

Protecting the EU budget through the statistical detection of anomalies in international trade data

Francesca Torti

European Commission, Joint Research Centre

Sofia, September 14th 2018

Statistics for the defense of the EU budget

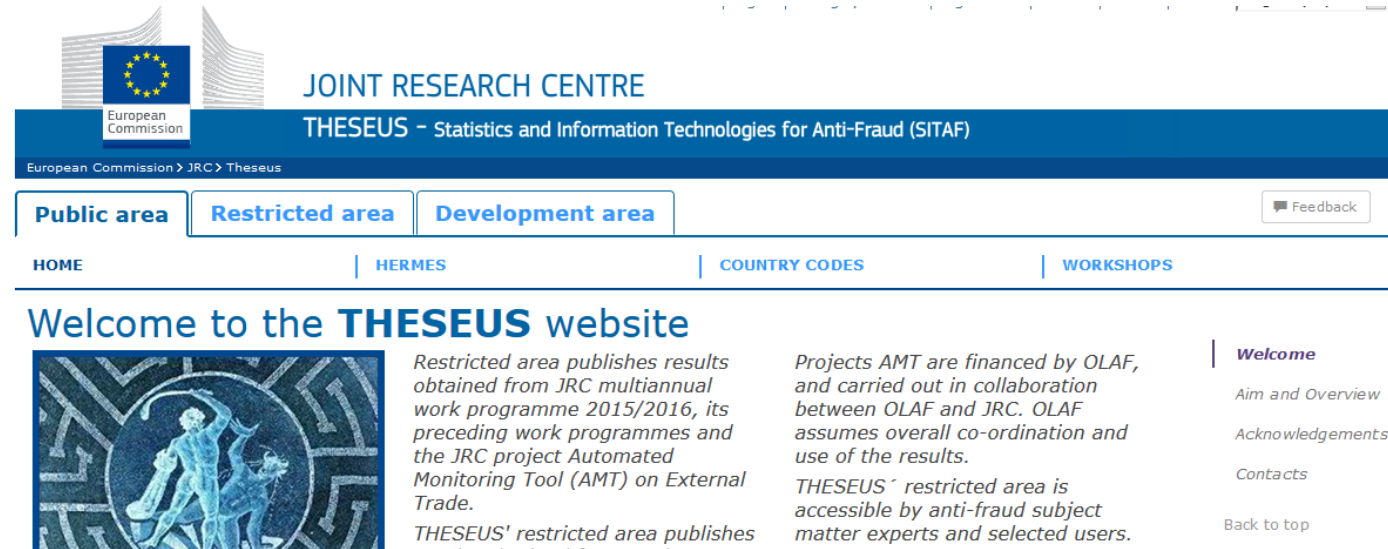
The Joint Research Centre (JRC) supports the Anti-Fraud Office of the European Union (OLAF) and its partners in the Member States in the **identification of anomalies in the trade** of goods between third countries and the European Union.

- Focus is mainly on imports.
- Fair prices and traded volumes estimates in time series are computed on data that are as homogeneous as possible in terms of product code, origin and destination.
- Flows with price significantly lower than the related JRC fair price are highlighted, as they may be linked to under-invoicing, and therefore import duties evasion.
- Spikes, level shifts and structural changes in time series of traded volumes are highlighted, as they may be linked to stockpiling or deflection of trade and therefore evasion of quotas etc.

Dissemination and processing tools

Results on suspect trade flows are made available to end users for further investigations.

<https://theseus.jrc.ec.europa.eu>



The screenshot shows the homepage of the THESEUS website. At the top, there is a header with the European Commission logo and the text 'JOINT RESEARCH CENTRE' and 'THESEUS - Statistics and Information Technologies for Anti-Fraud (SITAF)'. Below the header, there are navigation tabs for 'Public area', 'Restricted area', and 'Development area'. A secondary navigation bar includes links for 'HOME', 'HERMES', 'COUNTRY CODES', and 'WORKSHOPS'. The main content area features a large image of a classical figure holding a scale, with the text 'Welcome to the THESEUS website'. To the right of the image, there is a paragraph of text: 'Restricted area publishes results obtained from JRC multiannual work programme 2015/2016, its preceding work programmes and the JRC project Automated Monitoring Tool (AMT) on External Trade. THESEUS' restricted area publishes Projects AMT are financed by OLAF, and carried out in collaboration between OLAF and JRC. OLAF assumes overall co-ordination and use of the results. THESEUS' restricted area is accessible by anti-fraud subject matter experts and selected users.' On the far right, there is a 'Welcome' section with links for 'Aim and Overview', 'Acknowledgements', 'Contacts', and 'Back to top'.



The screenshot shows the header of the WebARIADNE website. It features the European Commission logo and the text 'JOINT RESEARCH CENTRE' and 'WebARIADNE - statistics for anti-fraud'. Below the header, there is a navigation bar with the text 'European Commission > EU Science Hub > WebAriadne'.

Welcome to WebARIADNE!

Username:

Password:

LOGIN

WebARIADNE at a glance

ARIADNE is an application for the detection of statistical anomalies and underlying structures in large scale data. It allows users to import, pre-process and analyze the data with standard SAS procedures, produce user-chosen descriptive statistics for insight into and data exploration, and run SITAFS-developed procedures.

WebARIADNE is the porting of ARIADNE to the web. It increases ARIADNE's potential by offering a comprehensive web-based execution environment for the statistical procedures developed within the SITAFS action.

To know more about **WebARIADNE**, please click [here](#).

Please take the time to read our [disclaimer](#) and our [privacy statement](#).

Customers can process data and generate statistical results using a web-based service, with user-friendly graphical interface

<https://webariadne.jrc.ec.europa.eu>

The “fair price” concept and its use

EN

2017

NO
19

Special Report

**Import procedures:
shortcomings in the legal
framework and an
ineffective implementation
impact the financial
interests of the EU**

(pursuant to Article 287(4), second subparagraph, TFEU)



40
1977-2017



EUROPEAN
COURT
OF AUDITORS

48. To overcome the risk of undervaluation, the Commission has developed a methodology to estimate "fair prices"²², applying a statistical procedure to COMEXT²³ data, in order to produce robust estimates for the prices of the imported goods²⁴. 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.

The “fair price” concept and its use



The **OLAF**
report **2017**

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

The detection of *relatively few patterns* underlies *numerous fraud-control problems*

	Spikes (in time series)	Outliers (in multi-variate data)	Systematic spices or outliers	Systematic associations in 2 way tables
Stockpiling	X		X	
Fraud in export refunds	X		X	
Evasion of import duties		X, LP outliers	X	X
Deflection of trade	X, partly		X, partly	
Trade based laundered money in Origin. Generation of black money at destination.		X, HP, LP outliers	X	X
VAT carousels		X, LP outliers	X	X

International trade data: source 1 - 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

International trade data: source 2 - Surveillance

Daily aggregates
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 records per year

