Optimal adaptive detection of small correlation functions

Alain Guay   Emmanuel Guerre   Stepana Lazarova

Vienna, 26 July 2008

Motivation

- **Descriptive statistic:** is $\{u_t\}$ is a (weak) white noise?
- **Model diagnostic**
  - *AR*: $\widehat{u}_t = u_t\left(\widehat{\theta}\right) = X_t - \widehat{\theta}_0 - \widehat{\theta}_1 X_{t-1} - \cdots - \widehat{\theta}_k X_{t-k}$
  - *ARCH*: $\widehat{u}_t = u_t\left(\widehat{\theta}\right) = \frac{X_t^2}{\widehat{\sigma}_t^2} - 1$, where $\widehat{\sigma}_t^2 = \widehat{\theta}_0 + \widehat{\theta}_1 X_{t-1}^2 + \cdots + \widehat{\theta}_k X_{t-k}^2$
  - If $\{u_t = u_t\left(\mathrm{plim}\,\widehat{\theta}\right)\}$ is not a white noise, the lag order k should be increased
- **Economics :** Market Efficiency and Rational Expectation Hypothesis

Framework

- $\{u_t\}$ stationary process with $\mathbb{E}[u_t] = 0$ and finite variance R_0
- Covariance function: $R_j = \operatorname{cov}(u_t, u_{t+j})$
- Sample covariances: $\widehat{R}_j = \frac{1}{n}\sum_{t=j+1}^{n} \widehat{u}_t \widehat{u}_{t-j}$
- **Hypotheses**
    - $H_0 : R_j = 0$ for all $j \neq 0$
    - $H_A : R_j \neq 0$ for some $j \neq 0$
- **Technical conditions**
    - Absolute summability Cumulants of $\{u_t\}$ up to 8th
    - $\widehat{u}_t(\theta)$ twice differentiable w.r.t to θ
    - $\widehat{\theta}$ $n^{1/2}$ consistent

Rest of the talk

Some testing procedures: Cramer von Mises tests and smooth tests
Main goals of the talk
The test
Null limit distribution
Rate consistency
Adaptive rate optimality: "sparse" alternatives
(Simulations)
Applications to financial squared returns
Final remarks

Cramer von Mises type tests

$$\text{CvM} = \frac{n}{\pi^2} \sum_{j=1}^{n-1} \frac{1}{j^2} \left(\frac{\widehat{R}_j}{\widehat{\sigma}_j} \right)^2 ,$$

$\{u_t\}$ observed, $\widehat{\sigma}_j = \widehat{R}_0$ or heteroscedasticity robust (Lobato et al, 2001)

- Does not use any smoothing parameters
- Detects $n^{1/2}$ Pitman local alternatives
- Not suitable for alternatives with small correlations at low order

Smooth test statistics

$$\widehat{BP}_p = n \sum_{j=1}^{p} \left(\frac{\widehat{R}_j}{\widehat{\sigma}_j} \right)^2 \text{ (Box and Pierce 1970)},$$

$$\widehat{S}_p = n \sum_{j=1}^{n-1} K^2 \left(\frac{j}{p} \right) \left(\frac{\widehat{R}_j}{\widehat{\sigma}_j} \right)^2 \text{ (Hong 1996)}$$

p = truncation/smoothing parameter

- Does not downweight large j, but p is difficult to choose in practice
- Asymptotically minimax (Ermakov, 1994) against smooth alternatives, but for a p dependent of the unknown alternative
- Does not detect $n^{1/2}$ Pitman local alternatives if p is too large

Data-driven choice of the smoothing parameter

- Fan and Yao (2005) propose

$$\max_{p\in\mathcal{P}} \frac{\widehat{S}_p - E(p)}{V(p)}$$

  - No theoretical study
  - Lack of proper critical values for usual sample sizes
- "Rule of thumb" for smooth tests (Hong (1996), Andrews (1991), Newey West (1994))
  - are in general "optimal" for estimation of the spectral density but not for testing
  - No clear limit distribution under the null due to a random choice of p

