0.0013	0.0107	0.0107	0.0102	0.0972	0.0967
	4000	0.0015	0.0088	0.0086	0.0065	0.0865	0.0864
	5000	0.0011	0.0083	0.0082	0.0111	0.0804	0.0797

The Standard MRS Model

- Hamilton (1989) investigates the impacts of business cycle on financial series and argue that the behaviour of financial series tends to be different depending on the states (regimes) of business cycles.
- MRS model is proposed by Hamilton (1988, 1989, 1994), which allows parameters of financial series to switch between states.
- Define $\{s_t\}$ be a stationary, irreducible Markov process with discrete state space $\{1, 2\}$ and transition matrix $P = [p_{jk}]$. A standard MRS model (regime-switching in mean) has the following specification.

Identify the States

- States of the MRS model can be identified from the smoothing probability defined below.

Smoothing Probability

$$P(s_t = 1|\theta, \Omega_T) = \omega_{1,t}[\frac{p_{11} P(s_{t+1} = 1|\theta, \Omega_T)}{P(s_{t+1} = 1|\theta, \Omega_t)} + \frac{p_{12} P(s_{t+1} = 2|\theta, \Omega_T)}{P(s_{t+1} = 2|\theta, \Omega_t)}]$$

where $\omega_{j,t-1} = P(s_{t-1} = j|\theta, \Omega_{t-1})$ is the update probability and can be obtained by an integrative algorithm given in Hamilton (1989). Since $P(s_T = 1|\theta, \Omega_T) = \omega_{1,T}$, the smoothing probability series $P(s_t = 1|\theta, \Omega_T)$ can be generated by iterating this equation backward from T to 1.

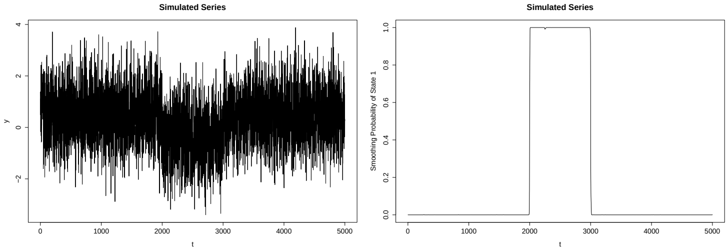
- As suggested by Hamilton (1989), when $P(s_T = 1|\theta, \Omega_T) \geq 0.5$, y_t lies in the state 1 and otherwise lies in state 2.

Which distribution to use?

- As noted by Klaassen (2002), Ardia (2009) and Haas (2009), if regimes (states) are not Normal but leptokurtic, the use of within-regime normality can seriously affect the identification of the regime process.
- The reason may be that Normal distribution cannot 'tolerate' large outliers.

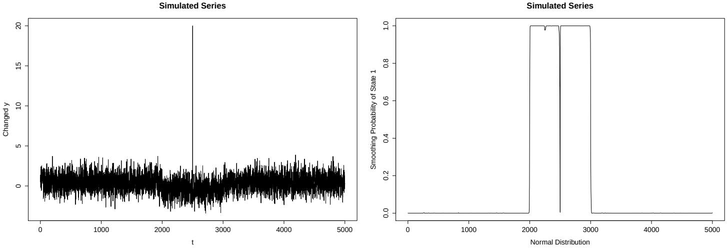
Which distribution to use?

- $\{y\}$ are simulated with $T = 5000$, $\mu = -0.5$ for t=2001, 2002,...,3000 and $\mu = 0.5$ otherwise.



Which distribution to use?

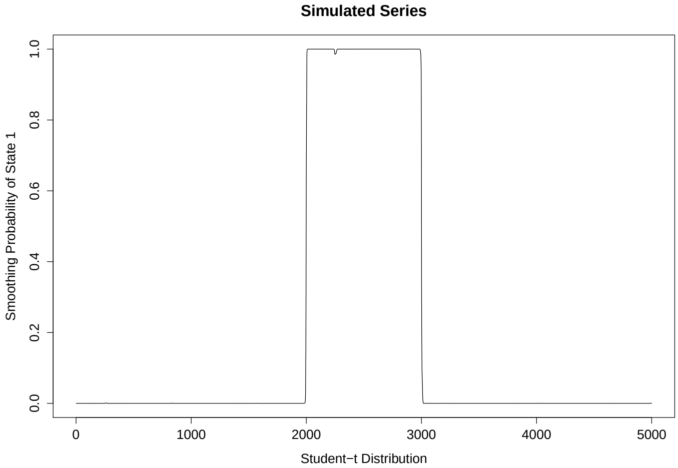
- Let $y_{2500} = 20$, MRS with Normal distribution leads to the following result.



- Haas and Paolella (2012) argue that QMLE based on Normal components does not provide a consistent estimator of the MRS model, if true distribution of innovations is not Normal.

Which distribution to use?

- Using Student-t distribution instead, the new result is more robust.



