Statistics and Information Technologies for Anti-Fraud, Security and Trade

Domenico Perrotta

EC Joint Research Centre

Benford's Law for fraud detection
Foundations, methods and applications

Stresa, 10-12 July 2019

Legal framework:
the EU mandate on trade, customs and anti-fraud

Treaty on the functioning of the EU:

The EU and the Member States shall counter fraud affecting the financial interests of the EU.

The Union shall have exclusive competence in the following areas: (a) customs union; <...> (e) common commercial policy.

Legal framework:
the EU mandate on trade, customs and anti-fraud

Council Regulations 515/97, 2988/95 and 2913/92 strengthen the cooperation between EC and Member States.

Directive COM(2012) 363 asks the EC to provide technical assistance to the Member States in the coordination of their investigations.

The current Commission with its recent **anti-fraud strategy** COM(2019) 196, acknowledge the role of the JRC in collecting and analyzing relevant data, in view to identify suspicious activities.

Data analysis in the international anti-fraud/trade context: recent position of the WCO

World Customs Organization
Organisation Mondiale des Douanes

World Customs Organization declares 2017 to be the year of Data Analysis

The Secretary General of the WCO, Kunio Mikuriya, announced today that 2017 will be dedicated to promoting data analysis under the slogan “Data Analysis for Effective Border Management.” WCO Members will thus be called upon to further promote their efforts and initiatives in a sector that is becoming a key element in Customs modernization process: collecting and analysing data.

European Commission

Data analysis in the international anti-fraud/trade context: JRC and OLAF (UCLAF) affirmed the need more than 20 years ago

From:

Special Report No 8/98 on the Commission's services specifically involved in the fight against fraud, notably the 'unité de coordination de la lutte anti- fraude' (UCLAF)

Official Journal C 230 , 22/07/1998 P. 0001 - 0044

Setting priorities

2.16. UCLAF needs a basis for deciding which cases should receive priority in the allocation of the limited resources for inspections to sensitive areas and 'high-risk` operators and/or recipients. This in turn requires criteria for optimising the use of the diverse information at the disposal of the unit. UCLAF concluded, on 8 December 1995, an administrative arrangement with the Joint Research Centre (JRC) to carry out a study on:

- the development and application of pattern recognition methods for fraud cases stored in the Commission's databases IRENE and PRE-IRENE;

- the assessment of risk parameters concerning transactions financed or co-financed from the Community budget and;

- the description of trends over a given period of time and detection of heterogeneity in the amounts of subsidies in reference transactions and IRENE cases.

Data analysis in the international anti-fraud/trade context: The Automated Monitoring Tool concept, more than 15 years ago

European Antifraud Office, Sixth Activity Report
for the period 1 July 2004 to 31 December 2005

3.1.2. Operational Intelligence

Operational intelligence includes specialist support and assistance to OLAF investigators on internal and external investigations. The team is sometimes also asked to assist the Member States.

• • • •

Major projects carried out by the operational intelligence analysts include:

- Gradual deployment of an automated monitoring tool. The Automated Monitoring Tool (AMT) automatically monitors a range of trade statistics in order to produce "alerts" which are triggered when changes in statistical patterns hit preset thresholds. For example

The defense of the EU budget:

fight against undervaluation

EUROPEAN COMMISSION

PRESS RELEASE

Brussels, 12 September 2014

Operation SNAKE: EU and Chinese customs join forces to target undervaluation of goods at customs

EU and national authorities prevented losses of over €80 million in customs duties, during a major joint customs operation (JCO) coordinated by the European Anti-Fraud Office (OLAF). This joint customs operation had particular significance as, for the first time ever, it also involved Chinese customs authorities. Operation "SNAKE" specifically targeted the undervaluation of imported goods, which causes huge losses to public budgets every year. Over a one month period, OLAF and the participating customs authorities detected more than 1,500 containers where the declared customs value was heavily undervalued. This included false descriptions of goods, false weights and quantities, and counterfeit goods. In addition, customs authorities succeeded in identifying several so-called missing traders and non-existent importers, triggering a number of criminal and administrative investigations in several countries.

The defense of the EU budget: fight against undervaluation

EU warns UK-centered China import scam may shift to Europe's 'Silk Road'

Francesco Guarascio

4 MIN READ

BRUSSELS (Reuters) - European Union anti-fraud investigators suspect Greece and Hungary may have become the main EU centers of a multi-million-euro scam involving imports of Chinese clothing and footwear that uses the infrastructure of China's new "Silk Road".

OLAF Uncovers Large-scale Frauds in Textile Imports, €300 Mn Lost in Slovakia

7. septembra 2018 19:54

Bratislava, September 7 (TASR) - During its investigation into textile and footwear imports from China to EU, the European Anti-fraud Office (OLAF) uncovered customs frauds totalling €2.2 billion, with goods worth roughly €300 million of the sum entering EU via Slovakia, Dennik N daily reported on Friday.

Greece faces €200M fine for failing to stop Chinese fraud network

Customs officials failed to halt big scheme to avoid import duties and tax, investigators say.

By **SIMON MARKS** | 1/14/19, 5:51 PM CET | Updated 4/19/19, 1:22 AM CET

EU anti-fraud investigators are demanding that Greek customs pay more than €200 million for failing to act against a major Chinese fraud network dumping ultra-cheap clothing and footwear in Europe.

European Commission

The defense of the EU budget: fight against undervaluation

3.1. Detecting and investigating revenue fraud: OLAF at the centre of large-scale investigations into the undervaluation of goods imported into the EU

To understand the phenomenon, OLAF carried out an extensive analysis of all customs declarations for all imports of textiles and shoes from China between 2013 and 2016. A "cleaned average price" was calculated for each category of textiles and shoes imported from China, based on the value of all import declarations in the EU between 2013 and

European Commission

Setting baseline values for the trade prices
References in legal documents

48. To overcome the risk of undervaluation, the Commission has developed a methodology to estimate "fair prices"[22], applying a statistical procedure to COMEXT[23] data, in order to produce robust estimates for the prices of the imported goods[24]. OLAF disseminates these estimates among Member States' customs authorities.

(**22**) Also known as Outlier-Free Average Prices. These are statistical estimates calculated for the prices of traded products on the basis of outlier-free data.

Data sources: COMEXT

Monthly aggregates
of quantities and values
for each Product, Origin and Destination

PRODUCT	PARTNER	DECLARANT	PERIOD	VALUE_1000EURO	QUANTITY_TON	SUP_QUANTITY
61045200	CN	GR	"2015-07-01"	1.4800	0.0500	300
61045200	CN	GR	"2016-02-01"	4.8900	0.5100	4191
61045200	CN	GR	"2016-11-01"	1	0	5
61045200	CN	HR	"2013-07-01"	0.1800	0.0100	28
61045200	CN	HR	"2013-08-01"	1.0300	0.0500	357
61045200	CN	HR	"2013-10-01"	0.7000	0.0300	268

A public EU database

Imports: about 6.000.000 records per year

Data sources: Surveillance

Daily aggregates or customs declarations
of quantities and values
for each Product, Origin and Destination

1 Date	2 Issuer	3 Procedure	4 Origin	5 CNCode	6 Volume	7 StatValue	8 SupplUnit	9 UnitPrice	10 PerKGPrice
'31/12/2016'	'NL'	'40'	'TR'	'6101201000'	1	34	2	17	34
'31/12/2016'	'PL'	'40'	'SG'	'6101201000'	4	85.0800	2	42.5400	21.2700
'01/12/2016'	'FR'	'42'	'CN'	'6101201000'	1.4740	52	8	6.5000	35.2782
'02/12/2016'	'FR'	'42'	'CN'	'6101201000'	1.6580	59	9	6.5600	35.5850
'02/12/2016'	'GB'	'42'	'CN'	'6101201000'	9933	5.8087e+03	15152	0.3800	0.5848
'03/12/2016'	'GB'	'42'	'CN'	'6101201000'	7299	2.5955e+03	12143	0.2100	0.3556
'04/12/2016'	'GB'	'42'	'CN'	'6101201000'	5315	4.3387e+03	4370	0.9900	0.8163
'05/12/2016'	'BE'	'42'	'PK'	'6101201000'	6239	5.6357e+04	10820	5.2100	9.0330

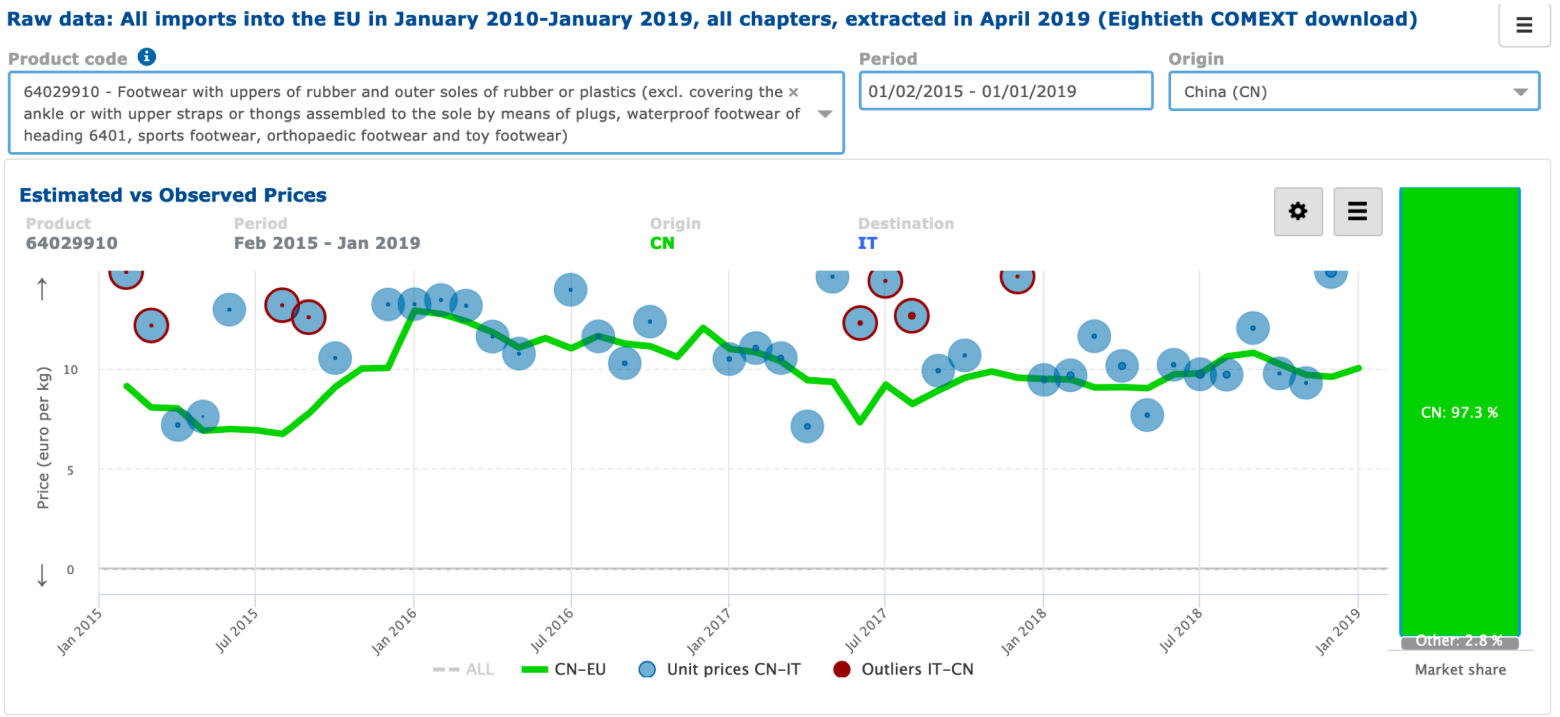
A restricted EU database

About 4.500.000 daily aggregates per year
only for imports of textiles

The defense of the EU budget:

fight against undervaluation

64029910 - footwear: import prices observed in **Italy** vs. estimated prices

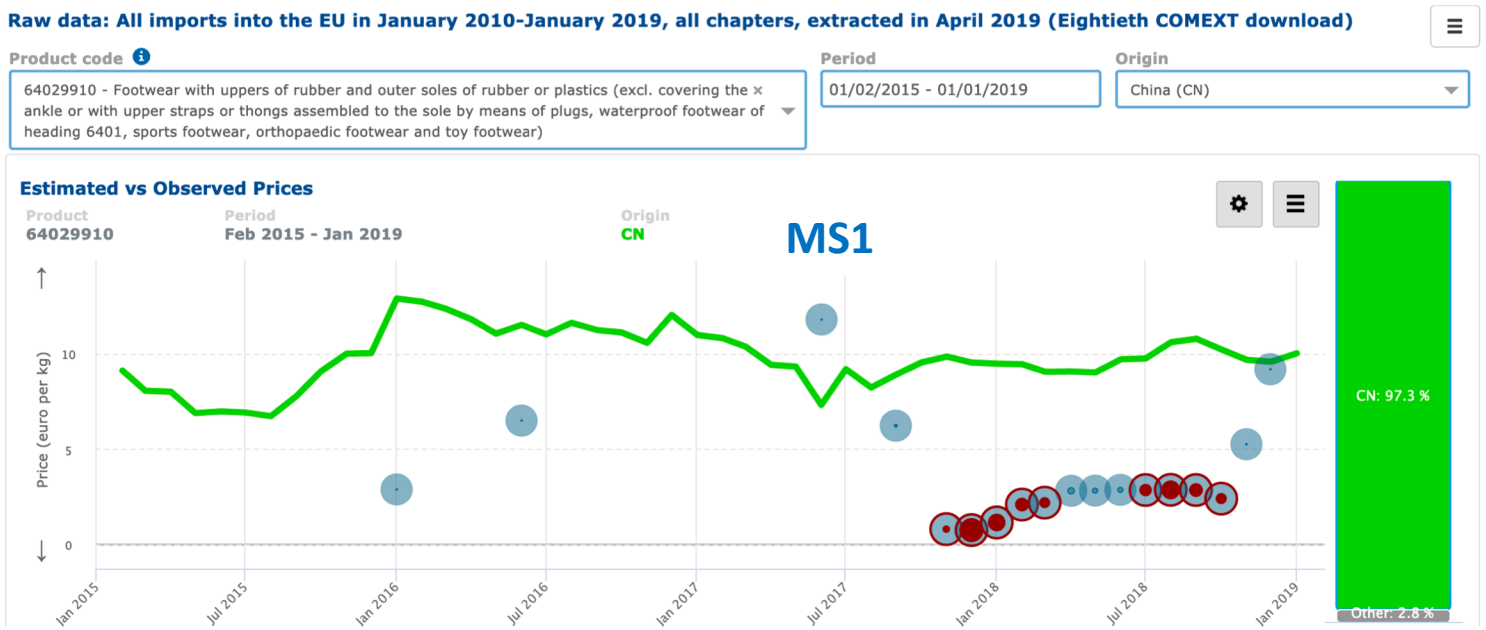


Nothing strange

European Commission

The defense of the EU budget:

fight against undervaluation



64029910 - footwear:
import prices
observed in **MS1**
vs.
estimated prices

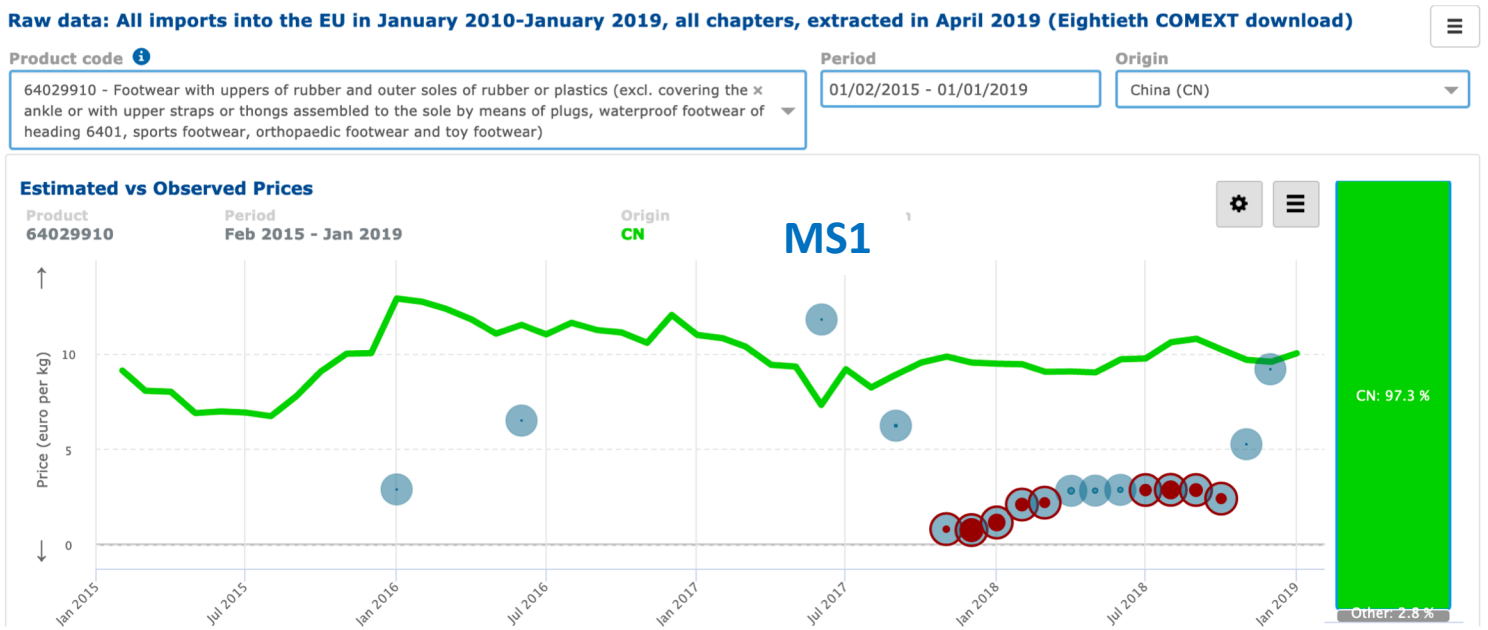
Almost no trade before October 2017
Then, persistent undervaluation

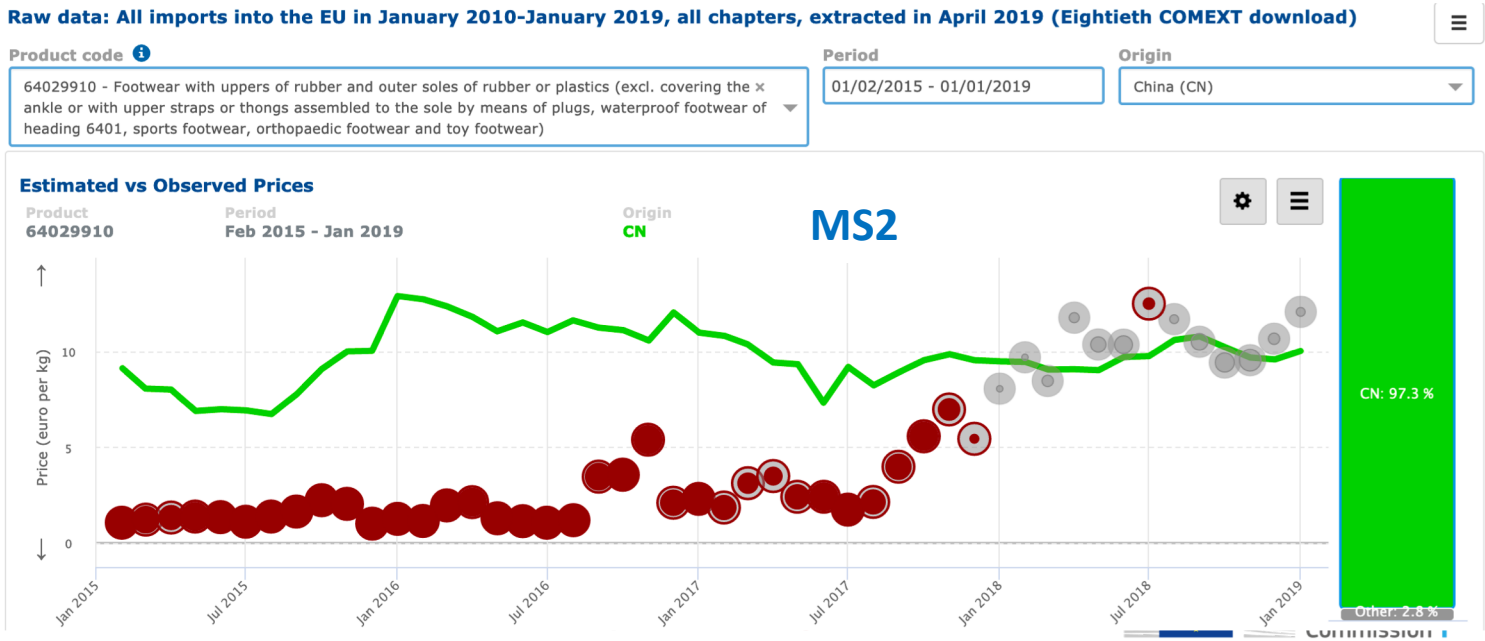
European Commission

The defense of the EU budget:

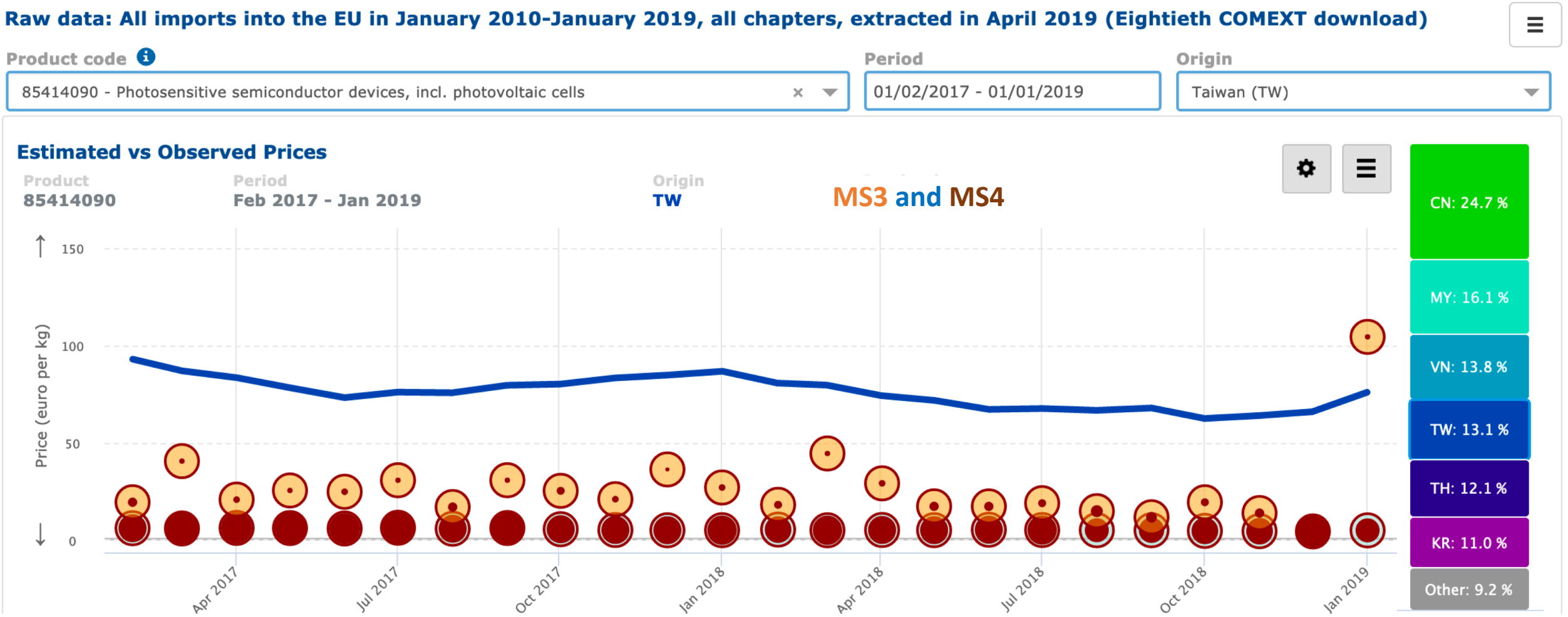
fight against undervaluation

64029910 - footwear: import prices observed in **MS1** and **MS2** vs. estimated prices





Undervaluation: “solar panels”



Undervaluation: “solar panels”

