Inference on breaks in weak location time series models with quasi-Fisher scores

Christian Francq, Lorenzo Trapani and Jean-Michel Zakoïan

1ères Journées d'Analyse des Séries Temporelles

Valenciennes, June 24, 2025

Objective

- The Estimating Function approach (Godambe (1960), Durbin (1960)) can be used to estimate weak location scale models (FZ, 2023).
  - revival in the last 20 years*
  - particularly efficient for weak time series models
  - based on solving estimating equations: $\boldsymbol{h}_n(\boldsymbol{\theta}) = \mathbf{0}$.
- **Aim:** use the EF approach to detect breaks in the conditional mean when the demeaned process may not be iid.
  - Offline detection based on CUSUM statistics† and one estimation of the (semi) parametric model.

*Book by Heyde (2008), surveys by Bera et al. (2006), Jacod and Sørensen (2018).

†see Horváth and Parzen (1994) CUSUM of Fisher's score; Aue and Horváth (2013) CUSUM of QMLE quasi-score for detecting breaks in conditional mean and variance; Horváth and Rice (2023) *Change point detection in time series.*

Objective

- The Estimating Function approach (Godambe (1960), Durbin (1960)) can be used to estimate weak location scale models (FZ, 2023).
  - revival in the last 20 years*
  - particularly efficient for weak time series models
  - based on solving estimating equations: $\boldsymbol{h}_n(\boldsymbol{\theta}) = \mathbf{0}$.

- **Aim:** use the EF approach to detect breaks in the conditional mean when the demeaned process may not be iid.
  - Offline detection based on CUSUM statistics† and one estimation of the (semi) parametric model.

*Book by Heyde (2008), surveys by Bera et al. (2006), Jacod and Sørensen (2018).

†see Horváth and Parzen (1994) CUSUM of Fisher's score; Aue and Horváth (2013) CUSUM of QMLE quasi-score for detecting breaks in conditional mean and variance; Horváth and Rice (2023) *Change point detection in time series*.

Parametric model for the conditional mean

Consider a real time series $(y_t)_{t\in\mathbb{Z}}$ and $\mathcal{F}_t = \sigma\{y_u : u \leq t\}$.

Write $E_t(\cdot) = E(\cdot \mid \mathcal{F}_t)$ and assume

$$m_t = m_t(\boldsymbol{\theta}_0) := E_{t-1}(y_t)$$

exists and depends on some parameter $\boldsymbol{\theta}_0 \in \Theta \subset \mathbb{R}^d$.

No specific assumptions on other conditional moments except the existence of a conditional variance.

Let $\epsilon_t = y_t - m_t$. This location model is said to be

weak as (ϵ_t) may not be an iid sequence.

Fisher and quasi-Fisher scores

If the conditional distribution of y_t depends on a time-varying parameter $m_t(\boldsymbol{\theta})$, Fisher's score is

$$\boldsymbol{h}_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} \frac{\partial}{\partial \boldsymbol{\theta}} \log f(y_t; \boldsymbol{\theta}, m_t(\boldsymbol{\theta})).$$

- If $\boldsymbol{h}_n$ is the Fisher score, the MLE $\widehat{\boldsymbol{\theta}}$ solves $\boldsymbol{h}_n(\widehat{\boldsymbol{\theta}}) = \mathbf{0}$.
- If $\boldsymbol{h}_n$ is a quasi-score, a solution of $\boldsymbol{h}_n(\widehat{\boldsymbol{\theta}}) = \mathbf{0}$ is called QMLE.

For a more general estimating function $\boldsymbol{h}_n$, a solution of $\boldsymbol{h}_n(\boldsymbol{\theta}) = \mathbf{0}$ is called Quasi-Likelihood Estimator (QLE) or Z-estimator.

An EF is called unbiased if $E\boldsymbol{h}_n(\boldsymbol{\theta}_0) = \mathbf{0}$.

A class of EF for the weak location model

Durbin and Godambe's theory of optimal unbiased EFs:

- extends the theory of unbiased estimation (BLUE) to EF;
- leads to a finite sample optimality concept. more on that theory

Godambe (1985) (see also Chandra and Taniguchi, 2001) showed that, within the class of the unbiased EFs of the form

$$\sum_{t=1}^{n} \boldsymbol{a}_{t-1}(\boldsymbol{\theta}) \left\{ y_t - m_t(\boldsymbol{\theta}) \right\},$$

$\boldsymbol{a}_{t-1}(\boldsymbol{\theta}) \in \mathcal{F}_{t-1}$ is a $d \times 1$ vector, an optimal EF in Godambe's sense is

$$\sum_{t=1}^{n} \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{1}{\sigma_t^2(\boldsymbol{\theta})} \left\{ y_t - m_t(\boldsymbol{\theta}) \right\}$$

where $\sigma_t^2(\boldsymbol{\theta})$ is the conditional variance (generally unknown, may depend on nuisance parameters).

Optimal EF in the strong case

Notation: If $X_t \in \mathcal{F}_t = \sigma(y_u, u < t)$ then $\widetilde{X}_t \in \mathcal{I}_t$ where $\mathcal{I}_t = \sigma(y_u, 1 \leq u < t)$ is the information available at t.

For a strong location model or when σ_t^2 is constant, the optimal EF is

$$\sum_{t=1}^{n} \frac{\partial \widetilde{m}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \widetilde{\epsilon}_t(\boldsymbol{\theta})$$

where $\widetilde{\epsilon}_t(\boldsymbol{\theta}) = y_t - \widetilde{m}_t(\boldsymbol{\theta})$, and the LS estimator is optimal among the QLEs.

A class of EF for the weak location model

The parameter $\boldsymbol{\theta}_0$ is estimated by solving

$$\sum_{t=1}^{n} \frac{\partial \widetilde{m}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\widetilde{\epsilon}_t(\boldsymbol{\theta})}{\widetilde{\kappa}_{2t}} = 0, \quad \widetilde{\epsilon}_t(\boldsymbol{\theta}) = y_t - \widetilde{m}_t(\boldsymbol{\theta}),$$

where

$$\widetilde{\kappa}_{2t} = \widetilde{\kappa}_{2t}(\boldsymbol{\theta}, \widehat{\gamma}_n) \text{ is a proxy of } \sigma_t^2(\boldsymbol{\theta}) = E_{t-1}\epsilon_t^2(\boldsymbol{\theta}),$$

with $\epsilon_t(\boldsymbol{\theta}) = y_t - m_t(\boldsymbol{\theta})$ and $\widehat{\gamma}_n$ a nuisance parameter estimate.

