Statistica Sinica: Supplement

ON CONDITIONALLY HETEROSCEDASTIC AR MODELS WITH THRESHOLDS

Kung-Sik Chan, Dong Li, Shiqing Ling and Howell Tong

University of Iowa, Tsinghua University, Hong Kong University of Science & Technology and London School of Economics & Political Science

Supplementary Material

S1 Proof of Theorem 2.1

Because P is irreducible, it admits a unique left eigenvector of 1, say u, such that u is a positive vector that sums to 1, and satisfies the condition

$$u^\tau P = u^\tau.$$

We may identify u with the stationary distribution of the Markov chain with probability transition matrix P. Let $w = (1, \ldots, 1)^\tau$ denote the right eigenvector of P of eigenvalue 1, so that $Pw = w$. The spectral decomposition of P gives

$$P = \{1wu^\tau/u^\tau w\} + Q = wu^\tau + Q. \tag{S1.1}$$

Note that zero is an eigenvalue of Q because $Qw = Pw - wu^\tau w = w - w = 0$, and Q shares the $m-1$ right eigenvectors of P of non-zero eigenvalue less than 1 in magnitude. Consequently, Eqn. (S1.1) implies that, for any positive integer k,

$$P^k = wu^\tau + Q^k. \tag{S1.2}$$

By the Cayley-Hamilton theorem we have

$$c(Q) = Q^{m-1} - c_1 Q^{m-2} - \ldots - c_{m-1} I = 0.$$

Without loss of generality, assume that $\mathbb{E}(h(X_t)) = 0$, otherwise we shall subtract the nonzero mean from h and argue as below. Let $\nu = (\nu_1, \ldots, \nu_m)^\tau$ and $\xi = (\xi_1, \ldots, \xi_m)^\tau$, where, for $1 \le j \le m$, $\nu_j = \mathbb{E}(h(\sigma_j \eta))$ and $\xi_j = \mathbb{E}(h(X_0) I\{X_0 \in R_j\})$. It can be seen that the vector of stationary probabilities u is orthogonal to the vector ν, by the following arguments.

$$0 = \mathbb{E}(h(X_t)) = \mathbb{E}(\mathbb{E}(h(X_t)|X_{t-1})) = \sum_{j=1}^m u_j \mathbb{E}(h(\sigma_j \eta_t)) = u^\tau \nu.$$

It can be checked that the lag-k autocovariance of $\{Y_t\}$ equals

$$\gamma_k = \mathbb{E}(Y_0 Y_k) = \mathbb{E}(Y_0 \mathbb{E}(Y_k | X_0, \ldots, X_{k-1})) = \xi^\tau P^{k-1} \nu = \xi^\tau Q^{k-1} \nu,$$

where the last equality follows from Eqn. (S1.2) and the orthogonality of u and ν. The validity of the Yule-Walker equation defined by Eqn. (3) for the $\{Y_t\}$ process then follows from the equation $c(Q) = 0$. This completes the proof of Theorem 2.1.

S2 Proof of Theorem 2.4

The proof is similar to that of Theorem 2 in Wu and Shao (2004), but for completeness, we outline it below. With no loss of generality, let $p = 1$ and $0 < \alpha < 1$ in conditions (i) and (ii), because the validity of conditions (i) and (ii) with some $\alpha = \alpha_0 > 1$ implies that they also hold for all smaller $0 < \alpha < \alpha_0$, by Hölder's inequality. It follows from condition (i) and (ii) that, for all integer $k \geq 1$,

$$\begin{aligned}
\mathbb{E}|Y_{t,k+1}(y_0) - Y_{t,k}(y_0)|^\alpha &\leq \mathbb{E}[\mathbb{E}\{|Y_{t,k}(X_{t-k-1} + H(y_0)) - Y_{t,k}(y_0)|^\alpha | \mathcal{F}_{t-k-1}\}] \\
&\leq C\zeta^k \mathbb{E}|X_{t-k-1} + H(y_0) - y_0|^\alpha \\
&\leq CD\zeta^k.
\end{aligned}$$

