Parameter stability and semiparametric inference in time-varying ARCH processes

Lionel Truquet,
lionel.truquet@ensai.fr

1. From stationary to locally stationary ARCH processes
2. Statistical inference. Asymptotic results
    - Asymptotic results for the parametric component
    - Estimation of the nonparametric component
3. Testing parameter constancy/second order dynamic
4. Implementation and a few simulations
5. Some illustrations on real data sets

Outline

1. From stationary to locally stationary ARCH processes
2. Statistical inference. Asymptotic results
   - Asymptotic results for the parametric component
   - Estimation of the nonparametric component
3. Testing parameter constancy/second order dynamic
4. Implementation and a few simulations
5. Some illustrations on real data sets

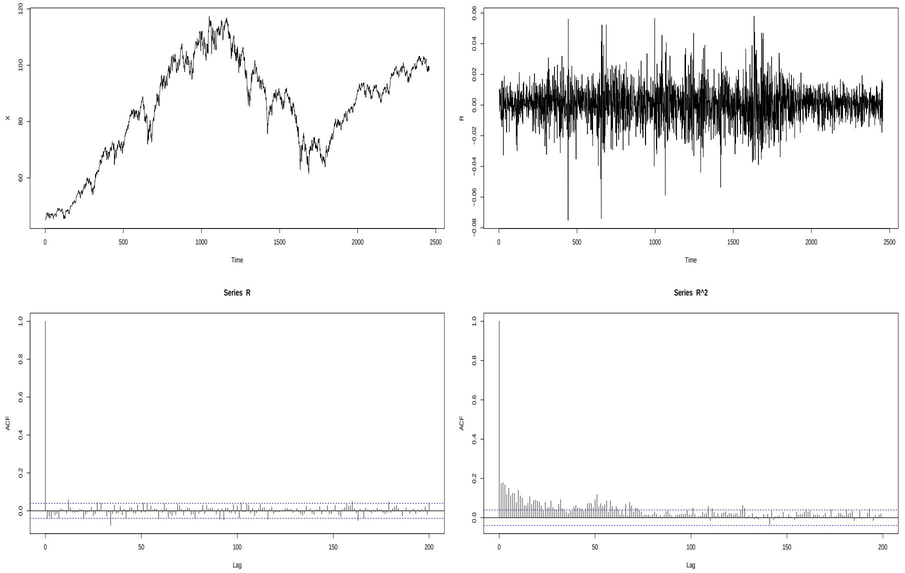
Classical stationary ARCH model of Engle (1982)

$$X_t = \xi_t \sqrt{a_0 + \sum_{j=1}^{p} a_i X_{t-i}^2}, \quad \mathbb{E}(\xi_0) = 0, \text{Var }(\xi_0) = 1, \ \xi \text{ i.i.d}$$

- A classical formulation used to model correlation of the squares (or absolute values) of financial returns $X_t = \frac{P_t - P_{t-1}}{P_{t-1}}$. $P_t =$ (daily) stock price, stock market index or currency exchange rates.

- But stationarity is realistic only on short periods,
  - unconditional variance can be time-varying,
  - stationarity and short memory is incompatible with slow decays in the ACF,
  - ARCH parameters are suspected to be time-varying (estimates of parameters on subsamples seem not the same).

Illustration with the index S&P500 from 1996 to 2005



First model. ARCH versus time-varying unconditional variance

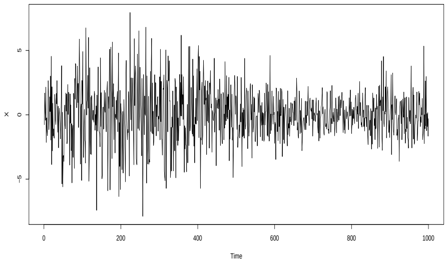
A simple model proposed by Starica & Granger (2005),