Examples of QLEs that are QMLEs

- If we assume $\widetilde{\kappa}_{2t} \propto m_t$ (with $m_t(\cdot) > 0$), the EE is

$$\sum_{t=1}^{n} \frac{\partial \widetilde{m}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{1}{\widetilde{m}_t(\boldsymbol{\theta})} \epsilon_t(\boldsymbol{\theta}) = 0.$$

The solution is the Poisson QMLE (even when $y_t \notin \mathbb{N}$):

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \sum_{t=1}^{n} y_t \log \widetilde{m}_t(\boldsymbol{\theta}) - \widetilde{m}_t(\boldsymbol{\theta}).$$

- If $\widetilde{\kappa}_{2t} \propto m_t^2$, then we end up with the EE

$$\sum_{t=1}^{n} \frac{\partial \widetilde{m}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{1}{\widetilde{m}_t^2(\boldsymbol{\theta})} \epsilon_t(\boldsymbol{\theta}) = 0,$$

and, when $\widetilde{m}_t(\boldsymbol{\theta}) > 0$, the solution is the exponential QMLE:

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{t=1}^{n} y_t / \widetilde{m}_t(\boldsymbol{\theta}) + \log \widetilde{m}_t(\boldsymbol{\theta}).$$

Example where the QLE is a new estimator

We can also consider the EE

$$\sum_{t=1}^{n} \frac{\partial \widetilde{m}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \widetilde{m}_t(\boldsymbol{\theta}) \epsilon_t(\boldsymbol{\theta}) = \mathbf{0}.$$

Solving this equation amounts to optimizing the objective function

$$\sum_{t=1}^{n} \widetilde{m}_t^2(\boldsymbol{\theta}) \left(\frac{\widetilde{m}_t(\boldsymbol{\theta})}{3} - \frac{y_t}{2} \right),$$

which does not seem to correspond to any standard criterion.

Case where the QLE is the MLE

Assume that the distribution of y_t given $\mathcal{F}_{t-1}$ belongs to the one-parameter exponential family, i.e. admits a density of the form

$$g_{m_t}(y) = k(y) \exp\left\{\eta(m_t)y - a(m_t)\right\},$$

for $k(\cdot) > 0$ and twice differentiable functions $\eta(\cdot)$ and $a(\cdot)$.

It is known that $\eta'(m_t) = a'(m_t)/m_t = 1/\sigma_t^2$. Thus

$$\frac{\partial \log g_{m_t(\boldsymbol{\theta})}(y_t)}{\partial \boldsymbol{\theta}} = \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\epsilon_t(\boldsymbol{\theta})}{\sigma_t^2(\boldsymbol{\theta})}.$$

The QLE is thus the MLE (only approximately when $m_t \neq \widetilde{m}_t$).

Intuition for CUSUM of quasi-scores

We use cumulative sums (CUSUM) of quasi-scores: if

$$\boldsymbol{h}_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} \boldsymbol{\Upsilon}_t(\boldsymbol{\theta}),$$

the Quasi-Likelihood Estimator (QLE) solves the Estimating Equation (EE)

$$\sum_{t=1}^{n} \boldsymbol{\Upsilon}_t(\widehat{\boldsymbol{\theta}}) = 0.$$

If $\{\boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_0)\}_t$ is stationary (no break) then a statistic like

$$\max_{k=1,\dots,n} \left| \sum_{t=1}^{k} \boldsymbol{\Upsilon}_t(\widehat{\boldsymbol{\theta}}) \right|$$

should not be too large (note that $\boldsymbol{\theta}$ is estimated once).

Consistency and Asymptotic Normality (CAN)

CAN of the QLEs (FZ, 2023)

Under regularity conditions A1-A8, for n large enough there exists a QLE $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}_0$ solving

$$\sum_{t=1}^{n} \widetilde{\boldsymbol{\Upsilon}}_t(\widehat{\boldsymbol{\theta}}) = \mathbf{0}, \quad \widetilde{\boldsymbol{\Upsilon}}_t(\boldsymbol{\theta}) = \frac{\partial \widetilde{m}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\widetilde{\epsilon}_t(\boldsymbol{\theta})}{\widetilde{\kappa}_{2t}(\boldsymbol{\theta})}.$$

Moreover, $\widehat{\boldsymbol{\theta}} \to \boldsymbol{\theta}_0$ a.s. as $n \to \infty$, and

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \overset{o_P(1)}{=} -\boldsymbol{J}^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}\right).$$

Asymptotic variance of the QLEs

Optimal QLEs in the asymptotic sense[‡]

The asymptotic variance is $\boldsymbol{\Sigma} = \boldsymbol{J}^{-1}\boldsymbol{I}\boldsymbol{J}^{-1}$ with

$$\boldsymbol{J} = E\left(\frac{-1}{\kappa_{2t}(\boldsymbol{\theta}_0)}\frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^\top}\right), \quad \boldsymbol{I} = E\left(\frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\kappa_{2t}^2(\boldsymbol{\theta}_0)}\frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^\top}\right).$$

If $\kappa_{2t}(\boldsymbol{\theta}_0) \propto \sigma_t^2(\boldsymbol{\theta}_0)$, then the asymptotic variance of the QLE

$$\boldsymbol{\Sigma}_{op} = \left\{E\frac{1}{\sigma_t^2(\boldsymbol{\theta}_0)}\frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^\top}\right\}^{-1},$$

is optimal in the sense that $\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_{op}$ is positive definite.

‡Godambe's sense of optimality is non-asymptotic

Test for breaks in the conditional mean

Assuming $E_{t-1}(y_t) = m_t(\boldsymbol{\theta}_t)$, where $\boldsymbol{\theta}_t \in \Theta$, we consider testing

$$\mathbf{H_0}: \quad \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \cdots = \boldsymbol{\theta}_n$$

against the alternative of at least one unknown breakpoint.
Inspired by CUSUM statistics used in changepoint problems, we consider the quasi-score process

$$\widetilde{\boldsymbol{T}}_n(u) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nu]} \widetilde{\boldsymbol{\Upsilon}}_t(\widehat{\boldsymbol{\theta}}), \qquad u \in [0, 1].$$

Note that $\widetilde{\boldsymbol{T}}_n(0) = 0$ and $\widetilde{\boldsymbol{T}}_n(1) = 0$.

Test for breaks in the conditional mean

A natural statistic for testing $\mathbf{H_0}$ is

$$\widetilde{S}_n = \sup_{u\in(0,1)} \widetilde{S}_n(u) = \max_{k\in\{1,\dots,n-1\}} \widetilde{S}_n(k/n),$$

where

$$\widetilde{S}_n(u) = \widetilde{\boldsymbol{T}}_n^\top(u)\boldsymbol{I}_n^{-1}\widetilde{\boldsymbol{T}}_n(u)$$

and $\boldsymbol{I}_n$ is a consistent estimator of

$$\boldsymbol{I} = E\boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_0)\boldsymbol{\Upsilon}_t^\top(\boldsymbol{\theta}_0), \qquad \boldsymbol{\Upsilon}_t(\boldsymbol{\theta}) = \frac{\partial m_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\epsilon_t(\boldsymbol{\theta})}{\kappa_{2t}(\boldsymbol{\theta})}.$$

Asymptotic behavior of the test statistic under the null

Under the previous assumptions, including $\mathbf{H_0}$, we have

$$\widetilde{S}_n \xrightarrow{d} S = \sup_{u \in (0,1)} \sum_{j=1}^{d} \{B_j(u)\}^2,$$

where $B(u) = (B_1(u), \ldots, B_d(u))^\top$ is a d-dimensional standard Brownian bridge.