May 2018

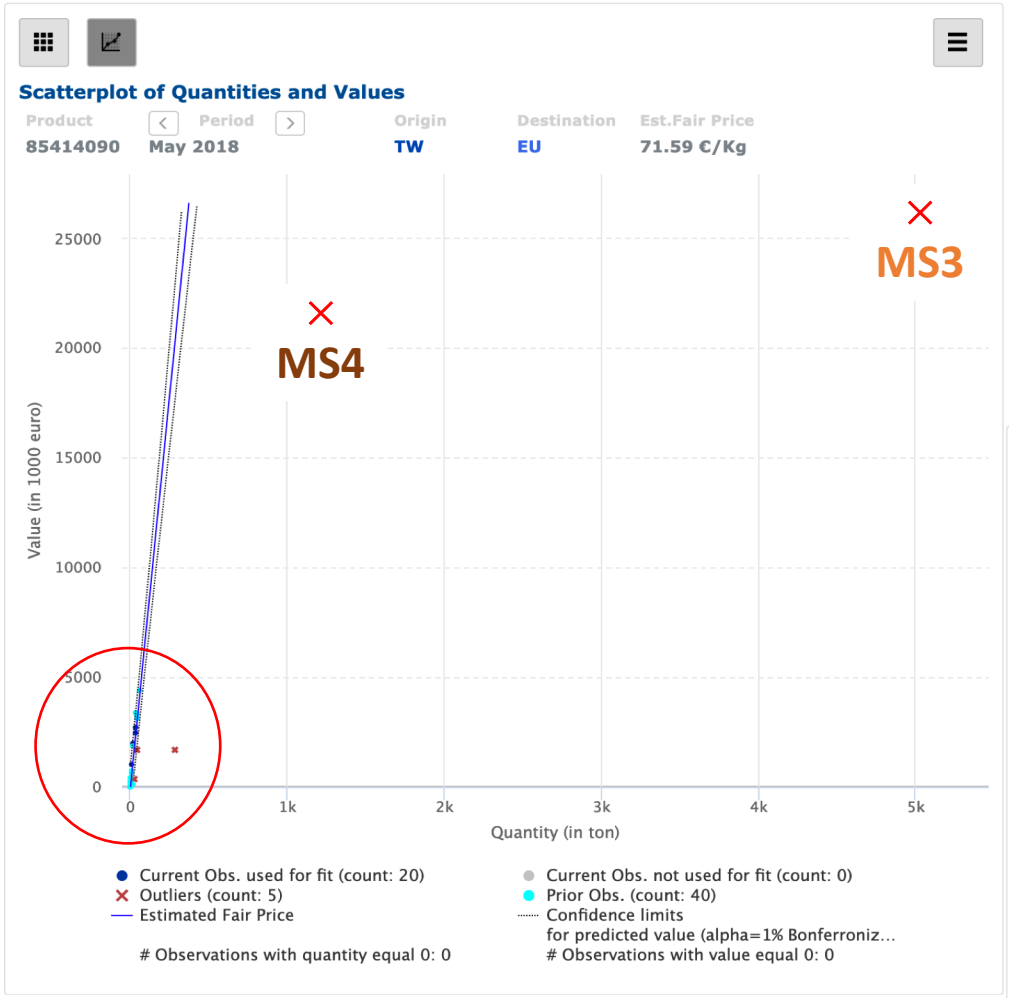
MS3 4.91 €/Kg

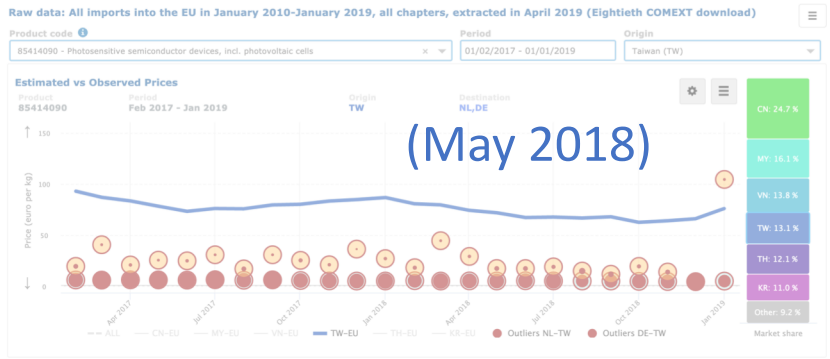
MS4 17.16 €/Kg

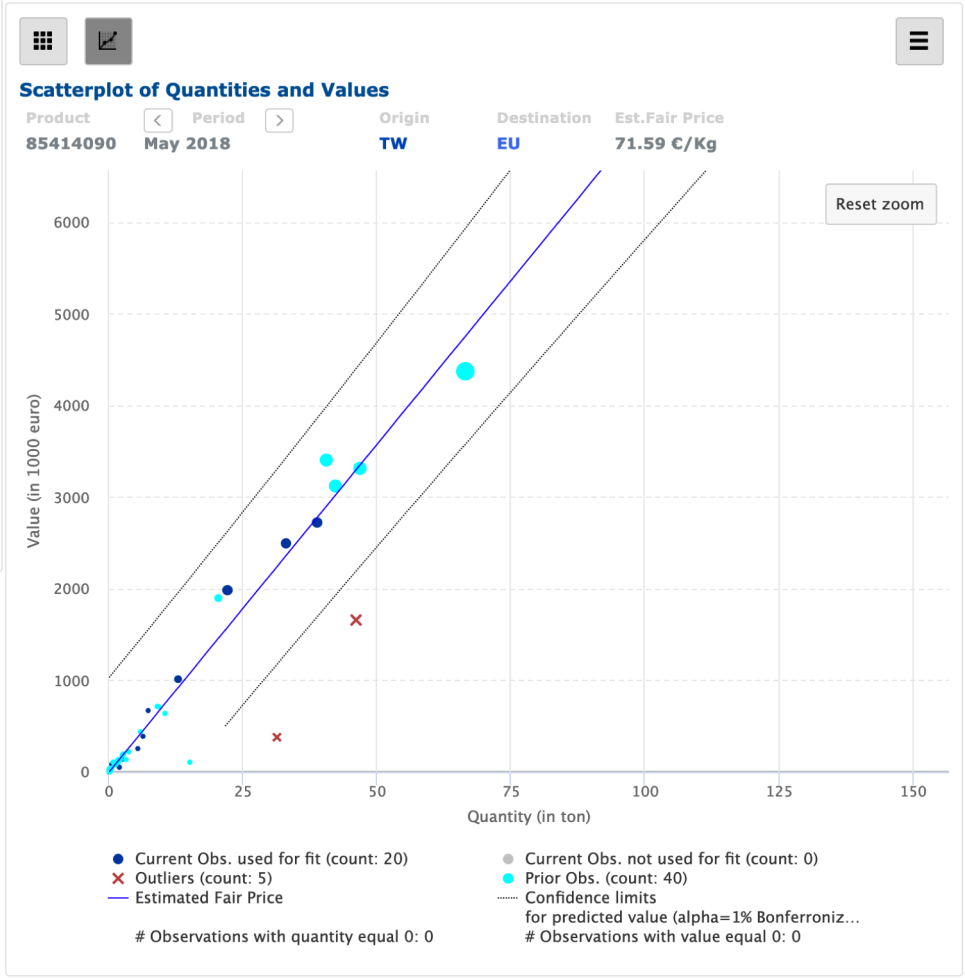
Vs

estimated

71.59 €/Kg





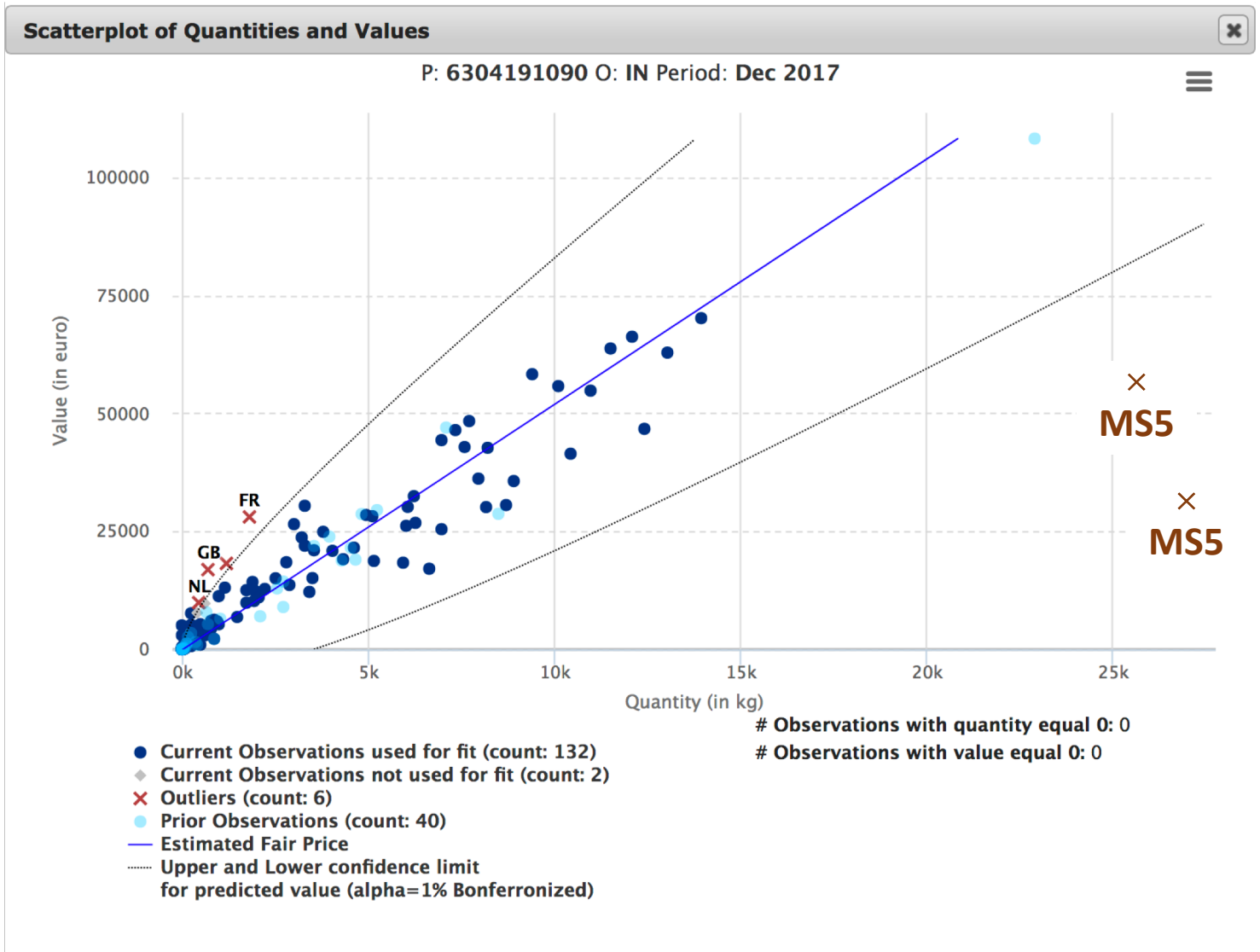


Undervaluation of “bedspreads of cotton”

heteroscedastic model on declaration-level data

estimated price: 5.16 €

MS5 declared price: 0.97 & 1.81 €



JRC approach to undervaluation: some references

Spyros A., Perrotta, D., Torti, F. (2015); *The* ***Estimation Of Fair Prices*** *Of Traded Goods From Outlier-Free Trade Data*; EUR27696 EN; doi:1 0.2 788/57125

Atkinson, A. C., Corbellini, A., Riani, M. (2017). ***Robust Bayesian regression*** *with the forward search: theory and data analysis*. Test, 26(4), 869-886.

Atkinson A.C., Riani M., Torti F., (2016). *Robust methods for* ***heteroskedastic regression***, Computational Statistics and Data Analysis 104, p. 209–222

Cerasa, A. (2016). *Combining* ***homogeneous groups*** *of pre-classified observations with application to international trade*. Statistica Neerlandica, 70(3), 229-259;

Cerasa, A., & Cerioli, A. (2017). *Outlier-free* ***merging of homogeneous groups*** *of pre-classified observations under contamination*. Journal of Statistical Computation and Simulation, 87(15), 2997-3020.

Anti-dumping and the deflection of trade: again “solar panels”

European Commission

English EN

Home › Trade › Policy › Accessing markets › Trade defence

Investigations

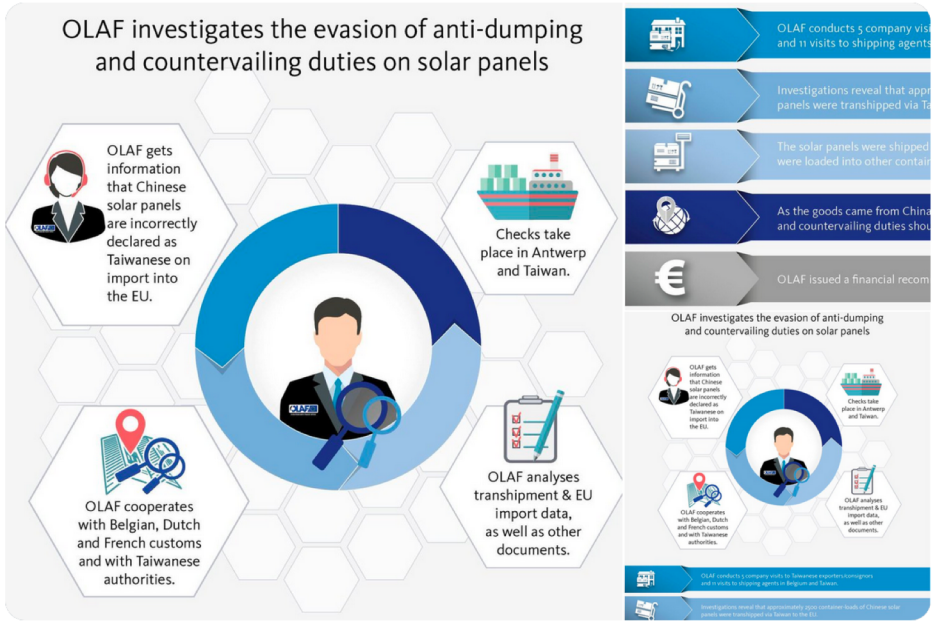
History of proceeding

Type of proceeding:	Anti-dumping
Product(s):	**Solar panels (Crystalline silicon photovoltaic modules and key components)**
CN(s):	Not available
Countries investigated:	People's Republic of China