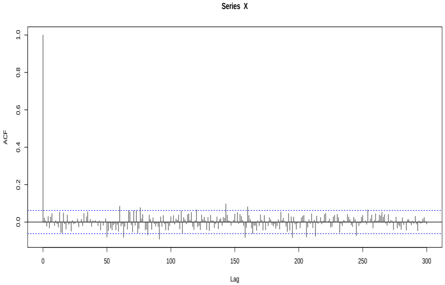
$$X_t = \sigma(t/T)\xi_t, \quad 1 \leq t \leq T,$$

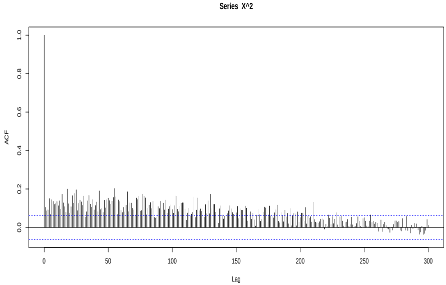
with ξ i.i.d and σ determistic function with smooth changes (locally i.i.d process).

- Compatible with the autocorrelograms of financial returns.
- This simple model can produce significantly better volatility forecasts than the GARCH$(1,1)$ (for $S\&P$ 500).
- According to Starica & Granger: "Most of the dynamics of the $S\&P$ 500 series $(1928 - 2000)$ seems to be concentrated in shifts of the unconditional variance".

$T = 1000$, $\xi_t \sim \mathcal{N}(0,1)$, $\sigma(t/T)^2 = 2 + sin(2\pi t/1000)$







Second model: time-varying unconditional variance and ARCH

"multiplicative volatility model" introduced by Engle & Rangel (2008)

$$X_t = \sigma\left(\frac{t}{T}\right) Y_t$$

where Y ARCH type stationary process (ARCH, GARCH$(1,1)$...) and σ is a deterministic smooth function.

- Nonstationary time series model. Extends the two previous models (ARCH and model of Starica & Granger).
- Inference: splines decomposition for σ (Engle & Rangel), kernel estimation of $\sigma^2(t/T) = \mathbb{E}X_t^2$ and parametric inference for the residuals (Hafner & Linton, 2010).

Third model: the time varying ARCH process

$$X_t = \xi_t \sigma_t = \xi_t \sqrt{a_0\left(\frac{t}{T}\right) + \sum_{i=1}^{p} a_i\left(\frac{t}{T}\right) X_{t-i,T}^2}, \quad t = p+1, \ldots, T.$$

$$\xi \text{ i.i.d }, \quad \mathbb{E}\xi_0 = 0, \quad \text{Var}\, \xi_0 = 1.$$

- Model introduced and studied by Dahlhaus and Subba Rao (AoS, 2006), Fryzlewicz, Sapatinas and Subba Rao (AoS, 2008).
- The a_j's are smooth functions (e.g Lipschitz continuity).
- Locally approximated by stationary ARCH processes

$$X_t(u) = \xi_t \sqrt{a_0(u) + \sum_{i=1}^{p} a_i(u) X_{t-i}(u)^2}, \quad \sum_{i=1}^{p} a_i(u) < 1.$$

Links between the three models

Model 1: Starica & Granger. Model 2: Engle & Rangel. Model 3: tv-ARCH. With non time-varying lag coefficients:

$$X_t = \xi_t \sqrt{a_0\left(\frac{t}{T}\right) + \sum_{i=1}^{p} a_i X_{t-i}^2}.$$

- We have $X_t = \sqrt{a_0\left(\frac{t}{T}\right)} Y_t$ where

$$Y_t^2 = \xi_t^2 \left(1 + \sum_{i=1}^{p} \frac{a_0\left(\frac{t-i}{T}\right)}{a_0\left(\frac{t}{T}\right)} a_i Y_{t-i}^2\right) \approx \xi_t^2 \left(1 + \sum_{i=1}^{p} a_i Y_{t-i}^2\right),$$

because $\frac{a_0\left(\frac{t-i}{T}\right)}{a_0\left(\frac{t}{T}\right)} = 1 + O\left(\frac{1}{T}\right)$. This restriction leads to the multiplicative volatility model with a stationary ARCH component.

Then: Model 1 $\subset$ Model 2 $\subset$ Model 3.

