European Commission

Pre-processing

RegArima modelling in Tramo and in X12/X13

ESTP Training

Outline

- Why pre-processing
- RegArima model
  – Definition, estimation
- Automatic model identification (AMI)
  – Overview, seasonality tests, log/level, calendar effects, outliers, ARIMA models, final checks
- Final remarks

1. Why pre-processing?

- Exercise:
  - Effects of log/level, trading days, outliers on the final SA series (trend, irregular…)
  - Impacts using Tramo-Seats and X12-Arima

2.1 RegArima model

- *Additive case:*

$$\Delta(B)\Phi(B)(y_c - X\beta) = \Theta(B)\varepsilon$$

- *Multiplicative case:*

$$\Delta(B)\Phi(B)(\ln y_c - X\beta) = \Theta(B)\varepsilon$$

- Notations:
  - y_c : series corrected for any pre-specified effect.
  - X: any regression variable (trend constant, calendar, outliers…)
  - $\Delta(B), \Phi(B), \Theta(B)$: differencing, auto-regressive and moving average polynomials

2.2 RegArima estimation (I)

- Exact estimation

$$\Phi(B)(\Delta(B)y_c - \Delta(B)X\beta) = \Theta(B)\varepsilon$$

$$\Phi(B)\left(\widetilde{y_c} - \tilde{X}\beta\right) = \Theta(B)\varepsilon$$

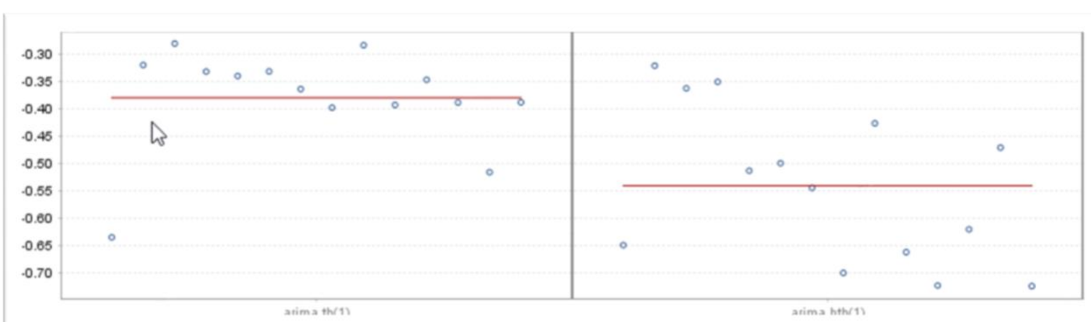
$$\widetilde{y_c} = \tilde{X}\beta + \xi, \qquad \xi \sim N(0, \sigma^2\Omega)$$

$$L^{-1}\widetilde{y_c} = L^{-1}\tilde{X}\beta + \varepsilon, \qquad \varepsilon \sim N(0, \sigma^2 I), \qquad LL' = \Omega$$

  - Differencing → Kalman filter → QR decomposition.
  - ML estimation of φ, θ by Levenberg-Marquardt (β, σ^2 concentrated out of the likelihood)
  - Residuals
    - « QR-residuals » (depends on the QR decomposition)
    - Full residuals $= L^{-1}(\widetilde{y_c} - \tilde{X}\hat{\beta})$