Main goals of the talk

To propose a test that

1. has a simple null limit distribution
2. achieves adaptive optimal detection of "sparse" alternatives

Construction of the test: improving a first initial test

Many applied works use a moderate deterministic $\underline{p}$ (typically, 5, 10, or $\ln n$). The test rejects H_0 if

$$\frac{\widehat{S}_{\underline{p}} - E(\underline{p})}{V(\underline{p})} \geq z_n(\alpha),$$

where

- $E(p) = \sum_{j=1}^{n-1} \left(1 - \frac{j}{n}\right) K^2 \left(\frac{j}{p}\right)$ and $V^2(p) = 2\sum_{j=1}^{n-1} \left(1 - \frac{j}{n}\right)^2 K^4 \left(\frac{j}{p}\right)$ are approximation of the mean and variance of $\widehat{S}_{\underline{p}}$ under the null of independence (Hong, 1996)
- $E(\underline{p}) + V(\underline{p}) z_n(\alpha)$ is a critical value:
  - Null of indepence: Chi square (Box-Pierce, 1970) or Fisher, Normal $z_n(\alpha)$ (Hong, 1996, $\underline{p} \to \infty$, estimated residuals)

When should I increase my favorite trunctation parameter?

- I should change my favorite $\underline{p}$ for a larger $\overline{p}$ if $\widehat{S}_{\overline{p}}$ strongly differs from $\widehat{S}_{\underline{p}}$, i.e. if for a level λ_n typically tending to 0

$$\frac{\left(\widehat{S}_{\overline{p}} - \widehat{S}_{\underline{p}}\right) - E(\overline{p}, \underline{p})}{V(\overline{p}, \underline{p})} \geq z_n(\lambda_n), \text{ where}$$

$$\begin{aligned} E(\overline{p}, \underline{p}) &= E(\overline{p}) - E(\underline{p}), \\ V^2(\overline{p}, \underline{p}) &= 2 \sum_{j=1}^{n-1} \left(1 - \frac{j}{n}\right)^2 \left(K^2\left(\frac{j}{\overline{p}}\right) - K^2\left(\frac{j}{\underline{p}}\right)\right)^2. \end{aligned}$$

- The final test uses the smoothing parameter

$$\widetilde{p} = \arg \max_{p \in \{\underline{p}, \overline{p}\}} \left(\left(\widehat{S}_p - \widehat{S}_{\underline{p}}\right) - E(p, \underline{p}) - z_n(\lambda_n) V(p, \underline{p})\right)$$

The proposed test

- $\gamma_n \geq 0$ penalty sequence, $\mathcal{P} = \left\{\underline{p}, 2\underline{p}, ..., 2^{Q-1}\underline{p} = \overline{p}\right\}$

$$\widehat{p} = \arg\max_{p \in \mathcal{P}} \left(\left(\widehat{S}_p - \widehat{S}_{\underline{p}} \right) - E(p, \underline{p}) - \gamma_n V(p, \underline{p}) \right).$$

- Rejects H_0 if

$$\frac{\widehat{S}_{\widehat{p}} - E(\underline{p})}{V(\underline{p})} \geq z_n(\alpha).$$

The test uses that $\mathbb{P}\left(\widehat{p} = \underline{p}\right) \to 1$ under the null:

- for the studentization with $E(\underline{p})$ and $V(\underline{p})$ (improves power)
- when choosing the critical value $z_n(\alpha) = z_n(\alpha; \underline{p})$

A penalty lower bound

Theorem

Suppose that u_t is identically distributed. Then, if $\overline{p} = o(n^{1/3})$, and if the selection sequence $\{\gamma_n, n \geq 1\}$ satisfies

$$\gamma_n \geq (2 \ln \ln n)^{1/2} + \epsilon \quad \text{for some } \epsilon > 0, \tag{1}$$

$\widehat{p} = \underline{p}$ with a probability tending to 1 under the null, and the test is asymptotically of level α.