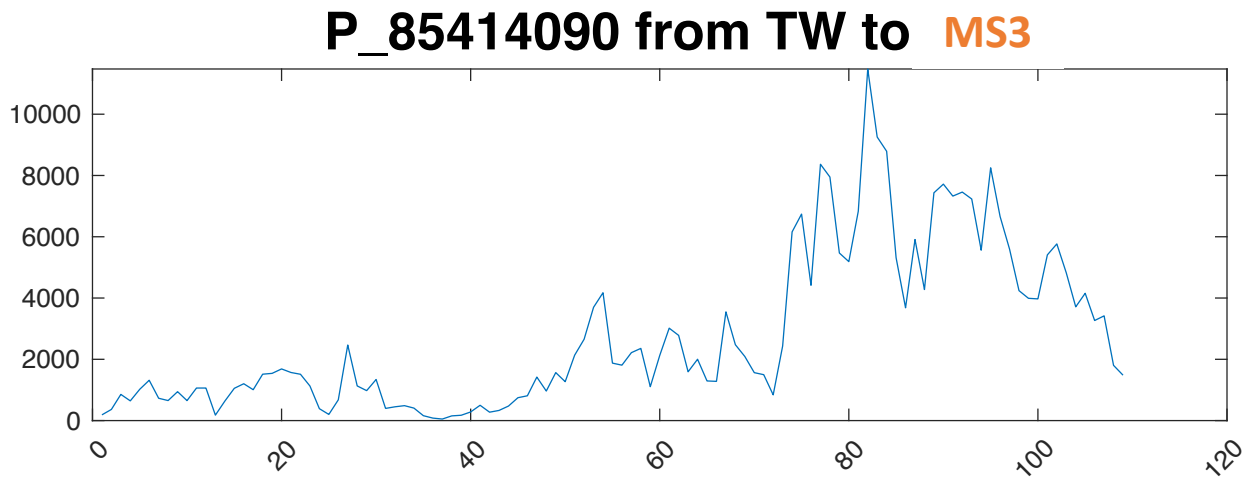
OLAF Press Office
@OLAFPress

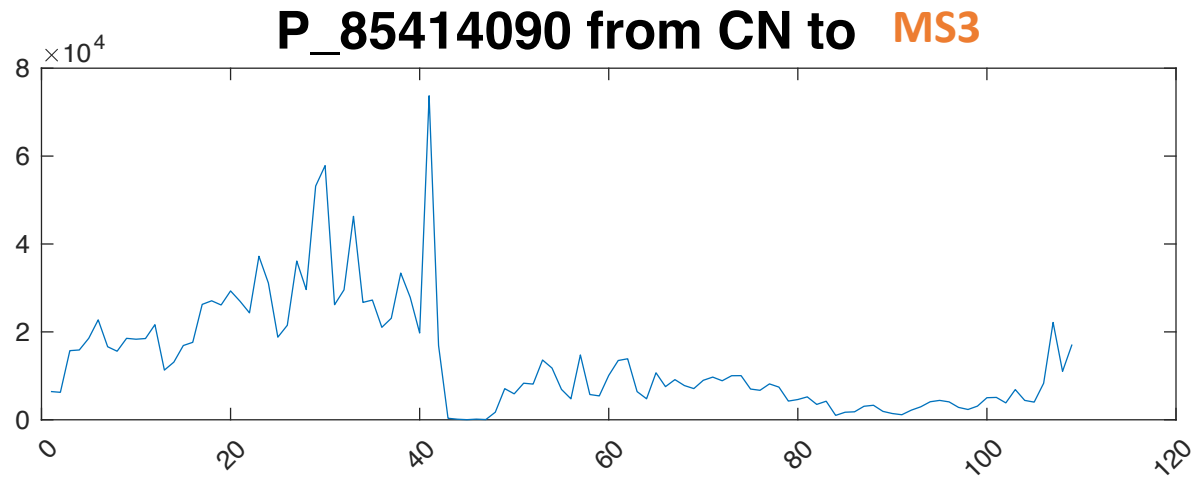
Follow

€228 million recommended recovery in cases of evasion of #anti-dumping duties on solar panels from #China. In one case, #OLAF uncovered misdeclared goods in cooperation with #Belgiancustoms, @Douane, @douane_france & #Taiwanese authorities



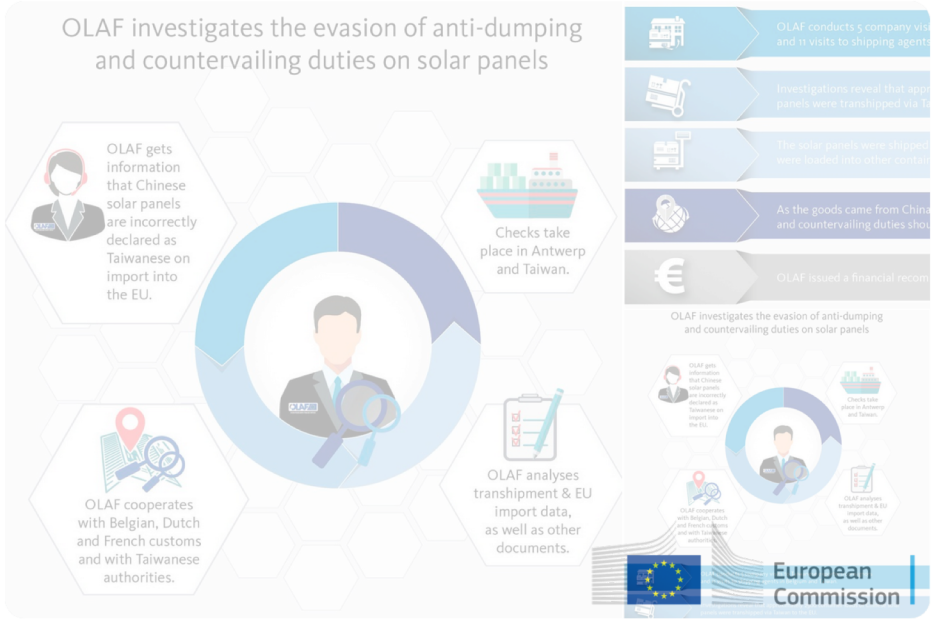
Anti-dumping and the deflection of trade: again “solar panels”





OLAF Press Office
@OLAFPress
Follow

€228 million recommended recovery in cases of evasion of #anti-dumping duties on solar panels from #China. In one case, #OLAF uncovered misdeclared goods in cooperation with #Belgiancustoms, @Douane, @douane_france & #Taiwanese authorities



Anti-dumping and the deflection of trade

Statistical purpose: provide a robust unified framework to treat simultaneously outliers, unknown level shifts and changes in the seasonal pattern

Anti-fraud purpose: identify situations in which a sudden reduction in trade volume for one country of origin matches an increase for another, which would indicate a potential **miss-declaration of origin or product** and a consequent **deflection of trade**.

Approach developed:

Rousseeuw, P.J., Perrotta, D., Riani, M., Hubert, M. (2019). Robust monitoring of time series *with application to fraud detection. Econometrics and Statistics, Volume 9, Pages 108-121.*

Model - we assume y_t with:

- A polynomial trend

$$y_t \sim \sum_{a=0}^{A} \alpha_a t^a$$

- A seasonal component

$$y_t \sim S_t = \sum_{b=1}^{B} \left(\beta_{b,1} \cos\left(\frac{2\pi b}{12} t\right) + \beta_{b,2} \sin\left(\frac{2\pi b}{12} t\right) \right)$$

- The amplitude of the seasonal component may vary over time in a polynomial way

$$y_t \sim \left(1 + \sum_{g=1}^{G} \gamma_g t^g\right) S_t$$

- A level shift in unknown time point

$$2 \leqslant \delta_2 \leqslant T, \text{ i.e. } y_t \sim \delta_1 \, I(t \geqslant \delta_2)$$

Model

- General equation

$$y_t = \sum_{a=0}^{A} \alpha_a t^a + \left[\sum_{b=1}^{B} \left(\beta_{b,1} \cos\left(\frac{2\pi b}{12} t\right) + \beta_{b,2} \sin\left(\frac{2\pi b}{12} t\right)\right)\right] \left(1 + \sum_{g=1}^{G} \gamma_g t^g\right) + \delta_1 I(t \geqslant \delta_2) + \varepsilon_t$$

- Parameter vector θ (of length p)

$$\{\alpha_0, \alpha_1, \ldots, \beta_{1,1}, \beta_{1,2}, \ldots, \gamma_1, \gamma_2, \ldots, \delta_1, \delta_2\}$$

Estimation method

- Non linear least trimmed squares estimator (our default: *h=0.75 T*)

$$\hat{\theta}_{\mathrm{NLTS}} = \underset{\theta}{\operatorname{argmin}} \sum_{t=1}^{h} r^2_{(t)}(\theta)$$

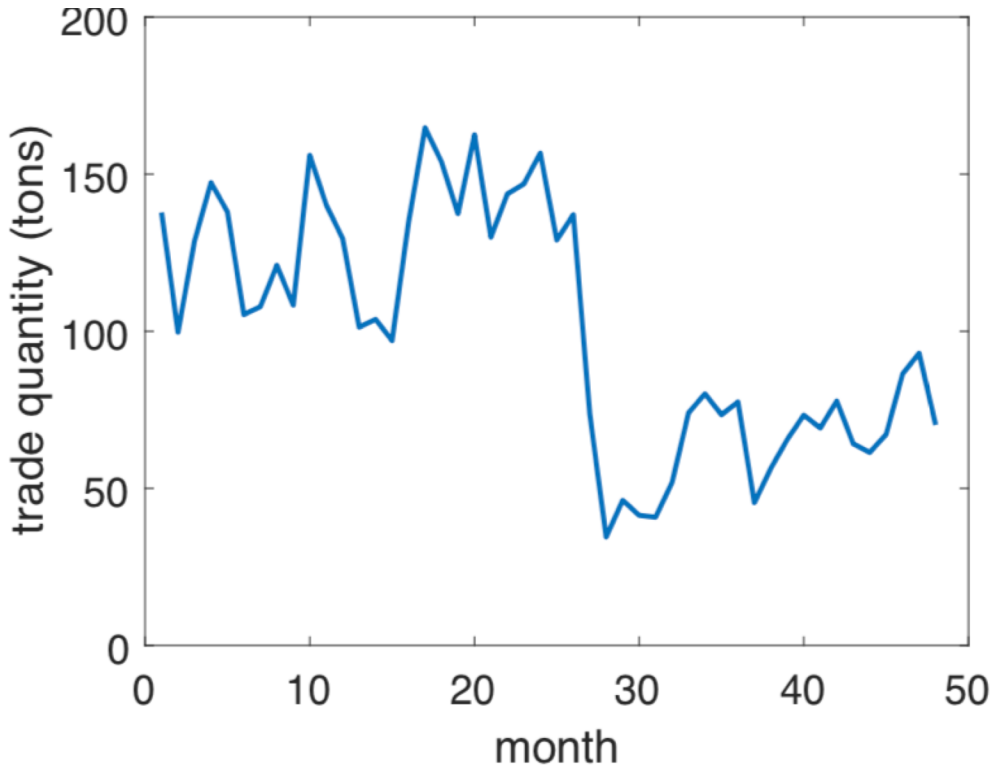
- We combine ideas from **Fast-LTS** for linear regression (Rousseeuw and Van Driessen, 1989) with alternating least squares (**ALS**)