Estimation: MRS with Student-t Distribution

Conditional Density in State s_t

$$\Omega_{t-1} = \left\{\varepsilon_{s_{t-1},t-1}, \varepsilon_{s_{t-2},t-2}, \ldots, \varepsilon_{s_1,1}\right\}$$
$$\theta = (\mu_1, \mu_2, p_{11}, p_{22}, v, \sigma_\varepsilon^2)'$$
$$f(\varepsilon_{s_t,t} | s_t = j, \theta, \Omega_{t-1}) = \frac{\Gamma[(v+1)/2]}{\Gamma(v/2)\sqrt{\pi(v-2)\sigma_\varepsilon^2}} \left[1 + \frac{\varepsilon_{j,t}^2}{(v-2)\sigma_\varepsilon^2}\right]^{\frac{v+1}{2}}$$

where Ω_{t-1} is the information set at time $t-1$. θ is the vector of parameters. $\Gamma(\cdot)$ is the Gamma function and $f(\varepsilon_{s_t,t}|s_t, \theta, \Omega_{t-1})$ is the conditional density of $\varepsilon_{s_t,t}$. This stems from the fact that at time t, $\varepsilon_{s_t,t}$ follows a Student-t distribution with mean 0, variance σ_ε^2 and degrees of freedom v given time $t-1$.

Estimation: MRS with Student-t Distribution

Overall Conditional Density

$$f(\varepsilon_{s_t,t}|\theta,\Omega_{t-1}) = \sum_{j=1}^{2}\sum_{k=1}^{2} p_{jk}\omega_{j,t-1} f(\varepsilon_{s_t,t}|s_t = j,\theta,\Omega_{t-1})$$

Recall that $\omega_{j,t-1} = P(s_{t-1} = j|\theta,\Omega_{t-1})$.

Log Likelihood Function

$$L(\theta|\varepsilon) = \sum_{t=2}^{T} \ln f(\varepsilon_{s_t,t}|\theta,\Omega_{t-1}) \text{ where } \varepsilon = (\varepsilon_{s_t,1},\varepsilon_{s_t,2},\ldots,\varepsilon_{s_t,T})'$$

- the MLE estimator $\hat{\theta}$ is obtained by maximizing $L(\theta|\varepsilon)$.

Simulation Study: QMLE of MRS Model

- 500 replicates of MRS model with $\mu_1 = -0.5$, $\mu_2 = 0.5$, $\sigma_\varepsilon^2 = 1$ and $v = 3$.

Table 3 : MRS Model with Normal Distribution

p_{11}	p_{22}	T	$Bias_{p_{11}}$	$RMSE_{p_{11}}$	$SE_{p_{11}}$	$Bias_{p_{22}}$	$RMSE_{p_{22}}$	$SE_{p_{22}}$
0.999	0.99	3000	-0.0178	0.1106	0.1093	-0.3866	0.6096	0.4718
		4000	-0.0254	0.1443	0.1422	-0.4047	0.6250	0.4767
		5000	-0.0136	0.1007	0.0999	-0.3934	0.6169	0.4757
0.99	0.999	3000	-0.4073	0.6286	0.4793	-0.0134	0.1001	0.0992
		4000	-0.3586	0.5889	0.4676	-0.0250	0.1484	0.1464
		5000	-0.3414	0.5754	0.4637	-0.0284	0.1611	0.1588
0.99	0.99	3000	-0.0308	0.1597	0.1569	-0.0159	0.1095	0.1085
		4000	-0.0203	0.1335	0.1321	-0.0189	0.1256	0.1243
		5000	-0.0146	0.1114	0.1106	-0.0192	0.1286	0.1272
0.999	0.999	3000	-0.0529	0.2227	0.2165	-0.0471	0.2089	0.2037
		4000	-0.0472	0.2112	0.2061	-0.0342	0.1795	0.1764
		5000	-0.0336	0.1789	0.1759	-0.0382	0.1896	0.1859

Simulation Study: QMLE of MRS Model

Table 4 : MRS Model with Student-t Distribution

p_{11}	p_{22}	T	$Bias_{p_{11}}$	$RMSE_{p_{11}}$	$SE_{p_{11}}$	$Bias_{p_{22}}$	$RMSE_{p_{22}}$	$SE_{p_{22}}$
0.999	0.99	3000	-0.0001	0.0007	0.0007	-0.0169	0.0795	0.0777
		4000	-0.0001	0.0006	0.0006	-0.0079	0.0472	0.0466
		5000	-0.0001	0.0005	0.0005	-0.0082	0.0622	0.0617
0.99	0.999	3000	-0.0148	0.0757	0.0744	-0.0002	0.0007	0.0007
		4000	-0.0073	0.0358	0.0351	-0.0001	0.0007	0.0007
		5000	-0.0042	0.0197	0.0193	-0.0001	0.0006	0.0006
0.99	0.99	3000	-0.0005	0.0031	0.0030	-0.0005	0.0031	0.0031
		4000	-0.0004	0.0027	0.0027	-0.0004	0.0026	0.0026
		5000	-0.0002	0.0020	0.0020	-0.0003	0.0021	0.0021
0.999	0.999	3000	-0.0007	0.0023	0.0022	-0.0008	0.0030	0.0028
		4000	-0.0008	0.0084	0.0084	-0.0004	0.0013	0.0012
		5000	-0.0002	0.0009	0.0008	-0.0014	0.0235	0.0235

Section Conclusion

- QMLE of MRS model is neither consistent nor efficient.
- When the true distribution is not Normal, we should always use fat-tailed distributions (e.g Student-t and GED).
- Therefore, we use Student-t distribution assumption for both ARFIMA and MRS models in the following sections.

Test for Long Memory: ARFIMA Simulations

- Overall, both R/S and V/S statistics indicate the significant existence of long memory at 5% level (critical values are 1.7470 and 0.1869, respectively).