Rate consistency

Theorem

Assume that γ_n diverges. Consider a sequence of alternatives $\{u_{t,n}\}$. Then the test is consistent, if, for some τ large enough,

$$n\sum_{j=1}^{\infty}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2 \geq \tau^2 \min_{p\in[\underline{p},\overline{p}]}\left(n\sum_{j=p}^{\infty}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2 + \gamma_n (2p)^{1/2}\right). \tag{2}$$

- The RHS of (2) is a "bias-variance" trade off when estimating $n\sum_{j=1}^{\infty}\left(\frac{R_j}{R_0}\right)^2$ between:
  - The "bias" of $\widehat{BP}_p - E(p)$, $n\sum_{j=p}^{\infty}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2$.
  - The penalisation term $\gamma_n V(p,\underline{p}) = O\left(\gamma_n (2p)^{1/2}\right)$, which plays the role of a variance

A framework for sparse alternatives

3 "ingredients" to describe the "sparsity" of $\{u_{t,n}\}$

1. A maximal lag index P_n, such that the correlations at lags larger than P_n are negligible:

$$\sum_{j=P_n+1}^{\infty}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2 = o\left(\sum_{j=1}^{\infty}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2\right)$$

2. A rate $\rho_n \to 0$ used to define "significant" correlation coefficients, $j \leq P_n$:

$$\frac{R_{j,n}}{R_{0,n}} \text{ "significant" if } \left|\frac{R_{j,n}}{R_{0,n}}\right| \geq \rho_n.$$

3. A lower bound N_n for the number of "significant" correlation coefficients, $j \leq P_n$:

$$\#\left\{\left|R_{j,n}/R_{0,n}\right| \geq \rho_n, j \in [1, P_n]\right\} \geq N_n.$$

Adaptive rate optimality for sparse alternatives

Theorem

Assume that γ_n diverges with $\gamma_n = o(\overline{p}^{1/2})$. Consider a sequence of alternatives $\{u_{t,n}\}$. Suppose that, for some unknown P_n in $[\underline{p}, \overline{p}]$ and $\rho_n \to 0$,

$$\sum_{j=P_n+1}^{\infty} \left(\frac{R_{j,n}}{R_{0,n}}\right)^2 = o\left(\sum_{j=1}^{P_n} \left(\frac{R_{j,n}}{R_{0,n}}\right)^2\right),$$

$$\#\left\{|R_{j,n}/R_{0,n}| \geq \rho_n, j \in [1, P_n]\right\} \geq N_n.$$

Then the test is consistent, if, for some τ large enough,

$$\rho_n \geq \frac{\tau}{n^{1/2}} \left(\frac{\gamma_n P_n^{1/2}}{N_n}\right)^{1/2}.$$

- Allows for detection of correlation coefficients of order $o\left(\frac{1}{n^{1/2}}\right)$ when $\frac{P_n^{1/2}}{N_n} \longrightarrow 0$.

- When $\gamma_n \asymp (2 \ln \ln n)^{1/2}$, the condition

$$\rho_n \geq \frac{\tau}{n^{1/2}} \left(\frac{\gamma_n P_n^{1/2}}{N_n} \right)^{1/2}$$

cannot be improved when $\frac{P_n^{1/2}}{N_n} \longrightarrow 0$, P_n being unknown.

Theorem

There is a τ in $[0, 1]$ and sequences of alternatives satisfying

$$\rho_n \geq \frac{\tau}{n^{1/2}} \left(\frac{(2 \ln \ln n)^{1/2} P_n^{1/2}}{N_n} \right)^{1/2}, \quad \frac{P_n^{1/2}}{N_n} \longrightarrow 0,$$

that cannot be detected by any test.