For textiles imports only

International trade data: source 3 - SAD

Customs declarations of each importer/exporter

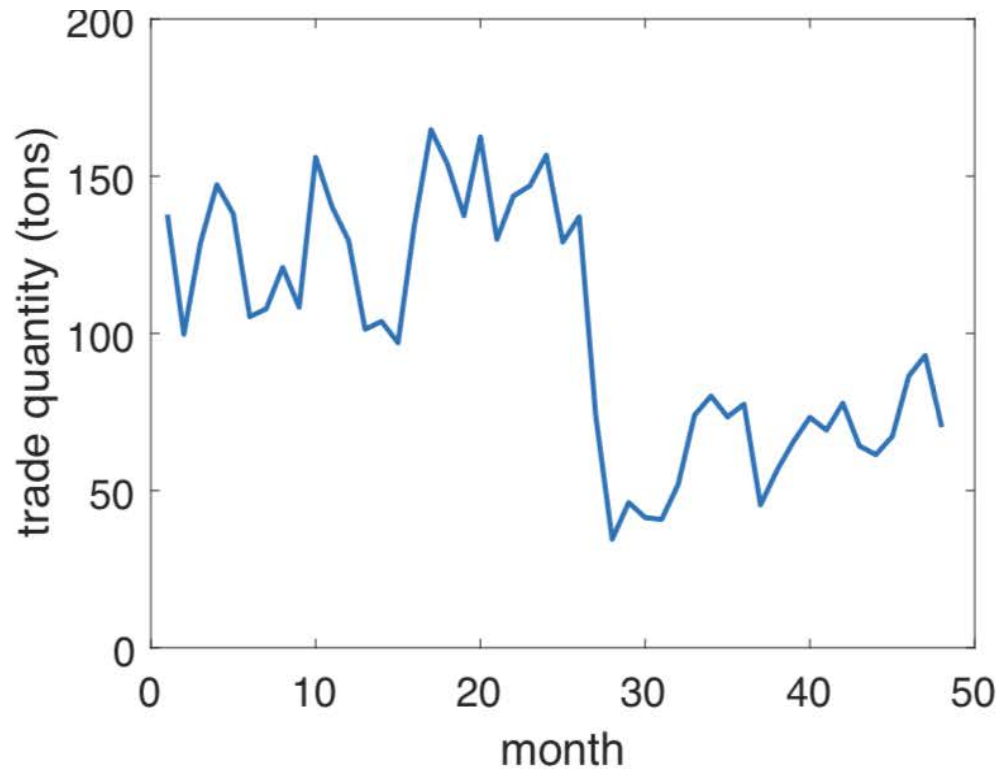
ID	DATE	BOX_16	BOX_17_a	BOX_33_1	BOX_38	BOX_41	BOX_46
31382639	2014-01-20	CN	IT	64011000	0.8	1	66.47
31385957	2014-01-23	CN	IT	64011000	1.6	1	89.48
31441658	2014-03-20	VN	IT	64011000	0.9	1	114.38
31449805	2014-03-28	CN	IT	64011000	19.2	25	467.03

Collected and analyzed
under bilateral agreements with Member States Customs

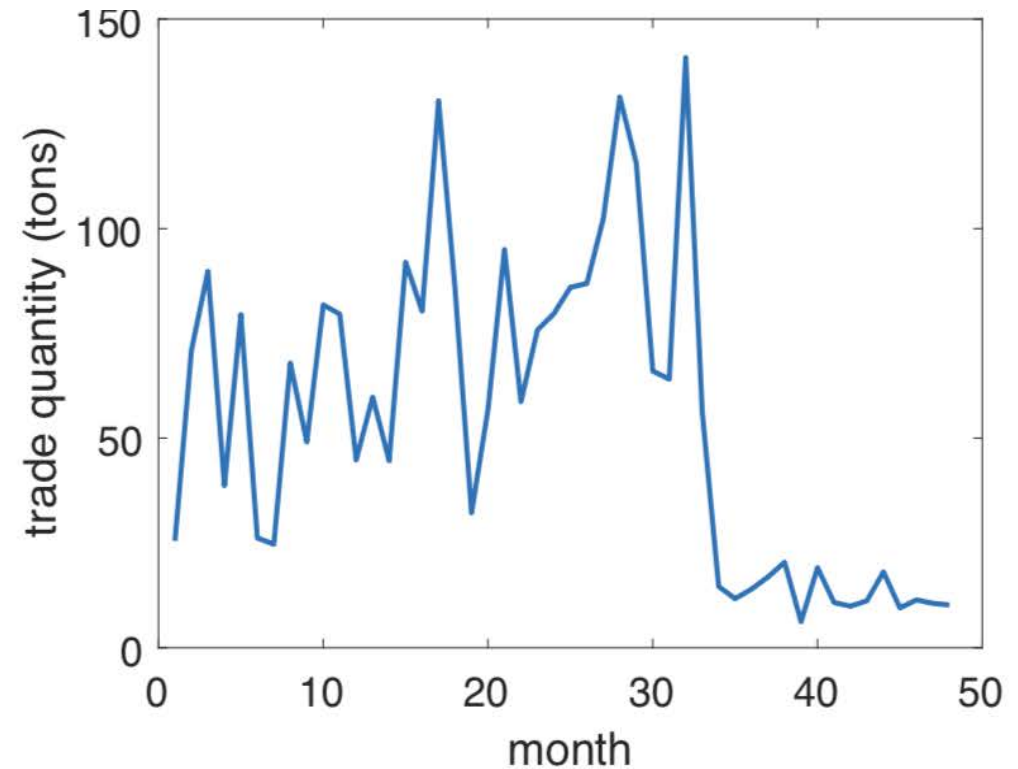
(8 millions import declarations per year for Italy)