Table 5 : Simulated ARFIMA(0,d,0) data

d	T	$Mean_{rs}$	SE_{rs}	$Mean_{vs}$	SE_{vs}
0.15	3000	1.8719	0.4988	0.2502	0.1773
	4000	1.9561	0.5101	0.2713	0.1961
	5000	2.0236	0.5057	0.2882	0.1958
0.25	3000	2.5235	0.6619	0.5093	0.3419
	4000	2.6501	0.7095	0.5643	0.4193
	5000	2.8214	0.7459	0.6284	0.4436
0.35	3000	3.2136	0.8510	0.8932	0.6123
	4000	3.5640	0.9057	1.0912	0.6907
	5000	3.7943	0.9738	1.2316	0.8006
0.45	3000	4.0282	0.9836	1.5033	0.8941
	4000	4.4719	1.1772	1.8706	1.1947
	5000	4.7476	1.1994	2.0684	1.3029

Test for Long Memory: MRS Simulations

- Overall, both R/S and V/S statistics also indicate the significant existence of long memory at 5% level (critical values are 1.7470 and 0.1869, respectively).

Table 6 : Simulated MRS data

p_{11}	p_{22}	T	$Mean_{rs}$	SE_{rs}	$Mean_{vs}$	SE_{vs}
0.999	0.99	3000	2.5694	0.9765	0.5432	0.4632
		4000	2.6592	0.9707	0.5691	0.4912
		5000	2.7345	0.9569	0.5812	0.4721
0.99	0.999	3000	2.5363	0.9832	0.5412	0.4852
		4000	2.5971	0.9551	0.5417	0.4539
		5000	2.7858	0.9537	0.6020	0.4800
0.99	0.99	3000	2.8894	0.6797	0.5587	0.3432
		4000	2.9862	0.7065	0.5934	0.3761
		5000	3.0085	0.6943	0.5811	0.3496
0.999	0.999	3000	5.0602	1.4217	2.1857	1.1843
		4000	5.7258	1.4849	2.7265	1.4501
		5000	6.2344	1.5608	3.1568	1.6942

Test for Long Memory: MRS Simulations

- The widely employed long memory tests cannot distinguish between the long memory and regime switching.
- What about the ARFIMA model?

MRS Data Fitted into ARFIMA Model

- MRS simulations are fitted into the ARFIMA(0,d,0) model with Student-t distribution. On average, the estimated d is significantly greater than 0 (Wilcox sign test) in all cases.

Table 7 : Estimates obtained from ARFIMA Model

p_{11}	p_{22}	T	$Mean_d$	SE_d	$Mean_{\sigma_\varepsilon^2}$	$SE_{\sigma_\varepsilon^2}$	$Mean_v$	SE_v
0.999	0.99	3000	0.0939	0.0481	1.0130	0.0927	3.2499	0.2560
		4000	0.0941	0.0424	1.0172	0.0765	3.2284	0.2128
		5000	0.0934	0.0403	1.0086	0.0662	3.2374	0.1988
0.99	0.999	3000	0.0929	0.0464	1.0178	0.0901	3.2329	0.2409
		4000	0.0905	0.0421	1.0072	0.0750	3.2357	0.2032
		5000	0.0967	0.0377	1.0123	0.0719	3.2363	0.2053
0.99	0.99	3000	0.1928	0.0106	1.0385	0.0785	3.5921	0.2489
		4000	0.1933	0.0090	1.0350	0.0695	3.6022	0.2385
		5000	0.1938	0.0077	1.0370	0.0595	3.5892	0.1908
0.999	0.999	3000	0.1550	0.0345	1.0163	0.0821	3.3626	0.2460
		4000	0.1579	0.0269	1.0174	0.0722	3.3465	0.2068
		5000	0.1612	0.0234	1.0116	0.0656	3.3717	0.1828

ARFIMA Data Fitted into MRS Model

- ARFIMA simulations are fitted into the MRS model with Student-t distribution. On average, the estimated p_{11} and p_{22} are both significantly greater than 0 (Wilcox sign test) in all cases.

Table 8 : Estimates obtained from MRS Model

d	T	$Mean_{p_{11}}$	$SE_{p_{11}}$	$Mean_{p_{22}}$	$SE_{p_{22}}$	$Mean_{\sigma^2_\varepsilon}$	$SE_{\sigma^2_\varepsilon}$	$Mean_v$	SE_v
0.15	3000	0.9138	0.0299	0.9128	0.0301	0.9793	0.1204	2.8860	0.1795
	4000	0.9147	0.0252	0.9146	0.0244	0.9871	0.1014	2.8652	0.1526
	5000	0.9157	0.0216	0.9143	0.0247	0.9836	0.0927	2.8719	0.1429
0.25	3000	0.9323	0.0176	0.9330	0.0163	0.9123	0.1094	3.1200	0.2218
	4000	0.9323	0.0151	0.9341	0.0144	0.9107	0.0852	3.1019	0.1785
	5000	0.9331	0.0126	0.9335	0.0133	0.9075	0.0781	3.1091	0.1660
0.35	3000	0.9431	0.0761	0.9442	0.0595	0.9307	0.1472	3.6446	0.3546
	4000	0.9214	0.1535	0.9328	0.1180	0.9396	0.1539	3.6898	0.3292
	5000	0.9302	0.1346	0.9350	0.1152	0.9294	0.1295	3.6762	0.2692
0.45	3000	0.8854	0.2460	0.8939	0.2453	1.2530	0.4405	4.7082	0.8185
	4000	0.9228	0.1727	0.9050	0.2328	1.2258	0.3706	4.7956	0.7629
	5000	0.9148	0.2015	0.9092	0.2230	1.2143	0.3430	4.7989	0.7076

Section Conclusion

- The large estimates of long memory parameter and transition probabilities suggest that data from long memory and regime-switching processes can be significantly confused with one another.
- This result is consistent with Diebold and Inoue (2001).
- They suggest that the cause of this confusion is the time-varying transition probabilities. However, both p_{11} and p_{22} are supposed to be constant in the original MRS model.
- Motivated by their proof, we give a refined version.