2.2 RegArima estimation (II)

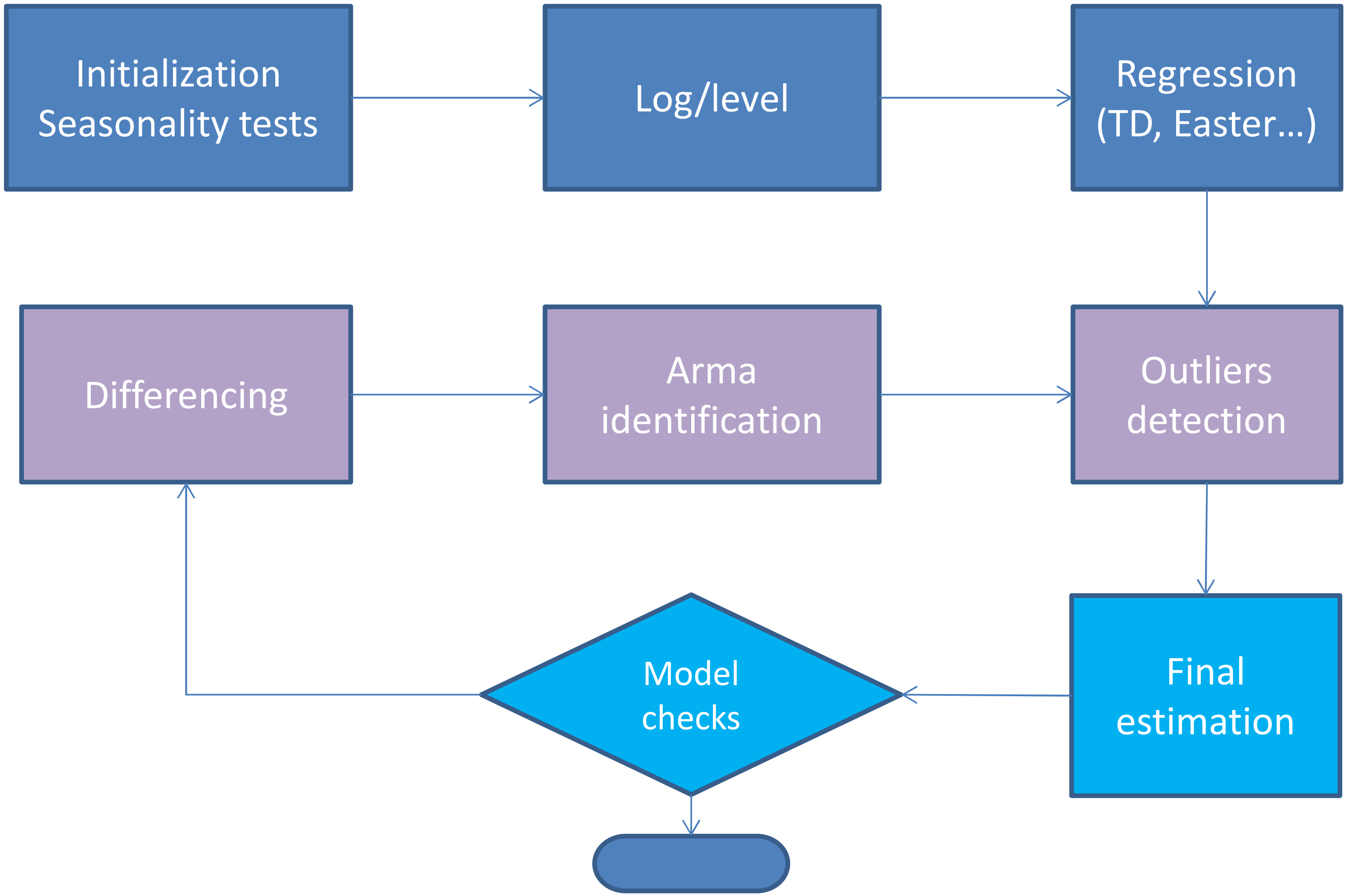
- Fast estimation
  - Differencing → OLS (QR decomposition) → Residuals (e_t)
  - Estimation of φ, θ by Hannan-Rissanen
    - $e_t = \sum_{i=1}^{n} e_{t-i} + a_t$
    - $e_t = -\sum_{i=1}^{p} \varphi_i e_{t-i} + \sum_{i=1}^{q} \theta_i a_{t-i} + \epsilon_t$
  - Not a real problem: why trying to estimate exact parameters when we know that they are unstable ?
    - Impact of the parameters on the regression coefficients, on the seasonally adjusted series ?