Monitoring trade volumes

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



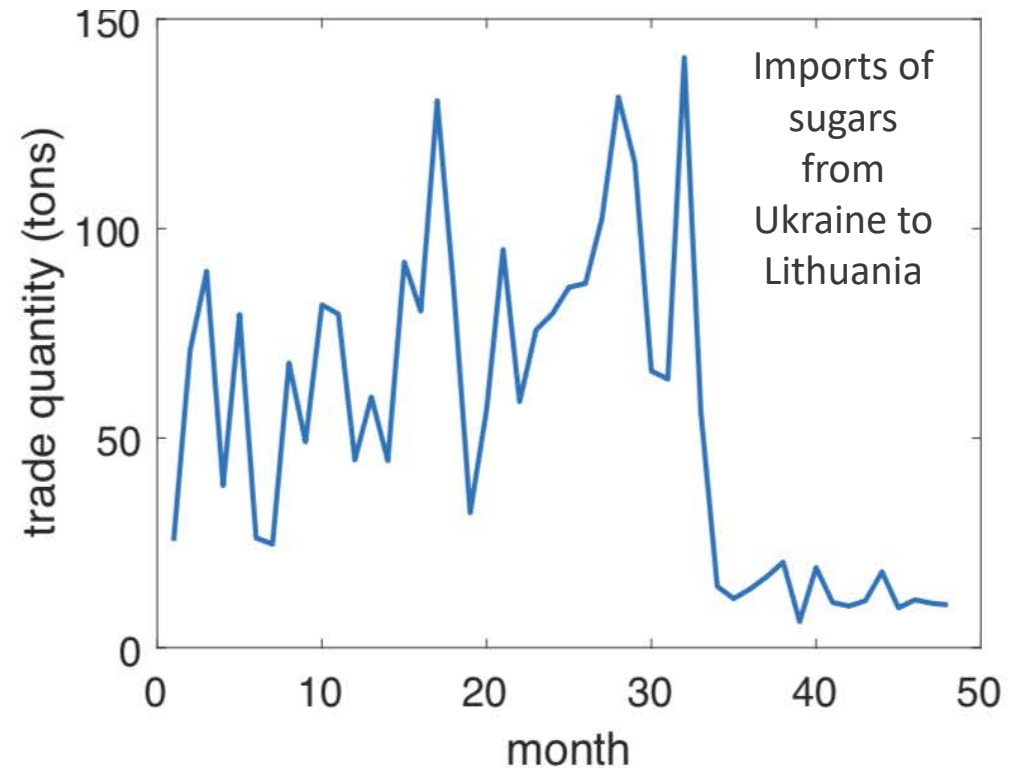
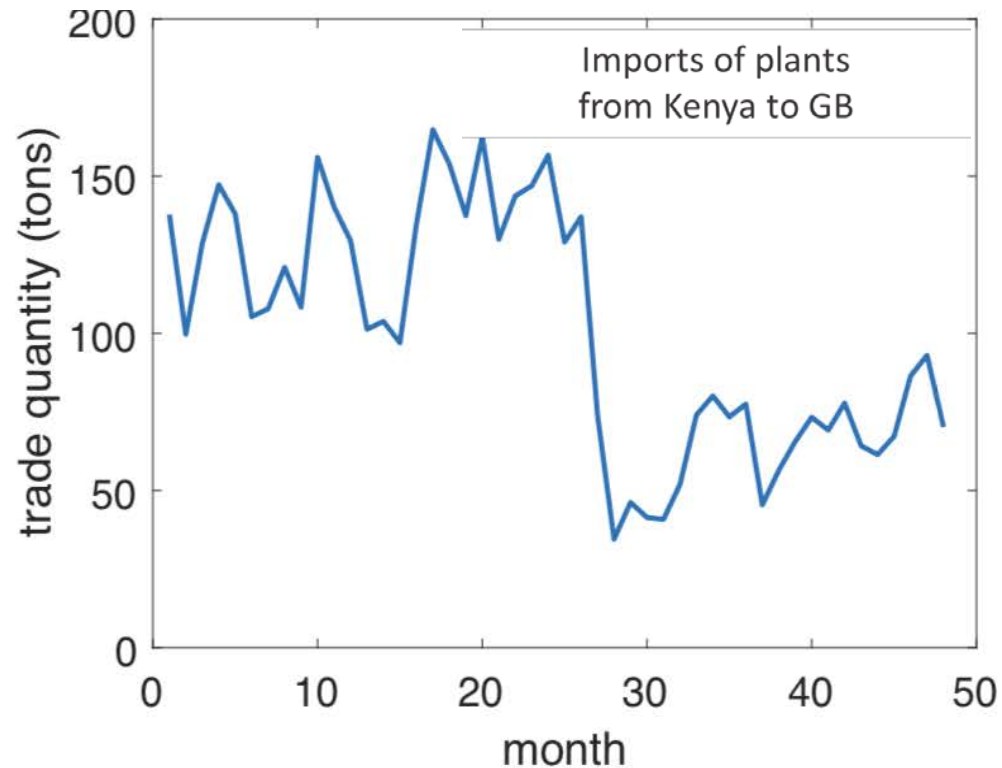
Imports of plants
from Kenya to GB



Imports of sugars
from Ukraine to Lithuania

Monitoring trade volumes

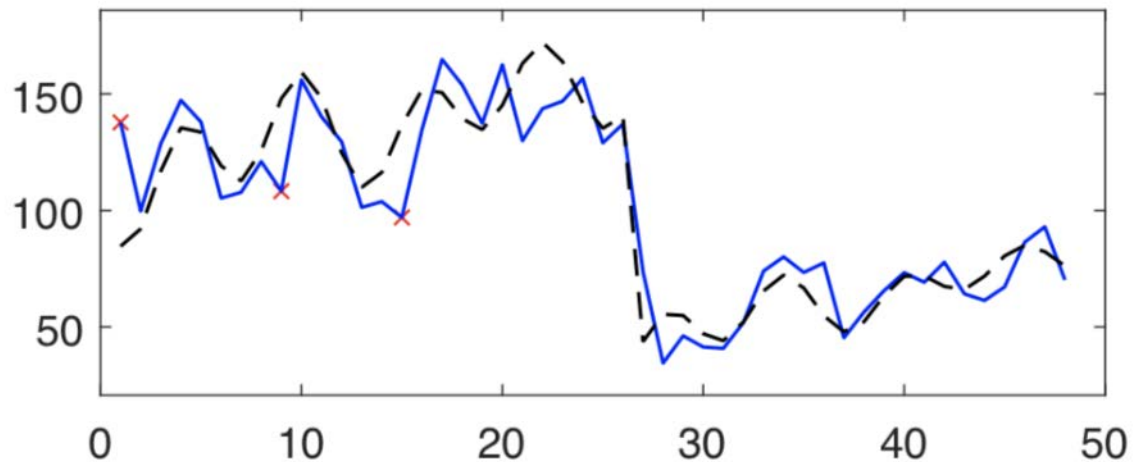
Statistical purpose: provide a robust unified framework to treat simultaneously outliers, unknown level shifts and changes in the seasonal pattern



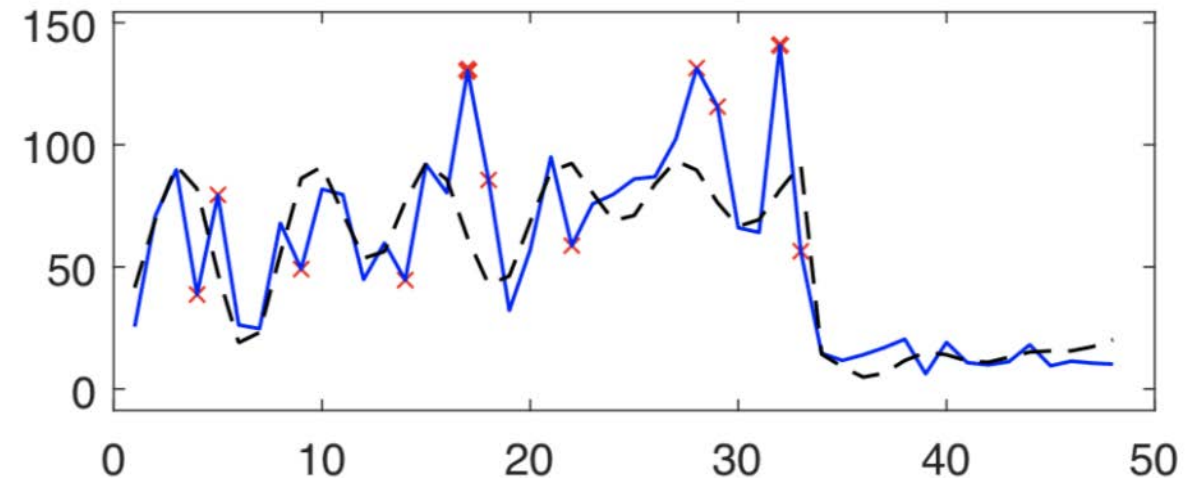
Rousseeuw, P.J., Perrotta, D., Riani, M., Hubert, M. (2018). Robust monitoring of time series with application to fraud detection. Econometrics and Statistics, in press.

