Testing Serial Correlation and ARCH Effect of High-Dimensional Time-Series Data: **Supplementary Appendix**

Shiqing Ling
Hong Kong University of Science and Technology

Ruey S. Tsay
University of Chicago

Yaxing Yang
Xiamen University

1 Data used

List of the tick symbols of 86 stocks used in Section 6.1: "AET", "FOX", "RTN" , "JNJ" , "REGN", "CCL," "SO", "ITW", "ADP", "AMAT", "NOC", "OXY", "XOM", "TJX", "CSX", "JPM", "WFC", "WMT", "BAC", "T", "PG", "GE", "CVX", "ORCL", "PFE", "VZ", "KO", "UNH", "CMCSA", "C", "HD", "INTC", "MRK", "DIS", "CSCO", "BA", "IBM", "AMGN", "MCD", "AAPL", "MO", "MMM", "MDT", "CELG", "HON", "NVDA", "NKE", "GILD", "UTX", "BMY", "SLB", "LLY", "USB", "MS", "WBA", "LMT", "ABT", "UNP", "TXN", "AGN", "CVS", "TWX", "MSFT", "DJI", "SBUX", "QCOM", "AXP", "ADBE", "DD", "COST", "CB", "TMO", "CAT", "LOW", "PNC", "CL", "BIIB", "AMZN", "DUK", "GD", "SCHW", "DHR", "BK", "COP", "FDX", "SYK".

List of the tick symbols of 24 daily stocks used in Section 6.2: "AAPL", "ABT", "ACN", "AIG", "ALL", "AMGN", "AMZN", "APA", "APC", "AXP", "BA", "BAC", "BAX", "BIIB", "BK", "BMY", "BRKB", "C", "CAT", "CL", "CMCSA", "COF", "COP", "COST".

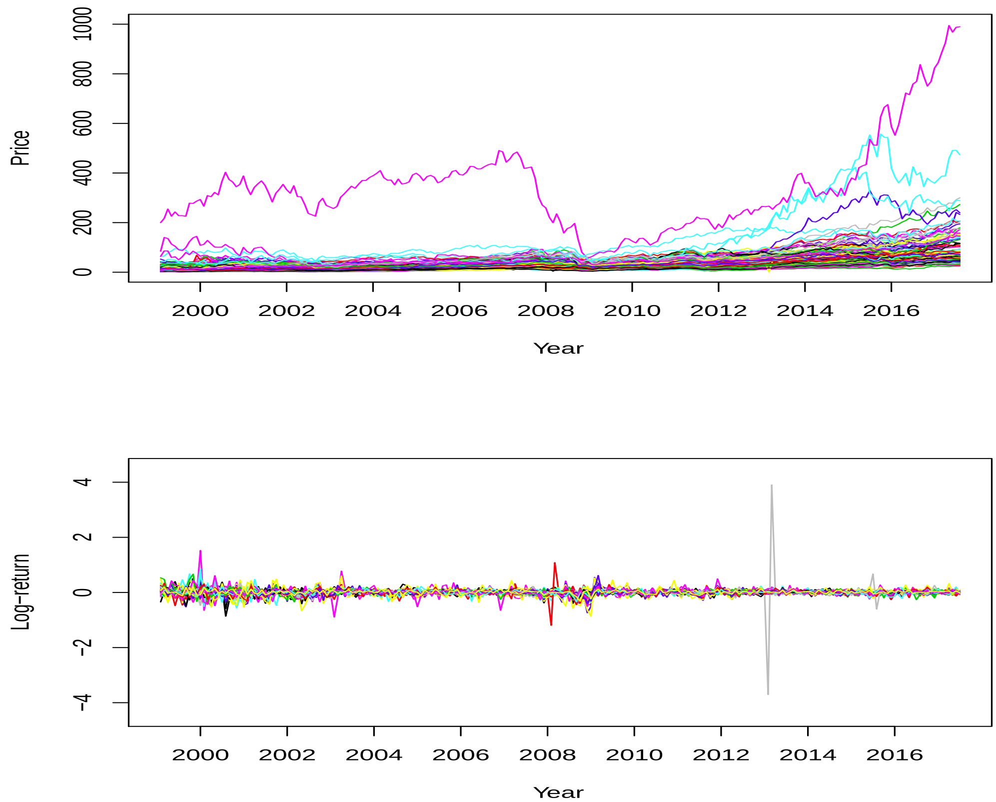

Figure 1: Time plots of monthly prices and theirs log-returns of 86 U.S. stocks.