At the nominal level $\alpha \in (0,1)$, critical region:

$$\left\{ \max_{1 \leq k \leq n} \widetilde{S}_n(k/n) > S_{1-\alpha} \right\}.$$

Alternative Nyblom test

The Nyblom-type test (based on Nyblom (1989)) rejects the parameter constancy for large values of

$$\widetilde{S}_n^N := \frac{1}{n}\sum_{k=1}^{n}\widetilde{S}_n(k/n)$$

which, by the continuous mapping theorem, has the asymptotic distribution $\int_0^1 \sum_{j=1}^d \{B_j(u)\}^2 du$ under $\mathbf{H_0}$.

- Enjoys some optimality properties under the alternative that the parameter process follows a martingale.
- The CUSUM test also has optimality properties, but for different types of alternatives (see Horváth and Rice, 2023).

Optimality of the QLE for testing?

The QLE with $\kappa_{2t}(\boldsymbol{\theta}_0)$ proportional to $\sigma_t^2(\boldsymbol{\theta}_0)$ is optimal within the class of EF estimators solving

$$\sum_{t=1}^{n} \boldsymbol{a}_{t-1}(\boldsymbol{\theta})\widetilde{\epsilon}_t(\boldsymbol{\theta}) = 0,$$

where $\boldsymbol{a}_{t-1}(\boldsymbol{\theta})$ is a $d \times 1$ vector belonging to $\mathcal{F}_{t-1}$.

Does this Godambe's optimal QLE lead to optimal tests?

We consider local asymptotic powers.

Example of "local breaks"

Let $u_0 \in (0,1)$. Assume $y_1, \dots, y_n$ are independent and Gaussian with variance σ^2, and that $y_t = y_{t,n}$ has mean

- $\theta_0 + \delta_1/\sqrt{[nu_0]}$ when $t \leq [nu_0]$;
- $\theta_0 + \delta_2/\sqrt{n - [nu_0]}$ when $t > [nu_0]$.

We then have

$$\frac{1}{\sqrt{[nu_0]}} \sum_{t=1}^{[nu_0]} (y_t - \theta_0) \sim \mathcal{N}\left(\delta_1, \sigma^2\right),$$

$$\frac{1}{\sqrt{n - [nu_0]}} \sum_{t=[nu_0]+1}^{n} (y_t - \theta_0) \sim \mathcal{N}\left(\delta_2, \sigma^2\right).$$

Example of "local breaks" (continued)

In this simple example, $\overline{y} = n^{-1} \sum_{t=1}^{n} y_t$ is the Q(M)LE of θ_0 (under the null $\delta_1 = \delta_2 = 0$ of no local break),

$$\widetilde{T}_n(u) = n^{-1/2} \sum_{t=1}^{[nu]} (y_t - \overline{y})$$

is the usual CUSUM process, and

$$\widetilde{S}_n = \sup_{u \in (0,1)} \frac{1}{n\widehat{\sigma}_y^2} \left\{ \sum_{t=1}^{[nu]} (y_t - \overline{y}) \right\}^2, \quad \widehat{\sigma}_y^2 = \frac{1}{n} \sum_{t=1}^{n} (y_t - \overline{y})^2,$$

is nothing else than the squared Kolmogorov-Smirnov test statistic.

General situation

Let a single break located at a fixed proportion u_0 of the observations, and

- $\widehat{\boldsymbol{\theta}}_{(1)}$ the QLE computed on $y_1, \dots, y_{[u_0 n]}$
- $\widehat{\boldsymbol{\theta}}_{(2)}$ the QLE computed on $y_{[u_0 n]+1}, \dots, y_n$
- $\widehat{\boldsymbol{\theta}}$ the QLE computed on $y_1, \dots, y_n$.

Let the local alternatives $H_{1,n}(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)$ such that

$$\sqrt{n u_0}\left(\widehat{\boldsymbol{\theta}}_{(1)} - \boldsymbol{\theta}_0\right) \xrightarrow{d} \mathcal{N}\left(\boldsymbol{\delta}_1, \boldsymbol{J}^{-1}\boldsymbol{I}\boldsymbol{J}^{-1}\right),$$
$$\sqrt{n(1-u_0)}\left(\widehat{\boldsymbol{\theta}}_{(2)} - \boldsymbol{\theta}_0\right) \xrightarrow{d} \mathcal{N}\left(\boldsymbol{\delta}_2, \boldsymbol{J}^{-1}\boldsymbol{I}\boldsymbol{J}^{-1}\right).$$

Local Asymptotic Power (LAP) of the tests

Under $H_{1,n}(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)$ and regularity conditions, for all $u \in (0, 1)$

$$\frac{\widetilde{S}_n(u)}{u(1-u)} \xrightarrow{d} \chi^2(d, \lambda)$$

where the noncentrality parameter is

$$\lambda = \frac{1}{u(1-u)} \boldsymbol{\delta}_{u_0}^{\top}(u) \boldsymbol{J} \boldsymbol{I}^{-1} \boldsymbol{J} \boldsymbol{\delta}_{u_0}(u),$$

When $\sqrt{1-u_0}\boldsymbol{\delta}_1 \neq \sqrt{u_0}\boldsymbol{\delta}_2$, we have $\lambda \neq 0$ and
the best LAP is obtained for the optimal QLE.

Comparing powers of alternative tests

Let us test for the existence of a local break in the mean of a sequence of independent Gaussian variables.

Consider 3 tests which reject for large values of $\widetilde{S}_n$, $\widetilde{S}_n^N$ and $\widetilde{S}_n^W$:

$$\widetilde{S}_n = \max_{1 \le k < n} \widetilde{S}_n\left(\frac{k}{n}\right), \quad \widetilde{S}_n^N = \frac{1}{n}\sum_{k=1}^{n} \widetilde{S}_n\left(\frac{k}{n}\right)$$

and

$$\widetilde{S}_n^W = \max_{1 \le k < n} \frac{n^2}{k(n-k)} \widetilde{S}_n\left(\frac{k}{n}\right)$$

with $\widetilde{S}_n(k/n) = \left\{\sum_{t=1}^{k}(y_t - \overline{y})\right\}^2 / (n\widehat{\sigma}_y^2)$.