What amount of nonstationarity/nonlinearity ?

$$X_t = \xi_t \sqrt{a_0\left(\frac{t}{T}\right) + \sum_{i=1}^{p} a_i\left(\frac{t}{T}\right) X_{t-i}^2}.$$

- Test: H_0 non time-varying lag coefficients vs H_1 time-varying lag coefficient.
- Test the second order dynamic (or ARCH effect): H_0 $a_{1:p} = 0$ vs H_1 one of the a_i's is time-varying.
- tv-ARCH is the most general but the most difficult to handle and to interpret...

Example: ARCH effect or no ARCH effect? (Figure: Fryzlewicz et al.)

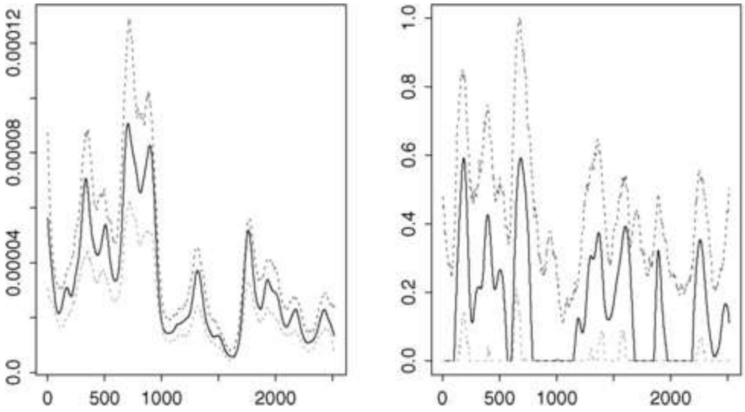


Figure : USD/GBP exchange rates $(1990 - 1999)$. Estimation of a_0 (left) and a_1 (right)

General problem:

$$X_t = \xi_t \sqrt{J_t' \alpha(t/T) + K_t' \beta},$$

where J_t (resp. K_t) contains lag squares or 1 and α (resp. β) contains some of the coefficients.

1. Test H_0: β non time-varying.
2. Inference in such semiparametric models (main focus on the case $a_{1:p}$ constant).
3. Test the second order dynamic in the semiparametric model.

$$X_t^2 = J_t' \alpha(t/T) + K_t' \beta + Z_t, \quad \mathbb{E}\left(Z_t | \mathcal{F}_{t-1}\right) = 0.$$

$\Rightarrow$ semiparametric inference in a time-varying regression model

Recent references for this problem: Chen & Hong (2012), Zhang & Wu (2012).

Outline

Inference in tv-ARCH using kernel estimation (Fryzlewicz et al. 2008)

Estimator of $(a_0(u), \ldots, a_p(u))$ minimizing

$$\alpha \mapsto \sum_{t=p+1}^{T} W_b\left(u - \frac{t}{T}\right) P_t \left(X_t^2 - \alpha_0 - \sum_{j=1}^{p} \alpha_j X_{t-j}^2\right)^2.$$

- Localization with a kernel W and a bandwidth b. Framework of the nonparametric estimation of the regression function with fixed design.
- Asymptotically normal estimates for a given choice of weights P_t (e.g $P_t = \left(1 + \sum_{j=1}^{p} X_{t-j}^2\right)^{-2}$).
- A two-step procedure for efficiency: the optimal weights (asymptotic variance) are $P_t = \sigma_t^{-4}$.
- A cross-validation procedure can be used for bandwidth selection.

Partially linear regression

$$X_t^2 = J_t'\alpha\left(\frac{t}{T}\right) + K_t'\beta + \sigma_t^2\left(\xi_t^2 - 1\right).$$

- Step 1: weighted approach

$$\sqrt{P_t}X_t^2 = \sqrt{P_t}J_t'\alpha\left(\frac{t}{T}\right) + \sqrt{P_t}K_t'\beta + \sqrt{P_t}\sigma_t^2\left(\xi_t^2 - 1\right).$$

- Step 2: projection onto $Vect_{\mathbb{L}^2}\left(\sqrt{P_t}J_t\right)$.