2 Plot of data

Time plots of the real data used are shown in Figures 1 and 2 of this supplement.

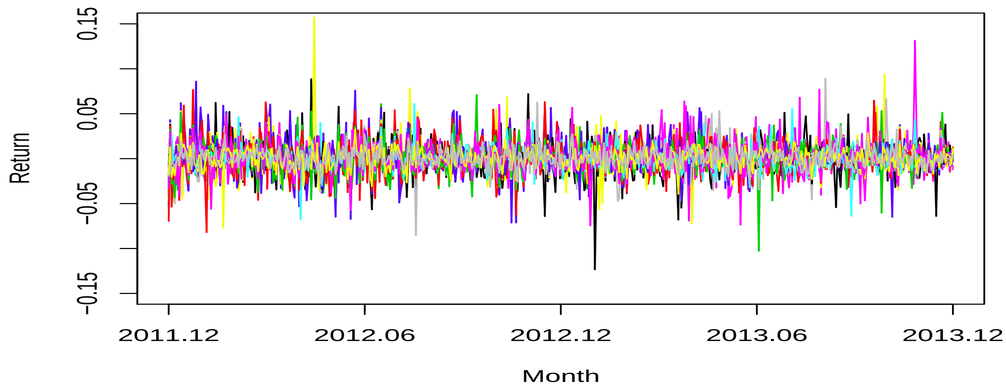

Figure 2: Time plots of daily log returns of 24 U.S. stocks.

3 Further Simulation

Simulation results of the performance of the proposed tests for large p are given in Tables 1 and 2 of this supplement.

4 Simulation Study for the Example in Section 6.1

To see if our tests remain valid when applied to residual series, we generate data from the estimated VAR(1) model with $e_t \sim N(0, I_p)$ and fit model (5.2) to the data, and then use our statistics to test the correlation in the residual series. We set $n = 222$ and use 1000 replications. Table 3 reports the rejection rates of these tests at the level $\alpha = 5\%$.

5 Simulation Study for Example in Section 6.2

To see if our application in Section 6.2 is reliable, we generate data from Model (6.4) with the estimated parameters and $e_t \sim N(0, I_{24})$. The sample size is $n = 518$ and the number of replications is 1000. We first use our tests to detect if any serial correlation or ARCH effect in $r_t = (r_{1t}, \cdots, r_{24t})'$ and then fit Model (6.4) to the

Table 1: Empirical Sizes of Various Test Statistics at the 5% Level for Case 1, where p is the dimension and m is the number of serial correlations used

			Norm			
p	T_n	T_r	W_n	W_r	T_{sn}	T_{sr}
			$m = 5$			
300	0.048	0.047	0.049	0.051	0.039	0.043
500	0.041	0.041	0.048	0.043	0.039	0.043
			$m = 10$			
300	0.066	0.053	0.057	0.050	0.059	0.056
500	0.065	0.062	0.062	0.044	0.069	0.056
			Skewed t_3			
p	T_n	T_r	W_n	W_r	T_{sn}	T_{sr}
			$m = 5$			
300	0.050	0.045	0.051	0.055	0.045	0.047
500	0.044	0.046	0.049	0.050	0.042	0.057
			$m = 10$			
300	0.053	0.059	0.049	0.064	0.054	0.055
500	0.053	0.057	0.054	0.056	0.048	0.056
			Cauchy			
p	T_n	T_r	W_n	W_r	T_{sn}	T_{sr}
			$m = 5$			
300	0.068	0.060	0.060	0.053	0.075	0.058
500	0.047	0.055	0.044	0.062	0.053	0.056
			$m = 10$			
300	0.062	0.048	0.052	0.045	0.072	0.043
500	0.072	0.034	0.062	0.041	0.081	0.036

Table 2: Empirical Sizes of Various Test Statistics at the 5% Level for Case 2, where p is the dimension and m is the number of serial correlations used.