Positions of signals: outliers and level shifts

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



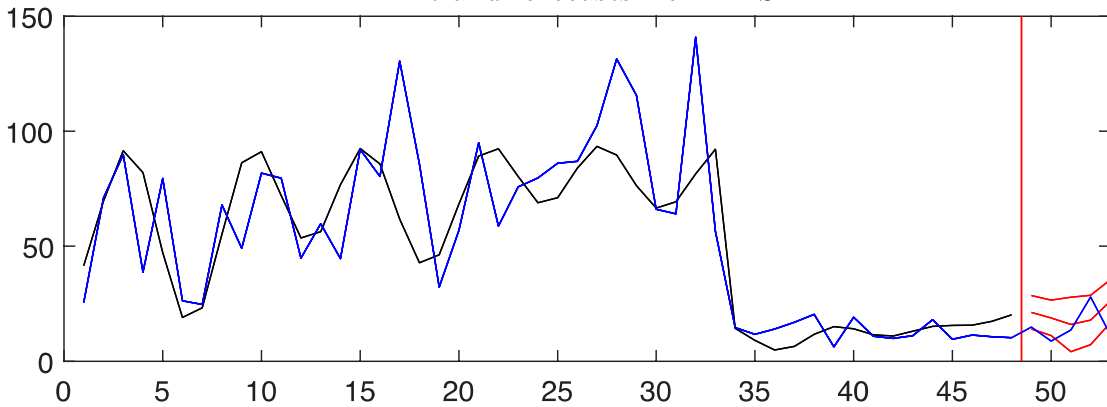
- Main outlier in position 32
- Local irregularities at pos. 4, 5, 17, 18
- level shift around position 35



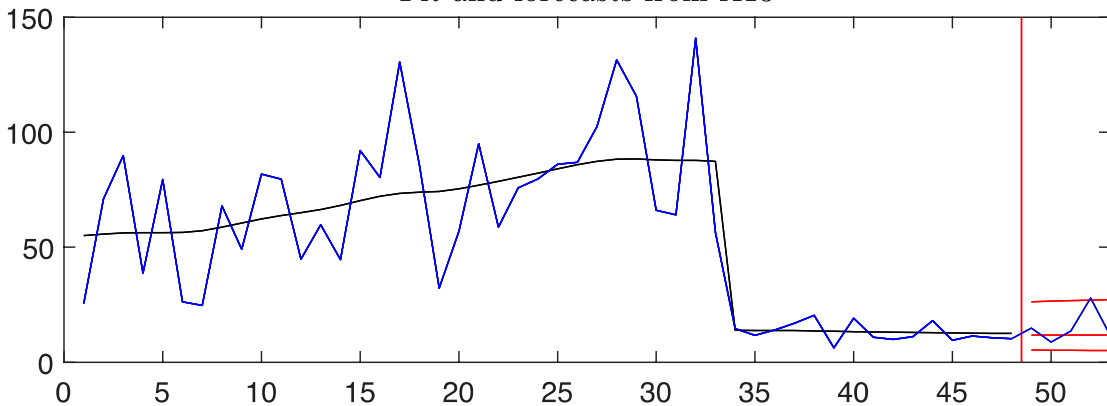
Short term predictions with two methods

Imports of plants from Kenya to GB

Fit and forecasts from LTS

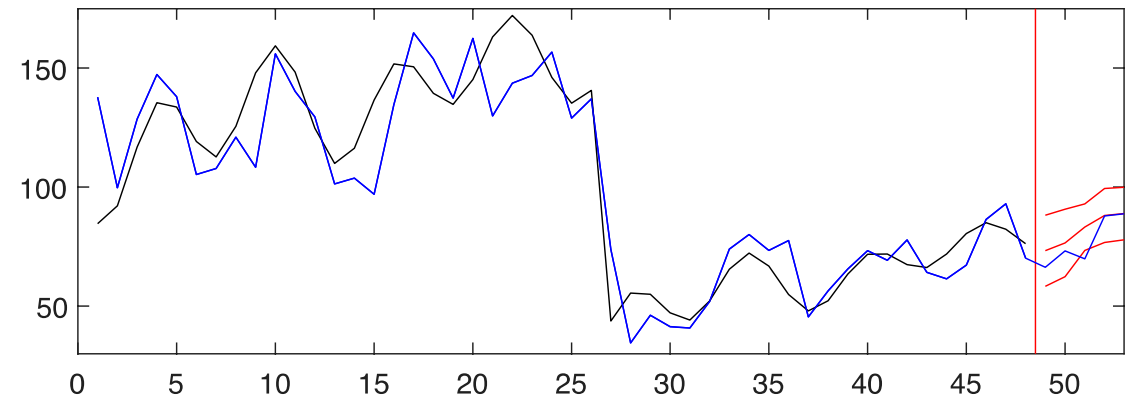


Fit and forecasts from X13

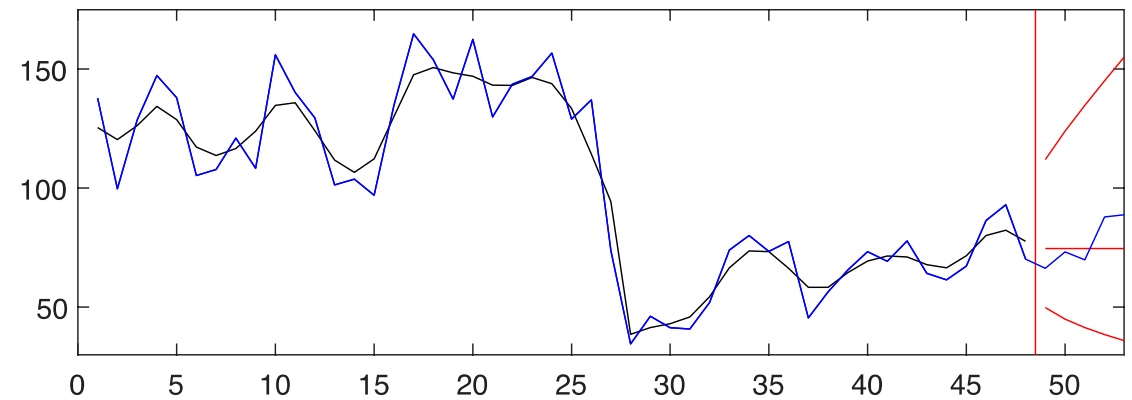


Imports of sugars from Ukraine to Lithuania

Fit and forecasts from LTS



Fit and forecasts from X13



Plants

interpretation of the 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*.

Sugars

interpretation of the 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 *mislabeled the product* with one not under surveillance.
- **Action:** check for upward level shifts in related products from the same country.

Another example: the trade of honey (CN 04090000)

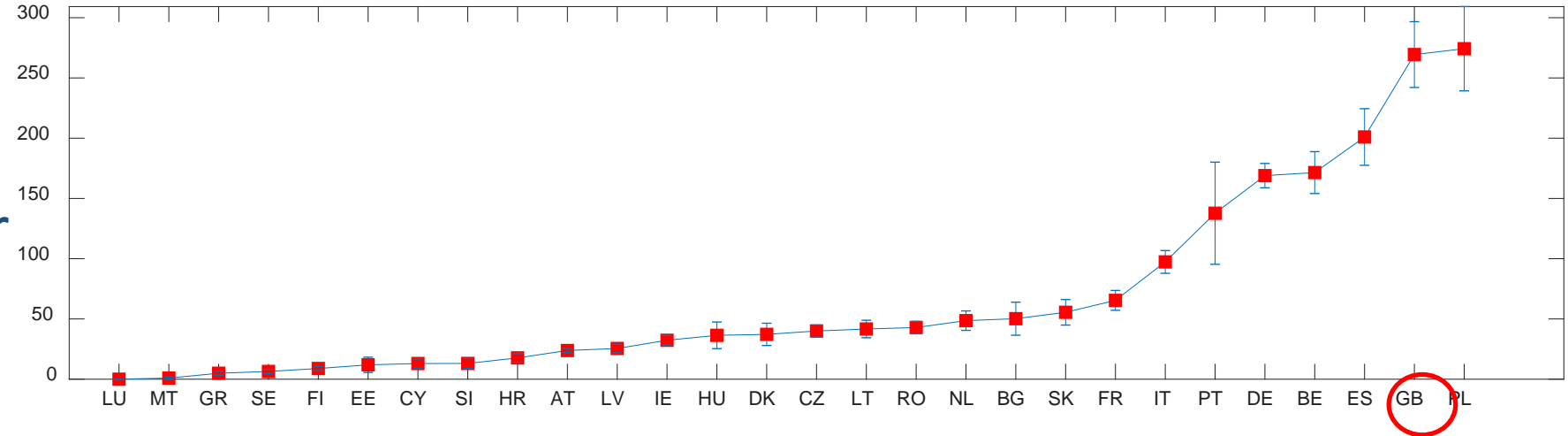
In the last 10 years:

GB is a heavy importer

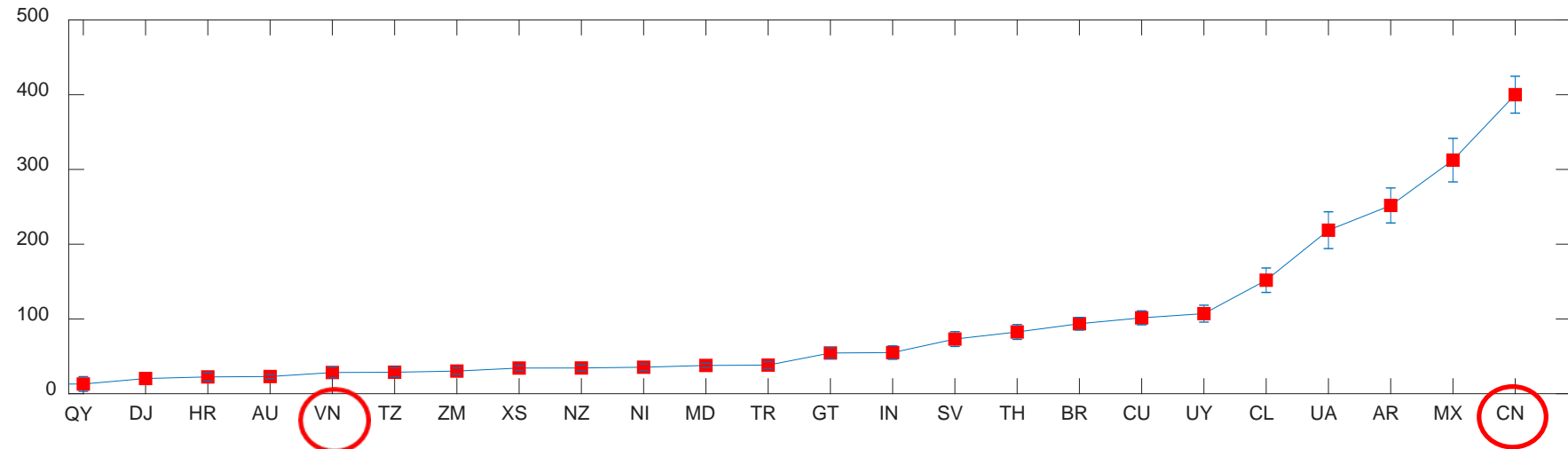
China is the biggest exporter

Vietnam is a small exporter

Mean quantities (in tons) imported in the period 2008-2018 by MS of destination with their 90% Confidence Intervals



Mean quantities (in tons) imported in the period 2008-2018 by country of origin with their 90% Confidence Intervals

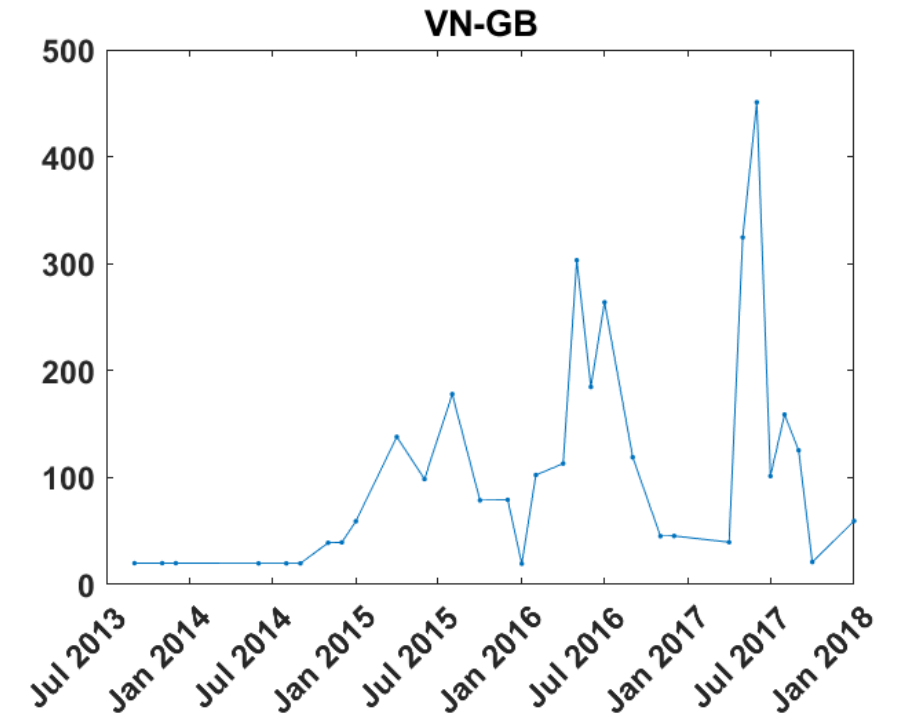
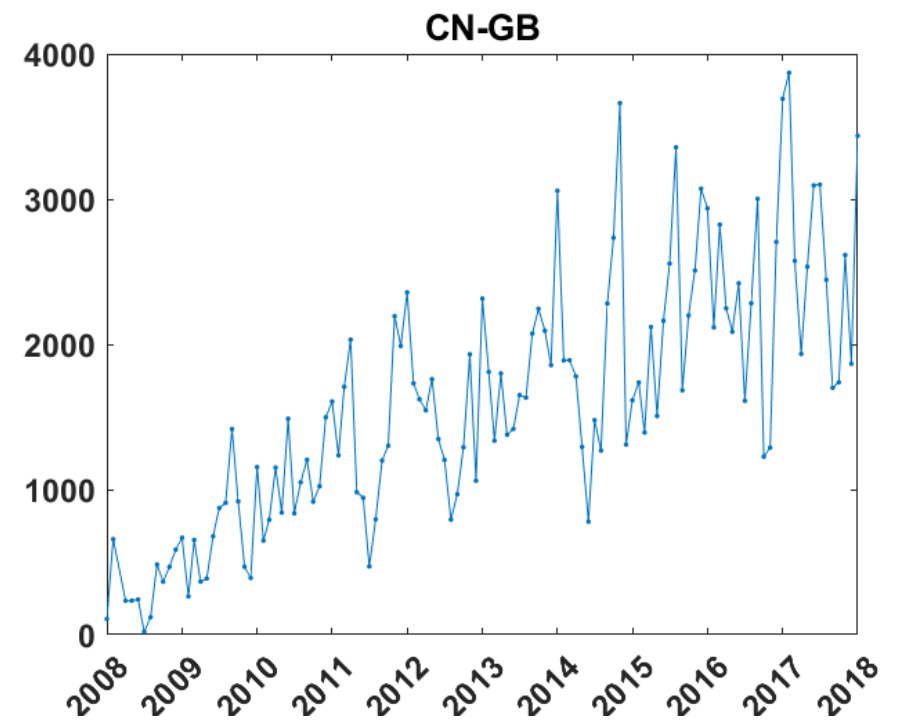


The volumes of honey imported in the last 10 years into GB from China and Vietnam

The volume of honey imported from China to GB is constantly increasing from 2008.