$r^2_{(t)}(\theta)$ the t-th smallest squared residual

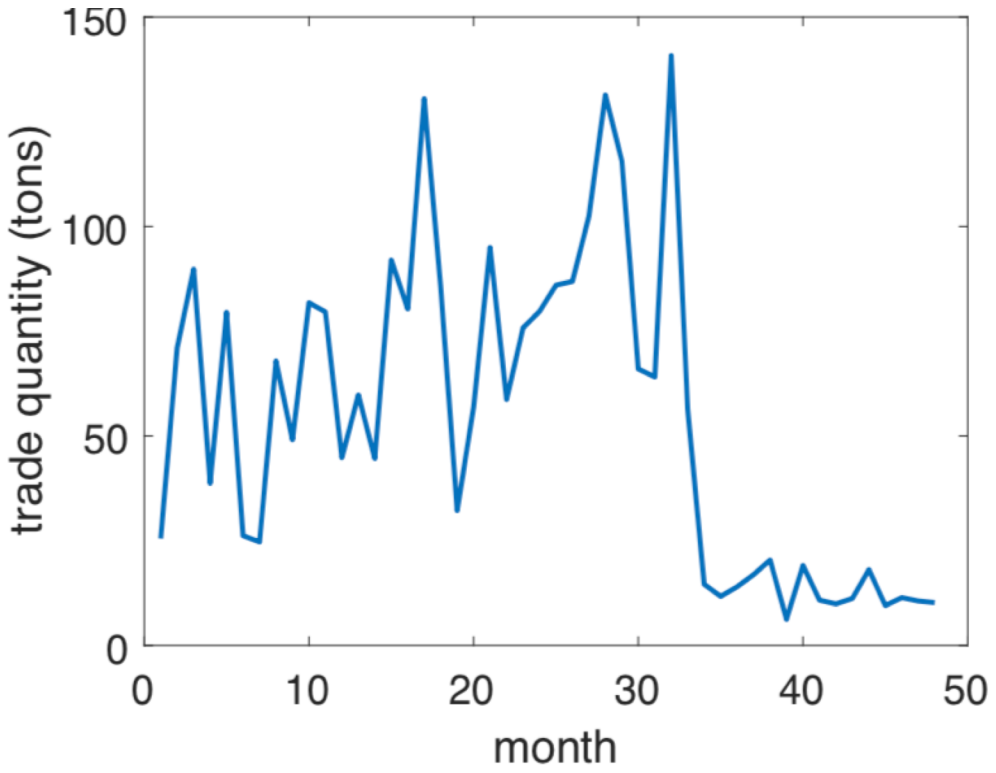
Two applications on COMEXT time series

***Potential* miss-declaration of origin or product** and consequent **deflection of trade:**

sudden reduction in trade volume for an exporting country, matching an increase for another



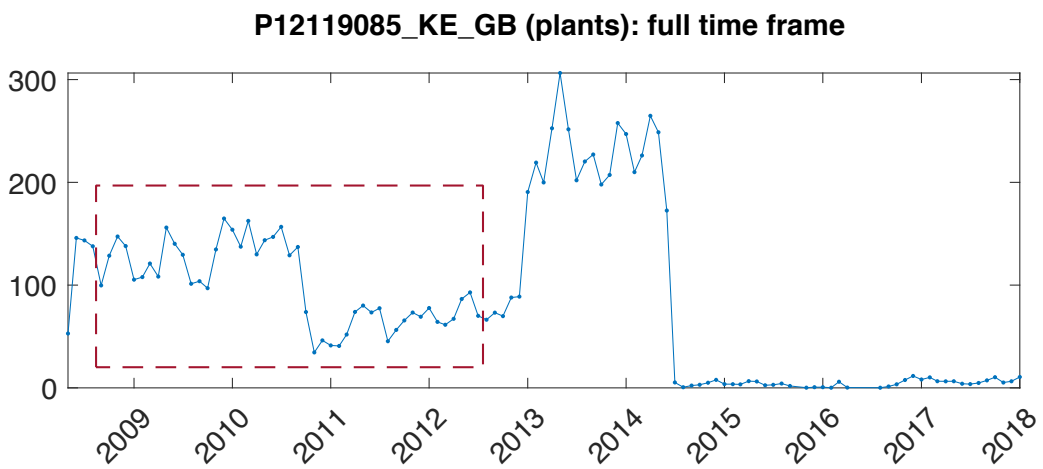
Imports of plants
from Kenya to UK

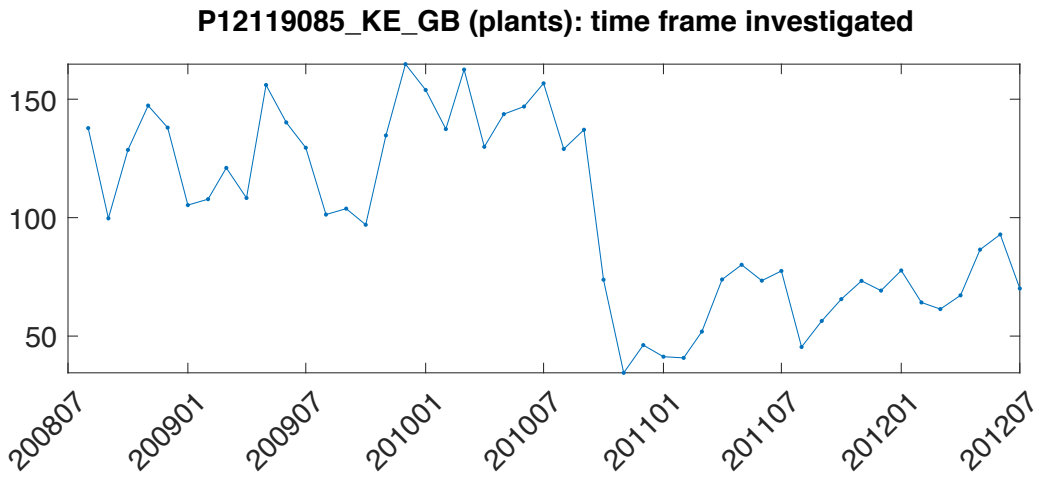


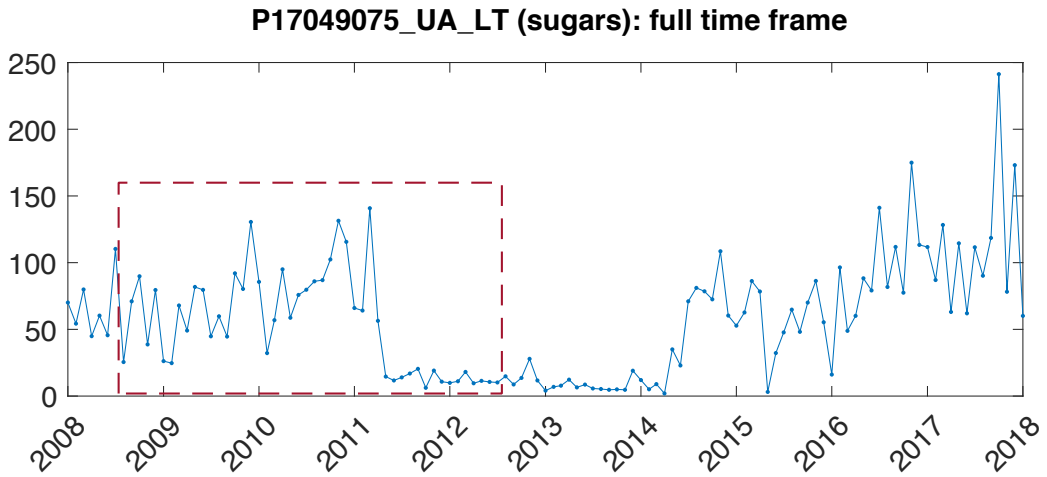
Imports of sugars
from Ukraine to Lithuania

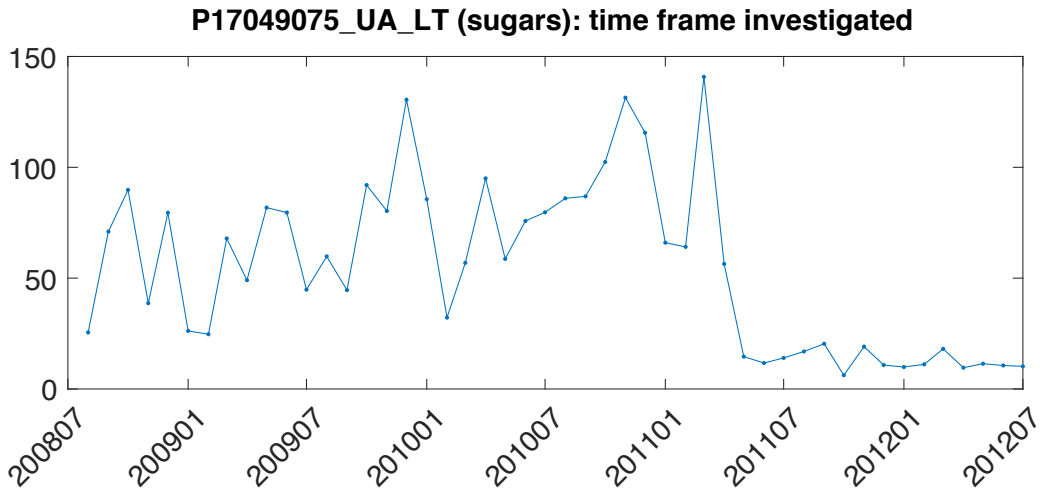
Application on two COMEXT time series

Need of early detection of signals and short term predictions, to be used as input for investigators and policy makers



