Mallows' Quasi-Likelihood Estimation for Log-linear Poisson Autoregressions

Konstantinos Fokianos

Department of Mathematics and Statistics, University of Cyprus

Joint work with S. Kitromilidou

CIRM
February 2016

Outline

Introduction to robust statistics

- Sensitivity of classical statistical procedures.

Robust Statistics

- Foundations of robust statistics:
  - Tukey (1960): A survey of sampling from contaminated distributions,
  - Huber (1964): Robust estimation of a location parameter,
  - Hampel (1968): Contributions to the theory of robust estimation.
- Aims of robust statistics:
  - Identification of possible outliers and decrease of their impact on estimation and testing,
  - Ability to fit well to the bulk of the data.
- Some textbooks:
  - Huber (1981), Hampel et al. (1986), Maronna et al. (2006), Huber and Ronchetti (2009).

Modeling count time series

- Examples of count time series

  monthly incidences of some disease,
  daily number of transactions of some stock,
  yearly number of fatalities in road accidents,
  monthly number of claims to an insurance agency etc.

- Theory of Generalized Linear Models (GLM)
  - independent data - McCullagh and Nelder (1989)
  - time series data - Kedem and Fokianos (2002)

- Distribution assumptions
  - Poisson - Davis et al. (2003), Fokianos et al. (2009), Fokianos and Tjøstheim (2011), Neumann (2011), Fokianos (2012), Fokianos and Tjøstheim (2012), Doukhan et al. (2012), Douc et al. (2013),
  - Negative Binomial - Davis and Wu (2009), Zhu (2011), Davis and Liu (2014), Christou and Fokianos (2014),

Linear Poisson autoregressive model - INGARCH

$$Y_t \parallel \mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t), \qquad \lambda_t = d + \sum_{i=1}^{p} a_i \lambda_{t-i} + \sum_{j=1}^{q} b_j Y_{t-j}, \qquad t \geq 1. \quad (1)$$

- $d, a_1, ..., a_p, b_1, ..., b_q$ take non-negative values,
- Rydberg and Shephard (2000), Streett (2000), Heinen (2003), Ferland et al. (2006), Fokianos et al. (2009), Fokianos (2012),
- Finite moments and second order stationarity under condition

$$0 < \sum_{i=1}^{p} a_i + \sum_{j=1}^{q} b_j < 1$$

  (Ferland et al. (2006)),
- class of observation driven models (Cox (1981)),
- geometric ergodicity - perturbation technique (Fokianos et al. (2009)).

Drawbacks:

- Negative correlation cannot be employed.
- Covariates can only be implemented if they result in a positive regression term since otherwise λ_t becomes negative.

Log-linear Poisson autoregressive model

$$Y_t \parallel \mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t), \quad \nu_t = d + \sum_{i=1}^{p} a_i \nu_{t-i} + \sum_{j=1}^{q} b_j \log(1 + Y_{t-j}), \quad t \geq 1 \tag{2}$$

- $\nu_t \equiv \log \lambda_t$ is the canonical link process,
- positive and negative correlation,
- time dependent covariates.

Log-linear Poisson autoregressive model

First order model:

$$Y_t \parallel \mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t), \quad \nu_t = d + a\nu_{t-1} + b\log(1 + Y_{t-1}), \quad t \geq 1.$$

- more parsimonious than a model which includes higher lags of $\log(1 + Y_{t-1})$ but does not include the feedback mechanism,
- class of observation driven models (Cox (1981)),
- geometric ergodicity, finite moments (Fokianos and Tjøstheim (2011))

  $|a + b| < 1$ when $|a| < 1$ and $b > 0$,

  $|a||a + b| < 1$ when $|a| < 1$ and $b < 0$,

- stationarity, consistency and asymptotic normality of the MLE (Fokianos and Tjøstheim (2011))

  $|a + b| < 1$ when a and b have the same sign,

  $a^2 + b^2 < 1$ when a and b have different signs.

Intervention effects

Fokianos and Fried (2010 and 2012) study the problem of estimation and detection of intervention effects for the first order linear and log-linear model respectively.

A sequence of covariates

$$X_t = \xi(\mathcal{B})\mathbb{1}(t = \tau), \quad t \geq 1$$

is introduced to the mean process that indicates an intervention happening at the time point τ.

Then a contaminated process Z_t for the log-linear model is observed:

$$Z_t \parallel \mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t^c), \quad \nu_t^c = d + \sum_{i=1}^{p} a_i \nu_{t-i}^c + \sum_{j=1}^{q} b_j \log(Z_{t-j} + 1) + \zeta X_t, \quad t \geq 1 \tag{3}$$

- $\xi(\mathcal{B}) = (1 - \delta\mathcal{B})^{-1}, \quad \delta \in [0, 1)$,
- $\mathcal{B}$ is a shift operator such that $\mathcal{B}^i X_t = X_{t-i}$,
- $I_t(\tau)$ is an indicator function that is equal to 1 if $t = \tau$ and 0 otherwise,
- ζ is the size of the intervention.

Intervention effects

Forms of interventions:

- Spiky Outlier (SO): $\delta = 0$,
- Transient Shift (TS): $\delta \in \{0.7, 0.8, 0.9\}$,
- Level Shift (LS): $X_t = \mathbb{1}(t \geq \tau)$.

Then X_t is deduced to $X_t = \delta^{t-\tau}\mathbb{1}(t = \tau)$.

Another form of outlier is an Additive Outlier (AO) of size $\zeta \in \mathbb{Z}$ at time τ

$$Z_t = \begin{cases} Y_t + \zeta, & t = \tau \\ Y_t, & \text{otherwise.} \end{cases}$$

Intervention effects

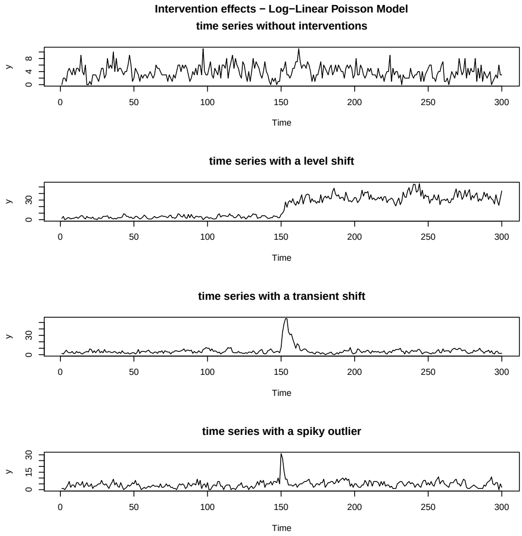

Figure 1: Example of a time series with various forms of interventions, based on the log-linear Poisson model.

Log-linear Poisson model without feedback

$$Y_t \parallel \mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t), \quad \nu_t = d + \sum_{j=1}^{q} b_j \log(Y_{t-j} + 1).$$

Methods of Estimation:

- Maximum Likelihood Estimation (MLE)
- Conditionally Unbiased Bounded-Influence Estimator (CUBIF) - Künsch et al. (1989)
- Mallows' Quasi-Likelihood Estimator (MQLE) - Cantoni and Ronchetti (2001)

Maximum Likelihood Estimator (MLE)

The conditional likelihood function for $\boldsymbol{\theta}$ is

$$\prod_{t=1}^{n} \frac{\exp(-\lambda_t(\boldsymbol{\theta}))\lambda_t(\boldsymbol{\theta})^{Y_t}}{Y_t!}.$$

The score function is defined by

$$S(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{t=1}^{n} \frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{t=1}^{n} \{Y_t - \exp(\nu_t(\boldsymbol{\theta}))\} \frac{\partial \nu_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \tag{4}$$

where $\partial \nu_t(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$ is the $q+1$ dimensional vector

$$\frac{\partial \nu_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = (1, \quad \log(1+Y_{t-1}), \ \dots \ \log(1+Y_{t-q}))^T \equiv X_{t-q}.$$

Conditionally Unbiased Bounded-Influence Estimator (CUBIF)

Idea - Künsch et al. (1989): To find a conditionally Fisher-consistent estimator that has small variance subject to a chosen bound on its influence function.