The volume of honey imported from Vietnam to GB started only in 2013.

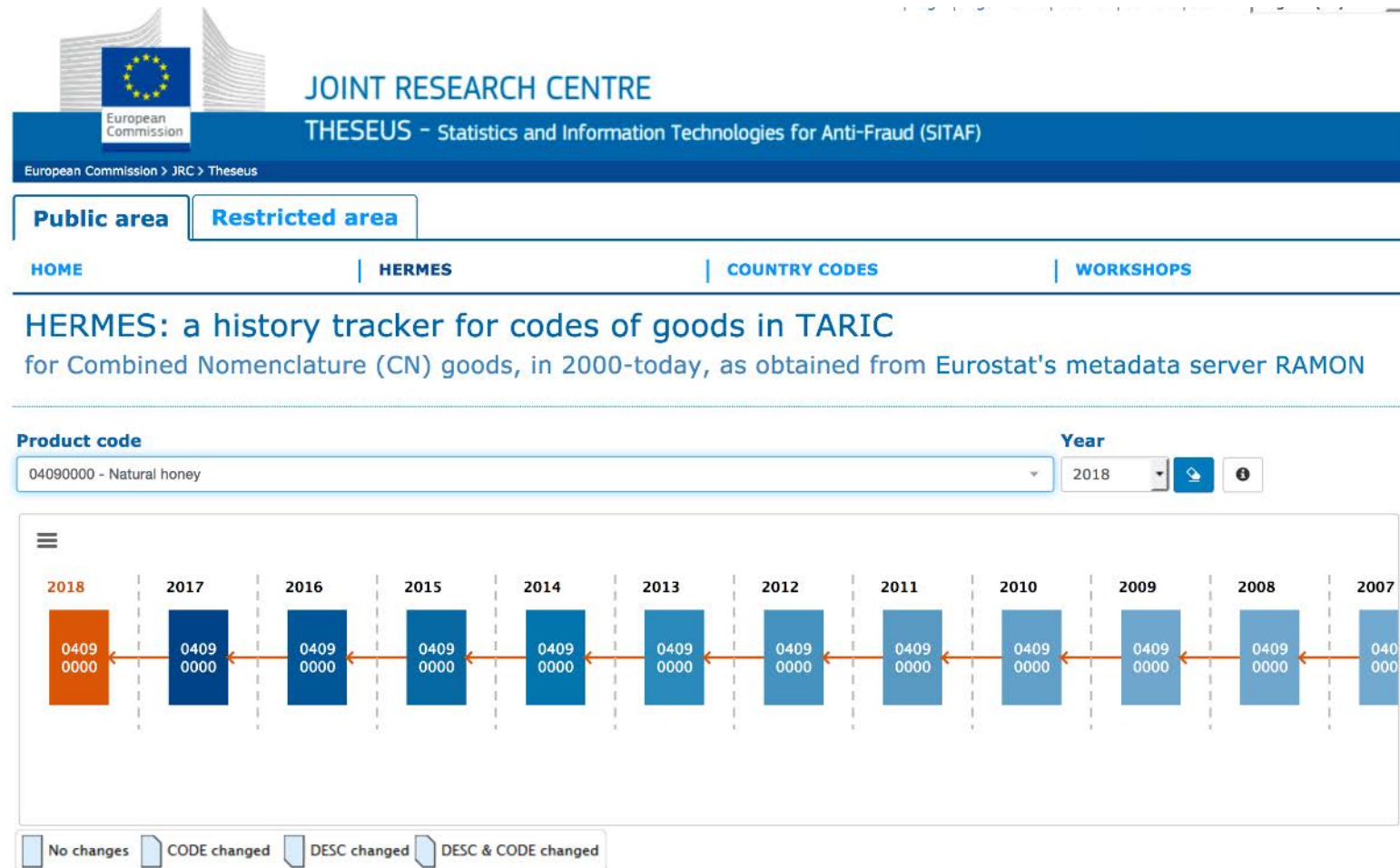
Seasonality: peaks in July.



History of the CN code of honey

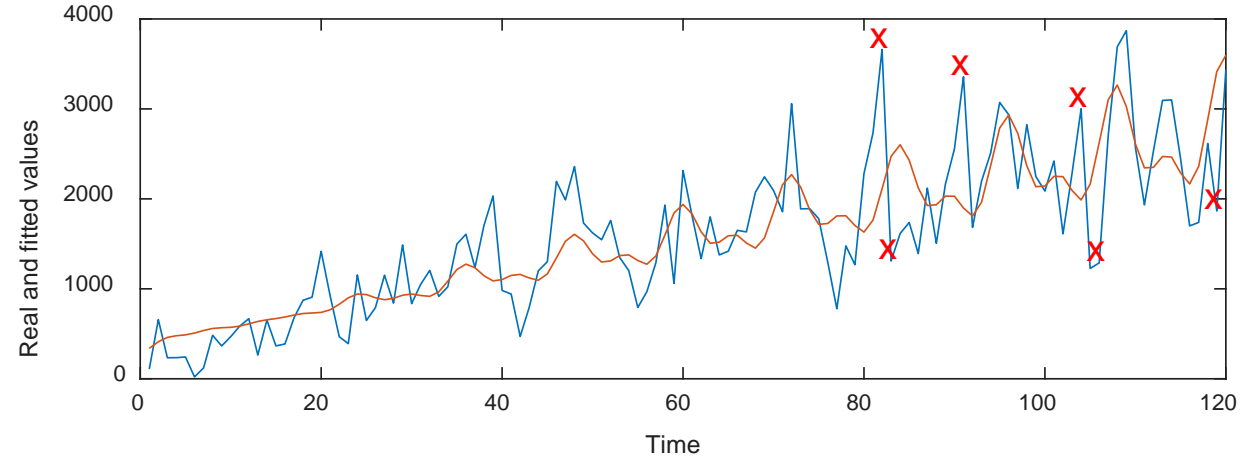
Changes in the product code could distort the analysis