What contributes to this confusion?

Proposition

For the following Markov Regime-Switching in mean model:

$$y_t = \mu_{s_t} + \varepsilon_t \text{ and } \varepsilon_t \overset{iid}{\sim} N(0, \sigma^2)$$

with transition probability matrix M defined as

$$M = \left\{ \begin{array}{cc} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{array} \right\}$$

Assume that $\mu_1 \neq \mu_2$, p_{11} and p_{22} are constant (non-time-varying) and long memory significantly exists for y_t. Then the significant long memory is caused by the time-varying smoothing probability series $P(s_t|\Omega_T)$.

What contributes to this confusion?

Proof

Let $\zeta_t = \begin{pmatrix} I(s_t = 1) & I(s_t = 2) \end{pmatrix}'$, following the proof in Diebold and Inoue (2001), we have

$$\begin{aligned} var\left(\sum\nolimits_{t=1}^{T} y_t\right) &= var\left[\sum\nolimits_{t=1}^{T} (\mu_{s_t} + \varepsilon_t)\right] \\ &= \mu' \left[\Gamma_0 T + \sum\nolimits_{j=1}^{T} (T-j)(\Gamma_j + \Gamma_j')\right] \mu + T\sigma^2 \end{aligned}$$

where $\mu = \begin{pmatrix} \mu_1 & \mu_2 \end{pmatrix}$ and $\Gamma_j = E(\xi_t \xi_{t-j}')$, and $\Gamma_0 = O(1)$.
When $j > 1$,

$$\Gamma_j = E(\xi_t \xi_{t-j}') = E\left(\begin{pmatrix} I(s_t = 1)I(s_{t-j} = 1) & I(s_t = 1)I(s_{t-j} = 2) \\ I(s_t = 2)I(s_{t-j} = 1) & I(s_t = 2)I(s_{t-j} = 2) \end{pmatrix}\right)$$

$$E\left(I(s_t = 1)I(s_{t-j} = 1)\right) = \begin{cases} 1 & P(s_t = 1, s_{t-j} = 1) \\ 0 & 1 - P(s_t = 1, s_{t-j} = 1) \end{cases}$$

$$P(s_t = 1, s_{t-j} = 1) = P(s_t = 1 | s_{t-j} = 1) P(s_{t-j} = 1)$$

$$P(s_t = 1 | s_{t-j} = 1) = p_{11}^{(j)} = \sum_{r=1,2} p_{1r} p_{r1}^{(j-1)}$$

What contributes to this confusion?

Proof cont'd

When j=2, $P(s_t = 1|s_{t-2} = 1) = p_{11}^{(2)} = \sum_{r=1,2} p_{1r}p_{r1} = p_{12}p_{21} + p_{11}^2 = (1 - p_{11})(1 - p_{22}) + p_{11}^2$ where $p_{11} = P(s_t = 1|s_{t-1} = 1)$ and $p_{22} = P(s_t = 2|s_{t-1} = 2)$ are the one step transition probabilities. Extended to the general case, $P(s_t = 1|s_{t-j} = 1) = p_{11}^{(j)} = f_{11}^j(p_{11}, p_{22})$, in other words, it is a defined function of p_{11} and p_{22}.

Furthermore, $E(I(s_t = 1)I(s_{t-j} = 1)) = P(s_t = 1, s_{t-j} = 1) = f_{11}^j(p_{11}, p_{22})P(s_{t-j} = 1)$.

Similarly, we have

$$\Gamma_j = \begin{pmatrix} f_{11}^j(p_{11}, p_{22})P(s_{t-j} = 1) & f_{12}^j(p_{11}, p_{22})P(s_{t-j} = 2) \\ f_{21}^j(p_{11}, p_{22})P(s_{t-j} = 1) & f_{22}^j(p_{11}, p_{22})P(s_{t-j} = 2) \end{pmatrix}$$

$$= \begin{pmatrix} f_{11}^j(p_{11}, p_{22})P(s_{t-j} = 1) & f_{12}^j(p_{11}, p_{22})[1 - P(s_{t-j} = 1)] \\ f_{21}^j(p_{11}, p_{22})P(s_{t-j} = 1) & f_{22}^j(p_{11}, p_{22})[1 - P(s_{t-j} = 1)] \end{pmatrix}$$

What contributes to this confusion?