Simulation setup

- $\gamma_n = (2 \times \ln(Q-1))^{1/2} + 3.2,\ Q = \#\mathcal{P}$
- n=200: $\mathcal{P} = \{2, 4, 8, 16, 32\}$
- n=1000: $\mathcal{P} = \{2, 4, 8, 16, 32, 64, 128, 256\}$
- Uniform kernel=Box Pierce statistics: critical value given by Chi Square (2)
- Parzen kernel

$$k(x) = \begin{cases} 1 - 6x^2 + 6|x|^3 & |x| \leq \frac{1}{2}, \\ 2(1-|x|)^3 & \frac{1}{2} \leq |x| \leq 1, \\ 0 & \text{otherwise}, \end{cases}$$

critical value given by Gamma approximations matching the two first moments $E(2)$ and $V(2)$.

Test statistics

- **GGL**: Data-driven $\widehat{p}$
  - Uniform Kernel (Box Pierce)
  - Parzen Kernel
- **IMSE** = "Rule of Thumb": p given by a data-driven procedure as in Andrews (1991), Newey West (1994)
- **CVM**: Cramer von Mises

Null hypothesis: 200 observations

Table 1:null, 200 obs

	GGL Box Pierce		GGL Parzen		IMSE Parzen		CVM CVM	
	10 %	5 %	10 %	5 %	10 %	5 %	10 %	5 %
Normal	9.45	5.00	9.77	4.94	10.32	4.92	9.36	4.70
Student(5)	9.58	5.10	9.54	4.83	10.01	4.81	9.24	4.60
Chi-square	9.18	4.79	9.77	4.74	10.29	4.74	9.14	4.48

Null hypothesis: 1000 observations

Table 2:null, 1000 obs

	GGL Box Pierce		GGL Parzen		IMSE Parzen		CVM CVM	
	10 %	5 %	10 %	5 %	10 %	5 %	10 %	5 %
Normal	10.30	5.04	10.20	5.00	11.21	5.43	10.12	4.91
Student(5)	10.10	4.93	10.05	4.89	10.82	5.36	9.54	4.92
Chi-square	9.62	5.08	10.29	5.03	11.25	5.37	9.88	4.82

Alternative hypothesis: Cramer von Mises alternatives

$$n\sum_{j=1}^{\infty}\frac{1}{j^2}\left(\frac{R_j}{R_0}\right)^2 = 3,\ \ n = 200$$

Table 3: 200 obs

	GGL Box Pierce		GGL Parzen		IMSE Parzen		CVM CVM	
	10 %	5 %	10 %	5 %	10 %	5 %	10 %	5 %
$MA(1)$	44.17	30.96	53.66	40.84	54.48	40.70	52.34	39.34
$MA(4)$	100.00	100.00	99.98	99.98	17.11	9.86	77.46	41.87
$AR(1)$	42.82	31.12	52.59	39.57	53.25	39.51	51.20	38.45
$AR(6)$	100.00	100.00	100.00	100.00	35.88	25.64	89.74	69.03

$$MA(1) \ : \ u_t = \varepsilon_t - .12\varepsilon_{t-1}, \quad MA(4) : u_t = \varepsilon_t - .82\varepsilon_{t-4}$$
$$AR(1) \ : \ u_t = .12u_{t-1} + \varepsilon_t, \quad AR(6) : u_t = .68u_{t-6} + \varepsilon_t$$

Alternative hypothesis: Cramer von Mises alternatives,

$$n\sum_{j=1}^{\infty}\frac{1}{j^2}\left(\frac{R_j}{R_0}\right)^2 = 3, n = 1000$$