Denote $\delta_k = CD\zeta^k$. Then, Markov's inequality entails that $\mathbb{P}(|Y_{t,k+1}(y_0) - Y_{t,k}(y_0)| > \delta_k^{1/(2\alpha)}) \leq \delta_k^{1/2}$. As $\sum_{k=1}^{\infty} \delta_k^{1/2} < \infty$, it then follows from the Borel-Cantelli lemma that $|Y_{t,k+1}(y_0) - Y_{t,k}(y_0)| \leq \delta_k^{1/2}$ eventually a.s., hence $\{Y_{t,k}(y_0), k \geq 1\}$ is a Cauchy sequence a.s. Consequently, there exists W_t which is $\mathcal{F}_t$-measurable and $Y_{t,k}(y_0) \to W_t$ a.s. Moreover, Fatou's lemma implies that

$$\begin{aligned}
\mathbb{E}|Y_{t,k}(y_0) - W_t|^\alpha &\leq \sum_{\ell=k}^{\infty} \mathbb{E}|Y_{t,\ell}(y_0) - Y_{t,\ell+1}(y_0)|^\alpha \\
&\leq \sum_{\ell=k}^{\infty} \delta_\ell \\
&= \delta_k/(1-\zeta),
\end{aligned}$$

which converges to 0 as $k \to \infty$.

Clearly, if $\{X_t\}$ is stationary ergodic, so is $\{W_t\}$. Furthermore, if $Y_t = W_t$ then $Y_{t+1} = W_{t+1}$, showing that the Y-process is stationary ergodic with the appropriate initial distribution.

Finally, the convergence of $Y_{t,k}(y)$ to W_t holds because

$$\begin{aligned}
\mathbb{E}|Y_{t,k}(y) - W_t|^\alpha &\leq \mathbb{E}|Y_{t,k}(y) - Y_{t,k}(y_0)|^\alpha + \mathbb{E}|Y_{t,k}(y_0) - W_t|^\alpha \\
&\leq C\zeta^k |y - y_0|^\alpha + \delta_k/(1-\zeta)
\end{aligned}$$

entailing that $Y_{t,k}(y) \to W_t$ a.s. This completes the proof.

S3 Proof of Theorem 3.2(b)

To simplify notations, let $\triangle_n = \{\mathbf{r} : \|\mathbf{r} - \mathbf{r}_0\| \leq B/n\}$. Note that

$$\begin{aligned}
& \mathbb{E}\Big(\sup_{\triangle_n} \frac{1}{n}\Big|\sum_{t=1}^{n} X_t^2 I_{it} - \sum_{t=1}^{n} X_t^2 I_{it}^0\Big|\Big) \\
\leq\ & \frac{1}{n}\sum_{t=1}^{n} \mathbb{E}\Big(\sup_{\triangle_n} X_t^2 |I_{it} - I_{it}^0|\Big) \\
\leq\ & \frac{1}{n}\sum_{t=1}^{n} \mathbb{E}(X_t^2[I\{|W_{t-1} - r_{i0}| \leq B/n\} + I\{|W_{t-1} - r_{i-1,0}| \leq B/n\}]) \\
\leq\ & \Big(\max_{1\leq i\leq m} \sigma_{i0}^2\Big)\{\mathbb{P}(|W_{t-1} - r_{i0}| \leq B/n) + \mathbb{P}(|W_{t-1} - r_{i-1,0}| \leq B/n)\}
\end{aligned}$$

for $i = 2, ..., m-1$, where $I_{it}^0 = I\{r_{i-1,0} < W_{t-1} \leq r_{i0}\}$, and

$$\begin{aligned}
\mathbb{E}\Big(\sup_{\triangle_n} \frac{1}{n}\Big|\sum_{t=1}^{n} X_t^2 I_{1t} - \sum_{t=1}^{n} X_t^2 I_{1t}^0\Big|\Big) &\leq \Big(\max_{1\leq i\leq m} \sigma_{i0}^2\Big)\mathbb{P}(|W_{t-1} - r_{10}| \leq B/n); \\
\mathbb{E}\Big(\sup_{\triangle_n} \frac{1}{n}\Big|\sum_{t=1}^{n} X_t^2 I_{mt} - \sum_{t=1}^{n} X_t^2 I_{mt}^0\Big|\Big) &\leq \Big(\max_{1\leq i\leq m} \sigma_{i0}^2\Big)\mathbb{P}(|W_{t-1} - r_{m-1,0}| \leq B/n).
\end{aligned}$$