The score function for the CUBIF estimator is given by

$$\begin{aligned}\psi(\boldsymbol{Y}_t^{(q)};\boldsymbol{\theta},c,\mathbf{B}) &= d(\boldsymbol{Y}_t^{(q)},\boldsymbol{\theta},c,\mathbf{B}) \\ &\times W_c(|d(\boldsymbol{Y}_t^{(q)},\boldsymbol{\theta},c,\mathbf{B})|(X_{t-q}^T\mathbf{B}^{-1}X_{t-q})^{-1/2})X_{t-q}^T \quad (5)\end{aligned}$$

with

$$\boldsymbol{Y}_t^{(q)} = (Y_t, Y_{t-1}, ..., Y_{t-q})^T$$

and

$$d(\boldsymbol{Y}_t^{(q)},\boldsymbol{\theta},c,\mathbf{B}) = Y_t - \lambda_t(\boldsymbol{\theta}) - C\left(\nu_t(\boldsymbol{\theta}), \frac{c}{(X_{t-q}^T\mathbf{B}^{-1}X_{t-q})^{-1/2}}\right)$$

Conditionally Unbiased Bounded-Influence Estimator (CUBIF)

$W_c(\alpha) = \psi_c(\alpha)/\alpha$ where $\psi_c(\alpha)$ is the Huber function

$$\psi_c(\alpha) = \begin{cases} \alpha, & |\alpha| \leq c \\ c\text{sign}(\alpha), & |\alpha| > c \end{cases}$$

c is the tuning constant of the Huber function.

The scalar function $C(\cdot)$ and the matrix $\mathbf{B}$ are chosen so that the sensitivity function is bounded:

$$\mathrm{E}(\psi(\boldsymbol{Y}_t^{(q)}; \boldsymbol{\theta}, c, \mathbf{B})\psi(\boldsymbol{Y}_t^{(q)}; \boldsymbol{\theta}, c, \mathbf{B})^T) = \mathbf{B}$$

and the estimating function to be unbiased:

$$\sum_{t=1}^{n} E(\psi(\boldsymbol{Y}_t^{(q)}; \boldsymbol{\theta}, c, \mathbf{B}) || \mathcal{F}_{t-1}) = 0.$$

Mallows' Quasi-Likelihood Estimator (MQLE)

Idea: Cantoni and Ronchetti (2001) robustified the quasi-likelihood by bounding and centering the quasi score function.

The MQLE is given as a solution of the following

$$Q(\boldsymbol{\theta}) = \sum_{t=1}^{n} \left\{ \psi_c\left(\frac{Y_t - \lambda_t(\boldsymbol{\theta})}{\sqrt{\lambda_t(\boldsymbol{\theta})}}\right) w_t \frac{1}{\sqrt{\lambda_t(\boldsymbol{\theta})}} \frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \alpha(\boldsymbol{\theta}) \right\} = 0 \qquad (6)$$

The sequence $\{w_t\}$ are suitable weights:

- $w_t = \sqrt{1 - h_t}$ where h_t is the t-th element of the hat matrix,
- the inverse of the robust Mahalanobis distance d_M where location and scatter are robustly estimated to have high breakdown properties using either
  - the Minimum Volume Ellipsoid (MVE) estimator or
  - the Minimum Covariance Determinant (MCD) algorithm.

  Rousseeuw and van Zomeren (1990), Rousseeuw and Driessen (1999).

Mallows' Quasi-Likelihood Estimator (MQLE)

The term $\alpha(\boldsymbol{\theta})$ is a bias correction term which is used to ensure Fisher-consistency and is given by

$$\alpha(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n} E\left\{\psi_c\left(\frac{Y_t - \lambda_t(\boldsymbol{\theta})}{\sqrt{\lambda_t(\boldsymbol{\theta})}}\right) w_t \frac{1}{\sqrt{\lambda_t(\boldsymbol{\theta})}} \frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\},$$

where

$$\begin{aligned} \mathrm{E}\left(\psi_c\left(\frac{Y_t - \lambda_t(\boldsymbol{\theta})}{\sqrt{\lambda_t(\boldsymbol{\theta})}}\right)|\mathcal{F}_{t-1}\right) &= c\{P(Y_t \geq j_2 + 1\|\mathcal{F}_{t-1}) - P(Y_t \leq j_1\|\mathcal{F}_{t-1})\} \\ &+ \sqrt{\lambda_t(\boldsymbol{\theta})}\{P(Y_t = j_1\|\mathcal{F}_{t-1}) - P(Y_t = j_2\|\mathcal{F}_{t-1})\}, \end{aligned}$$

$j_1 = \lfloor \lambda_t(\boldsymbol{\theta}) - c\sqrt{\lambda_t(\boldsymbol{\theta})} \rfloor$ and $j_2 = \lfloor \lambda_t(\boldsymbol{\theta}) + c\sqrt{\lambda_t(\boldsymbol{\theta})} \rfloor$.

Simulation study

- 1000 simulations,
- time series of length 500,
- estimators compared in terms of MSE, MAE and bias,
- the intervention occurs in the first quarter of the series.

Level Shift (LS) intervention

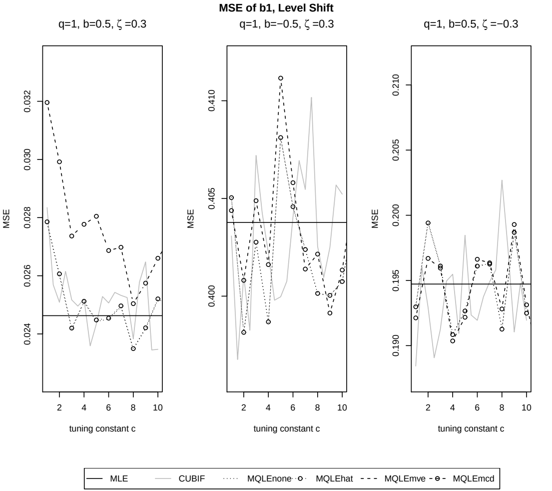

Figure 2: MSE of b_1, first order model with (a): $\theta = (d, b) = (0.2, 0.5)$, $\zeta = 0.3$, (b): $\theta = (d, b) = (0.2, -0.5)$, $\zeta = 0.3$ (c): $\theta = (d, b) = (0.2, 0.5)$, $\zeta = -0.3$.