The code 04090000 has not changed in the last 10 years

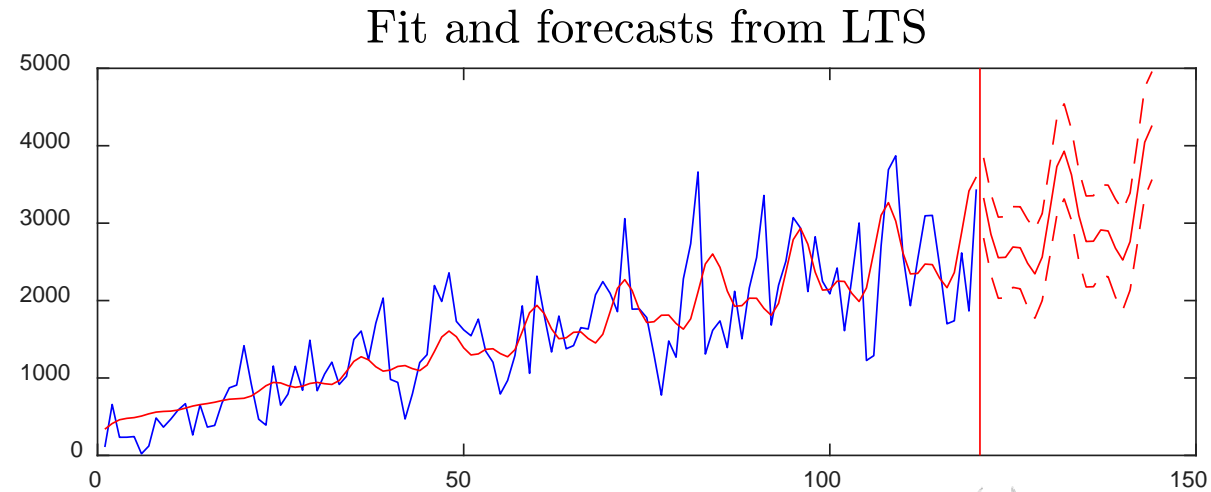


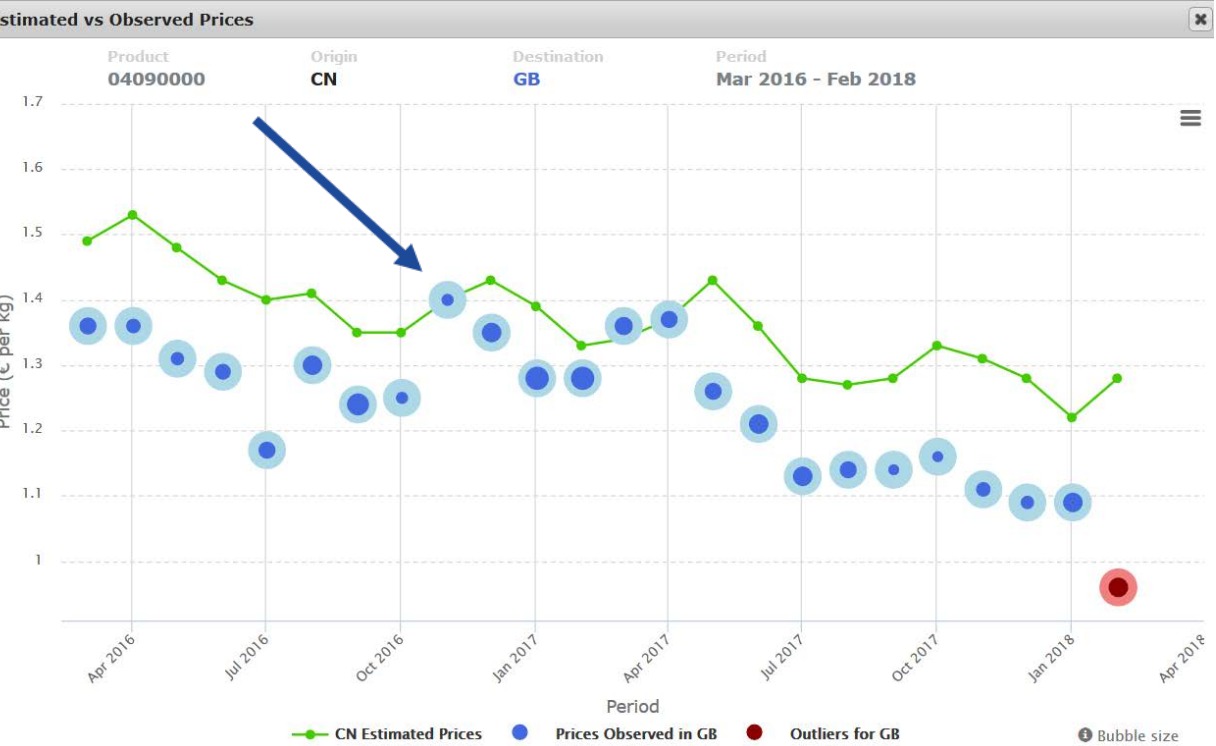
Example of estimation and prediction of the trade volumes of honey from China to GB

The volume of honey imported: blue curve
Its estimation: red curve
The detected outliers: red crosses



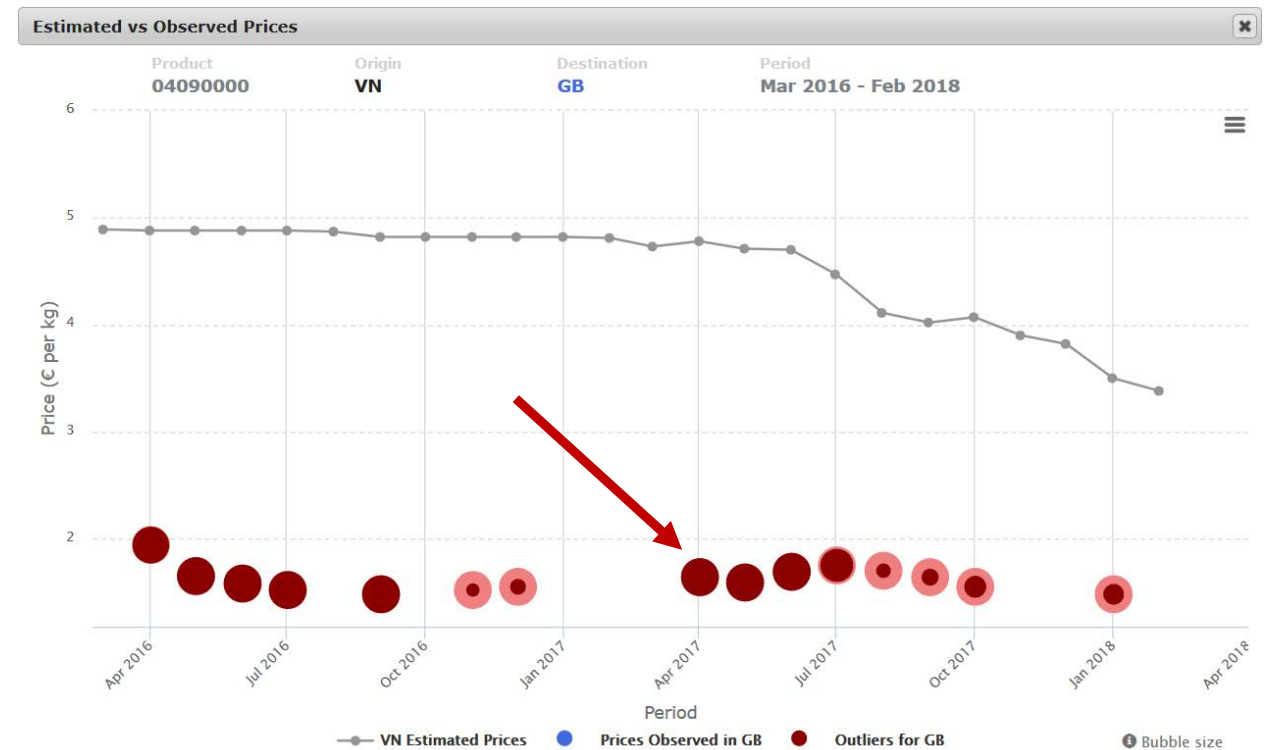
Same as first plot, with predictions
and related bands (in red)





The price of honey:
A case of systematic underpricing

Import prices
of “honey” (CN 04090000)
observed in **GB** from **China** and
Vietnam vs. estimated prices.
Period: 03-2016 – 02-2018