By Assumption 3.2, $f_w(\cdot)$ is bounded on the set $\bigcup_{i=1}^{m-1}\{x : |x - r_{i0}| \leq 1\}$. Hence, for any fixed $B > 0$ and $n \to \infty$, we have

$$\mathbb{E}\Big(\sup_{\triangle_n} \frac{1}{n}\Big|\sum_{t=1}^{n} X_t^2 I_{it} - \sum_{t=1}^{n} X_t^2 I_{it}^0\Big|\Big) = O(1/n),$$

which implies that

$$\sqrt{n}\sup_{\triangle_n} \frac{1}{n}\Big|\sum_{t=1}^{n} X_t^2 I_{it} - \sum_{t=1}^{n} X_t^2 I_{it}^0\Big| = o_p(1).$$

Similarly,

$$\sqrt{n}\sup_{\triangle_n} \frac{1}{n}\Big|\sum_{t=1}^{n} I_{it} - \sum_{t=1}^{n} I_{it}^0\Big| = o_p(1),$$

which yields that

$$\inf_{\triangle_n} \frac{1}{n}\sum_{t=1}^{n} I_{it} \geq \frac{1}{n}\sum_{t=1}^{n} I_{it}^0 - o_p(n^{-1/2}) = \{F_w(r_{i0}) - F_w(r_{i-1,0})\} - o_p(1).$$

Thus, by the expression of $\hat{\sigma}_{in}^2(\mathbf{r})$ and the law of large numbers, it follows that

$$\begin{aligned}
&\sqrt{n}\sup_{\Delta_n}\Big|\hat{\sigma}_{in}^2(\mathbf{r})-\hat{\sigma}_{in}^2(\mathbf{r}_0)\Big| \\
&\le \Big(\inf_{\Delta_n}\frac{1}{n}\sum_{t=1}^n I_{it}\Big)^{-1}\sqrt{n}\Big(\sup_{\Delta_n}\frac{1}{n}\Big|\sum_{t=1}^n X_t^2 I_{it}-\sum_{t=1}^n X_t^2 I_{it}^0\Big|\Big) \\
&\quad+\frac{\frac{1}{n}\sum_{t=1}^n X_t^2 I_{it}^0}{\Big(\inf_{\Delta_n}\frac{1}{n}\sum_{t=1}^n I_{it}\Big)\Big(\frac{1}{n}\sum_{t=1}^n I_{it}^0\Big)}\sqrt{n}\Big(\sup_{\Delta_n}\frac{1}{n}\Big|\sum_{t=1}^n I_{it}-\sum_{t=1}^n I_{it}^0\Big|\Big)=o_p(1).
\end{aligned}$$

Furthermore, by the law of large numbers,

$$\begin{aligned}
\sqrt{n}(\hat{\sigma}_{in}^2(\mathbf{r}_0)-\sigma_{i0}^2) &= \Big(\frac{1}{n}\sum_{t=1}^n I_{it}^0\Big)^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^n \sigma_{i0}^2(\eta_t^2-1)I_{it}^0 \\
&= \frac{\sigma_{i0}^2}{F_w(r_{i0})-F_w(r_{i-1,0})}\frac{1}{\sqrt{n}}\sum_{t=1}^n(\eta_t^2-1)I_{it}^0+o_p(1).
\end{aligned}$$

By the Cramer-Wold device and the martingale central limit theorem, (b) holds.

S4 Sketch of Proof on the Asymptotic Independence of the Maximum and its Location of an OU process over a Sufficiently Large Interval

We adapt some arguments in Berman (1971) who derived the asymptotic distribution of the maximum of a stationary Gaussian process. It is readily seen from the proof of Theorem 3.1 in Berman (1971) that, under some regularity conditions including a certain asymptotic rate of the autocorrelation function with lag approaching 0 for which the OU process holds, with a fixed $h>0$ and T tending to infinity, the maximum of such a process over the interval $[0,T]$ is asymptotically equivalent to the maximum of the maxima of the process over $[ih,(i+1)h], i=0,..,T-1$, with the maxima over the unit intervals being asymptotically independent. (With no loss of generality, T may be assumed to be a positive integer.) Hence, the maximum of the process over $[0,T]$ and the location of the maximum up to an uncertainty of h are asymptotically independent. By passing to the limit with $h\to 0$, the maximum of the process over $[0,T]$ can be shown to be asymptotically independent of the location of the maximum, with the latter uniformly distributed, for large T. The stationarity of the OU process then allows the result to hold for any sufficiently large, finite interval in lieu of $[0,T]$.