Additive Outliers (AO)

The observed contaminated series is of the form

$$Z_t = \begin{cases} Y_t + \zeta, & t = \tau_1, \tau_2, \ldots, \tau_k \\ Y_t, & \text{otherwise} \end{cases}$$

We examine the following three cases:

- Single outlier
- Patch of outliers
- Isolated outliers

Additive Outliers (AO)

Single Additive Outlier (AO)

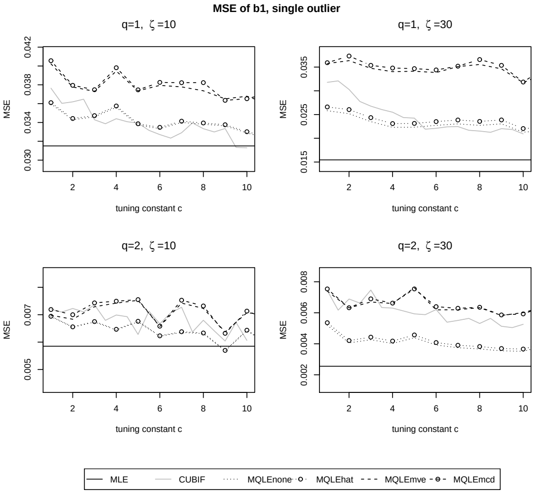

Figure 3: MSE of b_1: first order model with $\boldsymbol{\theta} = (d, b_1) = (0.2, 0.5)$ (upper plots) and second order model with $\boldsymbol{\theta} = (d, b_1, b_2) = (0.2, 0.3, 0.4)$ (bottom plots).

Additive Outliers (AO)
Patch of Outliers

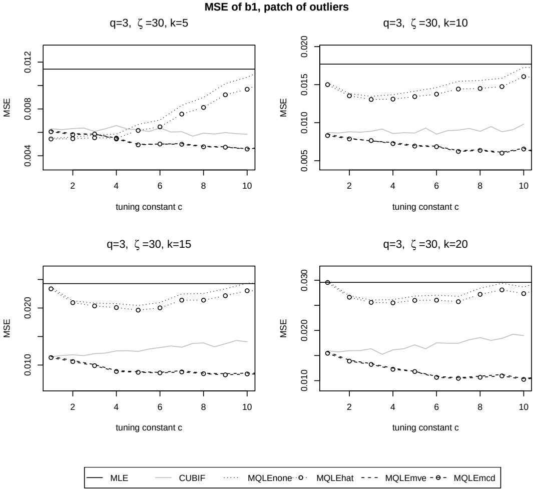

Figure 4: MSE of b_1, third order model with $\boldsymbol{\theta} = (d, b_1, b_2, b_3) = (0.2, 0.2, 0.3, 0.4)$ and $\zeta = 30$.

Additive Outliers (AO)

Isolated outliers

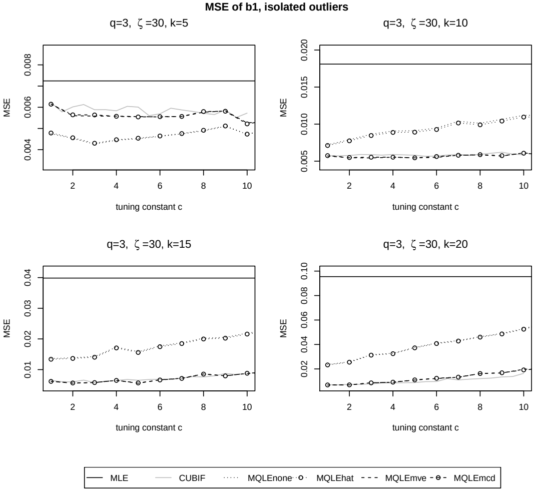

Figure 5: MSE of b_1, third order model with $\boldsymbol{\theta} = (d, b_1, b_2, b_3) = (0.2, 0.2, 0.3, 0.4)$ and $\zeta = 30$.

Simulation study conclusions

Conclusions:

- When a LS or TS exists then there do not exist noteworthy differences among the estimators.
- When AOs are added to the series, either as a patch of consecutive outliers or isolated, then the robustly weighted MQLE dominates all other proposed estimation methods.
- CUBIF is competitive only in the case of Isolated outliers.
- Similar results are obtained for
    - time series of length 200,
    - MSE, MAE, bias,
    - $\tau = n/4, n/2$,
    - first, second and third order models.

Polio data

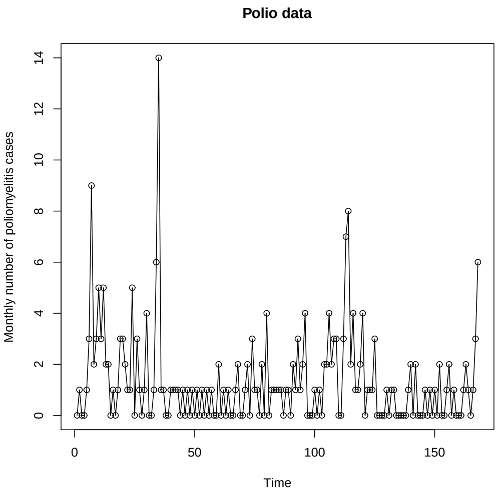

Figure 6: Monthly number of poliomyelitis cases during the years 1970 to 1983 in USA.

Polio data

- 168 observations,
- log-linear Poisson model without feedback of order $q = 6$,
- a long-term decrease of the incidence rate might exist so trend of the form t/n is included in the model,
- sinusoid terms to model annual seasonality,
- the last ten observations are excluded for prediction.

$$\nu_t = d + \sum_{j=1}^{q} b_j \log(1 + Y_{t-j}) + \beta t/n + \sum_{s=1}^{S} \{\beta_{1;s} \sin(\omega_s t) + \beta_{2;s} \cos(\omega_s t)\}$$

where S is the number of harmonics and $\omega_s = 2\pi s/12$ are the Fourier frequencies.

Polio data

- We fit different models and based on the AIC criterion we resort to the model

$$\nu_t = d + \sum_{j=1}^{5} b_j \log(1 + Y_{t-j}) + \beta t/n + \beta_{1;1} \sin(w_1 t) + \beta_{2;s} \cos(w_1 t)$$

- We fit the model and estimate the parameters using the MLE, CUBIF and MQLE for various values of the tuning constant c.
- Choose c for which the estimated MSE or MAE of the predicted values is smallest.

	MLE	CUBIF	MQLE	MQLE hat	MQLE mve	MQLE mcd
c	–	1.408	1.051	1.102	3.041	1.051
MSE	0.692	1.503	0.572	0.577	0.535	0.522
c	–	1.408	1.051	1.051	3.041	1.664
MAE	0.741	1.102	0.663	0.666	0.622	0.629

Table 1: Minimum MSE/MAE of the predicted values of the last ten observations of the polio data.

Polio data

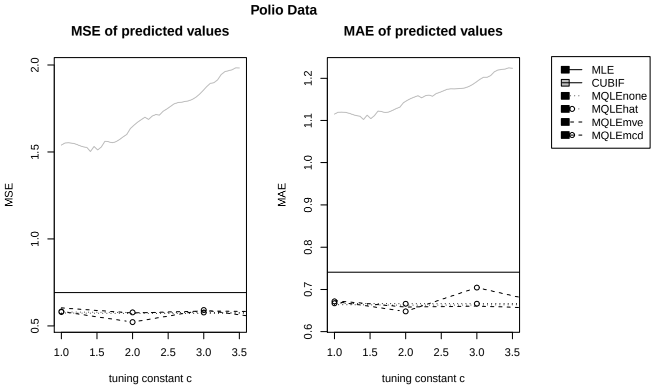

Figure 7: Estimated MSE (left plot) and MAE (right plot) of the predicted values of the last ten observations of the Polio data.