$$\mathcal{P}_{1,t} = \sqrt{P_t}J_t'\mathbb{E}^{-1}\left(P_tJ_tJ_t'\right)\mathbb{E}\left(P_tJ_tX_t^2\right),$$

$$\mathcal{P}_{2,t} = \sqrt{P_t}J_t'\mathbb{E}^{-1}\left(P_tJ_tJ_t'\right)\mathbb{E}\left(P_tJ_tK_t'\right).$$

$$\sqrt{P_t}X_t^2 - \mathcal{P}_{1,t} = \left(\sqrt{P_t}K_t - \mathcal{P}_{2,t}\right)'\beta + \sqrt{P_t}\sigma_t^2\left(\xi_t^2 - 1\right).$$

Partially linear regression

$$X_t^2 = J_t'\alpha\left(\frac{t}{T}\right) + K_t'\beta + \sigma_t^2\left(\xi_t^2 - 1\right).$$

- Step 3: projections estimation with a kernel (non-parametric regression), e.g

$$\hat{\mathcal{P}}_{1,t} = \sqrt{P}_t J_t'\left(\sum_{i=p+1}^{T} W_b\left(\frac{t-i}{Tb}\right) P_i J_i J_i'\right)^{-1} \sum_{i=p+1}^{T} W_b\left(\frac{t-i}{T}\right) P_i J_i X_i^2.$$

  $\hat{\beta}$ least squares estimator for β (without localization).

- Step 4: From Step 3, deduce an estimation of the optimal weights $P_t = \frac{1}{\sigma_t^4}$ and use a plug-in estimator.

$$\hat{\alpha}(\frac{t}{T}) = \left(\sum_{i=p+1}^{T} W_b\left(\frac{t-i}{T}\right) P_i J_i J_i'\right)^{-1} \sum_{i=p+1}^{T} W_b\left(\frac{t-i}{T}\right) P_i\left(J_i X_i^2 - J_i K_i'\hat{\beta}\right)$$

Sommaire

Asymptotic normality 1: assumptions

1. $\mathbb{E}\xi_0^{4+\delta} < \infty$,
2. Weights: $P_t = \frac{1}{\left(\gamma_0\left(\frac{t}{T}\right)+\sum_{\ell=1}^{p}\gamma_\ell\left(\frac{t}{T}\right)X_{t-j}^2\right)^2}$. $\gamma_\ell > 0$ are Lipschitz, $\ell = 0, 1, \ldots, p$.
3. The kernel W is continuously differentiable and supported on $[-1, 1]$.
4. the bandwidth parameter b (only one for simplicity):$b\sqrt{T} \to \infty$, $b^2\sqrt{T} \to 0$,
5. Lipschitz coefficients, $a_0 > 0$, $\sup_{u\in[0,1]} \sum_{\ell=1}^{p} a_\ell(u) < 1$.

Asymptotic normality 1

Theorem 1

$$\sqrt{T}\left(\hat{\beta}-\beta\right)\rightarrow_{T\rightarrow\infty}\mathcal{N}_k\left(0, \mathit{Var}\left(\xi_1^2\right)\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}\right),$$

with

$$\Sigma_1=\mathbb{E}\int_0^1 P_1(u)\left(K_1(u)-m(u)'J_1(u)\right)\cdot\left(K_1(u)-m(u)'J_1(u)\right)'du,$$

$$\Sigma_2=\mathit{Var}\left(\xi_1^2\right)\mathbb{E}\int P_1(u)^2\sigma_1(u)^4\left(K_1(u)-m(u)'J_1(u)\right)\cdot\left(K_1(u)-m(u)'J_1(u)\right)'du,$$

and

$$m(u)=\mathbb{E}^{-1}\left(P_1(u)J_1(u)J_1(u)'\right)\mathbb{E}\left(P_1(u)J_1(u)K_1(u)'\right).$$

If parameters are positive, a plug-in approach can be used to obtain a more efficient estimator of β (semiparametric asymptotic efficiency when $\xi_0\sim\mathcal{N}(0,1)$).