S5 Density Plots of M_-

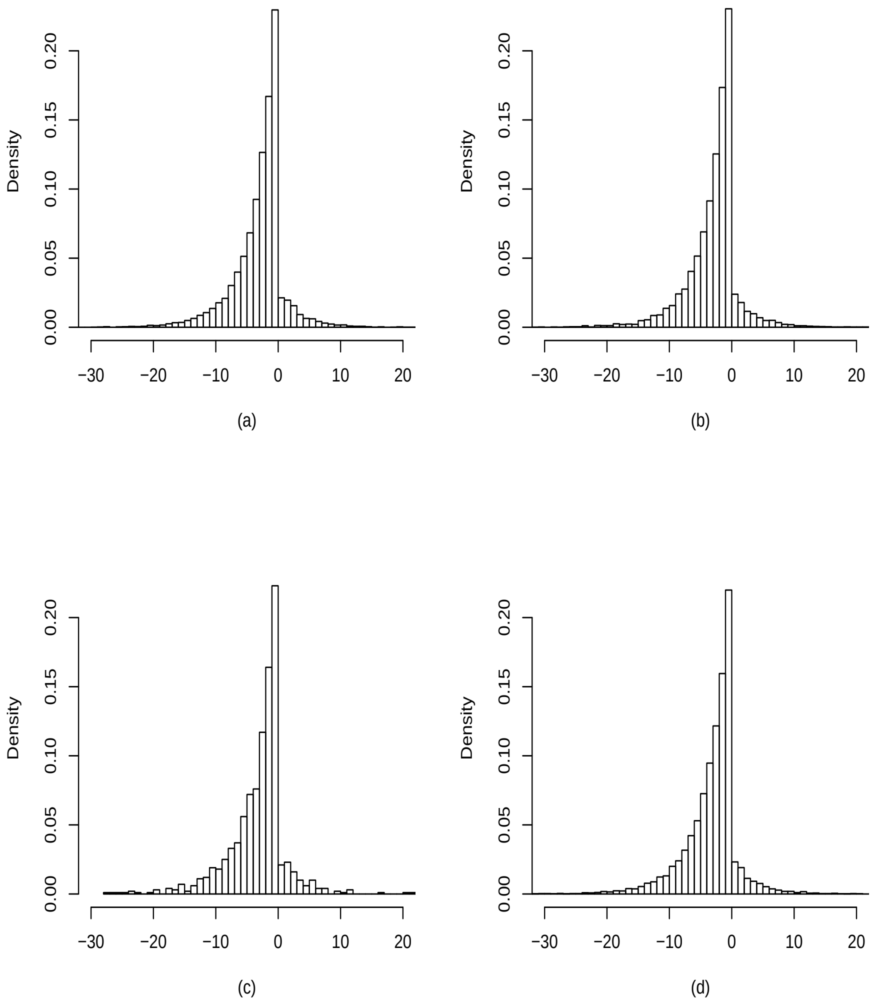

Figure 8: (a) The density of $n(\widehat{r}_n - r_0)$ when $n = 400$ and $\eta_t \sim \mathcal{N}(0, 1)$; (b) The density of M_- when θ_0, $f_w(r_0)$ and $f(x)$ are all known and 10,000 replications are used. (c) The density of $\widehat{M}_-$ when $\hat{\theta}_n$, $\widehat{f}_w(\widehat{r}_n)$ and $\widehat{f}(x)$ and 1,000 replications are used. (d) The density of $\widehat{M}_-$ when $\hat{\theta}_n$, $\widehat{f}_w(\widehat{r}_n)$ and $\widehat{f}(x)$ and 10,000 replications are used.