Numerical illustration

50,000 independent replications with $n = 1,000$; nominal level $\alpha = 1\%$, $\delta_1 = -\delta_2 = 3$

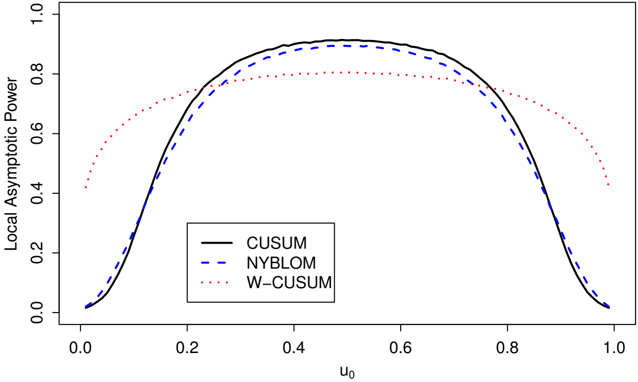


Figure: Powers of the CUSUM, Nyblom, and Weighted CUSUM tests as a function of the break date u_0.

Searching for the optimal QLE (and thus for an optimal test)

There are as many QLEs as there are choices of the weighting sequence $\widetilde{\kappa}_{2t}$. Under regularity conditions, all these QLEs are consistent, but their performance depends on the chosen weights.

In practice, two situations:

1. The model at hand suggests one or several possible values of $\widetilde{\kappa}_{2t}$, which can be selected from the data.
2. The statistician has no idea of a reasonable $\widetilde{\kappa}_{2t}$.

In case 1, we suggest minimizing an empirical QLIK loss. In case 2, we suggest using GARCH-type estimators.

Examples with "natural" weights

- For count time series—the benchmark model being the Poisson INGARCH—a natural choice is $\kappa_{2t}(\boldsymbol{\theta}) = m_t(\boldsymbol{\theta})$.
- If one believes in a standard additive model, such as an ARMA, it is natural to consider constant weights $\kappa_{2t}(\cdot) = 1$.
- For positive data, such as durations or volumes, Multiplicative Error Models (MEM) being often used, it is natural to consider the weights $\kappa_{2t}(\boldsymbol{\theta}) = m_t^2(\boldsymbol{\theta})$.

In practice, the DGP is obviously unknown:
$\Longrightarrow$ data driven procedure for choosing between several weighting schemes.

Optimal theoretical QLIK

For a stationary weighting sequence $\{\kappa_{2t}(\boldsymbol{\theta})\}$, let the theoretical Quasi-Likelihood loss function (QLIK)

$$\mathsf{QLIK}(\kappa_{2t}(\boldsymbol{\theta})) = \min_{c>0} E\left\{ \frac{\{y_t - m_t(\boldsymbol{\theta})\}^2}{c\kappa_{2t}(\boldsymbol{\theta})} + \log\left(c\kappa_{2t}(\boldsymbol{\theta})\right) \right\}.$$

Note that

$$\sigma_t^2(\boldsymbol{\theta}_0) = \arg\min_{\kappa_2 \in \mathcal{F}_{t-1}} \mathsf{QLIK}(\kappa_2).$$

Weights can be selected by minimizing the empirical QLIK over a finite set of potential weighting sequences.

Minimizing the empirical QLIK "loss"

For a set of weighting sequences, $\left\{\widetilde{\kappa}_{2t}^{(i)}(\boldsymbol{\theta}), i \in \{1, \dots, I\}\right\}$, weights are selected by minimizing over i the empirical QLIK loss function

$$\mathrm{QLIK}_n\left(\widetilde{\kappa}_{2\cdot}^{(i)}(\widehat{\boldsymbol{\theta}})\right) = \frac{1}{n}\sum_{t=1}^{n}\left\{\frac{\widetilde{\epsilon}_t^2(\widehat{\boldsymbol{\theta}})}{\widehat{c}_n^{(i)}\widetilde{\kappa}_{2t}^{(i)}(\widehat{\boldsymbol{\theta}})} + \log\left(\widehat{c}_n^{(i)}\widetilde{\kappa}_{2t}^{(i)}(\widehat{\boldsymbol{\theta}})\right)\right\},$$

$$\widehat{c}_n^{(i)} = \frac{1}{n}\sum_{t=1}^{n}\frac{\widetilde{\epsilon}_t^2(\widehat{\boldsymbol{\theta}})}{\widetilde{\kappa}_{2t}^{(i)}(\widehat{\boldsymbol{\theta}})},$$

where $\widehat{\boldsymbol{\theta}}$ is a first step estimator of $\boldsymbol{\theta}_0$. ▸ consistency of the method

GARCH estimation of the optimal weights

If there is no natural set of candidate weights, a simple solution consists in estimating the conditional variance

$$\sigma_t^2(\boldsymbol{\theta}_0) = E\left(\epsilon_t^2(\boldsymbol{\theta}_0) \mid \mathcal{F}_{t-1}\right)$$

by fitting a GARCH-type model on the sequence $\{\widetilde{\epsilon}_1(\widehat{\boldsymbol{\theta}}), \dots, \widetilde{\epsilon}_n(\widehat{\boldsymbol{\theta}})\}$, where $\widehat{\boldsymbol{\theta}}$ is a first step (in general non optimal) estimator of $\boldsymbol{\theta}_0$.

For instance, fitting a simple GARCH(1,1) by QMLE leads to a weighting sequence of the form

$$\widetilde{\kappa}_{2t} = \widehat{\omega} + \widehat{\alpha}\widetilde{\epsilon}_{t-1}^2(\widehat{\boldsymbol{\theta}}) + \widehat{\beta}\widetilde{\kappa}_{2,t-1}.$$

Other GARCH-type estimation of the optimal weights

In order to allow weights that are proportional to the conditional mean, or its square, we can also fit GARCH-X models by QMLE, as

$$\widetilde{\kappa}_{2t} = \widehat{\omega} + \widehat{\alpha}\widetilde{\epsilon}_{t-1}^2(\widehat{\boldsymbol{\theta}}) + \widehat{\beta}\widetilde{\kappa}_{2,t-1} + \widehat{\pi}_1|\widetilde{m}_t(\widehat{\boldsymbol{\theta}})|$$

or

$$\widetilde{\kappa}_{2t} = \widehat{\omega} + \widehat{\alpha}\widetilde{\epsilon}_{t-1}^2(\widehat{\boldsymbol{\theta}}) + \widehat{\beta}\widetilde{\kappa}_{2,t-1} + \widehat{\pi}_1|\widetilde{m}_t(\widehat{\boldsymbol{\theta}})| + \widehat{\pi}_2\widetilde{m}_t^2(\widehat{\boldsymbol{\theta}}).$$

Remark: Not necessary to be very accurate when estimating $\widetilde{\kappa}_{2t}$.

Change-point estimation

Assume that, for $u_0 \in (0, 1]$

$$y_t = y_{t,n} = \begin{cases} m_t(\boldsymbol{\theta}_1) & \text{if } t \leq [nu_0] \\ m_t(\boldsymbol{\theta}_2) & \text{if } t > [nu_0] \end{cases} + \epsilon_t,$$

where (ϵ_t) is such that $E_{t-1}(\epsilon_t) \equiv 0$.