Sommaire

Asymptotic normality 2

Theorem 2

If $b' \to 0, \quad Tb' \to \infty$, $\sqrt{Tb'}\left(\hat{\alpha}(u) - \alpha(u) - S_u^{-1} D_u\right)$ *is asymptotically Gaussian with mean* 0 *and variance*

$$\mathit{Var}\left(\xi_1^2\right) \times \int_{-1}^{1} W^2(v)dv \times \Sigma_1(u)^{-1}\Sigma_2(u)\Sigma_1(u)^{-1},$$

with $\Sigma_1(u) = \mathbb{E}\left(P_1(u)J_1(u)J_1(u)'\right), \quad \Sigma_2(u) = \mathbb{E}\left(P_1(u)^2\sigma_1(u)^4 J_1(u)J_1(u)'\right),$

$$S_u = \mathbb{E}\left(P_1(u)J_1(u)J_1(u)'\right),$$

$$D_u = \sum_{i=p+1}^{T} W_{b'}\left(u - \frac{i}{T}\right)\left(P_i J_i X_i^2 - P_i(u)J_i(u)X_i(u)^2\right) + \ldots = O_{\mathbb{P}}(b').$$

Analogue to the asymptotic normality result of Fryzlewicz et al. Possibility to get a more efficient estimation by plug-in (if coefficients of volatility are all positive).

Outline

H_0: $\beta(\cdot)$ is constant v.s $H1$: $\beta(\cdot)$ non-constant.

For weights (P_t), set

$$S_T = \int_0^1 \left(\hat{\beta}(u) - \hat{\beta}\right)' \left(\hat{\beta}(u) - \hat{\beta}\right) du.$$

Notations

- $\mathcal{O}(u) = \kappa_u^{-1}\zeta_u\kappa_u^{-1}$, $\kappa_u = \mathbb{E}\left(P_t(u)\mathcal{X}_t(u)\mathcal{X}_t(u)'\right)$,
  $\zeta_u = \mathbb{E}\left(P_t(u)^2\sigma_t(u)^2\mathcal{X}_t(u)\mathcal{X}_t(u)'\right), \quad \mathcal{X}_t(u) = \begin{pmatrix} J_t(u) \\ K_t(u) \end{pmatrix}.$
- $\omega_n = \int_0^1 \text{Trace}\left[A\mathcal{O}(u)A'\right]^n du$, $A\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \beta$.
- $\|f\|_2^2 = \int_{-1}^1 f(v)^2 dv$, $W^*(x) = \int_{-1}^{1-2|x|} W(v)W(v+2|x|)dv$.

Theorem 3

Assume $\mathbb{E}\xi_0^{8(1+\delta)} < \infty$, $Tb^2 \to \infty$ *and* $Tb^{3.5} \to 0$. *Then*

$$T\sqrt{b}\left\{S_T - \frac{\|W\|_2^2 \mathit{Var}\left(\xi_1^2\right)\omega_1}{Tb}\right\} \to \mathcal{N}\left(0, 4\|W^*\|_2^2 \mathit{Var}^2\left(\xi_1^2\right)\omega_2\right).$$

Testing $H_0 : a_{1:p} = 0$ in sp-tv (constant lag coefficients)

We test a parameter is on the boundary.

- Use the asymptotic behavior of $\sqrt{T}\hat{a}_{1:p}$ and the χ^2 limiting distribution of $T\|\hat{\Sigma}^{-1/2}\hat{a}_{1:p}\|^2$: technically possible (least squares objective function does not require positivity) but unnatural "bilateral" test (loss of power). Truncated least squares cannot give a pivotal statistic.
- Instead, use $P_t = 1$ for which we have,

Proposition 1

If $Tb^2 \to \infty$, $Tb^4 \to 0$, $\mathbb{E}\left(\xi_0^8\right) < \infty$, then under H_0,

$$\hat{a}_{1:p} = \arg\min_a \sum_{t=p+1}^{T} \left(X_t^2 - \hat{\mathcal{P}}_{t,1} - \sum_{j=1}^{p} a_j \left(X_{t-j}^2 - \hat{\mathcal{P}}_{2,t,j} \right) \right)^2$$

satisfies $T\sum_{j=1}^{p} \max\left(\hat{a}_j, 0\right)^2 \to_{\mathcal{D}} \sigma^2 \sum_{j=1}^{p} \max\left(Z_j, 0\right)^2$, *where Z standard Gaussian vector and*

$$\sigma^2 = \frac{\int_0^1 \mathrm{Var}\left(X_0(u)^2\right)^2 du}{\left(\int_0^1 \mathrm{Var}\left(X_0(u)^2\right) du\right)^2}.$$