2.2 RegArima estimation (III)

- X12-Arima
  - Exact ML estimation
- Tramo
  - Fast estimation in most intermediary steps
  - Exact ML estimation in final steps

3.1 AMI: simplified schema

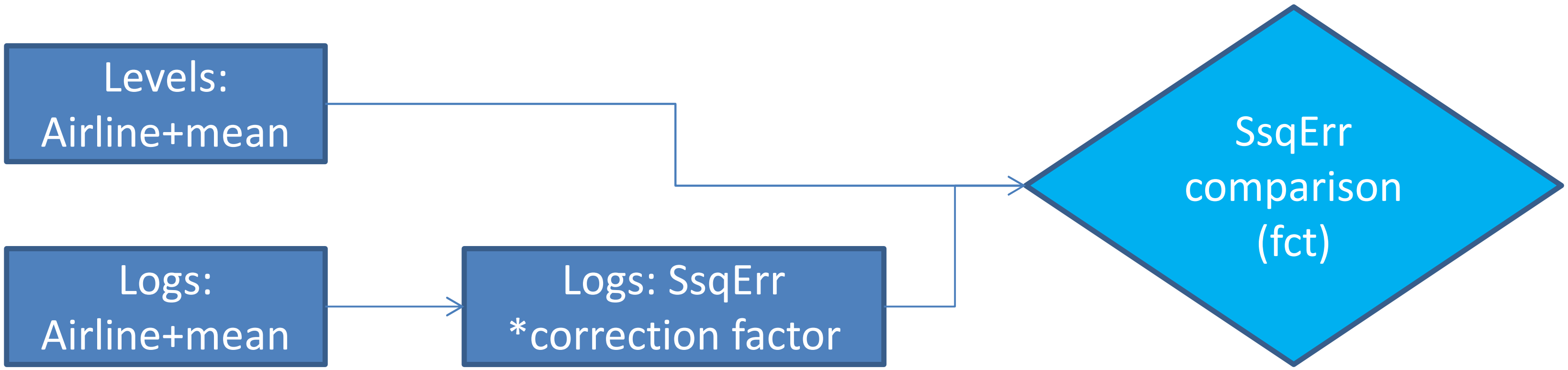


3.2 Seasonality tests

- Only in the last versions of Tramo and in JD+
- Initial tests
  - Ljung-Box test: auto-correlations at seasonal lags
    - Correlations between (y(t), y(t-s)), (y(t), y(t-2*s))
  - Friedman (non parametric) test

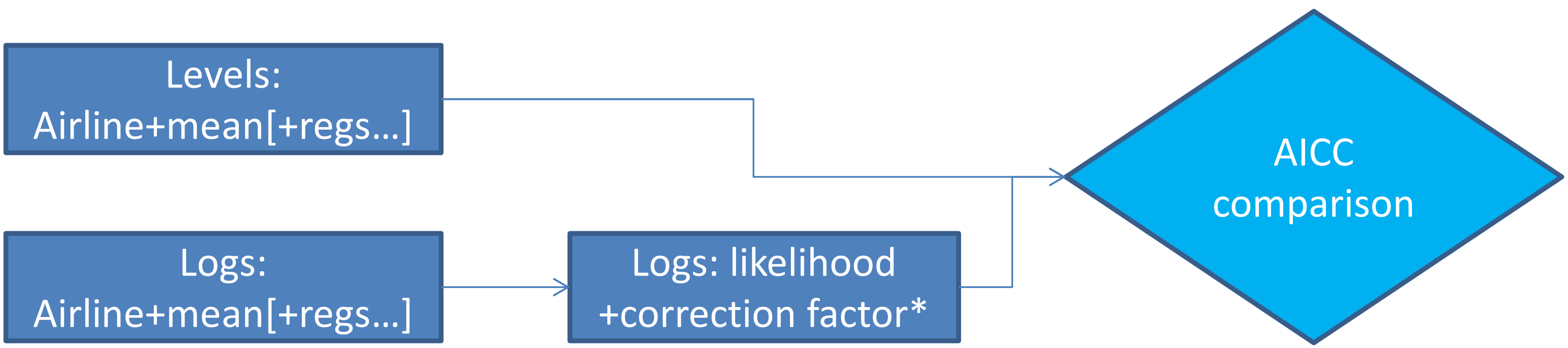
3.3 Log-level test (I)