Norm						
p	T_n	T_r	W_n	W_r	T_{sn}	T_{sr}
			$m = 5$			
300	0.056	0.053	0.061	0.066	0.056	0.054
500	0.042	0.053	0.046	0.054	0.041	0.053
			$m = 10$			
300	0.062	0.059	0.061	0.062	0.059	0.064
500	0.042	0.048	0.051	0.051	0.043	0.048
Skewed t_3						
p	T_n	T_r	W_n	W_r	T_{sn}	T_{sr}
			$m = 5$			
300	0.052	0.064	0.051	0.068	0.053	0.060
500	0.046	0.053	0.043	0.049	0.049	0.054
			$m = 10$			
300	0.049	0.043	0.049	0.050	0.049	0.043
500	0.063	0.054	0.066	0.051	0.066	0.053
Cauchy						
p	T_n	T_r	W_n	W_r	T_{sn}	T_{sr}
			$m = 5$			
300	0.054	0.057	0.047	0.056	0.057	0.057
500	0.052	0.057	0.050	0.056	0.058	0.054
			$m = 10$			
300	0.045	0.046	0.043	0.048	0.054	0.045
500	0.059	0.054	0.055	0.054	0.073	0.051

Table 3: Rejection Rates for Serial Correlation When Tests are Applied to the VAR(1) Residuals, where m is the number of serial correlations used.

	$m = 1$	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 10$
T_n	.064	.066	.056	.050	.056	.070
T_r	.074	.073	.078	.061	.071	.072
W_n	.065	.058	.060	.046	.062	.067
W_r	.074	.071	.065	.068	.075	.072
T_{sn}	.062	.060	.043	.044	.047	.067
T_{sr}	.061	.063	.069	.072	.063	.068

data set $\{r_t\}$ and apply our tests to detect if any correlation or ARCH effect exists in the standardized residual series $\{\hat{\eta}_t\}$. Table 4 reports the rejection rates of these tests for $\{r_t\}$ and $\{\hat{\eta}_t\}$ at the level $\alpha = 5\%$.

Table 4: Rejection Rates for Bootstrap Series and Standardized Residuals. Sample size is 518 and the number of iterations is 1000.

Series Simulated from EGARCH(1,1) Model						
	$m = 1$	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 10$
T_n	.699	.782	.816	.832	.852	.869
T_r	.667	.748	.786	.814	.838	.847
W_n	.699	.780	.807	.827	.847	.871
W_r	.669	.749	.783	.812	.829	.859
T_{sn}	.486	.542	.572	.595	.601	.611
T_{sr}	.451	.522	.546	.566	.572	.569
Standardized Residuals of a Fitted EGARCH(1,1) Model						
	$m = 1$	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 10$
T_n	.070	.049	.054	.062	.063	.067
T_r	.068	.054	.051	.064	.057	.062
W_n	.071	.065	.064	.063	.061	.070
W_r	.068	.062	.062	.068	.063	.067
T_{sn}	.070	.049	.052	.063	.062	.066
T_{sr}	.068	.054	.052	.064	.057	.062

6 Proof of Theorem 4.1

Proof of Theorem 4.1. Let $\tilde{\gamma}_{sl}$ be defined as $\hat{\gamma}$ in Section 2.1 with X_t and $\bar{Y}$ replaced by $\tilde{X}_t$ and $\tilde{Y} = \sum_{t=1}^{n} \tilde{X}_t/n$, respectively. By Theorem 2.1, it is sufficient to show that