S6 CREF Analysis

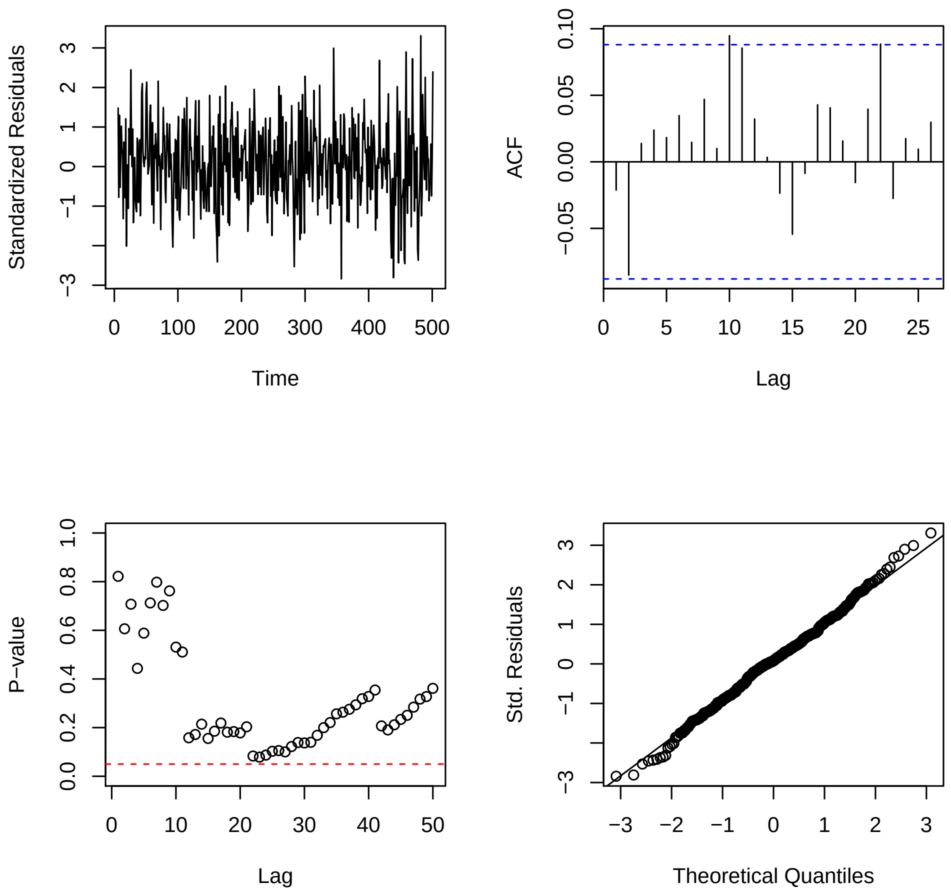

Figure 9: Upper left diagram shows the time plot of the standardized residuals from the fitted T-CHARM. Upper right diagram is the sample ACF of the standardized residuals. Lower left diagram plots the p-values of the McLeod-Li test for residual ARCH effects in the residuals, based on the first k lags of the autocorrelations of the squared standardized residuals, where $k = 1, \cdots, 50$; the dotted horizontal line shows the 5% level. Lower right diagram is the quantile-quantile normal score plot for the standardized residuals.

Figure 9 shows that the model provides a good fit to the data. In particular, the standardized residuals from the fitted T-CHARM appear not to show any conditional heteroscedasticity as judged by the sample ACF of their absolute values. This conclusion is corroborated by the McLeod-Li test (see Li and Li, 1996) as well as a (unreported)

Lagrange multiplier test for residual ARCH effects; see Li (2004) and Ling and Tong (2011) for surveys on goodness-of-fit tests in time series modelling. The standardized residuals also appear to be normally distributed, as its quantile-quantile normal score plot is quite straight.

Figure 10 shows the out-of-sample 1-step-ahead predictive performance of both models with fifty new observations collected from August 16 to October 24 in the year 2006. It plots the average cumulative predictive log-likelihood of the new observations; their per-

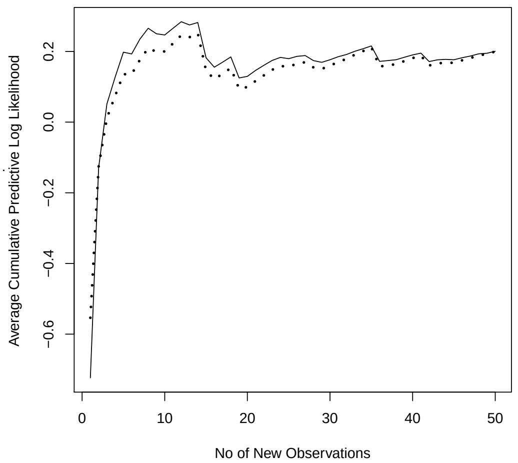

Figure 10: Average cumulative 1-step ahead predictive log-like likelihood, from the T-CHARM (solid line) and the GARCH model (dotted line).