First order log-linear model with feedback

$$Y_t \parallel \mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t), \quad \nu_t = d + a\nu_{t-1} + b\log(1 + Y_{t-1}), \tag{7}$$

- Introduce at each time point t a Poisson process $N_t(\cdot)$ of unit intensity

$$Y_t = N_t(\lambda_t), \quad \nu_t = d + a\nu_{t-1} + b\log(1 + Y_{t-1}).$$

- Introduce a perturbed chain (Y_t^m, ν_t^m)

$$Y_t^m = N_t(\lambda_t^m), \quad \nu_t^m = d + a\nu_{t-1}^m + b\log(1 + Y_{t-1}^m) + \epsilon_{t,m} \tag{8}$$

- $\epsilon_{t,m} = c_m \mathbb{1}(Y_{t-1}^m = 1)\mathcal{U}_t, \quad c_m > 0, \quad c_m \to 0$ as $m \to \infty$,
- $\mathcal{U}_t$ is a sequence of iid uniform random variables on $(0, 1)$ such that $\mathcal{U}_t$ is independent of $N_t(\cdot)$

MQLE

- Fisher consistent M-estimator,
- solution of the quasi-score equation $S_n(\boldsymbol{\theta}) = 0$ where

$$S_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} \left(m_t(\boldsymbol{\theta}) - E\left(m_t(\boldsymbol{\theta}) \parallel \mathcal{F}_{t-1} \right) \right) = \sum_{t=1}^{n} s_t(\boldsymbol{\theta}), \qquad (9)$$

with

$$m_t(\boldsymbol{\theta}) = \psi\left(r_t(\boldsymbol{\theta}) \right) w_t e^{\nu_t(\boldsymbol{\theta})/2} \frac{\partial \nu_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}.$$

- Perturbed score:

$$S_n^m(\boldsymbol{\theta}) = \sum_{t=1}^{n} \left(m_t^m(\boldsymbol{\theta}) - E\left(m_t^m(\boldsymbol{\theta}) \parallel \mathcal{F}_{t-1}^m \right) \right) = \sum_{t=1}^{n} s_t^m(\boldsymbol{\theta}),$$

with

$$m_t^m(\boldsymbol{\theta}) = \psi_c(r_t^m) w_t e^{\nu_t^m/2} \frac{\partial \nu_t^m}{\partial \boldsymbol{\theta}}.$$

MQLE

- $r_t = (Y_t - \lambda_t)/\sqrt{\lambda_t}$ are the Pearson residuals,
- ψ is a suitable weight function that depends on a tuning constant chosen to ensure a desired level of asymptotic efficiency,
- $\partial\nu_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ is a three dimensional vector with components

$$\frac{\partial\nu_t(\boldsymbol{\theta})}{\partial d} = 1 + a\frac{\partial\nu_{t-1}(\boldsymbol{\theta})}{\partial d},$$
$$\frac{\partial\nu_t(\boldsymbol{\theta})}{\partial a} = \nu_{t-1} + a\frac{\partial\nu_{t-1}(\boldsymbol{\theta})}{\partial a},$$
$$\frac{\partial\nu_t(\boldsymbol{\theta})}{\partial b} = \log(1 + Y_{t-1}) + a\frac{\partial\nu_{t-1}(\boldsymbol{\theta})}{\partial b}.$$

- $\{w_t\}$ is an appropriate sequence of weights, $0 < w_t < 1$.

MQLE

- Maximization problem:

$$M_t(\boldsymbol{\theta}) = \int_{\tilde{s}}^{\lambda_t(\boldsymbol{\theta})} \left(\psi\left(\frac{Y_t - z}{\sqrt{z}}\right) - E\left[\psi\left(\frac{Y_t - z}{\sqrt{z}}\right) \parallel \mathcal{F}_{t-1}\right] \right) w_t \frac{1}{\sqrt{z}} dz,$$

with $\tilde{s}$ such that $\left(\psi\left(\frac{Y_t - \tilde{s}}{\sqrt{\tilde{s}}}\right) - E\left[\psi\left(\frac{Y_t - \tilde{s}}{\sqrt{\tilde{s}}}\right) \parallel \mathcal{F}_{t-1}\right] \right) w_t \frac{1}{\sqrt{\tilde{s}}} = 0$, we obtain that

$$\frac{\partial}{\partial \boldsymbol{\theta}} M_t(\boldsymbol{\theta}) = m_t(\boldsymbol{\theta}) - E\Big(m_t(\boldsymbol{\theta}) \parallel \mathcal{F}_{t-1} \Big).$$

- Twice differentiable penalty function $M_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} M_t(\boldsymbol{\theta})$.
- Existence, consistency, asymptotic normality of $\hat{\boldsymbol{\theta}}_{\text{MQLE}}$, (Taniguchi and Kakizawa (2000, Thm 3.2.23).

MQLE

Taniguchi and Kakizawa (2000, Thm 3.2.23)

$$\begin{aligned} M_n(\boldsymbol{\theta}) &= M_n(\boldsymbol{\theta}_0) + (\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T \frac{\partial}{\partial \boldsymbol{\theta}} M_n(\boldsymbol{\theta}_0) + \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T \frac{\partial^2}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T} M_n(\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_0) \\ &+ \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T \left\{ \frac{\partial^2}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T} M_n(\boldsymbol{\theta}^*) - \frac{\partial^2}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T} M_n(\boldsymbol{\theta}_0) \right\} (\boldsymbol{\theta}-\boldsymbol{\theta}_0) \\ &= M_n(\boldsymbol{\theta}_0) + (\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T \frac{\partial}{\partial \boldsymbol{\theta}} M_n(\boldsymbol{\theta}_0) + \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T V_n(\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_0) \\ &+ \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T T_n(\boldsymbol{\theta}^*)(\boldsymbol{\theta}-\boldsymbol{\theta}_0) \end{aligned}$$

Assumptions:

- A1 $\frac{1}{n}\frac{\partial}{\partial \boldsymbol{\theta}} M_n(\boldsymbol{\theta}_0) \xrightarrow{\text{a.s}} 0$,
- A2 $\frac{1}{n} V_n(\boldsymbol{\theta}_0) \xrightarrow{\text{a.s}} V$, V a positive definite matrix,
- A3 for $j,k = 1,2,3$, $\lim_{n\to\infty} \sup_{\delta \to 0} \frac{1}{n\delta} |T_n(\boldsymbol{\theta}^*)_{jk}| < \infty$ a.s

Then, there exists a sequence of estimates $\hat{\boldsymbol{\theta}}_n$ such that $\hat{\boldsymbol{\theta}}_n \xrightarrow{\text{a.s}} \boldsymbol{\theta}_0$,

- A4 $\frac{1}{\sqrt{n}}\frac{\partial}{\partial \boldsymbol{\theta}} M_n(\boldsymbol{\theta}_0) \xrightarrow{\text{d}} N(0, W)$

$$\Rightarrow \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\text{d}} N(0, V^{-1} W V^{-1})$$

MQLE - Asymptotic theory

Theorem 1

Consider model (7). Let $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^3$ which is assumed compact and suppose that the true value $\boldsymbol{\theta}_0$ belongs to the interior of $\boldsymbol{\Theta}$. Assume further that ψ is two times continuously differentiable bounded function. Introduce lower and upper values of each component of $\boldsymbol{\theta}_0 = (d_0, a_0, b_0)^T$ such that $d_L < d_0 < d_U$, $-1 < a_L < a_0 < a_U < 1$ and $b_L < b_0 < b_U$ and suppose that at the true value $\boldsymbol{\theta}_0$, $|a_0 + b_0| < 1$ if a_0 and b_0 have the same sign, and $a_0^2 + b_0^2 < 1$ if a_0 and b_0 have different sign. Then, there exists a fixed open neighborhood $O(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$

$$O(\boldsymbol{\theta}_0) = \{\boldsymbol{\theta} | d_L < d < d_U, -1 < a_L < a < a_U < 1, b_L < b < b_U\}$$

such that with probability tending to 1 as $n \to \infty$, the equation $S_n(\boldsymbol{\theta}) = 0$ has a unique solution, say $\hat{\boldsymbol{\theta}}_{\text{MQLE}}$. Furthermore, $\hat{\boldsymbol{\theta}}_{\text{MQLE}}$ is strongly consistent and asymptotically normal,

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{\text{MQLE}} - \boldsymbol{\theta}_0) \xrightarrow{\text{d}} N(0, V^{-1} W V^{-1})$$

where the matrices W and V are defined in the following Lemmas.