Tramo



$$SsqErr \propto likelihood$$

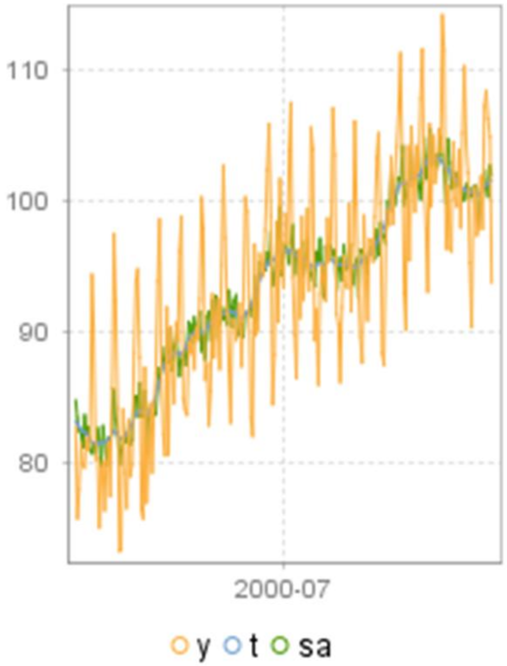
X13



$$AICC = -2 \cdot (loglikelihood - n \cdot np/(n - np - 1))$$

3.3 Log-level test (II)

- Tramo
  - Levels: mean + airline
    - $Sslevels$ =7.736395
  - Logs: mean + airline
    - $Sslogs \cdot gmean(levels)^2$ = 7.768457
  - $logs - levels = -0.03206 > \log(.95) = -0.05129$



- X13
  - Levels: lp + td + mean + airline
    - Loglikelihood = -463.920346919396
    - AICC= 938.1264081245064
  - Logs (+lp adjust) : td + mean + airline
    - Adjusted loglikelihood = -469.5936870686463
    - AICC = 947.3769475970083

3.4 Regression (calendar) tests (I)

- Tramo
  - Legacy code:
    - T-tests on trading days, leap year, Easter variable
  - New tests:
    - F-tests on trading/working days,
    - T-tests on trading days, leap year, Easter variable
- JD+
  - Also Wald tests on trading/working days

```
double fdel = (td1Stats.SsqErr - td6Stats.SsqErr) / (5 * sigma);
if (fdel > 0) {
    fstat.setDFNum(5);
    pdel = fstat.getProbability(fdel, ProbabilityType.Upper);
}
```