Assume there exist stationary processes, $(y_t^{(1)})_{t\in\mathbb{Z}}$ and $(y_t^{(2)})_{t\in\mathbb{Z}}$, approximating the observed process before and after the break, respectively.

For all $\boldsymbol{\theta} \in \Theta$, let

$$m_t^{(i)}(\boldsymbol{\theta}) = m(\boldsymbol{\theta}; y_{t-1}^{(i)}, y_{t-2}^{(i)}, \dots), \quad \kappa_{2t}^{(i)}(\boldsymbol{\theta}) = \kappa_2(\boldsymbol{\theta}; y_{t-1}^{(i)}, y_{t-2}^{(i)}, \dots)$$

be stationary approximations of the conditional mean and weight sequence before and after the break.

The QLE converges to a pseudo-true value

It can be shown that, under general conditions, $\widehat{\boldsymbol{\theta}}$ converges to the assumed unique solution $\boldsymbol{\theta}_0^\star = \boldsymbol{\theta}_0^\star(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ of the equation

$$u_0 E\left\{\boldsymbol{\Upsilon}_t^{(1)}(\boldsymbol{\theta})\right\} + (1-u_0) E\left\{\boldsymbol{\Upsilon}_t^{(2)}(\boldsymbol{\theta})\right\} = 0,$$

where

$$\boldsymbol{\Upsilon}_t^{(i)}(\boldsymbol{\theta}) = \frac{\partial m_t^{(i)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{y_t^{(i)} - m_t^{(i)}(\boldsymbol{\theta})}{\kappa_{2t}^{(i)}(\boldsymbol{\theta})}.$$

The break fraction is consistently estimated

Let the change-point estimator

$$\widetilde{k} = \arg\max_{k \in \{1,\ldots,n-1\}} \widetilde{S}_n(k/n), \qquad \widetilde{S}_n(u) = \widetilde{\boldsymbol{T}}_n^{\top}(u) \boldsymbol{I}_n^{-1} \widetilde{\boldsymbol{T}}_n(u).$$

Under regularity conditions, when $u_0 \in (0,1)$ and $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$ we have

$$\frac{\widetilde{k}}{n} \to u_0, \quad \text{in probability as } n \to \infty.$$

Case where $m_t(\cdot)$ is misspecified

The intuition is that, even if the conditional mean is not correctly specified, its estimated value should not vary too much when the DGP is stable.

Let $\boldsymbol{\Upsilon}_t(\boldsymbol{\theta}) = \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{y_t - m_t(\boldsymbol{\theta})}{\kappa_{2t}(\boldsymbol{\theta})}$.

Assume

A3*: If $E\{\boldsymbol{\Upsilon}_t(\boldsymbol{\theta})\} = 0$ for some $\boldsymbol{\theta} \in \Theta$, then $\boldsymbol{\theta} = \boldsymbol{\theta}_0^*$, where the pseudo-true value $\boldsymbol{\theta}_0^* \in \overset{\circ}{\Theta}$.

A5*: We have $\sigma_t^2(\boldsymbol{\theta}_0^*) > 0$, a.s. Moreover, if $\boldsymbol{\lambda}^\top \frac{\partial m_t(\boldsymbol{\theta}_0^*)}{\partial \boldsymbol{\theta}} = 0$ a.s. then $\boldsymbol{\lambda} = \mathbf{0}_d$.

Example: conditional mean approximated by an AR(1)

Assume, perhaps wrongly, that $m_t(\boldsymbol{\theta}) = a + by_{t-1}$ with $\boldsymbol{\theta} = (a, b)^\top$. We then have

$$\boldsymbol{\Upsilon}_t(\boldsymbol{\theta}) = \begin{pmatrix} 1 \\ y_{t-1} \end{pmatrix} \frac{1}{\kappa_{2t}}(y_t - a - by_{t-1})$$

Then **A3*** is satisfied with

$$\boldsymbol{\theta}_0^* = \boldsymbol{A}^{-1}\boldsymbol{b}, \quad \boldsymbol{b} = \begin{pmatrix} E\frac{y_t}{\kappa_{2t}} \\ E\frac{y_t y_{t-1}}{\kappa_{2t}} \end{pmatrix}, \quad \boldsymbol{A} = \begin{pmatrix} E\frac{1}{\kappa_{2t}} & E\frac{y_{t-1}}{\kappa_{2t}} \\ E\frac{y_{t-1}}{\kappa_{2t}} & E\frac{y_{t-1}^2}{\kappa_{2t}} \end{pmatrix}$$

when $\boldsymbol{b}$ and $\boldsymbol{A}$ exist and $\boldsymbol{A}$ is invertible (which is for instance the case when κ_{2t} is constant and Var$(y_t) > 0$).

Asymptotics for CUSUM of misspecified quasi-scores

Let $\boldsymbol{\Upsilon}_t^* = \boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_0^*) = \frac{\partial m_t(\boldsymbol{\theta}_0^*)}{\partial \boldsymbol{\theta}} \frac{\epsilon_t(\boldsymbol{\theta}_0^*)}{\kappa_{2t}(\boldsymbol{\theta}_0^*)}$.

Under for instance mixing and moment conditions, we have the CLT

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \boldsymbol{\Upsilon}_t^* \xrightarrow{d} \mathcal{N}(0, \boldsymbol{I}^*)$$

for some long-run nonsingular variance matrix $\boldsymbol{I}^*$.

Let $\boldsymbol{I}_n^*$ be a consistent HAC estimator of $\boldsymbol{I}^*$, and let the statistic

$$\widetilde{S}_n^* = \sup_{u \in (0,1)} \widetilde{S}_n^*(u), \qquad \widetilde{S}_n^*(u) = \widetilde{\boldsymbol{T}}_n^\top(u) \boldsymbol{I}_n^{*-1} \widetilde{\boldsymbol{T}}_n(u).$$

Under regularity conditions including $\mathbf{H_0}$ (no break), we have
$\widetilde{S}_n^* \xrightarrow{d} S = \sup_{u \in (0,1)} \sum_{j=1}^{d} \{B_j(u)\}^2$.

Monte Carlo design

$N = 1,000$ simulations of size $n = 2,000$ of

$$y_t \mid \mathcal{F}_{t-1} \sim \text{Gamma}_t, \quad E_{t-1}(y_t) = m_t, \quad \text{Var}_{t-1}(y_t) = \sigma_t^2,$$

where $m_t = c + ay_{t-1} + bm_{t-1}$ and

DGP A: $\sigma_t^2 = 1$; DGP B: $\sigma_t^2 = m_t$;
DGP C: $\sigma_t^2 = m_t^2$; DGP D: $\sigma_t^2 = m_t^{3/2}$.

We considered 8 different QLEs:

QLE A: $\widetilde{\kappa}_{2t} \propto 1$; QLE B: $\widetilde{\kappa}_{2t} \propto m_t$; QLE C: $\widetilde{\kappa}_{2t} \propto m_t^2$;
QLE D: $\widetilde{\kappa}_{2t} \propto m_t^{3/2}$; QLIK; GARCH; X1; X2,

where the last 4 QLE are optimal QLEs estimated by the QLIK$_n$-method or by fitting GARCH or two different GARCH-X.