MQLE - Asymptotic theory

An approximation lemma:

1. $E|\nu_t^m - \nu_t| \to 0$ and $|\nu_t^m - \nu_t| < \delta_{1,m}$ almost surely for m large.
2. $E(\nu_t^m - \nu_t)^2 \leq \delta_{2,m}$,
3. $E|\lambda_t^m - \lambda_t| \leq \delta_{3,m}$,
4. $E|Y_t^m - Y_t| \leq \delta_{4,m}$,
5. $E(\lambda_t^m - \lambda_t)^2 \leq \delta_{5,m}$,
6. $E(Y_t^m - Y_t)^2 \leq \delta_{6,m}$,
7. $E|r_t^m - r_t| \to 0$,
8. $E(r_t^m - r_t)^2 \leq \delta_{7;m}$,

where $\delta_{i,m} \to 0$ as $m \to \infty$ for $i = 1, \ldots, 7$. Furthermore, almost surely, with m large enough

$$|\lambda_t^m - \lambda_t| \leq \delta, \quad |r_t^m - r_t| \leq \delta \quad \text{and} \quad |Y_t^m - Y_t| \leq \delta, \quad \text{for any} \quad \delta > 0.$$

MQLE - Asymptotic theory

Lemma 1

Define the matrices

$$W^m(\boldsymbol{\theta}) = E\left(s_t^m(\boldsymbol{\theta}) s_t^m(\boldsymbol{\theta})^T\right) \quad \text{and} \quad W(\boldsymbol{\theta}) = E\left(s_t(\boldsymbol{\theta}) s_t(\boldsymbol{\theta})^T\right).$$

Under the assumptions of Theorem 1, the above matrices evaluated at the true value $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, satisfy $W^m \to W$, as $m \to \infty$.

Outline of the proof:

- $E\left(m_t^m(\boldsymbol{\theta})\left(m_t^m(\boldsymbol{\theta})\right)^T\right) - E\left(m_t(\boldsymbol{\theta})\left(m_t(\boldsymbol{\theta})\right)^T\right) \to 0$ and $E\left(m_t^m(\boldsymbol{\theta})\right) E^T\left(m_t^m(\boldsymbol{\theta})\right) - E\left(m_t(\boldsymbol{\theta})\right) E^T\left(m_t(\boldsymbol{\theta})\right) \to 0$ as $m \to \infty$.
- Consider for each $\theta_i = d, a, b$ the differences $$E\left|(Z_t^m)^2\left(\frac{\partial \nu_t^m}{\partial \theta_i}\right)^2 - Z_t^2\left(\frac{\partial \nu_t}{\partial \theta_i}\right)^2\right| \text{ and } \left|E^2\left(Z_t^m \frac{\partial \nu_t^m}{\partial \theta_i}\right) - E^2\left(Z_t \frac{\partial \nu_t}{\partial \theta_i}\right)\right| \text{ with}$$ $Z_t = \psi_c(r_t) w_t e^{\nu_t/2}$ and Z_t^m defined analogously,
- $|\partial \nu_t^m / \partial \theta_i - \partial \nu_t / \partial \theta_i|$, $|r_t^m - r_t|$, $|\lambda_t^m - \lambda_t| \to 0$, as $m \to \infty$.

MQLE - Asymptotic theory

Lemma 2

Under the assumptions of Theorem 1, the score functions for the perturbed (8) and unperturbed model (7) evaluated at the true value $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ satisfy the following:

1. $S_n^m/n \xrightarrow{\text{a.s}} 0$,
2. $S_n^m/\sqrt{n} \xrightarrow{\text{d}} S^m := N(0, W^m)$,
3. $S^m \xrightarrow{\text{d}} N(0, W)$, as $m \to \infty$,
4. $\lim_{m\to\infty} \limsup_{n\to\infty} P(||S_n^m - S_n|| > \epsilon\sqrt{n}) = 0, \quad \forall \epsilon > 0$.

MQLE - Asymptotic Theory

Proof of Lemma 2

Outline of the proof:

- Strong LLN for martingales (Chow (1967))
  S_n^m square integrable martingale sequence
- CLT for martingales (Hall and Heyde (1980, Cor. 3.1))
  conditional Lindeberg condition and conditional variance condition
- Lemma 1 by Prop. 6.4.9 of Brockwell and Davis (1991)
-

$$\frac{1}{\sqrt{n}}(S_n^m - S_n) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[W_t^m\left(\frac{\partial \nu_t^m}{\partial \boldsymbol{\theta}} - \frac{\partial \nu_t}{\partial \boldsymbol{\theta}}\right) + (W_t^m - W_t)\frac{\partial \nu_t}{\partial \boldsymbol{\theta}}\right]$$

where $W_t = Z_t - E[Z_t \parallel \mathcal{F}_{t-1}]$ and W_t^m defined analogously

$P\left(\left\|\sum_{t=1}^{n} W_t^m\left(\frac{\partial \nu_t^m}{\partial \boldsymbol{\theta}} - \frac{\partial \nu_t}{\partial \boldsymbol{\theta}}\right)\right\| > \delta\sqrt{n}\right) \to 0$ and

$P\left(\| \sum_{t=1}^{n}(W_t^m - W_t)\frac{\partial \nu_t}{\partial \boldsymbol{\theta}}\| > \delta\sqrt{n}\right) \to 0$ as $m \to \infty$

MQLE - Asymptotic theory

Lemma 3

Define the matrices

$$V^m(\boldsymbol{\theta}) = -E\left[\frac{\partial}{\partial\boldsymbol{\theta}} s_t^m(\boldsymbol{\theta})\right], \qquad V(\boldsymbol{\theta}) = -E\left[\frac{\partial}{\partial\boldsymbol{\theta}} s_t(\boldsymbol{\theta})\right].$$

Under the assumptions of Theorem 1, the above matrices evaluated at the true value $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, satisfy $V^m \to V$, as $m \to \infty$.