Proof cont'd

For finite T, we know that smoothing probability $P(s_t = 1|\Omega_T)$ is a good proxy of $P(s_t = 1)$. As a result,

$$\Gamma_j \approx \begin{pmatrix} f_{11}^j(p_{11}, p_{22})P(s_{t-j} = 1|\Omega_T) & f_{12}^j(p_{11}, p_{22})[1 - P(s_{t-j} = 1|\Omega_T])) \\ f_{21}^j(p_{11}, p_{22})P(s_{t-j} = 1|\Omega_T) & f_{22}^j(p_{11}, p_{22})[1 - P(s_{t-j} = 1|\Omega_T])) \end{pmatrix}$$

Finally, from Diebold and Inoue (2001) we have $var\left(\sum_{t=1}^T y_t\right) = O(T) + \sum_{j=1}^T (T-j)(\Gamma_j + \Gamma_j')$. Since long memory exists for y_t, we would assume that $var(\sum_{t=1}^T y_t) = O(T^{2d+1})$ is true, where $0 < d < 1$. Then, since in standard regime-switching model, p_{11} and p_{22} are non-time-varying constant, it can be seen from the above equations that the smoothing probability $P(s_{t-j} = 1|\Omega_T)$ is the only time-varying term and can lead $\sum_{j=1}^T (T-j)(\Gamma_j + \Gamma_j')$ to be $O(T^{2d+1})$, which completes the proof.

Section Conclusion

- In this part, we show that the time-varying $P(s_{t-j} = 1|\Omega_T)$ is the cause of the significant long memory of regime switching.
- It is expected that if the effect of $P(s_{t-j} = 1|\Omega_T)$ can be properly controlled for, the long memory of the regime switching process should disappear.

2S-ARFIMA Model

- Adopting the idea of the standard MRS model, we allow mean and ARMA terms of y_t to be time dependent and propose a two-stage two-state ARFIMA (2S-ARFIMA) model.
  - A MRS model is fitted for the data to estimate the smoothing probability series $P(s_t = 1|\Omega_T)$.
  - Using the criteria of Hamilton (1989), when $P(s_t = 1|\Omega_T)$ is greater than 0.5, y_t is assumed to lie in the state 1 and otherwise in the state 2.

2S-ARFIMA Model

- The second stage of the 2S-ARFIMA(p,d,q) model is specified as follows:

2S-ARFIMA(p,d,q) model

$$\varphi_{s_t}(L)(1-L)^d(y_t-\mu_{s_t})=\theta_{s_t}(L)\varepsilon_t$$

$$\text{where } \varphi_{s_t}(L)=1-\sum_{i=1}^{p}\varphi_{s_t,i}L^i,\ \ \theta_{s_t}(L)=1-\sum_{j=1}^{p}\theta_{s_t,j}L^j \text{ and } \varepsilon_t=\eta_t\sqrt{\sigma_\varepsilon^2}$$

$\varphi_{s_t,i}$, $\theta_{s_t,j}$ and μ_{s_t} are set to $\varphi_{1,i}$, $\theta_{1,j}$ and μ_1 respectively if y_t lies in the state 1 and are set to $\varphi_{2,i}$, $\theta_{2,j}$ and μ_2 if y_t lies in the state 2. Moreover, we require $\mu_1<\mu_2$. Therefore, the overall mean in state 2 is greater than that in state 1. η_t is an iid series with mean 0 and variance 1 following a specific distribution.

2S-ARFIMA Model

- The time-varying mean μ_{s_t}(and/or the ARMA terms, if any) should capture the variation of $P(s_t|\Omega_T)$.
- If long memory is purely caused by regime switching, when the data are fitted into the 2S-ARFIMA model, we would expect the estimate of d should be close to or smaller than 0.
- Note that d is not allowed to change. Due to the definition of long memory ($var(\sum_{t=1}^{T} y_t) = O(T^{2d+1})$), allowing d to be time-varying can lead to problematic interpretation.

Simulation Study: 2S-ARFIMA Model

- For ARFIMA data, estimated d is significantly greater than 0 (Wilcox sign test) in all cases.

Table 9 : Estimates of 2S-ARFIMA(0,d,0) Model: ARFIMA Data

d	T	$Mean_d$	SE_d	$Mean_{\sigma_\varepsilon^2}$	$SE_{\sigma_\varepsilon^2}$	$Mean_v$	SE_v
0.15	3000	0.0296	0.0206	0.9481	0.1426	2.7756	0.1853
	4000	0.0304	0.0164	0.9549	0.1200	2.7586	0.1594
	5000	0.0310	0.0166	0.9472	0.1049	2.7652	0.1466
0.25	3000	0.1252	0.0240	0.8627	0.1104	2.9731	0.2054
	4000	0.1268	0.0197	0.8630	0.0889	2.9517	0.1643
	5000	0.1255	0.0183	0.8629	0.0776	2.9512	0.1496
0.35	3000	0.2442	0.0280	0.8388	0.0957	3.1312	0.2153
	4000	0.2491	0.0292	0.8381	0.0866	3.1382	0.1951
	5000	0.2479	0.0248	0.8327	0.0757	3.1452	0.1724
0.45	3000	0.3821	0.0330	0.8659	0.0972	3.2019	0.2166
	4000	0.3817	0.0300	0.8597	0.0837	3.2143	0.2015
	5000	0.3810	0.0271	0.8606	0.0760	3.1955	0.1803

Simulation Study: 2S-ARFIMA Model

- For MRS data, estimated d is not significantly greater than 0 (Wilcox sign test) in all cases.