Empirical size of the tests $(n = 2,000,\ N = 1,000)$

$m_t = 0.01 + 0.1y_{t-1} + 0.89m_{t-1}$

	A	B	C	D	QLIK	GARCH	X1	X2
				DGP A				
1%	1.3	2.5	7.2	6.6	1.2	0.9	0.7	0.7
5%	4.7	8.3	15.4	13.0	4.3	4.5	4.7	4.7
10%	10.3	14.5	22.8	20.9	9.9	9.3	9.2	9.3
				DGP B				
1%	1.7	0.7	1.3	1.1	0.7	0.4	0.6	0.7
5%	6.7	5.8	5.9	5.4	5.7	4.7	5.1	5.2
10%	11.7	10.1	13.6	12.0	10.1	8.5	9.3	9.3
				DGP C				
1%	5.3	1.0	1.0	0.6	1.0	0.6	0.5	0.8
5%	13.5	5.5	5.5	5.7	5.5	5.0	4.7	5.5
10%	19.9	10.8	10.1	10.6	10.1	9.6	10.0	9.7
				DGP D				
1%	2.5	1.0	1.3	0.8	0.8	0.5	0.9	0.9
5%	8.5	5.8	6.3	6.1	6.2	4.9	5.2	5.6
10%	15.8	10.6	11.1	10.7	10.8	10.9	11.3	10.9

Empirical powers ($n = 2,000,\ N = 1,000$)

(break at $t = 800$, but no change of the marginal mean)

α	A	B	C	D	QLIK	GARCH	X1	X2
				DGP A*				
1%	85.5	38.5	26.7	33.4	78.0	81.8	81.3	81.2
5%	95.7	57.7	40.7	51.4	90.5	94.8	94.0	94.0
10%	98.3	70.4	50.3	62.2	94.7	97.9	97.5	97.5
				DGP B*				
1%	66.8	86.0	34.8	77.0	85.7	83.2	88.8	88.8
5%	84.1	97.7	57.8	91.6	97.6	96.3	98.3	98.2
10%	92.3	99.1	71.0	96.7	99.1	98.8	99.5	99.3
				DGP C*				
1%	54.0	75.5	86.3	89.1	86.7	87.4	88.9	91.7
5%	67.0	88.9	97.3	96.8	97.4	96.8	97.2	98.8
10%	74.5	93.4	99.3	99.4	99.3	98.5	99.1	99.4
				DGP D*				
1%	55.7	87.4	70.9	92.6	91.0	89.0	88.9	91.1
5%	72.4	97.4	89.6	98.5	98.1	98.3	98.0	98.3
10%	81.1	99.1	95.8	99.4	99.2	99.1	99.1	99.3

Change point estimates ($nu_0 = 800$, $n = 2000$)

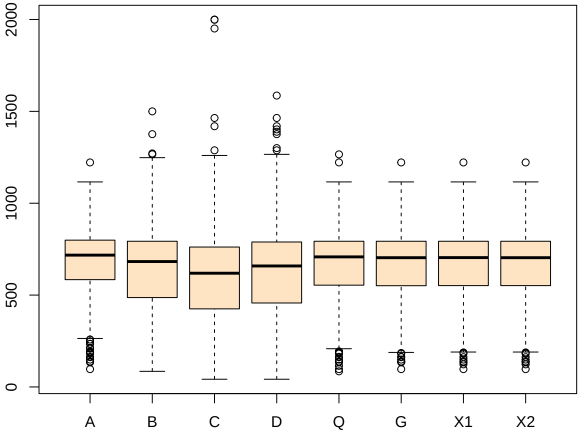


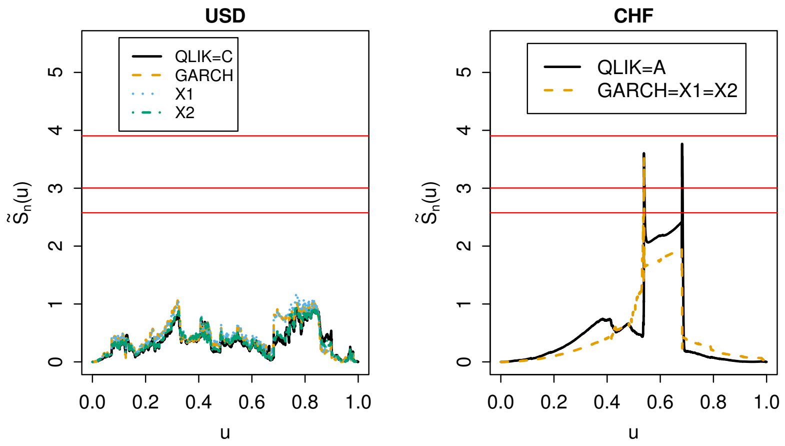
Figure: Dist. of the change point estimates (simulated DGP: A*)

Illustration on exchange rates

- Returns series of daily exchange rates of the USD and CHF with respect to the Euro.
- 1999-01-04 to 2022-07-12 (6025 observations).
- GARCH(1,1) (i.e. ARMA(1,1) on the squares) estimated by QLEs.
- tests for breaks performed using the statistic $\widetilde{S}_n$.

No evidence of breaks for USD but breaks for CHF

Swiss franc was pegged to the euro between Sept. 6, 2011 and Jan. 15, 2015



An alternative approach to CUSUM tests for break detection

- Involves minimizing the OLS sum of squared residuals in a linear model where the beta coefficient is constrained to remain constant over $m+1$ subperiods of some minimum length.
- Bai and Perron (2003) showed that this is feasible, even for $m>1$, using a dynamic programming algorithm.
- The computation time for this method is substantial ($m \leq 5$ breaks, and $p \leq 7$ for the AR order).
- The number of breaks, m, was estimated by BIC minimization.

Bai and Perron approach based on AR(p) for squared returns

The estimated number and timing of breaks vary considerably with p.

	USD								CHF							
p	0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7
m	3	3	3	3	3	3	2	2	0	1	2	2	0	0	1	1
Dates	04-05-17			05-06-14			05-06-14			15-01-13	11-09-06	11-09-06			11-09-06	
	08-08-07			08-12-19			08-12-19				15-01-12	15-01-09				
	12-02-02			12-06-29												

Table: Estimation of the number m of breaks and of their dates.

- For the CHF, a break on September 6, 2011 is often detected, but unlike the CUSUM test, the break around January 15, 2015 is not clearly identified.
- Most tricky output: for the USD series, 3 breaks are often detected while the CUSUM test does not detect any break.