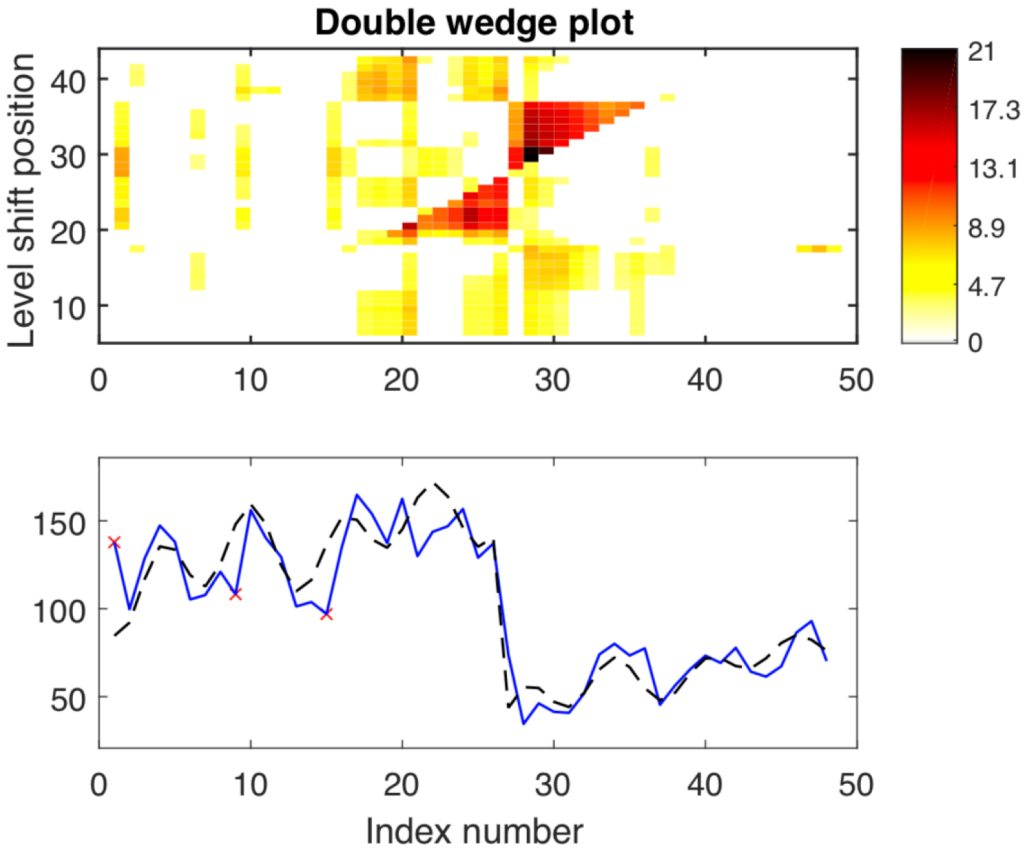




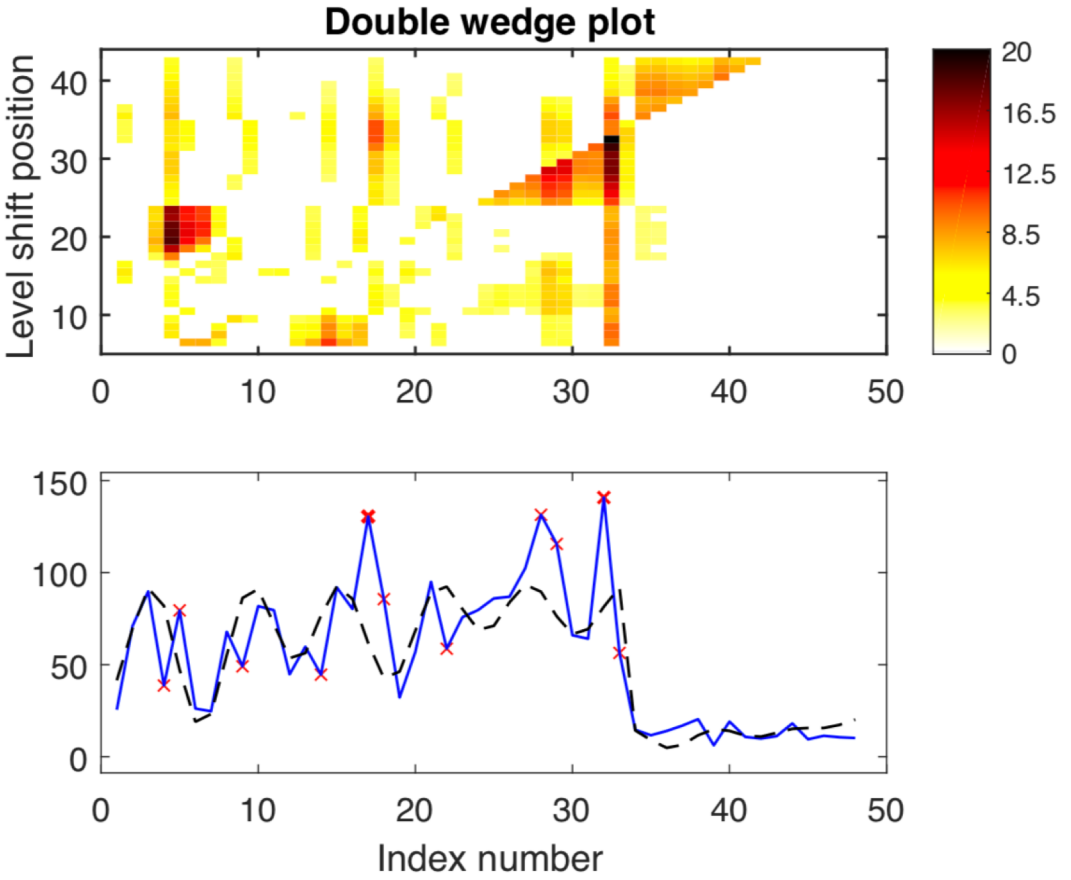


Patterns detected and Double Wedge Plot

- Not relevant outliers
- Level shift position around 27-28

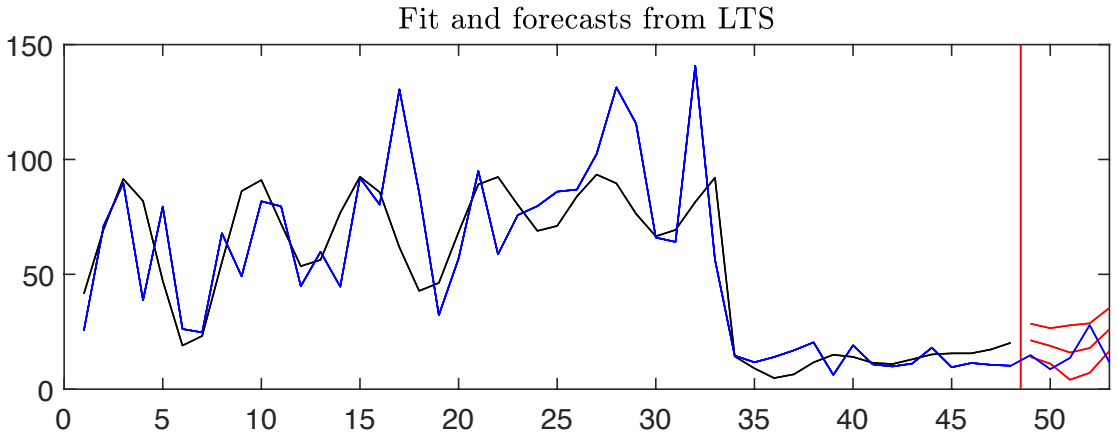


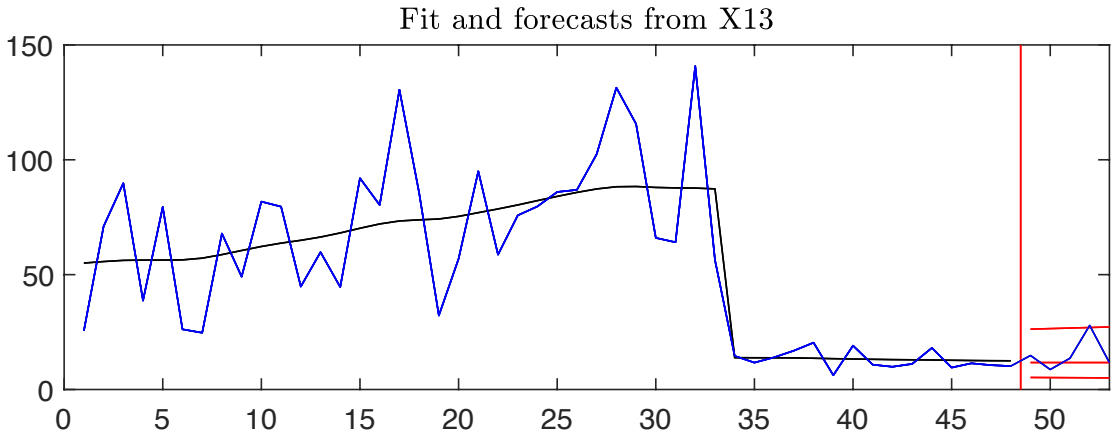
- Main outlier in position 32
- Local irregularities at pos. 4, 5, 17, 18
- level shift around position 35



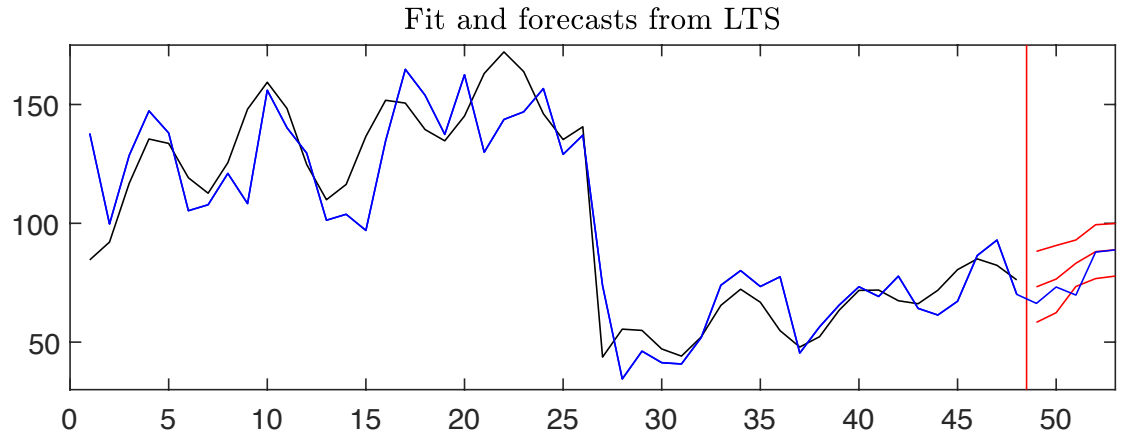
Short term predictions and comparison with X13

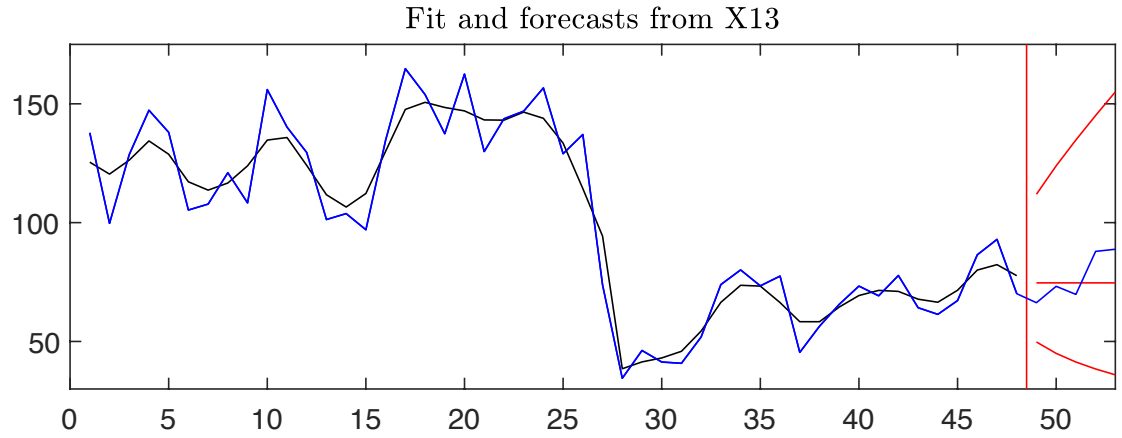
Imports of plants from Kenya to UK





Imports of sugars from Ukraine to Lithuania





Flowers: relevant t-statistics and interpretation

t-statistic of the height of the level shift is quite large: |t| = 14.7

Anomalous drop.