Outline

Implementation

- We choose $P_t = 1/(\hat{\mu} + \sum_{j=1}^{p} X_{t-j}^2)^2$ with $\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} X_t^2$.
- Bandwidth selection for tv-ARCH: CV (leave one out) used in Fryzlewicz et al.
- Bandwidth selection for sp-ARCH (constant lag coefficients): minimize over a grid of bandwidths,

$$(b, a_{1:p}) \mapsto \sum_{t=p+1}^{T} \left(\sqrt{P_t} X_t^2 - \hat{\mathcal{P}}_{1,t}^{(-t)}(b) - \sum_{j=1}^{p} a_j \left(\sqrt{P_t} X_{t-j}^2 - \hat{\mathcal{P}}_{2,t,j}^{(-t)}(b) \right) \right)^2 .$$

- For statistical testing: get pivotal statistics and use Monte-Carlo simulation to fix the critical value (as in Zhang & Wu, 2012): we simulate B realizations of the pivotal statistic using T i.i.d $\mathcal{N}(0,1)$.
- Selection of order p: minimize $\{0, 1, \ldots, p_{max}\} \mapsto \log(RSS(p)) + \frac{p+1}{T^{2/3}}$.

Example of semiparametric inference

$$X_t = \xi_t \sqrt{a_0(t/T) + 0.3X_{t-1}^2 + 0.3X_{t-2}^2},$$

$$a_0(u) = 0.0001 + 0.006 \cdot (1 + sin(6\pi u)), \quad \xi_0 \sim t(5), \quad T = 1500.$$

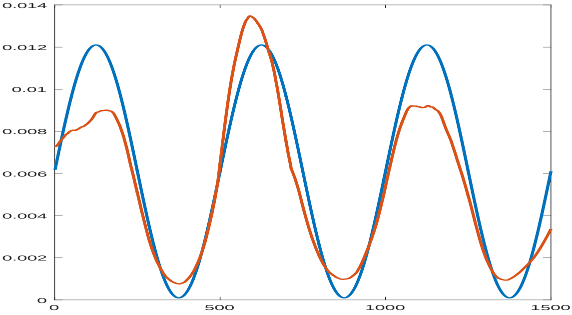


Figure : Estimation of a_0, $\hat{a}_1 = 0.3531$ (s.e 0.051), $\hat{a}_2 = 0.2904$ (s.e 0.0588)

Testing parameter stability, $p = 2$, $T = 2500$, $\alpha = 10\%$

H_0 : a_0 constant (resp. a_1 constant, a_2 constant, $(a_1, a2)$ constant) when

$$a_0(u) = 2\left(1 + \theta \sin(2\pi u)\right), \quad a_1(u) = 0.2 + \frac{\theta}{2}\sin(2\pi u), \quad a_2(u) = 0.2 + \frac{\theta}{2}\cos(2\pi u),$$

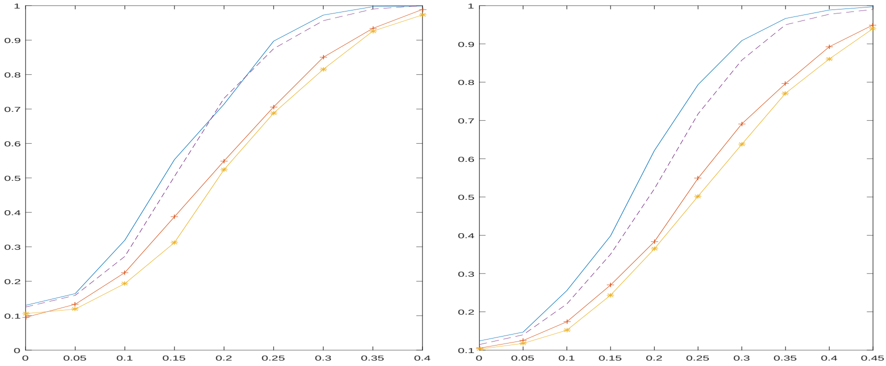


Figure : Power curves when the noise is Gaussian (on the left) or follows a $t-$distribution (on the right). Legend: $-$ for a_0 constant, $--$ for (a_1, a_2) constant, $+$ for a_1 constant and $*$ for a_2 constant.