QMLE estimates of GARCH(1,1) models for the USD series

4 subperiods defined by the 3 breaks identified in the AR(p) models with $p \in \{2, \dots, 5\}$

Period	ω	α	β
1999-01-04 to 2005-06-07	0.021 (0.027)	0.048 (0.048)	0.937 (0.056)
2005-06-21 to 2008-12-12	0.000 (0.001)	0.036 (0.017)	0.972 (0.013)
2008-12-30 to 2012-06-22	0.013 (0.006)	0.000 (0.007)	0.980 (0.012)
2012-07-06 to 2022-07-11	0.001 (0.002)	0.061 (0.044)	0.960 (0.024)

- The estimated GARCH parameters do not show substantial variation
- Wald test of the null hypothesis of identical GARCH coefficients: p-value $= 7.8\%$.

Additional simulations

BP test on simulations of a GARCH model (without breaks)

	Simulation 1								Simulation 2							
p	0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7
m	3	2	2	2	2	2	2	2	4	3	3	3	3	3	3	0
Dates	914				1858				904	904			904		904	
	1858				2763				2349	2349			2349		2349	
	2763								3330	3304			3280		3317	
									4760							

- Failure of the BP test: can be explained by the irrelevance of the AR(p) for squared returns (i.e. ARCH(p) for returns).
- On the other hand, extensions of the BP dynamic programming algorithm to persistent models don't seem to exist.

Summary

CUSUM of quasi-score for detecting breaks

- obviously requires less strong assumptions than CUSUM of Fisher's score (semi-parametric method);
- will result in an infinite number of break tests (as many as there are time-varying weights);
- can be more efficient than CUSUM of QMLE-score;
- can even work when m_t is misspecified (with the HAC version);
- is easy to implement (just one optimization to compute $\widehat{\boldsymbol{\theta}}$);
- possible extension: weighted versions to detect early or late breaks.

Thank you!

Regularity conditions

A1: The process $(y_t)_{t\in\mathbb{Z}}$ is strictly stationary and ergodic.

A2: There exists $\rho \in [0,1)$ such that, a.s. $\sup_{\boldsymbol{\theta}\in\Theta} |m_t(\boldsymbol{\theta}) - \widetilde{m}_t(\boldsymbol{\theta})| \leq K_t \rho^t$, where K_t is a generic $\mathcal{F}_{t-1}$-measurable r.v. such that $\sup_t EK_t^r < \infty$ for some $r > 0$.

A3: Let $\boldsymbol{\Upsilon}_t(\boldsymbol{\theta}) = \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\epsilon_t(\boldsymbol{\theta})}{\kappa_{2t}(\boldsymbol{\theta})}$. If $E\{\boldsymbol{\Upsilon}_t(\boldsymbol{\theta})\} = 0$ for some $\boldsymbol{\theta} \in \Theta$, then $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. The parameter $\boldsymbol{\theta}_0$ belongs to the interior of the compact set Θ.

A4: The function $\boldsymbol{\theta} \mapsto m_t(\boldsymbol{\theta})$ is continuously differentiable, and

$$\sup_{\boldsymbol{\theta}\in\Theta} \left\| \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{\partial \widetilde{m}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\| \leq K_t \rho^t, \qquad a.s.$$

where K_t is as in **A2**, $\|\cdot\|$ denotes any norm on $\mathbb{R}^d$. Moreover, assume $E|y_t|^s < \infty$ and $E \sup_{\boldsymbol{\theta}\in\Theta} \left\{ |m_t(\boldsymbol{\theta})|^s + \left\| \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\|^s \right\} < \infty$, for some $s > 0$.

A5: We have $\sigma_t^2(\boldsymbol{\theta}_0) > 0$, a.s. Moreover, if $\boldsymbol{\lambda}^\top \frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} = 0$ a.s. then $\boldsymbol{\lambda} = \mathbf{0}_d$.

A6: There exists a constant $\underline{\kappa} > 0$ such that $\inf_{\boldsymbol{\theta}\in\Theta} \kappa_{2t}(\boldsymbol{\theta}) \geq \underline{\kappa}$ a.s.

Regularity conditions (continued)

A7: For all $\boldsymbol{\theta} \in \Theta$ the sequence $\{\kappa_{2t}(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ is stationary, ergodic and $\mathcal{F}_{t-1}$-measurable, the function $\boldsymbol{\theta} \mapsto \kappa_{2t}(\boldsymbol{\theta})$ admits continuous derivatives, there exist $\rho \in [0,1)$ and K_t as in **A2** such that, a.s.,

$$\sup_{\boldsymbol{\theta}\in\Theta} \left\{ |\kappa_{2t}(\boldsymbol{\theta}) - \widetilde{\kappa}_{2t}(\boldsymbol{\theta})| + \left\| \frac{\partial \kappa_{2t}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{\partial \widetilde{\kappa}_{2t}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\| \right\} \leq K_t \rho^t$$

for n large enough. Moreover $E \sup_{\boldsymbol{\theta}\in\Theta} |\kappa_{2t}(\boldsymbol{\theta})|^s < \infty$ for some $s > 0$.

A8: We have

$$E \sup_{\boldsymbol{\theta}\in\Theta} \|\boldsymbol{\Upsilon}_t(\boldsymbol{\theta})\|^2 < \infty \quad \text{and} \quad E \sup_{\boldsymbol{\theta}\in\Theta} \left\| \frac{\partial \boldsymbol{\Upsilon}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \right\| < \infty.$$

return

Unbiased EF and motivating example

An EF is said to be unbiased when $E\boldsymbol{h}_n(\boldsymbol{\theta}_0) = \mathbf{0}$.

Example (Durbin (1960))

In the AR(1) model $y_t = \theta y_{t-1} + \eta_t$, η_t iid $(0, \sigma^2)$, the OLS solves the unbiased estimating equation $\sum_{t=2}^{n} y_t y_{t-1} - \theta \sum_{t=2}^{n} y_{t-1}^2 = 0$ and has the smallest variance among the linear unbiased estimating functions of the form $\sum_{t=2}^{n} a_{t-1}(y_t - \theta y_{t-1})$ where a_{t-1} is a function of $y_1, \ldots, y_{t-1}$ and satisfies some identifiability conditions (a kind of BLUE property).

A natural class of EF for the weak location model

Notation convention: $X_t \in \mathcal{F}_t = \sigma(y_u, u < t)$ and $\widetilde{X}_t \in \mathcal{I}_t = \sigma(y_u, 1 \leq u < t)$ ($\mathcal{I}_t$ is the information available at t).

Extending Durbin's EF for the AR(1), consider EFs of the form

$$\widetilde{\boldsymbol{h}}_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} \widetilde{\boldsymbol{a}}_{t-1}(\boldsymbol{\theta})\widetilde{\epsilon}_t(\boldsymbol{\theta}), \quad \widetilde{\epsilon}_t(\boldsymbol{\theta}) = y_t - \widetilde{m}_t(\boldsymbol{\theta}),$$

where, for all $\boldsymbol{\theta} \in \Theta$, the variable $\widetilde{m}_t(\boldsymbol{\theta})$ denotes a $\mathcal{I}_{t-1}$-measurable approximation of $m_t(\boldsymbol{\theta})$ and the $d \times 1$ vector $\widetilde{\boldsymbol{a}}_t(\boldsymbol{\theta}) \in \mathcal{I}_t$.