3.4 Regression (calendar) tests (II)

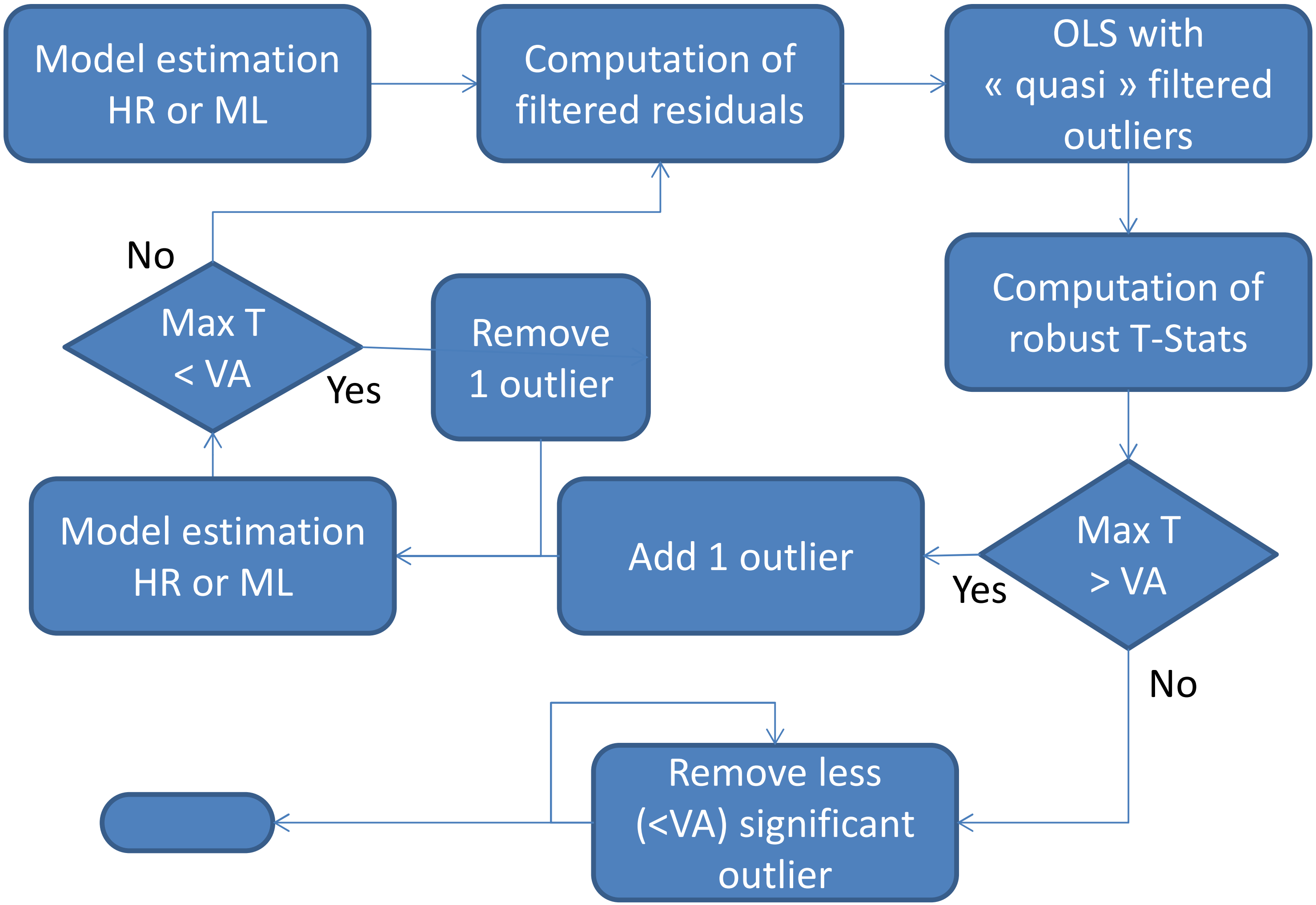
- X12
  - AIC comparison between:
    - Models with and without trading days
    - Models with different Easter variables (length=1, 8, 15)
    - Models with other regression variables

```
public void calc() {
    double n = effectiveObservationsCount;
    double np = estimatedParametersCount;
    double ll = adjustedLogLikelihood;
    double nll = logLikelihood;
    AIC = -2 * (ll - np);
    HannanQuinn = -2 * (ll - np * Math.log(Math.log(n)));
    AICC = -2 * (ll - (n * np) / (n - np - 1));
    BIC = -2 * ll + np * Math.log(n);
    BIC2 = (-2 * nll + np * Math.log(n)) / n;
    BICC = Math.log(SsqErr / n) + (np - 1) * Math.log(n) / n; // TRAMO-like
}
```

