Prediction of quantiles by statistical learning and application to GDP forecasting

Joint work with Pierre Alquier, CREST and University Paris 7

Xiaoyin LI

Xiaoyin.li@u-cergy.fr

UMR 8088 AGM, University of Cergy-Pontoise
Institut Universitaire de France

Luminy April 2012

Outline

A problem of statistical inference

We observe $X_1, \ldots, X_n$ from a $\mathbb{R}^p$-valued stationnary time series $X = (X_t)_{t \in \mathbb{Z}}$ defined on $(\Omega, \mathcal{A}, \mathbb{P})$.

A problem of statistical inference

We observe $X_1, \ldots, X_n$ from a $\mathbb{R}^p$-valued stationnary time series $X = (X_t)_{t \in \mathbb{Z}}$ defined on $(\Omega, \mathcal{A}, \mathbb{P})$.
We have a family of predictors $\left\{ f_\theta : (\mathbb{R}^p)^k \to \mathbb{R}^p, \theta \in \Theta \right\}$

$$\hat{X}_t^\theta = f_\theta(X_{t-1}, \ldots, X_{t-k}).$$

A problem of statistical inference

We observe $X_1, \ldots, X_n$ from a $\mathbb{R}^p$-valued stationnary time series $X = (X_t)_{t\in\mathbb{Z}}$ defined on $(\Omega, \mathcal{A}, \mathbb{P})$.
We have a family of predictors$\left\{ f_\theta : (\mathbb{R}^p)^k \to \mathbb{R}^p, \theta \in \Theta \right\}$

$$\hat{X}_t^\theta = f_\theta(X_{t-1}, \ldots, X_{t-k}).$$

We consider a model-selection type approach:

$$\Theta = \bigcup_{j=1}^{m} \Theta_j.$$

A problem of statistical inference

We observe $X_1, \ldots, X_n$ from a $\mathbb{R}^p$-valued stationnary time series $X = (X_t)_{t\in\mathbb{Z}}$ defined on $(\Omega, \mathcal{A}, \mathbb{P})$.
We have a family of predictors$\left\{ f_\theta : (\mathbb{R}^p)^k \to \mathbb{R}^p, \theta \in \Theta \right\}$

$$\hat{X}_t^\theta = f_\theta(X_{t-1}, \ldots, X_{t-k}).$$

We consider a model-selection type approach:

$$\Theta = \bigcup_{j=1}^{m} \Theta_j.$$

Objective: find a $\theta \in \Theta$ such that $f_\theta(X_{t-1}, \ldots, X_{t-k})$ is a good prediction of X_t.

Familles of classical predictor

Definition

For any $\theta \in \Theta$,

$$\hat{X}_t^{\theta} = f_{\theta}(X_{t-1}, \ldots, X_{t-k}).$$

Familles of classical predictor

Definition

For any $\theta \in \Theta$,

$$\hat{X}_t^{\theta} = f_{\theta}(X_{t-1}, \ldots, X_{t-k}).$$

Linear auto-regressive class of predictors:

$$f_{\theta}(X_{t-1}, \ldots, X_{t-k}) = \theta_0 + \sum_{j=1}^{k} \theta_j X_{t-j}.$$

Familles of classical predictor

Definition

For any $\theta \in \Theta$,

$$\hat{X}_t^{\theta} = f_{\theta}(X_{t-1}, \ldots, X_{t-k}).$$

Linear auto-regressive class of predictors:

$$f_{\theta}(X_{t-1}, \ldots, X_{t-k}) = \theta_0 + \sum_{j=1}^{k} \theta_j X_{t-j}.$$

Non-parametric auto-regression predictors:

$$f_{\theta}(X_{t-1}, \ldots, X_{t-k}) = \sum_{i=1}^{j} \theta_i \varphi_i(X_{t-1}, \ldots, X_{t-k}).$$

Loss and risk

Let ℓ be a loss function: $\ell(\hat{X}_t^{\theta}, X_t) \geq 0$.