Table 10 : Estimates of 2S-ARFIMA(0,d,0) Model: MRS Data

p_{11}	p_{22}	T	$Mean_d$	SE_d	$Mean_{\sigma^2_\varepsilon}$	$SE_{\sigma^2_\varepsilon}$	$Mean_v$	SE_v
0.999	0.99	3000	-0.0018	0.0107	1.0048	0.1070	3.0246	0.1949
		4000	-0.0012	0.0087	1.0099	0.0880	3.0026	0.1579
		5000	-0.0011	0.0082	1.0019	0.0766	3.0111	0.1446
0.99	0.999	3000	-0.0016	0.0103	1.0129	0.1067	3.0051	0.1867
		4000	-0.0026	0.0097	1.0000	0.0882	3.0172	0.1571
		5000	-0.0012	0.0083	1.0051	0.0818	3.0048	0.1480
0.99	0.99	3000	-0.0042	0.0103	1.0126	0.1033	2.9907	0.1802
		4000	-0.0045	0.0091	1.0036	0.0913	3.0041	0.1724
		5000	-0.0037	0.0077	1.0079	0.0797	2.9885	0.1397
0.999	0.999	3000	-0.0020	0.0107	1.0108	0.0990	3.0047	0.1799
		4000	-0.0013	0.0089	1.0135	0.0883	2.9917	0.1583
		5000	-0.0015	0.0078	1.0064	0.0816	3.0060	0.1455

Section Conclusion

- When fitted into the 2S-ARFIMA model, ARFIMA data still lead to significant long memory.
- For MRS data, estimated d are insignificant.
- 2S-ARFIMA can appropriately control for the effect of $P(s_t|\Omega_T)$ and can distinguish between the pure long memory and pure regime switching.
- Can we do it in one stage?

MRS-ARFIMA Model

- We integrate both MRS and ARFIMA models and propose the following MRS-ARFIMA(p, d, q) framework with Student-t distribution:

MRS-ARFIMA(p, d, q) model

$$\varphi_{s_t}(L)(1-L)^d(y_t-\mu_{s_t})=\theta_{s_t}(L)\varepsilon_{s_t,t}$$

$$\text{where } \varphi_{s_t}(L)=1-\sum_{i=1}^{p}\varphi_{s_t,i}L^i,\ \ \theta_{s_t}(L)=1-\sum_{j=1}^{p}\theta_{s_t,j}L^j\ ,$$

$$\varepsilon_{s_t,t}=\eta_t\sqrt{\sigma_\varepsilon^2} \text{ and } \eta_t\overset{iid}{\sim}t(0,1,v)$$

where the ARMA parameters φ and θ and mean term μ can switch between states, while d is not allowed to change.

- To estimate the parameters, the same MLE procedures for estimating MRS model are employed, where it is also constrained that $\mu_1 < \mu_2$.

Simulation Study: MRS-ARFIMA Model

- For ARFIMA data, estimated d is significantly greater than 0 (Wilcox sign test) and is close to the true value in all cases.

Table 11 : Estimates of MRS-ARFIMA(0,d,0) Model: ARFIMA Data

d	T	$Mean_{p_{11}}$	$SE_{p_{11}}$	$Mean_{p_{22}}$	$SE_{p_{22}}$	$Mean_d$	SE_d
0.15	3000	0.9210	0.1383	0.8564	0.1642	0.1436	0.0126
	4000	0.9060	0.1752	0.8726	0.1445	0.1461	0.0109
	5000	0.9072	0.1704	0.8595	0.1652	0.1454	0.0091
0.25	3000	0.8829	0.1938	0.8225	0.2230	0.2447	0.0120
	4000	0.8878	0.1866	0.8299	0.2147	0.2451	0.0104
	5000	0.8811	0.2025	0.8391	0.2087	0.2459	0.0091
0.35	3000	0.8018	0.2710	0.6670	0.2664	0.3483	0.0112
	4000	0.8013	0.2730	0.6777	0.2723	0.3481	0.0099
	5000	0.8084	0.2673	0.6940	0.2625	0.3478	0.0091
0.45	3000	0.8258	0.2761	0.7668	0.2472	0.4499	0.0112
	4000	0.8370	0.2761	0.8028	0.2126	0.4503	0.0091
	5000	0.8507	0.2479	0.7719	0.2426	0.4499	0.0087

Simulation Study: MRS-ARFIMA Model

- For MRS data, estimated d is still not significantly greater than 0 (Wilcox sign test), while estimated p_{11} and p_{22} are close to their corresponding true values in all cases.

Table 12 : Estimates of MRS-ARFIMA(0,d,0) Model: MRS Data

p_{11}	p_{22}	T	$Mean_{p_{11}}$	$SE_{p_{11}}$	$Mean_{p_{22}}$	$SE_{p_{22}}$	$Mean_d$	SE_d
0.999	0.99	3000	0.9989	0.0007	0.9801	0.0352	-0.0009	0.0107
		4000	0.9989	0.0010	0.9805	0.0598	-0.0004	0.0085
		5000	0.9989	0.0005	0.9828	0.0474	-0.0004	0.0078
0.99	0.999	3000	0.9692	0.0864	0.9981	0.0085	-0.0007	0.0122
		4000	0.9811	0.0404	0.9981	0.0194	-0.0018	0.0087
		5000	0.9839	0.0378	0.9989	0.0008	-0.0007	0.0078
0.99	0.99	3000	0.9881	0.0317	0.9883	0.0254	-0.0004	0.0132
		4000	0.9896	0.0027	0.9896	0.0026	-0.0008	0.0077
		5000	0.9898	0.0020	0.9897	0.0021	-0.0004	0.0063
0.999	0.999	3000	0.9952	0.0489	0.9977	0.0095	-0.0006	0.0136
		4000	0.9982	0.0079	0.9986	0.0012	-0.0007	0.0076
		5000	0.9988	0.0008	0.9986	0.0029	-0.0007	0.0066

Simulation Study: MRS-ARFIMA Data

- We simulate 500 replicates of MRS-ARFIMA(0,d,0) data, where $\mu_1 = -0.5$, $\mu_2 = 0.5$, $\sigma_\varepsilon^2 = 1$ and $v = 3$.