Table 4: 1000 obs

	GGL truncated		GGL Parzen		IMSE Parzen		CVM CVM	
	10 %	5 %	10 %	5 %	10 %	5 %	10 %	5 %
$MA(1)$	43.65	31.41	53.14	40.14	53.94	40.51	52.01	39.58
$MA(4)$	99.98	99.98	98.92	98.92	12.08	6.07	75.88	41.12
$AR(1)$	44.72	32.38	54.24	41.35	54.93	41.80	52.86	40.82
$AR(6)$	100.00	100.00	100.00	100.00	16.18	9.07	84.89	47.92

$$MA(1) : u_t = \varepsilon_t - .06\varepsilon_{t-1}, \quad MA(4) : u_t = \varepsilon_t - .23\varepsilon_{t-4}$$

$$AR(1) : u_t = .05u_{t-1} + \varepsilon_t, \quad AR(6) : u_t = .32u_{t-6} + \varepsilon_t$$

Alternative hypothesis: Small MA coefficients

$$u_t = \varepsilon_t + \frac{(3\gamma_n)^{1/2}}{n^{1/2}P^{1/4}}\left(\zeta_1\varepsilon_{t-1} + \cdots + \zeta_P\varepsilon_{t-P}\right),$$
$$\{\varepsilon_t\}, \{\zeta_t\} \text{ i.i.d. } \mathcal{N}(0,1).$$

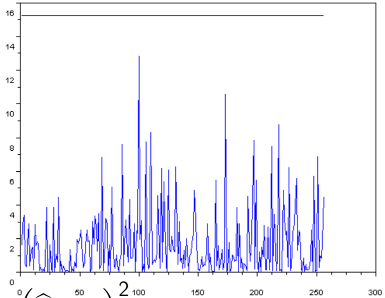
Table 5: 200 obs

	GGL Box Pierce		GGL Parzen		IMSE Parzen		CVM CVM	
	10 %	5 %	10 %	5 %	10 %	5 %	10 %	5 %
P=15	83.11	79.96	68.11	64.38	50.20	39.64	59.39	46.54
P=30	78.45	75.25	54.21	49.15	42.58	31.69	49.17	36.95

Table 6: 1000 obs.

	GGL Box Pierce		GGL Parzen		IMSE Parzen		CVM CVM	
	10 %	5 %	10 %	5 %	10 %	5 %	10 %	5 %
P=75	94.60	93.75	92.90	92.33	40.44	29.14	42.75	30.55
P=150	94.03	93.26	79.71	77.84	32.56	21.66	33.13	22.24

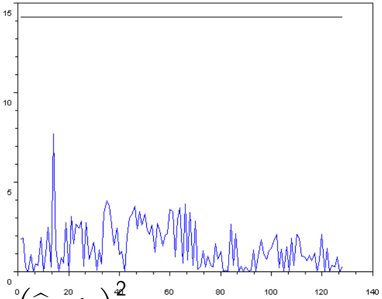
Squared returns, DJI (monthly)



$$n\left(\widehat{R}_j/\widehat{\sigma}_j\right)^2,\ j=1,\dots,256,\ n=699$$

- CvM, IMSE and $\mathrm{Max}_{j\in[1,128]}\, n\frac{\widehat{R}_j^2}{\widehat{\sigma}_j^2}$ accepts H_0 at 5% and 10% (P value CvM=12%)
- Adaptive test rejects H_0 at any level ($\widehat{p}=256$, $\mathcal{P}=\{4,\dots,256\}$).

Squared returns, Coca (monthly)



$n\left(\widehat{R}_j/\widehat{\sigma}_j\right)^2$, $j = 1, \ldots, 128$, $n = 555$

- CvM, IMSE and $\mathrm{Max}_{j\in[1,128]}\, n\frac{\widehat{R}_j^2}{\widehat{\sigma}_j^2}$ accepts H_0 at any reasonable significant level.
- Adaptive test rejects H_0 at any level ($\widehat{p} = 64$, $\mathcal{P} = \{4, \ldots, 128\}$).

To conclude:

The adaptive test

- Has simple critical values, that seems to be accurate in our simulation experiments
- Can detect correlation coefficients smaller than $1/n^{1/2}$
- Is adaptive rate optimal for detection of smooth alternatives
- Can detect Pitman alternatives which goes to the null with a rate close to $1/n^{1/2}$
- Succeeds to detect correlations where other tests failed