Example

- $\ell(x, x') = \|x - x'\|$.
- $\ell(x, x') = \|x - x'\|^2$.

Loss and risk

Let ℓ be a loss function: $\ell(\hat{X}_t^{\theta}, X_t) \geq 0$.

Example

- $\ell(x, x') = \|x - x'\|$.
- $\ell(x, x') = \|x - x'\|^2$.

The (prevision) risk :

$$R(\theta) = \mathbb{E}\left[\ell\left(\hat{X}_t^{\theta}, X_t\right)\right] \textit{ is unknown}.$$

The empirical risk:

$$r_n(\theta) = \frac{1}{n-k} \sum_{i=k+1}^{n} \ell\left(\hat{X}_i^{\theta}, X_i\right).$$

Outline

Prior and Estimators

Let $\mathcal{M}_+^1(\Theta)$ denote the set of all probability measures on $(\Theta, \mathcal{T})$. Let us take $\pi \in \mathcal{M}_+^1(\Theta)$, the prior.

Prior and Estimators

Let $\mathcal{M}_+^1(\Theta)$ denote the set of all probability measures on $(\Theta, \mathcal{T})$. Let us take $\pi \in \mathcal{M}_+^1(\Theta)$, the prior.

Definition (Gibbs Estimators)

We put, for any $\lambda > 0$,

$$\hat{\theta}_\lambda = \int_\Theta \theta \hat{\rho}_\lambda(\mathrm{d}\theta)$$

where

$$\hat{\rho}_\lambda(\mathrm{d}\theta) = \frac{e^{-\lambda r_n(\theta)}\pi(\mathrm{d}\theta)}{\int e^{-\lambda r_n(\theta')}\pi(\mathrm{d}\theta')}.$$

PAC-Bayesian Bound

The idea is that the risk of the Gibbs estimator will be close to $\inf_\theta R(\theta)$ up to a small remainder. More precisely, we upper-bound it by

$$\inf_\rho \left\{ \int R(\theta)\rho(\mathrm{d}\theta) + \text{ remainder}(\rho, \pi) \right\}$$

where the inf is taken upon all the probability distributions on Θ.

Hypothesis

We assume:

- $(X_t)_{t\in\mathbb{Z}}$ is bounded, *ie* a.s. $\|X_0\|_\infty \leq \mathcal{B} < \infty$

Hypothesis

We assume:

- $(X_t)_{t\in\mathbb{Z}}$ is bounded, *ie* a.s. $\|X_0\|_\infty \leq \mathcal{B} < \infty$
- $\ell(x, x') = g(x - x')$ for some convex function g and g is K-Lipschitz.

Hypothesis

We assume:

- $(X_t)_{t\in\mathbb{Z}}$ is bounded, *ie* a.s. $\|X_0\|_\infty \leq \mathcal{B} < \infty$
- $\ell(x,x') = g(x-x')$ for some convex function g and g is K-Lipschitz.
- any f,

$$\|f_\theta(x_1,\ldots,x_k) - f_\theta(y_1,\ldots,y_k)\| \leq \sum_{j=1}^{k} a_j(\theta)\|x_j - y_j\|.$$

$$L := \sup_{\theta\in\Theta}\sum_{j=1}^{k} a_j(\theta)$$

Hypothesis

We assume:

- $(X_t)_{t\in\mathbb{Z}}$ is bounded, *ie* a.s. $\|X_0\|_\infty \leq \mathcal{B} < \infty$
- $\ell(x, x') = g(x - x')$ for some convex function g and g is K-Lipschitz.
- any f,

$$\left\| f_\theta\left(x_1, \ldots, x_k\right) - f_\theta\left(y_1, \ldots, y_k\right)\right\| \leq \sum_{j=1}^{k} a_j\left(\theta\right)\left\|x_j - y_j\right\|.$$

$$L := \sup_{\theta\in\Theta} \sum_{j=1}^{k} a_j\left(\theta\right)$$

- $k \leq n/2$.