Table 13 : Estimates of ARFIMA(0,d,0) Model

p_{11}	p_{22}	d	$Mean_d$	SE_d	$Mean_{\sigma_\varepsilon^2}$	$SE_{\sigma_\varepsilon^2}$	$Mean_v$	SE_v
0.999	0.99	0.15	0.1913	0.0201	1.0036	0.0712	3.1133	0.1595
		0.25	0.2697	0.0119	1.0012	0.0739	3.0736	0.1522
		0.35	0.3599	0.0089	1.0037	0.0729	3.0414	0.1414
0.99	0.999	0.15	0.1903	0.0186	1.0022	0.0689	3.1096	0.1489
		0.25	0.2696	0.0120	1.0089	0.0738	3.0512	0.1436
		0.35	0.3603	0.0089	1.0060	0.0815	3.0433	0.1563
0.99	0.99	0.15	0.2536	0.0088	1.0029	0.0609	3.3652	0.1714
		0.25	0.3113	0.0098	1.0066	0.0677	3.2172	0.1634
		0.35	0.3854	0.0096	1.0035	0.0711	3.1414	0.1556
0.999	0.999	0.15	0.2189	0.0157	0.9967	0.0699	3.1661	0.1599
		0.25	0.2827	0.0110	1.0004	0.0756	3.0888	0.1520
		0.35	0.3641	0.0088	1.0021	0.0755	3.0416	0.1459

Simulation Study: MRS-ARFIMA Data

- Compared to ARFIMA model, where d is generally overestimated, MRS-ARFIMA model can give estimates that are close to the true values for all parameters.

Table 14 : Estimates of MRS-ARFIMA(0,d,0) Model

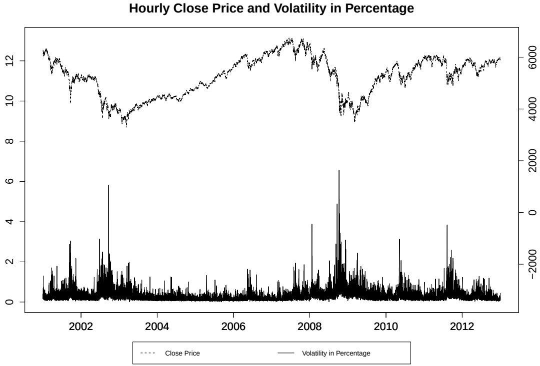
p_{11}	p_{22}	d	$Mean_{p_{11}}$	$SE_{p_{11}}$	$Mean_{p_{22}}$	$SE_{p_{22}}$	$Mean_d$	SE_d
0.999	0.99	0.15	0.9986	0.0008	0.9833	0.0153	0.1475	0.0088
		0.25	0.9977	0.0040	0.9727	0.0676	0.2500	0.0092
		0.35	0.9889	0.0329	0.9682	0.0632	0.3507	0.0090
0.99	0.999	0.15	0.9834	0.0143	0.9985	0.0036	0.1475	0.0085
		0.25	0.9747	0.0448	0.9974	0.0054	0.2500	0.0090
		0.35	0.9553	0.1056	0.9870	0.0369	0.3512	0.0095
0.99	0.99	0.15	0.9868	0.0036	0.9868	0.0031	0.1391	0.0107
		0.25	0.9830	0.0062	0.9833	0.0057	0.2538	0.0116
		0.35	0.9771	0.0198	0.9753	0.0209	0.3627	0.0123
0.999	0.999	0.15	0.9984	0.0018	0.9984	0.0022	0.1473	0.0095
		0.25	0.9977	0.0034	0.9974	0.0064	0.2522	0.0095
		0.35	0.9945	0.0153	0.9951	0.0092	0.3520	0.0085

Section Conclusion

- Via a series of simulation studies, we have demonstrated that the proposed MRS-ARFIMA framework can distinguish between the pure ARFIMA and pure MRS processes.
- Therefore, if long memory is purely caused by regime switching, MRS-ARFIMA would give an insignificant d.
- If estimated d in MRS-ARFIMA model is significantly greater than 0, it suggests that at least long memory is not 'spurious'.
- MRS-ARFIMA framework can provide consistent estimates of all parameters such as the long memory parameter, the transition probability, variance and parameters of the selected distribution.