Outline of the proof:

- $V_n(\boldsymbol{\theta}) = -\frac{1}{n}\sum_{t=1}^{n} E\left(\frac{\partial}{\partial\boldsymbol{\theta}} s_t(\boldsymbol{\theta}) || \mathcal{F}_{t-1}\right)$ and $V_n^m(\boldsymbol{\theta}) = -\frac{1}{n}\sum_{t=1}^{n} E\left(\frac{\partial}{\partial\boldsymbol{\theta}} s_t^m(\boldsymbol{\theta}) || \mathcal{F}_{t-1}\right)$ are consistent estimators of the matrices V and V^m respectively

MQLE - Asymptotic Theory

Proof of Lemma 3

- Because $S_n(\boldsymbol{\theta}) = 0$ is an unbiased estimating function, it holds that

$$-E\left(\frac{\partial}{\partial \boldsymbol{\theta}} s_t(\boldsymbol{\theta}) || \mathcal{F}_{t-1}\right) = E\left(s_t(\boldsymbol{\theta}) \frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} || \mathcal{F}_{t-1}\right) \tag{10}$$

where $\ell_t(\boldsymbol{\theta}) = (Y_t \nu_t(\boldsymbol{\theta}) - \exp(\nu_t(\boldsymbol{\theta})))$, is the logarithm of the conditional probability of $Y_t || \mathcal{F}_{t-1}$ under the Poisson assumption.

- Examine the difference $s_t^m(\partial \ell_t^m / \partial \theta_i) - s_t(\partial \ell_t / \partial \theta_i)$ for $\theta_i = d, a, b$

- $|[\psi(r_t^m) - E\left(\psi(r_t^m) || \mathcal{F}_{t-1}^m\right)] - [\psi(r_t) - E\left(\psi(r_t) || \mathcal{F}_{t-1}\right)]|$,

$$\left|\left(\frac{\partial \nu_t^m}{\partial \theta_i}\right)^2 - \left(\frac{\partial \nu_t}{\partial \theta_i}\right)^2\right|, \; |r_t^m - r_t|, \; |\lambda_t^m - \lambda_t| \to 0, \text{ as } m \to \infty.$$

MQLE - Asymptotic theory

Lemma 4

Denote by

$$H_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n} s_t(\boldsymbol{\theta})\frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}},$$

where $\ell_t(\boldsymbol{\theta}) = Y_t\nu_t(\boldsymbol{\theta}) - \exp(\nu_t(\boldsymbol{\theta}))$, is the t'th component of the Poisson log-likelihood function. Define analogously $H_n^m(\boldsymbol{\theta})$. Then, under the assumptions of Theorem 1,

1. $H_n^m \xrightarrow{p} V^m$ as $n \to \infty$
2. $\lim_{m\to\infty} \limsup_{n\to\infty} P(||H_n^m - H_n|| > \epsilon n) = 0, \quad \forall \epsilon > 0.$

MQLE - Asymptotic Theory

Proof of Lemma 4

Outline of the proof:

- LLN,
- $$H_n = \frac{1}{n}\sum_{t=1}^{n}\left\{w_t e^{\nu_t} r_t \left[\psi(r_t) - E\left(\psi(r_t)||\mathcal{F}_{t-1}\right)\right]\left(\frac{\partial \nu_t}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \nu_t}{\partial \boldsymbol{\theta}}\right)^T\right\}$$

  and H_n^m is defined analogously.
- Examine the difference $H_n^m - H_n$
- $E\left\|\left(\partial\nu_t/\partial\boldsymbol{\theta}\right)\left(\partial\nu_t/\partial\boldsymbol{\theta}\right)^T\right\| < \infty$
- $|[\psi(r_t^m) - E\left(\psi(r_t^m)||\mathcal{F}_{t-1}^m\right)] - [\psi(r_t) - E\left(\psi(r_t)||\mathcal{F}_{t-1}\right)]|$,

  $|\lambda_t^m - \lambda_t|$, $\left\|\left(\frac{\partial \nu_t^m}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \nu_t^m}{\partial \boldsymbol{\theta}}\right)^T - \left(\frac{\partial \nu_t}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \nu_t}{\partial \boldsymbol{\theta}}\right)^T\right\| \to 0$, as $m \to \infty$.

MQLE - Asymptotic theory

Lemma 5

Under the assumptions of Theorem 1,

$$\max_{i,j,k=1,2,3} \sup_{\boldsymbol{\theta} \in O(\boldsymbol{\theta}_0)} \left| \frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 s_{ti}(\boldsymbol{\theta})}{\partial \theta_k \partial \theta_j} \right| \leq \tilde{M}_n := \frac{1}{n} \sum_{t=1}^{n} \tilde{m}_t$$

where θ_i for $i = 1, 2, 3$ refers to $\theta_i = d, a, b$ respectively. Define analogously $\tilde{M}_n^m$. Then

1. $\tilde{M}_n^m \xrightarrow{p} \tilde{M}^m$, as $n \to \infty$ for each $m = 1, 2, \ldots$,
2. $\tilde{M}^m \to \tilde{M}$, as $m \to \infty$, where $\tilde{M}$ is a finite constant,
3. $\lim_{m\to\infty} \limsup_{n\to\infty} P(|\tilde{M}_n^m - \tilde{M}_n| > \epsilon n) = 0, \quad \forall \epsilon > 0.$

Empirical results

In our calculations:

- 1000 simulations,
- generate 800 and discard the first 300 observations,
- initialize $\nu_0 = 1$, $\partial\nu_0/\partial\boldsymbol{\theta} = 1$, interventions occur at $n/4$,
- Huber function

$$\psi_c(x) = \begin{cases} x, & |x| \leq c \\ c\mathrm{sign}(x), & |x| > c \end{cases}$$

For this choice of $\psi(\cdot)$, the bias term can be calculated (Cantoni and Ronchetti (2001)):

$$E\Big(m_t(\boldsymbol{\theta}) \parallel \mathcal{F}_{t-1}\Big) = E\Big(\psi_c(r_t(\boldsymbol{\theta}) \parallel \mathcal{F}_{t-1}\Big) w_t e^{\nu_t(\boldsymbol{\theta})/2} \frac{\partial\nu_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}$$

where

$$\begin{aligned} E\left(\psi_c\left(\frac{Y_t - \lambda_t(\boldsymbol{\theta})}{\sqrt{\lambda_t(\boldsymbol{\theta})}}\right) \parallel \mathcal{F}_{t-1}\right) &= c\left\{P(Y_t \geq j_2 + 1 \| \mathcal{F}_{t-1}) - P(Y_t \leq j_1 \| \mathcal{F}_{t-1})\right\} \\ &+ \sqrt{\lambda_t(\boldsymbol{\theta})}\left\{P(Y_t = j_1 \| \mathcal{F}_{t-1}) - P(Y_t = j_2 \| \mathcal{F}_{t-1})\right\}, \end{aligned}$$

with $j_1 = \lfloor \lambda_t(\boldsymbol{\theta}) - c\sqrt{\lambda_t(\boldsymbol{\theta})} \rfloor$ and $j_2 = \lfloor \lambda_t(\boldsymbol{\theta}) + c\sqrt{\lambda_t(\boldsymbol{\theta})} \rfloor$.

Robust weights

Implementation of robustly weighted methods:

- Method A:
  Approximate ν_t by

$$\hat{\nu}_t = d + a\hat{\nu}_{t-1} + b_1 \log(1 + Y_{t-1}),$$

  where $\hat{\nu}_t$ is computed by employing $\hat{\theta}_{MQLE}$ calculated without weights.