PAC-Bayesian Oracle Inequality

Theorem (Alquier Li 2012)

for any $\lambda > 0$, *for any* $\varepsilon > 0$,

$$\mathbb{P}\left\{ \int R(\hat{\theta}_\lambda) \leq \inf_\rho \left[\int R \mathrm{d}\rho + \frac{2\lambda\kappa_n^2}{n\left(1-\frac{k}{n}\right)^2} + \frac{2\mathcal{K}(\rho,\pi) + 2\log\left(\frac{2}{\varepsilon}\right)}{\lambda} \right] \right\} \geq 1-\varepsilon$$

where $\kappa_n := \sqrt{2}K(1+L)(\mathcal{B}+\theta_{\infty,n}(1))$.

θ weak dependent ceofficient

Introduced by Doukhan, θ coefficient is as follow:

$$\theta_\infty(\mathfrak{S}, Z) = \sup_{h \in \Lambda_1^q} \left\| \mathbb{E}\left[f(Z)|\mathfrak{S}\right] - \mathbb{E}\left[f(Z)\right] \right\|_\infty$$

where

$$\Lambda_1^q = \left\{ f : (\mathbb{R}^p)^q \to \mathbb{R}, \quad \frac{|f(z_1, \dots, z_q) - f(z'_1, \dots, z'_q)|}{\sum_{j=1}^q \|z_j - z'_j\|} \leq 1 \right\},$$

and that

$$\theta_{\infty,k}(1) := \sup_{p<j_1<\dots<j_\ell, 1\leq \ell \leq k} \left\{ \theta_\infty(\sigma(X_t, t \leq p), (X_{j_1}, \dots, X_{j_\ell})) \right\}.$$

Some preliminary lemmas

Lemma (Rio,2000)

Let h be a function $(\mathbb{R}^p)^n \to \mathbb{R}$ such that h is $1-$Lipschitz. Then for any $t > 0$ we have

$$\mathbb{E}\left(e^{t\{\mathbb{E}[h(X_1,\ldots,X_n)]-h(X_1,\ldots,X_n)\}}\right) \leq e^{\frac{t^2 n\left(\mathcal{B}+\theta_{\infty,n}(1)\right)^2}{2}}.$$

Note that $h(X_1, \ldots, X_n) = \frac{n-k}{K(1+L)} r_n(\theta)$ and we choose $t = K(1+L)\lambda/(n-k)$, we obtain:

$$\mathbb{E}\left(e^{\lambda[R(\theta)-r_n(\theta)]}\right) \leq e^{\frac{\lambda^2 \kappa^2}{n\left(1-\frac{k}{n}\right)^2}}.$$

Some proofs

$$\mathbb{E}\left(e^{\lambda\left[R(\theta)-r_n(\theta)-\frac{\lambda\kappa^2}{n\left(1-\frac{k}{n}\right)^2}\right]}\right)\leq 1,$$

$$\mathbb{E}\left(\int e^{\lambda[R(\theta)-r_n(\theta)]-\frac{\lambda^2\kappa^2}{n\left(1-\frac{k}{n}\right)^2}-\log\left(\frac{2}{\varepsilon}\right)}\right)\leq \frac{\varepsilon}{2}.$$

$$\mathbb{E}\left(e^{\sup_\rho\left\{\lambda\int[R(\theta)-r_n(\theta)]\rho(\mathrm{d}\theta)-\frac{\lambda^2\kappa^2}{n\left(1-\frac{k}{n}\right)^2}-\log\left(\frac{2}{\varepsilon}\right)-\mathcal{K}(\rho,\pi)\right\}}\right)\leq \frac{\varepsilon}{2}.$$

$$\mathbb{P}\left\{\sup_\rho\left\{\lambda\int[R(\theta)-r_n(\theta)]\,\rho(\mathrm{d}\theta)-\frac{\lambda^2\kappa^2}{n\left(1-\frac{k}{n}\right)^2}-\log\left(\frac{2}{\varepsilon}\right)-\mathcal{K}(\rho,\pi)\right\}\geq 0\right\}\leq\frac{\varepsilon}{2}.$$

Outline

The context

Objective: at each quarter t, predict the flash estimate of GDP growth: $\Delta\mathrm{GDP}_t$.