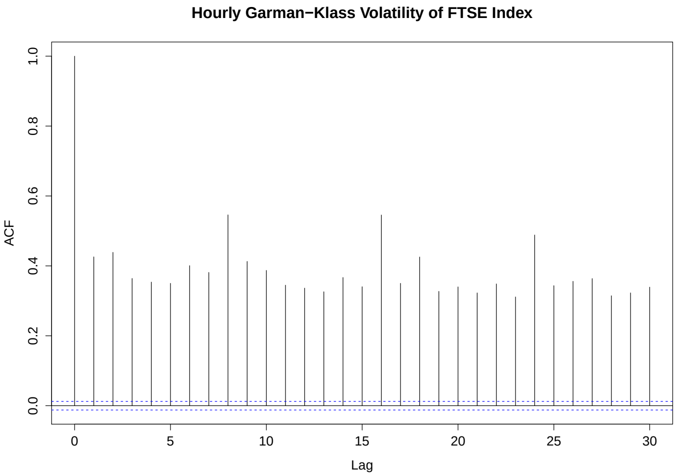
FTSE Hourly Prices and Volatility

- We obtain the hourly prices for FTSE over the period from January 1, 2001 to December 31, 2012 from Thomson Reuters Tick History (TRTH) database
- The volatility is estimated via the Garman-Klass method (Garman and Klass, 1980) from the hourly open, high, low and closing FTSE prices.



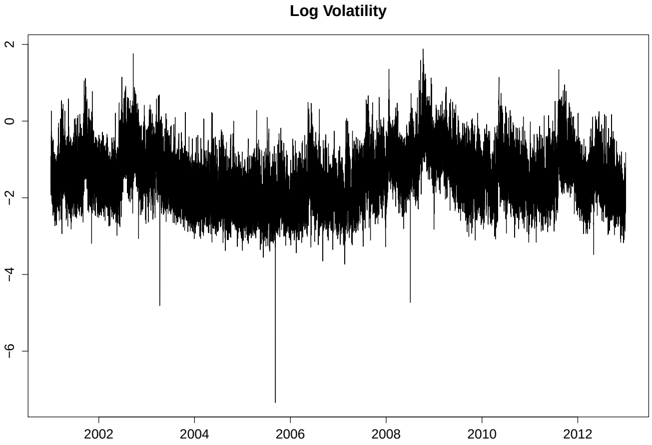
FTSE Hourly Prices and Volatility

- ACF of FTSE gives preliminary evidence of the long memory of the volatility.



FTSE Hourly Prices and Volatility

- Due to the large skewness of the original volatility (4.9662), we focus on the log volatility.



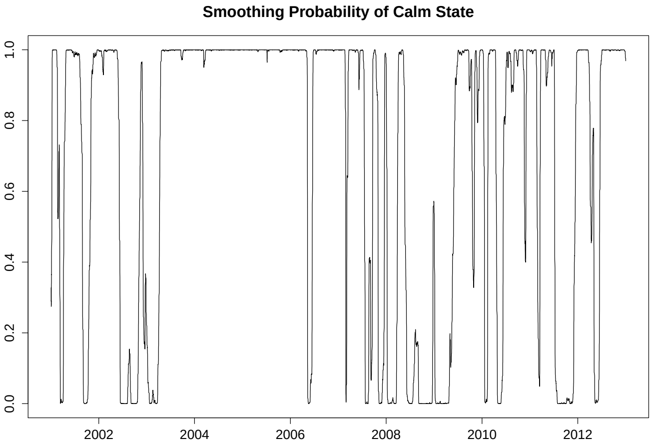
Model Fitting Results

Table 15 : Hourly Volatility of FTSE Index

	ARFIMA		MRS		MRS-ARFIMA
μ	-1.6130 (0.0000)	μ_1	-1.8624 (0.0000)	μ_1	-1.9037 (0.0000)
		μ_2	-1.0282 (0.0000)	μ_2	-0.8333 (0.0000)
		p_{11}	0.9926 (0.0000)	p_{11}	0.9987 (0.0000)
		p_{22}	0.9888 (0.0000)	p_{22}	0.9970 (0.0000)
d	0.2655 (0.0000)			d	0.2058 (0.0000)
σ_ε^2	0.2594 (0.0058)	σ_ε^2	0.2949 (0.0000)	σ_ε^2	0.2525 (0.0000)
v	13.9404 (0.0000)	v	16.3423 (0.0000)	v	12.9235 (0.0000)
$log.lik$	-18668	$log.lik$	-21065	$log.lik$	-18423
AIC	37344	AIC	42142	AIC	36861
BIC	37377	BIC	42191	BIC	36918

Smoothing Probability

- The extracted smoothing probability of MRS-ARFIMA model is consistent with the real macro-economic situation.



Section Conclusion

- Our empirical results demonstrate that MRS-ARFIMA framework is capable of estimating the true long memory parameter and identifying the volatility states.
- Compared with ARFIMA model, it can control for the effects of regime switching and generate more reliable estimate of long memory.
- In terms of model evaluations, MRS-ARFIMA framework outperforms both ARFIMA and MRS models.
- MRS-ARFIMA could be a widely useful tool for modelling the first moment of high-frequency financial series in other contexts.

Conclusion

- Via a serial of simulation studies, we firstly demonstrate that QMLE of AFIMA model is consistent but not efficient, and QMLE of the MRS model is neither consistent nor efficient.
- The confusion between long memory and regime switching is evidenced by long memory tests and model fitting results.
- A theoretical proof is provided suggesting that the existing long memory in regime-switching process is caused by the time-varying smoothing probability series.

Conclusion

- Simulation study based 2S-ARFIMA indicates that when the effect of smoothing probability can be properly controlled for, the long memory of regime-switching process would disappear.
- We further propose a MRS-ARFIMA model which can effectively distinguish between the pure long-memory and pure regime-switching processes.
- An empirical study for the hourly Garman-Klass volatility of FTSE demonstrates that MRS-ARFIMA framework outperforms both the ARFIMA and MRS models and can provide reliable estimates of the long memory parameter and identify the volatility states.