Testing $H0$: $(a_1, a_2) = (0, 0)$ vs $H1$: $(a_1, a_2) \neq (0, 0)$

$$a_0(u) = 2 + sin(6\pi u), \quad a_1 = a_2 = \theta, \quad T = 1000, \quad \alpha = 10\%.$$

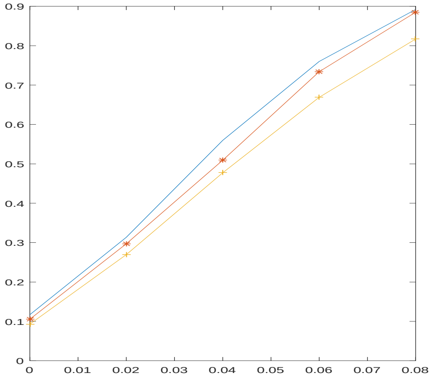
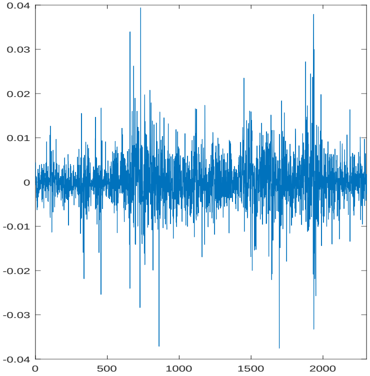
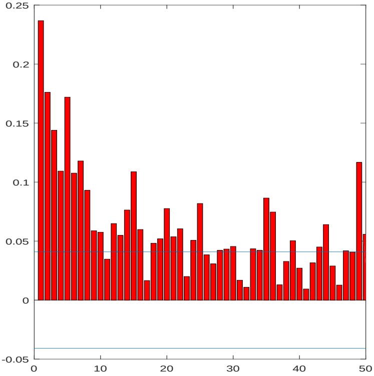


Figure : Power curves: $\xi_0 \sim \mathcal{N}(0, 1)$ (blue), $\xi_0 \sim t(9)$ (red), $\xi_0 \sim t(5)$ (yellow)

Outline

USD/RUP (19/12/2005-18/02/2015), $T = 2303$

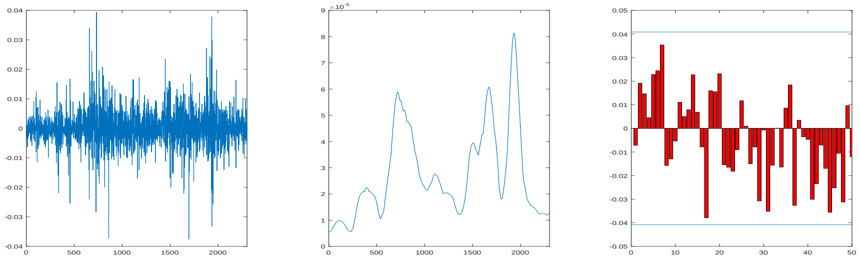




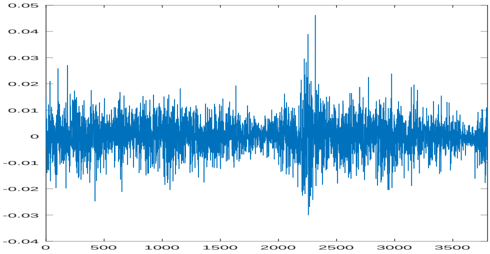
USD/RUP (19/12/2005-18/02/2015), $T = 2303$

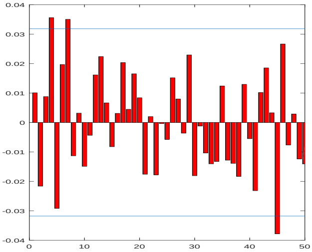
Estimation of p: **1**.

