Detection of breaks in weak location time series models with quasi-Fisher scores

Christian Francq, Lorenzo Trapani and Jean-Michel Zakoïan

ECODEP Conference

IHP, Sept. 30 - Oct. 1, 2024

Estimating Function (EF) approach

- Originally proposed in two seminal papers by Godambe (1960), for fully parametric estimation, and by Durbin (1960), for estimating a simple AR. *
- Has been applied to estimate particular time series models†
- Based on estimating a finite dimensional parameter $\boldsymbol{\theta}$, by solving the equation $h_n(\boldsymbol{\theta}) = 0$, where $h_n(\cdot)$ is a function of the observations.

*see the book by Heyde (2008) a series of papers by Godambe, review papers by Bera, Bilias, Simlai (2006), Jacod and Sørensen (2018).

†in particular Li and Turtle (2000), Chandra and Taniguchi (2001) and Kanai, Ogata and Taniguchi (2010) for ARCH, RCA and CHARN models

Fisher and quasi-Fisher scores

- If $y_1, \ldots, y_n$ iid with distribution $f(y; \boldsymbol{\theta})$, Fisher's score is

$$\boldsymbol{h}_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} \frac{\partial}{\partial \boldsymbol{\theta}} \log f(y_i; \boldsymbol{\theta}).$$

- 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 $y_t \geq 0$ and $m_t(\boldsymbol{\theta})$ is the conditional mean, the Poisson quasi-score is

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

MLE, QMLE and QLEs

- 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 EF $\boldsymbol{h}_n$, a solution of the estimating equation (EE) $\boldsymbol{h}_n(\boldsymbol{\theta}) = \mathbf{0}$ is called Quasi-Likelihood Estimator (QLE) or Z-estimator.

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.

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

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

A strong time series model is driven by a strong white noise.

Objective

- The EF approach can be used to estimate weak location scale models (FZ, 2023).
- **Aim:** use the EF approach to detect breaks in the conditional mean when the demeaned process may not be iid.
  - Procedure based on a CUSUM process depending on a sequence of weights. Properties of tests based on optimal QLE.
  - Data driven selection of the weights.
  - Estimation of the breakdate.
  - Case where the conditional mean is misspecified.
- **Main related references:**
  - 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.*

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

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).

- Which statistic has a nondegenerate asymptotic distribution?
- Is there an optimal choice of the EF (of the $\boldsymbol{\Upsilon}_t$s)?

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

this is better than GMM

A class of EF for the weak location model

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

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 $\widetilde{\kappa}_{2t} = \widetilde{\kappa}_{2t}(\boldsymbol{\theta}, \widehat{\gamma}_n)$ is an *assumed* $\mathcal{I}_n$-measurable 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.

Optimal EF in the strong case

If we assume a strong location model or σ_t^2 constant, the optimal EF is

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

and the LS estimator is optimal among the QLEs.

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: the conditional distribution admits a density of the form

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

for some positive function k and twice differentiable functions $\eta(\cdot)$ and $a(\cdot)$. It is known that $\eta'(m_t) = a'(m_t)/m_t = 1/\sigma_t^2$. It follows that

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

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

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 $y_1, \ldots, y_n$ satisfy $E_{t-1}(y_t) = m_t(\boldsymbol{\theta}_t)$, 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, defined for $u \in [0,1]$ by

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

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 $\boldsymbol{I}_n$ denotes 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 $\alpha \in (0,1)$, rejection region of $\mathbf{H_0}$ of the form:

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

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

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

$$\sqrt{nu_0}\left(\widehat{\boldsymbol{\theta}}_{(1)} - \boldsymbol{\theta}_0\right) \xrightarrow{d} \mathcal{N}\left(\boldsymbol{\delta}_1, \boldsymbol{J}^{-1}\boldsymbol{I}\boldsymbol{J}^{-1}\right),$$

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

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

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

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 LAPs 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 $\widetilde{S}_n^W$ defined by

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

and

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

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 $\alpha = 1\%$, $\delta_1 = -\delta_2 = 3$

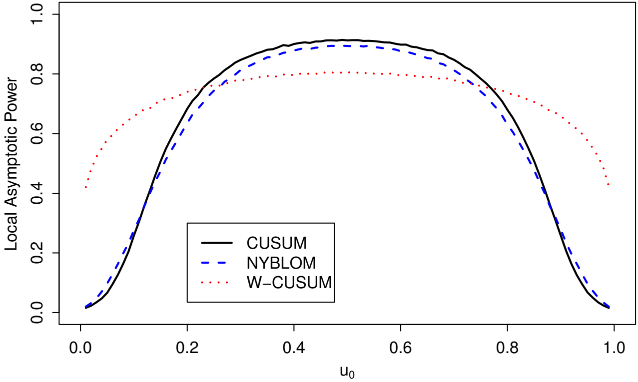


Figure: Powers of the CUSUM, Nyblom, and Weighted CUSUM tests as a function of the break date u_0.