Available information:

- $\Delta\mathrm{GDP}_{t'}$, for all $t' < t$
- $I_{t'}$, for all $t' < t$, I_{t-1} is the climate indicator available to the INSEE at time t.
- The observation period is 1988-Q1 (1st quarter of 1988) to 2011-Q3.

Quntile loss function

We define $X_t = (\Delta\mathrm{GDP}_t, I_t)' \in \mathbb{R}^2$. Following Cornec(CIRET conference 2010), we consider predictors of the form:

$$f_\theta(X_{t-1}, X_{t-2}) = \theta_0 + \theta_1 \Delta\mathrm{GDP}_{t-1} + \theta_2 I_{t-1} + \theta_3 (I_{t-1} - I_{t-2})|I_{t-1} - I_{t-2}|$$

These family of predictors allow to obtain a forecasting as precise as the INSEE one.

We use the quantile loss function :

$$\ell_\tau((\Delta\mathrm{GDP}_t, I_t), (\Delta'\mathrm{GDP}_t, I'_t)) = \begin{cases} \tau\left(\Delta\mathrm{GDP}_t - \Delta'\mathrm{GDP}_t\right), & \text{if } \Delta\mathrm{GDP}_t - \Delta'\mathrm{GDP}_t > 0 \\ -(1-\tau)\left(\Delta\mathrm{GDP}_t - \Delta'\mathrm{GDP}_t\right), & \text{otherwise.} \end{cases}$$

Results: GDP forecasting

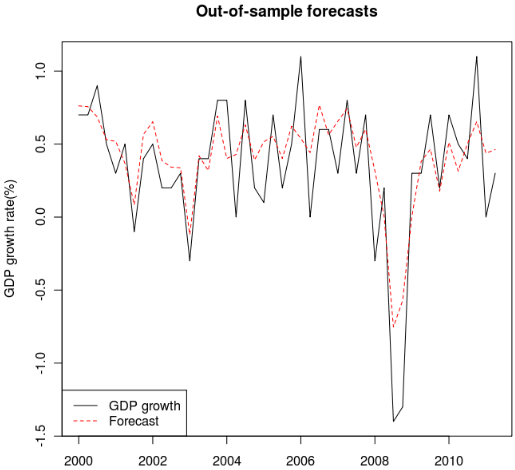
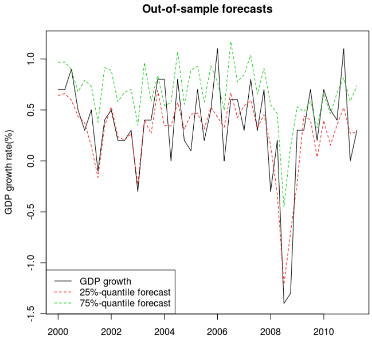


Figure: French GDP online prediction using the quantile loss function with $\tau = 0.5$.

Results: Confidence intervals



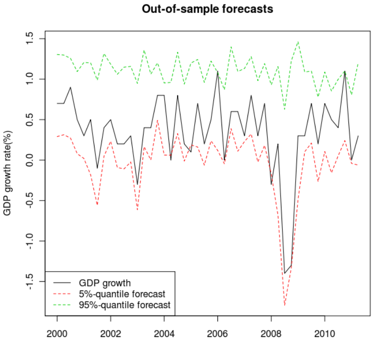


Figure: French GDP online 50%-confidence intervals (left) and 90%-confidence intervals (right).

Conclusion

- We proposed some theoretical results to extend learning theory to the context of weakly dependent time series. The method showed good results on an application to GDP forecasting.
- It would now be interesting to extend this result to a more general class of processes, some other weakly dependent ones.

Merci!