- Method B:
  Approximate ν_t by

$$\hat{\nu}_t = d^* + \sum_{i=1}^{M} a_i^* \log(1 + Y_{t-i}),$$

  for some truncation point M and some regression parameters $\{d^*, a_1^*, \ldots, a_M^*\}$. This choice is motivated by the fact that repeated substitution in the log intensity process ν_t shows that

$$\nu_t = d\frac{1-a^t}{1-a} + a^t\nu_0 + b\sum_{i=0}^{t-1} a^i \log(1 + Y_{t-i-1})$$

Empirical results - Patch of outliers

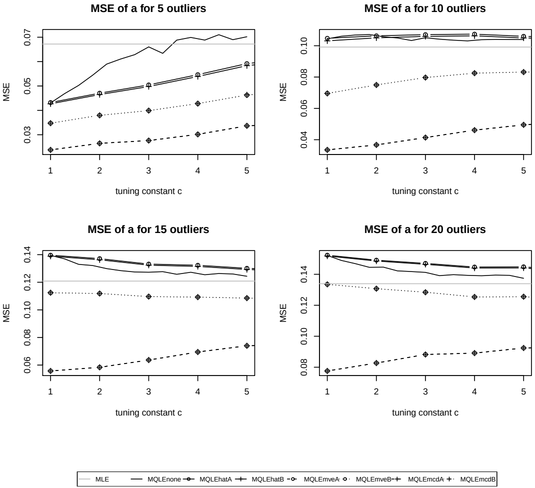

Figure 8: MSE of $\hat{a}_{\mathrm{MQLE}}$, $\boldsymbol{\theta} = (d, a, b) = (0.2, 0.3, 0.5)$, patch of outliers of size $\zeta = 20$.

Empirical results - Patch of outliers

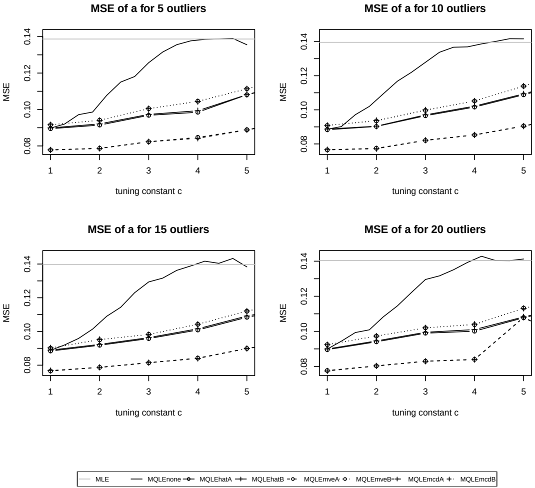

Figure 9: MSE of $\hat{a}_{\mathrm{MQLE}}$, $\boldsymbol{\theta} = (d, a, b) = (0.2, 0.3, 0.65)$, patch of outliers of size $\zeta = 20$.

Empirical results - Level Shift (LS) and Transient Shift (TS)

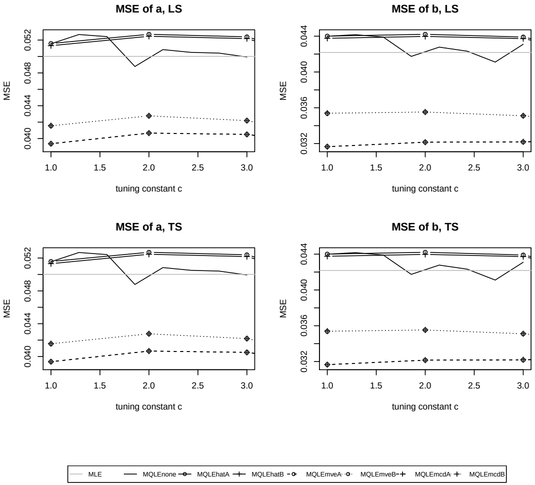

Figure 10: MSE of $\hat{a}_{\text{MQLE}}$ (left plots) and $\hat{b}_{\text{MQLE}}$ (right plots), $\boldsymbol{\theta} = (d, a, b) = (0.2, 0.3, 0.5)$ with LS of size $\zeta = 0.2$ or TS of size $\zeta = 1$.

Simulation study conclusions

Conclusions:

- In all cases, the robustly weighted MQLE outperforms the non-robustly weighted MQLE and MLE,
- Method A is superior to method B,
- Similar results in terms of MSE, MAE, bias,
- Similar results when the intervention/outliers occur in the first quarter, middle and end of the series ($\tau = [n \cdot 85\%]$).

Robust score test

$$H_0 : a = 0 \quad \text{vs.} \quad H_1 : a \neq 0$$

- Consider the partition $\boldsymbol{\theta} = (\theta^{(1)}, \theta^{(2)})$ where $\theta^{(1)} = (d, b)$ and $\theta^{(2)} = a$.
- Then, the score test is defined by

$$ST_n = [S_n^{(2)}(\tilde{\boldsymbol{\theta}}_{\text{MQLE}})]^2 / \tilde{\sigma}^2, \tag{11}$$

  where $S_n = (S_n^{(1)}, S_n^{(2)})$ and $\tilde{\boldsymbol{\theta}}_{\text{MQLE}}$ is the constrained MQLE under the null hypothesis (49), given by $\tilde{\boldsymbol{\theta}}_n = (\tilde{\boldsymbol{\theta}}_n^{(1)}, 0)$,
- Breslow (1990), Harvey (1990), Francq and Zakoïan (2010), Christou and Fokianos (2015),
- $\tilde{\sigma}^2$ is a consistent estimator of

$$\sigma^2 = W_{22} - V_{21} V_{11}^{-1} W_{12} - W_{21} V_{11}^{-1} V_{12} + V_{21} V_{11}^{-1} W_{11} V_{11}^{-1} V_{12}$$

  where V_{ij}, W_{ij}, $i, j = 1, 2$ correspond to partitions of the matrices V and W.

Robust score test

Theorem 2

Consider model (7) and assume the conditions of Theorem 1. Then, under the null hypothesis (49) we have the following:

1. Define the score test for the perturbed model (8) by ST_n^m. Then

$$ST_n^m \xrightarrow{\mathrm{d}} \chi_1^2$$

where χ_d^2 denotes the chi–square distribution with d degrees of freedom.

2. The score statistic for the perturbed model (8) and unperturbed model (7) satisfy

$$\lim_{m\to\infty} \limsup_{n\to\infty} P(|ST_n^m(\tilde{\theta}_n) - ST_n(\tilde{\theta}_n)| > \epsilon n) = 0, \quad \forall \epsilon > 0.$$

Outline of the proof:

- Francq and Zakoïan (2010, Prop. 8.3)
- Show that the following differences tend to 0: $ST_n^m(\tilde{\theta}_n) - ST_n(\tilde{\theta}_n)$, $W_{22}^m - W_{22}$, $W_{12}^m - W_{12}$, $W_{21}^m - W_{21}$, $W_{11}^m - W_{11}$, $V_{11}^{m-1} V_{12}^m - V_{11}{}^{-1} V_{12}$, $V_{21}^m V_{11}^{m-1} - V_{21} V_{11}^{-1}$ and $V_{11}^{m-1} V_{12}^m - V_{11}^{-1} V_{12}$

Robust score test

Empirical results - Size of the test

		Patch of Outliers			Isolated Outliers		
Number of outliers	Weights	Significance level			Significance level		
		$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
no outliers	none	0.003	0.047	0.082	0.009	0.055	0.113
	hat	0.003	0.046	0.084	0.006	0.055	0.114
	mve	0.008	0.049	0.102	0.012	0.056	0.104
	mcd	0.007	0.036	0.090	0.007	0.054	0.102
10 outliers	none	0.281	0.528	0.683	0.018	0.072	0.127
	hat	0.250	0.497	0.649	0.014	0.069	0.133
	mve	0.006	0.048	0.103	0.011	0.049	0.095
	mcd	0.008	0.056	0.108	0.009	0.045	0.084
20 outliers	none	0.825	0.947	0.977	0.021	0.106	0.165
	hat	0.825	0.945	0.978	0.023	0.103	0.168
	mve	0.015	0.061	0.119	0.012	0.054	0.101
	mcd	0.026	0.084	0.155	0.013	0.053	0.106

Table 2: Empirical size of the test for the case of a patch of outliers and the case of isolated outliers based on 1000 samples, $c = 1.571$.