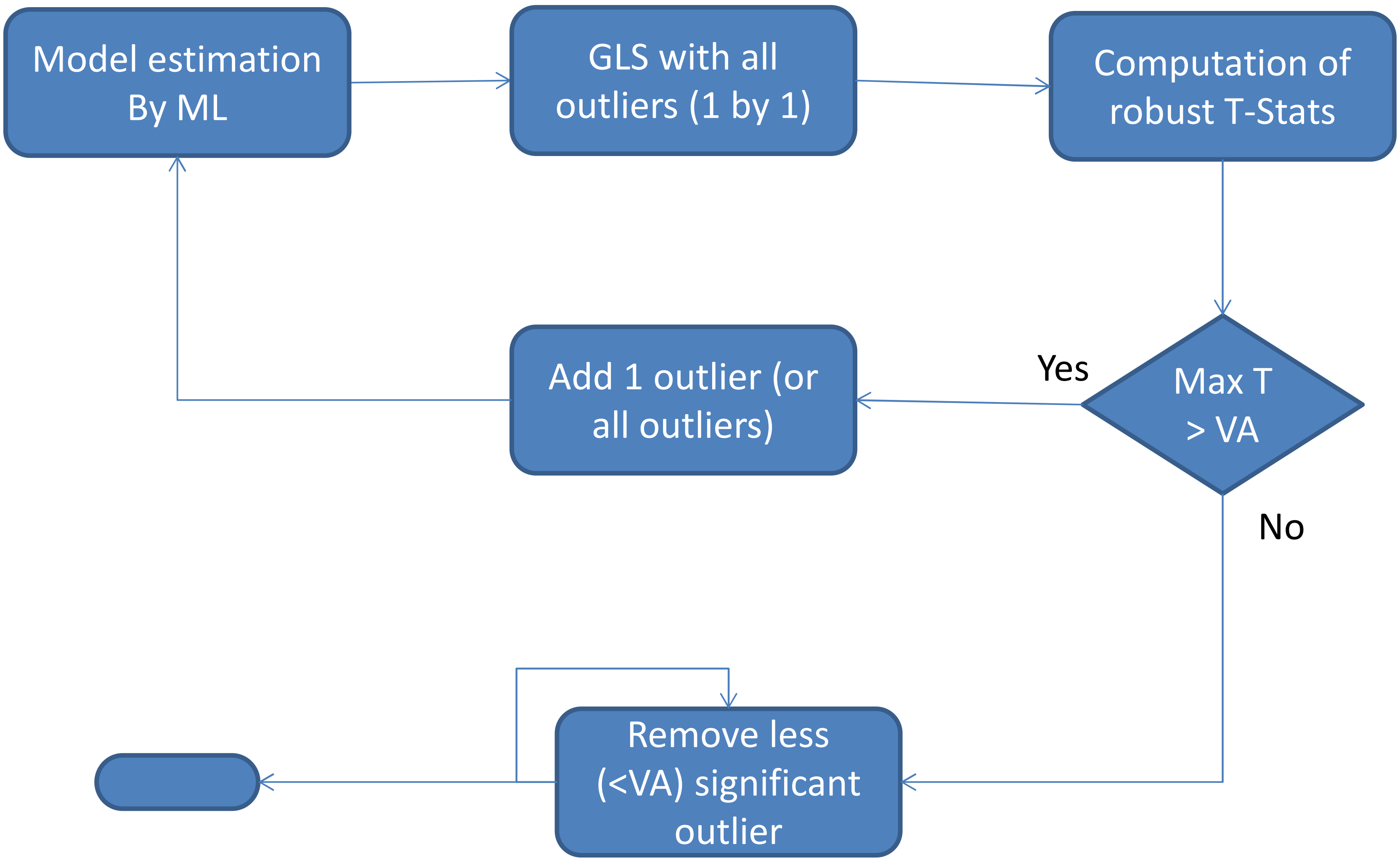
3.5 Outliers detection

- Types of outliers
  - AO, LS, TC, SO (SLS and IO not supported)
- Impact of outliers
  - Estimated parameters
    - Forecasts…
  - Specific SA issues
- Solutions
  - ML-based solutions
  - [Robust estimates]
- Issues
  - Unstable algorithms
  - « Masking » effects

3.5 Outliers detection in Tramo



3.5 Outliers detection in X12



3.5 Outliers detection. Key points

- ML or HR estimation
- Computation of robust stdev on the residuals
- Regression on residuals or on complete models
- Back calculation to identify masking effects
- Exact or approximate GLS estimation

3.61 Differencing

- Tramo ~ X12
- Step 1:
  - Estimate (2 0 0)(1 0 0)+mean
  - Select root(s) : 1/|r| > initial UR (0.97)
- Step 2:
  - Estimate (1 x 1)(1 y 1)+mean
  - Cancel similar AR, MA roots (cancel)
  - Select root(s) : 1/|r| > final UR (0.91)

3.62 ARMA identification

- Tramo ~ X12
- Computes BIC (Tramo-like) for different ARMA orders
  - Estimation of the models:
    - Tramo: Hannan-Rissanen
    - X12: exact ML
- Best (acceptable) model selected

3.7 Final checks

- **Suppression of non-significant ARIMA parameters**
- **Regular/seasonal under-differencing (quasi-unit roots)**
- Seasonality control
- Residual trading days (F-test on the residuals)
- Benchmarking with reference (airline) models
  - Criterions: BIC , outliers, LjungBox, Seasonal LjungBox, Skewness

4. Final (personal) remarks

- Wonderful “expert system”
- But…
  - Several known weaknesses
    - Log/level in case of large outliers
    - Trading days selection (Tramo)
    - Outliers (not robust enough)
    - Selection of non decomposable models
    - …
- But…
  - Not easy to outperform the current algorithms