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 several possible values of $\widetilde{\kappa}_{2t}$, which must be chosen 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—it seems natural to consider the weights $\kappa_{2t}(\boldsymbol{\theta}) = m_t(\boldsymbol{\theta})$.
- 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 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

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

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

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 proportional to the conditional mean, its square or its inverse, as suggested previously, we can also fit GARCH-X models by QMLE, leading to

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

or

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

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 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

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

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

	A	B	C	D	QLIK	GARCH	X1	X2
				DGP A				
1%	1.300	2.500	7.200	6.600	1.200	0.900	0.700	0.700
5%	4.700	8.300	15.400	13.000	4.300	4.500	4.700	4.700
10%	10.300	14.500	22.800	20.900	9.900	9.300	9.200	9.300
				DGP B				
1%	1.700	0.700	1.300	1.100	0.700	0.400	0.600	0.700
5%	6.700	5.800	5.900	5.400	5.700	4.700	5.100	5.200
10%	11.700	10.100	13.600	12.000	10.100	8.500	9.300	9.300
				DGP C				
1%	5.300	1.000	1.000	0.600	1.000	0.600	0.500	0.800
5%	13.500	5.500	5.500	5.700	5.500	5.000	4.700	5.500
10%	19.900	10.800	10.100	10.600	10.100	9.600	10.000	9.700
				DGP D				
1%	2.500	1.000	1.300	0.800	0.800	0.500	0.900	0.900
5%	8.500	5.800	6.300	6.100	6.200	4.900	5.200	5.600
10%	15.800	10.600	11.100	10.700	10.800	10.900	11.300	10.900

Empirical powers (break at $t = 800$)

α	A	B	C	D	QLIK	GARCH	X1	X2
				DGP A*				
1%	85.500	38.500	26.700	33.400	78.000	81.800	81.300	81.200
5%	95.700	57.700	40.700	51.400	90.500	94.800	94.000	94.000
10%	98.300	70.400	50.300	62.200	94.700	97.900	97.500	97.500
				DGP B*				
1%	66.800	86.000	34.800	77.000	85.700	83.200	88.800	88.800
5%	84.100	97.700	57.800	91.600	97.600	96.300	98.300	98.200
10%	92.300	99.100	71.000	96.700	99.100	98.800	99.500	99.300
				DGP C*				
1%	54.000	75.500	86.300	89.100	86.700	87.400	88.900	91.700
5%	67.000	88.900	97.300	96.800	97.400	96.800	97.200	98.800
10%	74.500	93.400	99.300	99.400	99.300	98.500	99.100	99.400
				DGP D*				
1%	55.700	87.400	70.900	92.600	91.000	89.000	88.900	91.100
5%	72.400	97.400	89.600	98.500	98.100	98.300	98.000	98.300
10%	81.100	99.100	95.800	99.400	99.200	99.100	99.100	99.300

Change point estimates ($nu_0 = 3200$, $n = 8000$, DGP A*)

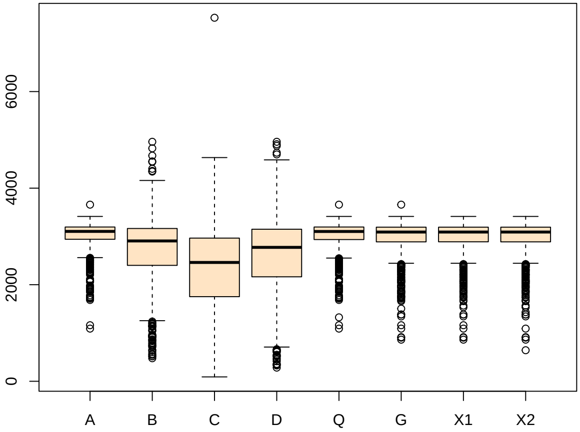


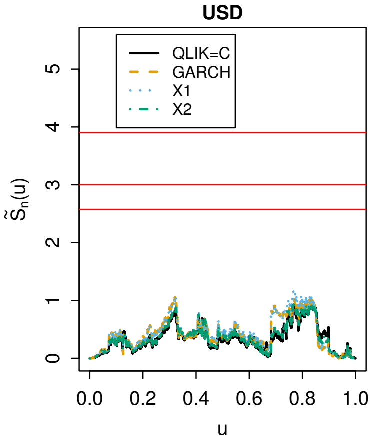
Figure: Distributions of the change point estimes

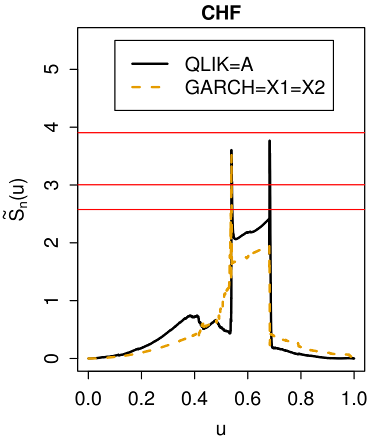
Illustration on exchange rates

- 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





Summary

CUSUM of quasi-score for detecting breaks

- obviously requires less strong assumptions than CUSUM of Fisher's score (semi-parametric method);
- leads to an infinite number of break tests (as many as time-varying weights $\widetilde{\kappa}_{2t}$);
- can be more efficient than CUSUM of QMLE-score (data-driven choice of $\widehat{\boldsymbol{\theta}}$);
- can even work when m_t is misspecified (with HAC version);
- is easy to implement (just one optimization to compute $\widehat{\boldsymbol{\theta}}$);
- this work is still in progress (needs weighted versions to detect early or late breaks, more applications, ...)

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