$$\sqrt{n}p(\hat{\gamma}_{sl} - \tilde{\gamma}_{sl}) = o_p(1), \tag{6.1}$$

$$p(\hat{\gamma}_{s0} - \tilde{\gamma}_{s0}) = o_p(1). \tag{6.2}$$

for $l > 0$. We first have

$$\begin{aligned}
\Delta_{nt} &\equiv \|\hat{X}_t\| - \|\tilde{X}_t\| = \frac{1}{p}\sum_{i=1}^{p} |x_{it}| \left| \frac{1}{s_{in}} - \frac{1}{\sigma_i} \right| \\
&= \frac{1}{p}\sum_{i=1}^{p} |x_{it}| \left| \frac{1}{s_{in}} - \frac{1}{\sigma_i} \right| [I\{s_{in} > \sigma_0/2\} \\
&\quad + \frac{1}{p}\sum_{i=1}^{p} |x_{it}| \left| \frac{1}{\sigma_0} - \frac{1}{\sigma_i} \right| I\{s_{in} \le \sigma_0/2\}] \\
&\equiv A_{1n} + A_{2n}. \qquad (6.3)
\end{aligned}$$

Since $E(s_{in}^2 - \sigma_i^2)^2 \le CE(\tilde{x}_{it} - \sigma_i^2)^2/n$, we have

$$\begin{aligned}
EA_{1n} &\le O(1)\frac{1}{p}\sum_{i=1}^{p} E|x_{it}||s_{in}^2 - \sigma_i^2| = O(\frac{1}{\sqrt{n}}), \qquad (6.4) \\
EA_{2n} &\le O(1)\frac{1}{p}\sum_{i=1}^{p} E|x_{it}|I\{s_{in} \le \sigma_0/2\}] \\
&\le O(1)\frac{1}{p}\sum_{i=1}^{p} \sqrt{E|x_{it}|^2 P\{s_{in} \le \sigma_0/2\}}] \\
&\le O(1)\frac{1}{p}\sum_{i=1}^{p} \sqrt{E|x_{it}|^2}\sqrt{P\{|s_{in}^2 - \sigma_i^2| > \sigma_0/2\}}] \\
&\le O(1)\frac{1}{p}\sum_{i=1}^{p} \sqrt{E|x_{it}|^2}\sqrt{E|s_{in}^2 - \sigma_i^2|^2} = O(\frac{1}{\sqrt{n}}). \qquad (6.5)
\end{aligned}$$

By (6.3)-(6.5), we have $|\bar{Y}_s - \tilde{Y}_s| \le \sum_{t=1}^{n} |\Delta_{nt}|/n$, and

$$(\bar{Y}_s - \mu)^2 \le \left[\frac{1}{n}\sum_{t=1}^{n} \Delta_{nt}\right]^2 + (\bar{Y} - \mu)^2 = O_p(\frac{1}{n}).$$

By previous equation and (6.3)-(6.5), we can show that

$$\begin{aligned}
\hat{\gamma}_{sl} &= \frac{1}{n}\sum_{t=l+1}^{n} (\|\hat{X}_{t-l}\| - \mu)(\|\hat{X}_t\| - \mu) - (\bar{Y}_s - \mu)^2 \\
&= \frac{1}{n}\sum_{t=l+1}^{n} (\|\tilde{X}_{t-l}\| - \mu + \Delta_{nt-l})(\|\tilde{X}_t\| - \mu + \Delta_{nt}) + O_p(\frac{1}{n}) \\
&= \tilde{\gamma}_{sl} + \frac{2}{n}\sum_{t=l+1}^{n} ((\|\tilde{X}_{t-l}\| - \mu)\Delta_{nt}) + \frac{1}{n}\sum_{t=l+1}^{n} \Delta_{nt}^2 + O_p(\frac{1}{n}) \\
&= \tilde{\gamma}_{sl} + \frac{2}{np}\sum_{i=1}^{p} \left[\sum_{t=l+1}^{n} (\|\tilde{x}_{it-l}\| - \mu_i)\right] \Delta_{nt} + O_p(\frac{1}{n}) \\
&= \tilde{\gamma}_{sl} + O_p(\frac{1}{n}).
\end{aligned}$$

Thus, (6.1) holds when $p/\sqrt{n} \to 0$. Similarly, we can show that (6.2) holds. This completes the proof. $\square$