formances seem reasonably comparable with the T-CHARM performing perhaps slightly better. Interestingly, all the fifty new observations fall into the first regime. If we extend the out-of-sample comparison further, the GARCH(1,1) model begins to outperform the T-CHARM, but then the dependence structure of the CREF series also experienced an unprecedented change as the market was heading into the financial crisis in 2007–2010 (Figure 11). Since financial markets are almost invariably nonstationary, it is wise to note the limitations of *stationary* models for financial time series. At best, stationary models such as T-CHARM and GARCH models merely serve to capture the approximate dynamics of the conditional variance over a relatively stationary (therefore limited) period. A really challenging research problem is to model the nonstationarity in the volatilities based on past data; this is a daunting task since market collapses are often triggered by extraneous circumstances, e.g. the credit crisis from 2007 to 2009.

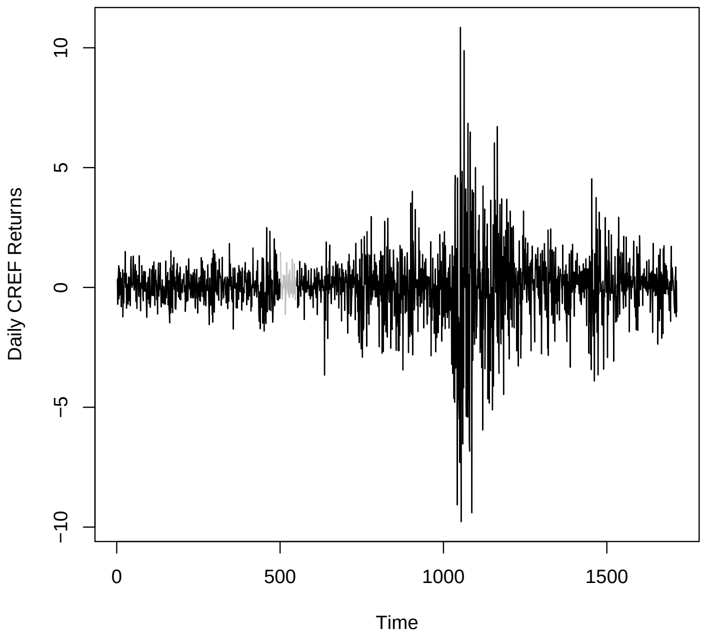

Figure 11: Time plot of an extended daily CREF returns. The initial black solid line shows the data initially analysed, the middle gray line corresponds to the period of out-of-sample forecast and final black solid line draws the more recent data, which shows the dramatic changes in the dependence structure of the CREF series as the market went through the 2007-2010 financial crisis.

S7 The Tree-ring Analysis

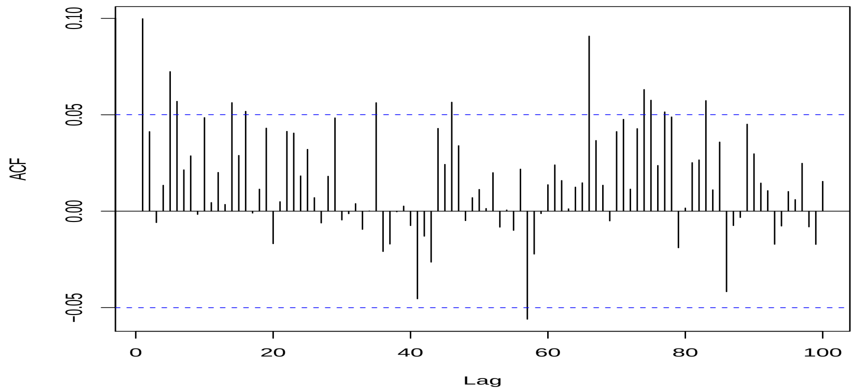

Figure 12: ACF of the absolute residuals from the IMA(1,1) model fitted to the tree ring data.