Robust score test

Empirical results - Power of the test

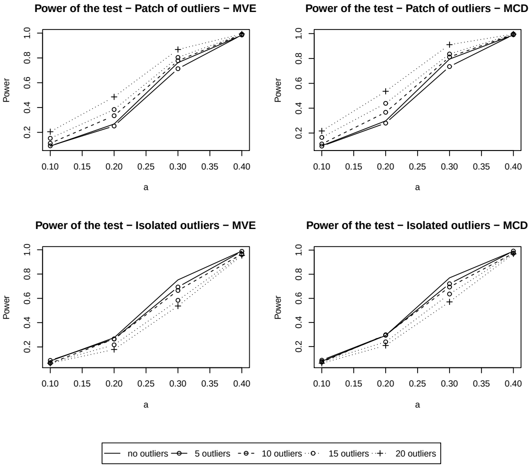

Figure 11: Power of the test statistic, $c = 1.571$ and $\alpha = 0.05$.

Measles data

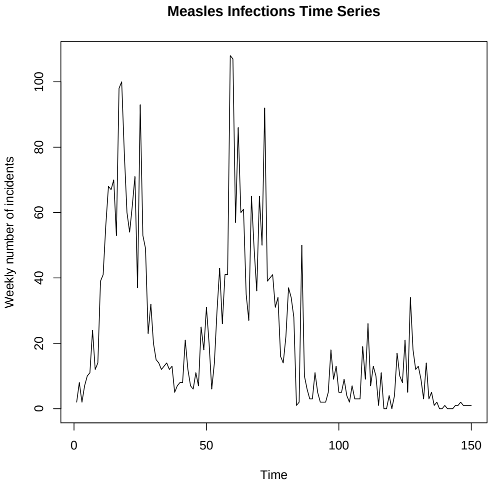

Figure 12: Weekly number of measles infections reported in North Rhine-Westphalia, Germany from January 2001 to November 2003.

Measles data

- 150 observations,
- For method B we choose M=10,
- intervention and outlier detection:
  - Chen and Liu (1993): 27 interventions including 7 AOs, 2 consecutive,
  - Fokianos and Fried (2012): 3 TSs and 5 SOs.
- log-linear model without feedback of order $q = 13$,
- we choose $c \geq 3$.

Measles data

Estimation procedure	d	a	b
MLE	0.242(0.001)	0.435(0.010)	0.500(0.009)
MQLE no weights	0.077(0.002)	0.379(0.001)	0.587(0.001)
MQLE hat (A)	0.076(0.002)	0.378(0.001)	0.588(0.001)
MQLE hat (B)	0.024(0.002)	0.309(0.001)	0.665(0.001)
MQLE mve (A)	-0.005(0.003)	0.359(0.001)	0.628(0.001)
MQLE mcd (A)	-0.035(0.003)	0.358(0.001)	0.636(0.001)
MQLE mve (B)	0.049(0.002)	0.268(0.001)	0.697(0.001)
MQLE mcd (B)	0.067(0.002)	0.255(0.001)	0.706(0.001)

Table 3: Estimates (standard errors) of the parameters of model (7) when applied to the measles infection time series. Fitting is done by employing (43) with $c = 3$.

- MQLE gives estimates with smaller standard deviation than MLE,
- method A and method B give similar results,
- the sum $a + b$ is close to 1.
- All test procedures reject the null hypothesis of non-existence of the feedback term (p-value < 0.001).

Further research

- Make our R code on the MQLE procedure for the log-linear model with feedback available through an R package.
- Investigation of MQLE for a higher order log-linear Poisson model.
- Alternative distributional assumptions.
- Other options for the ψ function in the calculations of MQLE than the Huber function.
- Study of MQLE for a linear Poisson model.

References I

Tukey, J. W. (1960)
A survey of sampling from contaminated distributions.
Contributions to probability and statistics, Stanford Univ. Press, Stanford, Calif., 448–485.

Huber, P. J. (1964)
Robust estimation of a location parameter.
Ann. Math. Statist. 35, 73–101.

Lebedev, N. N. (1972)
Special functions and their applications.
Dover Publications, Inc., New York. Revised edition, translated from the Russian and edited by Richard A. Silverman, Unabridged and corrected republication.

Hampel, F. R. (1974)
The influence curve and its role in robust estimation.
J. Amer. Statist. Assoc. 69, 383–393.

Chihara, T. S. (1978)
An introduction to orthogonal polynomials.
Gordon and Breach Science Publishers, New York-London-Paris Mathematics and its Applications, Vol. 13

Chow, Y. S. (1967).
On a strong law of large numbers for martingales.
Ann. Math. Statist. 38, 610.

Hampel, F. R. and Ronchetti, E. M. and Rousseeuw, P. J. and Stahel, W. A. (1986)
Robust statistics. The approach based on influence functions.
Wiley, Hoboken, NJ.

References II

Künsch, H. R. and Stefanski, L. A. and Carroll, R. J. (1989).
Conditionally unbiased bounded-influence estimation in general regression models, with applications to generalized linear models.
Journal of the American Statistical Association 84, 460–466.

Chen, C and Liu, LM (1993)
Joint estimation of model parameters and outlier effects in time-series.
J. Amer. Statist. Assoc. 88, 284-297.

Cantoni, E. and Ronchetti, E. M. (2001).
Robust inference for generalized linear models.
J. Amer. Statist. Assoc. 96, 1022–1030.

Kedem, B. and Fokianos, K. (2002)
Regression models for time series analysis.
Wiley, Hoboken, NJ

Ferland, R. and Latour, A. and Oraichi, D. (2006)
Integer-valued GARCH process.
J. Time Ser. Anal. 27, 923–942.

Hall, P. and Heyde, C. C. (1980)
Martingale Limit Theory and its Application.
Academic Press, New York.

Huber, P. J. and Ronchetti, E. M. (2009)
Robust statistics (Second ed.).
Wiley, Hoboken, NJ.

References III

Maronna, R. A. and Martin, R. D. and Yohai, V. J. (2006)
Robust statistics.
Wiley, Hoboken, NJ.

Fokianos, K. and Rahbek, A. and Tjøstheim, D. (2009).
Poisson autoregression.
J. Amer. Statist. Assoc. 104, 1430–1439.

Fokianos, K. and Fried, R. (2010).
Interventions in INGARCH processes.
J. Time Ser. Anal. 31, 210–225.

Fokianos, K. and Tjøstheim, D. (2011).
Log-linear Poisson autoregression.
J. Multivariate Anal. 102, 563–578.

Fokianos, K. and Fried, R. (2012).
Interventions in Log-linear Poisson Autoregression.
Stat. Model. 12, 299–322.

Szegő, G. (1975)
Orthogonal polynomials (Fourth ed.)
American Mathematical Society, Providence, R.I. Colloquium Publications, Vol. XXIII.

Taniguchi, M. and Kakizawa, Y. (2000)
Asymptotic theory of statistical inference for time series.
Springer-Verlag.

Thank you!