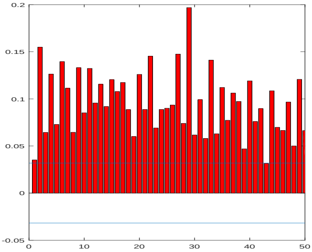
Non t-v a_0	Non t-v a_1	$\hat{b}_{NP}$	$\hat{a}_1$	$\hat{b}_{SP}$
$< 10^{-4}$	0.4215	0.035	0.1527 (s.e 0.0688)	0.028



USD/Euro (03/01/2000-13/02/2015), $T = 3799$



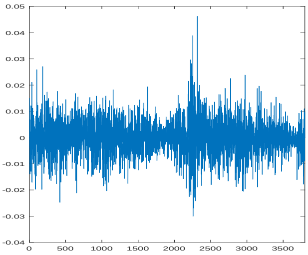


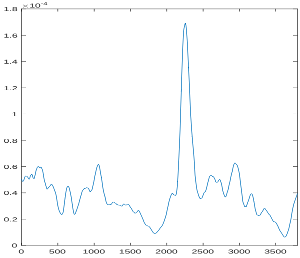


Estimation of p: $\mathbf{0}$.

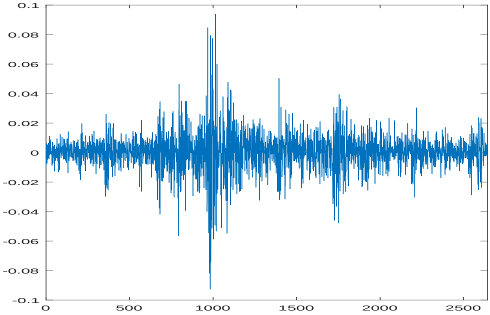
Non t-v a_0	Non t-v a_1	Non t-v a_2	Non t-v (a_1, a_2)	$\hat{b}$
0.0005	0.138	0.1645	0.6415	0.028

Model with non time-varying coefficients does not give significant lag estimates.





.



Estimation of p: 5 (a bit large for accuracy in tv modeling).

$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_3$
0.0547 (s.e 0.0321)	0.1155 (s.e 0.0320)	0.1204 (s.e 0.0311)
$\hat{a}_4$	$\hat{a}_5$	$\hat{b}_{SP}$
0.0942 (s.e 0.0367)	0.1201 (s.e 0.0324)	0.063

The effect of adding a time-varying unconditional variance on the second order dynamic

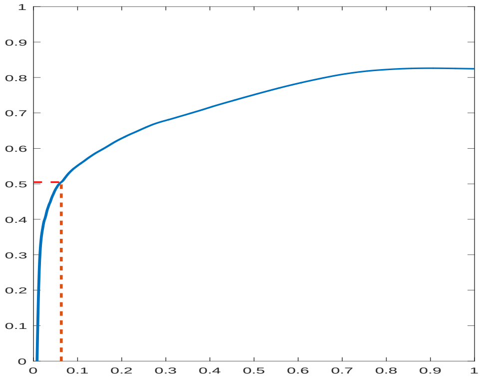


Figure : Sum of the first five lag coefficients with respect to the value of bandwidth

Limitations and directions for future works

- Cross-Validation and hypothesis testing do not have the same focus (bandwidth selection for testing?).
- A global bandwidth could be too restrictive and local bandwidths more adapted.
- Break points are not considered.
- Recent paper in connection with this work: Dalla, Giraitis, Phillips (2016) apply CUSUM type tests for testing mean and variance stability and find that "Changes in the volatility of S&P and IBM returns seems to be initiated by economic events and news, rather than a form of stationary conditional heteroscedasticity. Although short transition periods still might hide GARCH type effects, modeling returns as independent variables with piecewise constant unconditional variance seems to be an attractive alternative."