- Kenya was the only country of the East African Community (EAC) paying high European import duties on flowers.
- On the other hand, Kenya is the third largest exporter of cut flowers in the world.
- **Action:** check for a simultaneous upward level shift in an EAC country not paying import duties, which could point to a *misdeclaration of origin*.

	P12119085_KE_GB		
	Coeff	t-stat	p-values
$\hat{\alpha}_0$	115.27	25.6	0
$\hat{\alpha}_1$	1.59	5.80	0
$\hat{\beta}_{11}$	-2.83	-0.72	0.47
$\hat{\beta}_{12}$	-12.42	-2.65	0.012
$\hat{\beta}_{21}$	-9.07	-1.95	0.059
$\hat{\beta}_{22}$	-22.60	-4.80	0
$\hat{\gamma}_1$	-0.016	-3.72	0.00061
$\hat{\delta}_1$	-112.62	-14.7	0

Sugars: relevant t-statistics and interpretation

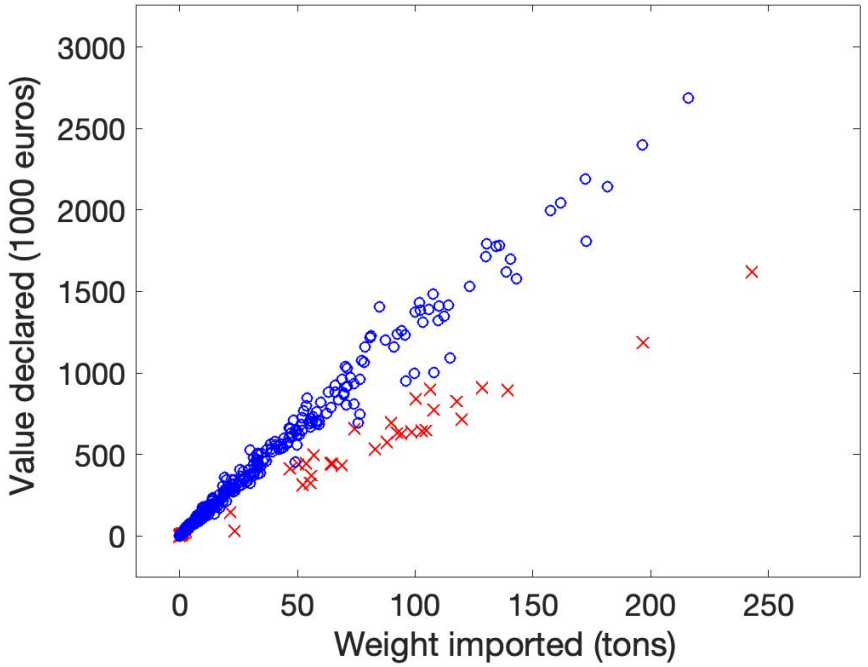
t-statistic of the height of the level shift is quite large: |t| = 13.9

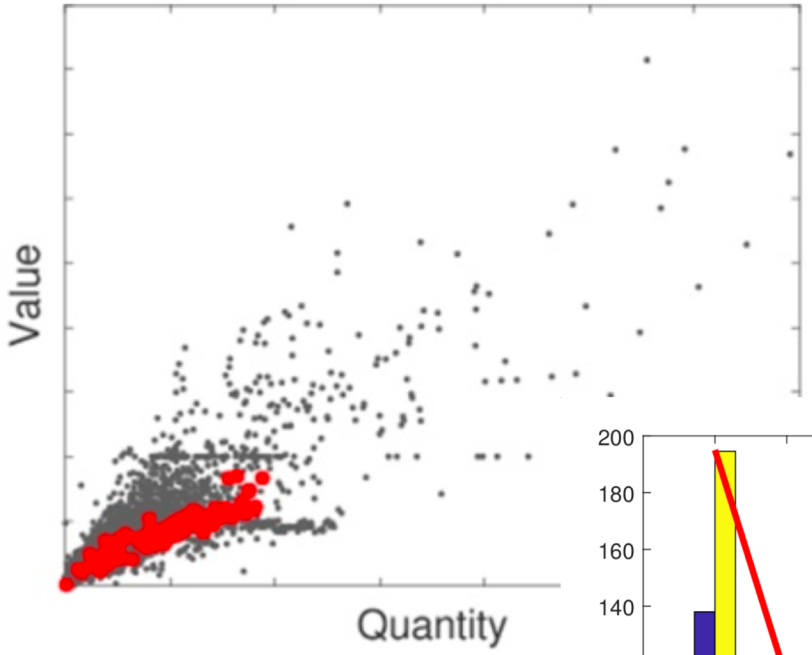
Anomalous drop.

- Sugar market is very restricted and regulated.
- Country-specific quotas, with higher duty for imports beyond the quota (tariff rate quotas).
- Fraud incentive: circumvent the quota by *mislabeling the product* with one not under surveillance.
- **Action:** check for upward level shifts in related products from the same country.

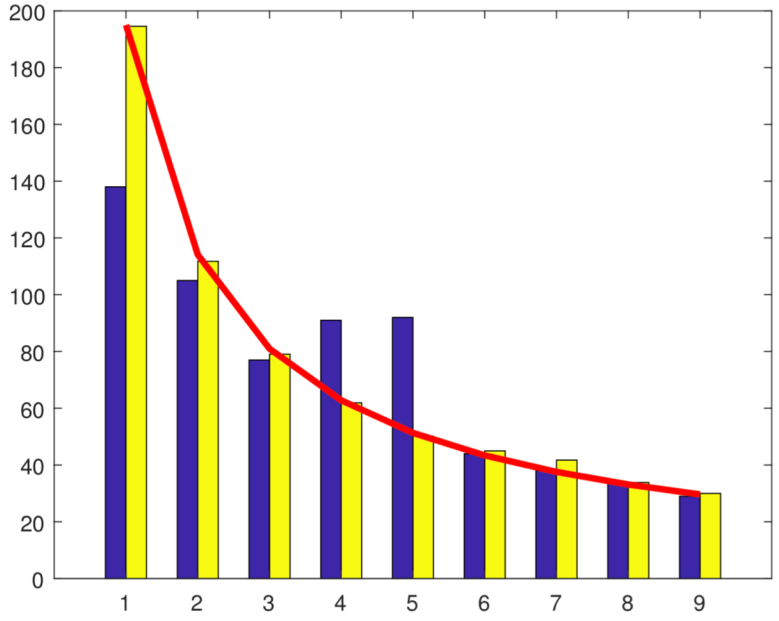
	P17049075_UA_LT		
	Coeff	t-stat	p-values
$\hat{\alpha}_0$	55.14	14.3	0
$\hat{\alpha}_1$	0.90	4.52	0
$\hat{\beta}_{11}$	15.55	3.75	0.00056
$\hat{\beta}_{12}$	3.61	0.85	0.40
$\hat{\beta}_{21}$	-32.50	-7.64	0
$\hat{\beta}_{22}$	-16.06	-3.72	0.00061
$\hat{\gamma}_1$	-0.023	-12.1	0
$\hat{\delta}_1$	-79.41	-13.9	0

Operational limitations of robust methods



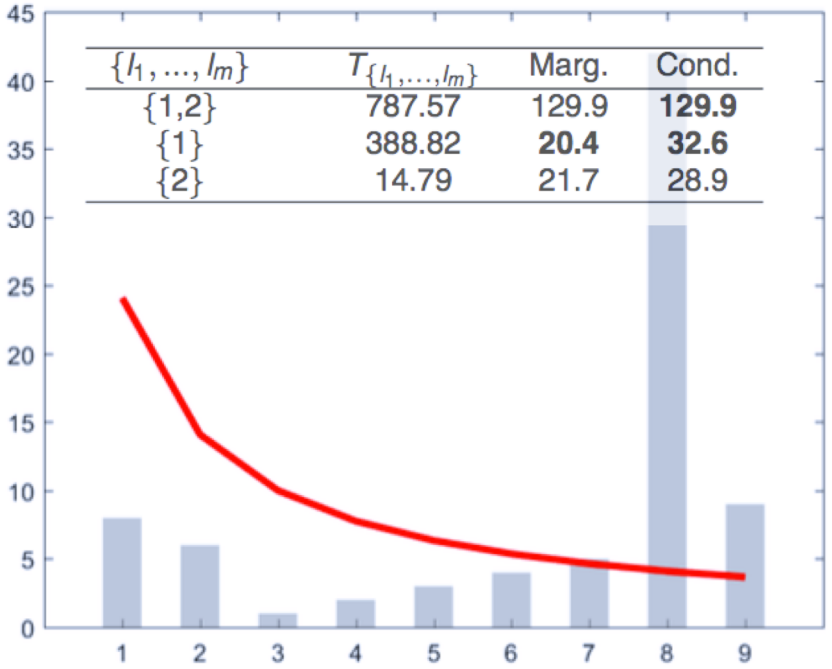


Newcomb-Benford analysis




```
load fishery;
X  = fishery.data(:,1);
y  = fishery.data(:,2);
n  = length(y);
cl = 0.001/n;
outFSr = FSRr(y,X,'alpha',cl,'R2th',0.85);
```

Detection of data manipulations



$\{I_1, \ldots, I_m\}$	$T_{\{I_1,\ldots,I_m\}}$	Marg.	Cond.
{1,2}	787.57	129.9	**129.9**
{1}	388.82	**20.4**	**32.6**
{2}	14.79	21.7	28.9

A serial fraudster detected in SAD data by our (two-stage)

Newcomb-Benford analysis

European Commission

Benford's Law Conference

10-12 July 2019 - Stresa, Italy