S8 Lagged Regression Plot for the Old Faithful Geyser Data

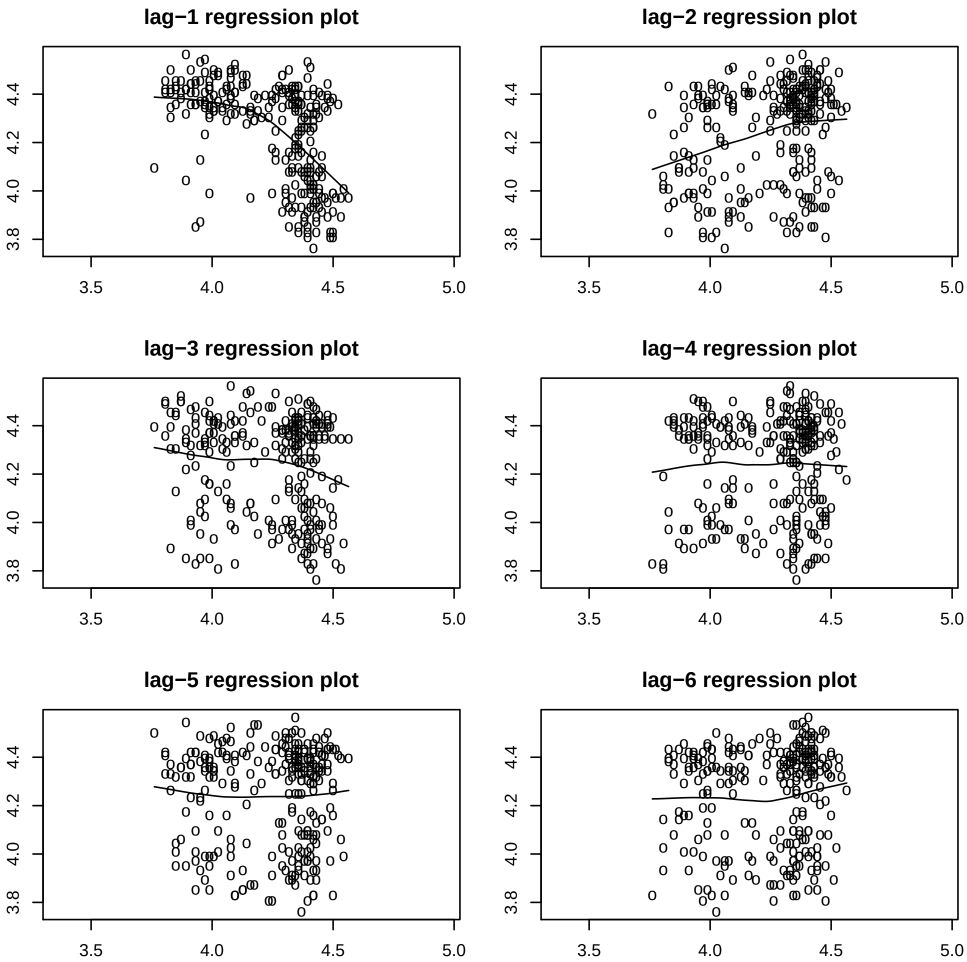

Figure 13: Lagged regression plot of the log waiting time against its lag $k, k = 1, ..., 6$. Open circles are data and solid lines are nonparametric curve fits.

References

Berman, S.M. (1971). Maxima and high level excursions of stationary Gaussian processes. *Trans. Amer. Math. Soc.* **160**, 65–85.

Li, C.W. and Li, W.K. (1996). On a double threshold autoregressive heteroscedastic time series model. *J. Appl. Econometrics* **11**, 253–74.

Li, W.K. (2004). *Diagnostic Checks in Time Series.* Chapman and Hall, London.

Ling, S. and Tong, H. (2011). Score based goodness-of-fit tests for time series. *Statist. Sinica* **21**, 1807–1829.

Wu, W. B. and Shao, X. (2004). Limit theorems for iterated random functions. *J. Appl. Probab.* **41**, 425-436.

Department of Statistics & Actuarial Science, University of Iowa, IA, 52242, USA.
E-mail: kung-sik-chan@uiowa.edu

Mathematical Sciences Center, Tsinghua University, Beijing, China.
E-mail: li20072010@gmail.com

Department of Mathematics, Hong Kong University of Science and Technology,
Clear Water Bay, Hong Kong.
E-mail: maling@ust.hk

Department of Statistics, London School of Economics and Political Science.
E-mail: howell.tong@gmail.com