Consider QLEs obtained by solving the EE $\widetilde{\boldsymbol{h}}_n(\boldsymbol{\theta}) = 0$.

Optimal EF in Godambe's sense

Godambe (1985) introduced the notion of optimal estimating function. Let $\mathcal{H}$ the class of unbiased EFs satisfying some regularity conditions. An estimating function $\boldsymbol{h}_n^*$ is said to be optimal in $\mathcal{H}$ if

$$\left\{E\left[\frac{\partial \boldsymbol{h}_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}\right]\right\}^{-1} E\{\boldsymbol{h}_n(\boldsymbol{\theta}_0)\boldsymbol{h}_n'(\boldsymbol{\theta}_0)\}\left\{E\left[\frac{\partial \boldsymbol{h}_n'(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right]\right\}^{-1}$$

is minimized at $\boldsymbol{h}_n(\boldsymbol{\theta}_0) = \boldsymbol{h}_n^*(\boldsymbol{\theta}_0)$ in the sense of semi-positive definite matrices.

- Intuition: small variance at $\boldsymbol{\theta}_0$ (numerator) and high sensitivity to parameter change (denominator).
- Godambe's justification: $\boldsymbol{h}^*$ is the score when $\mathcal{H}$ allows it.

Optimal unbiased EF for the weak location model

Godambe (1985) (see also Chandra and Taniguchi, 2001) showed that, within the class $\mathcal{H}$ of the unbiased EFs of the form $\sum_{t=1}^{n} \boldsymbol{a}_{t-1}(\boldsymbol{\theta})\epsilon_t(\boldsymbol{\theta})$, an optimal EF in Godambe's sense is

$$\sum_{t=1}^{n} \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{1}{\sigma_t^2(\boldsymbol{\theta})} \{y_t - m_t(\boldsymbol{\theta})\}$$

where $\sigma_t^2(\boldsymbol{\theta})$ is the conditional variance (which is generally unknown and depends on nuisance parameters).

- Require that $m_t(\boldsymbol{\theta})$ and $\sigma_t^2(\boldsymbol{\theta})$ be $\mathcal{I}_{t-1}$-measurable (which is generally not the case). return

GMM

We have moment restrictions of the form $E\boldsymbol{g}_t(\boldsymbol{\theta}) = \mathbf{0}$ iff $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, where $\boldsymbol{g}_t(\boldsymbol{\theta}) = \boldsymbol{z}_t \epsilon_t(\boldsymbol{\theta})$ with a vector of instruments $\boldsymbol{z}_t \in \mathcal{F}_{t-1}$ valued in $\mathbb{R}^m$, $m \geq d$.

Let $\overline{\boldsymbol{g}}_n(\boldsymbol{\theta}) = n^{-1} \sum_{t=1}^n \widetilde{\boldsymbol{g}}_t(\boldsymbol{\theta})$, where $\widetilde{\boldsymbol{g}}_t(\boldsymbol{\theta})$ is an $\mathcal{I}_t$-measurable approximation of $\boldsymbol{g}_t(\boldsymbol{\theta})$.

The GMM estimators minimize

$$\overline{\boldsymbol{g}}_n'(\boldsymbol{\theta}) \widehat{\boldsymbol{S}}^{-1} \overline{\boldsymbol{g}}_n(\boldsymbol{\theta}),$$

where $\widehat{\boldsymbol{S}}$ is a positive definite weight matrix.

QL and GMM estimators

The first order conditions yield the EF

$$\widehat{\boldsymbol{h}}_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} \widehat{\boldsymbol{\Omega}}_t(\boldsymbol{\theta})\widehat{\boldsymbol{S}}^{-1}\overline{\boldsymbol{g}}_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} \widehat{\boldsymbol{\Omega}}(\boldsymbol{\theta})\widehat{\boldsymbol{S}}^{-1}\widetilde{\boldsymbol{z}}_t\widetilde{\epsilon}_t(\boldsymbol{\theta})$$

where $\widehat{\boldsymbol{\Omega}}_t(\boldsymbol{\theta}) = \frac{\partial}{\partial\boldsymbol{\theta}}\widetilde{\boldsymbol{g}}_t'(\boldsymbol{\theta})$ and $\widehat{\boldsymbol{\Omega}}(\boldsymbol{\theta}) = n^{-1}\sum_{t=1}^{n}\widehat{\boldsymbol{\Omega}}_t(\boldsymbol{\theta})$.
Therefore the GMM estimators are QLEs, and

the optimal QLE $\succeq$ the optimal GMM

(in the Godambe's sense and asymptotically).
Christensen, Posch and van der Wel (JoE, 2016) showed that in general

the optimal QLE $\succ$ the optimal GMM. return

Justification of the selection method

Consistency of the weights selected by QLIK

Let two sequences $\left\{\kappa_{2t}^{(1)}(\boldsymbol{\theta}), \widetilde{\kappa}_{2t}^{(1)}(\boldsymbol{\theta})\right\}_t$ and $\left\{\kappa_{2t}^{(2)}(\boldsymbol{\theta}), \widetilde{\kappa}_{2t}^{(2)}(\boldsymbol{\theta})\right\}_t$ such that the regularity conditions of the CAN of the QLEs are satisfied. For $\boldsymbol{\theta} \in \Theta_\vartheta$, if

$$\mathsf{QLIK}(\kappa_{2t}^{(1)}(\boldsymbol{\theta})) < \mathsf{QLIK}(\kappa_{2t}^{(2)}(\boldsymbol{\theta})),$$

then, almost surely

$$\mathsf{QLIK}_n\left(\widetilde{\kappa}_{2\cdot}^{(1)}(\boldsymbol{\theta})\right) = \min\left\{\mathsf{QLIK}_n\left(\widetilde{\kappa}_{2\cdot}^{(1)}(\boldsymbol{\theta})\right), \mathsf{QLIK}_n\left(\widetilde{\kappa}_{2\cdot}^{(2)}(\boldsymbol{\theta})\right)\right\}$$

for n large enough.

return

Related literature

- Horváth and Parzen (1994) CUSUM of Fisher's score.
- Lee *et al.* (2003) CUSUM of $\widehat{\boldsymbol{\theta}}_k - \widehat{\boldsymbol{\theta}}_n$.
- Berkes, Horváth and Kokoszka (2004) for GARCH models.
- Shao and Zhang (2010) self-normalized K-S test.
- Aue and Horváth (2013) CUSUM of unconditional and conditional mean and variance.
- Kutoyants (2016) CUSUM of Fisher's score.
- Negri and Nishiyama (2017) with applications to diffusions.
- Truong et al. (2020) overview of change point detection.
- ... return