The “monthly fair price” of honey

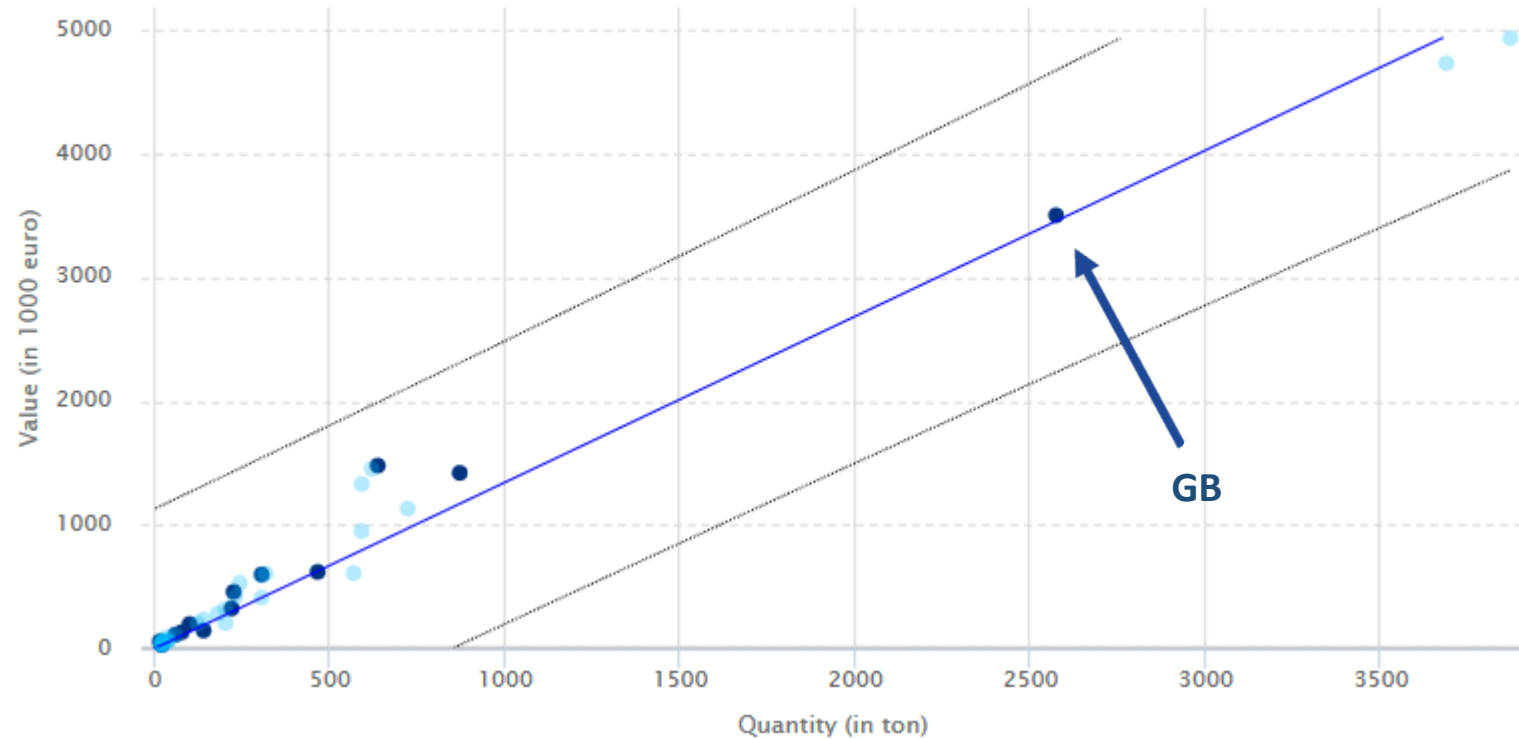
EU prices and estimated price
of “honey” (CN 04090000)
from China in a specific
month (March 2017)

GB unit price is in line with EU
estimated import price from
China

(1.34 €/Kg. vs. 1.36 €/Kg)

Scatterplot of Quantities and Values

P: 04090000 O: CN Period: Mar 2017

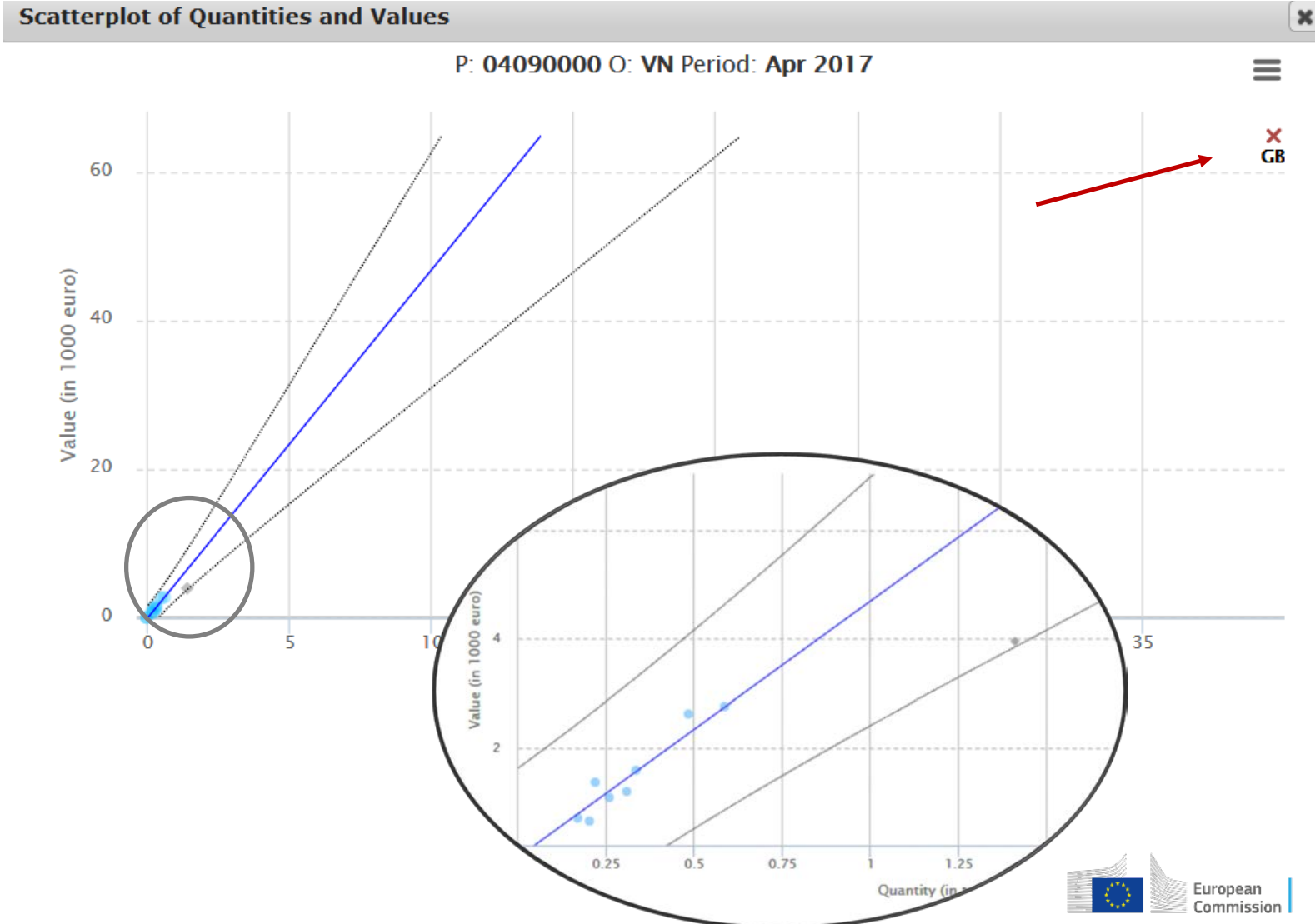


The “monthly fair price” of honey

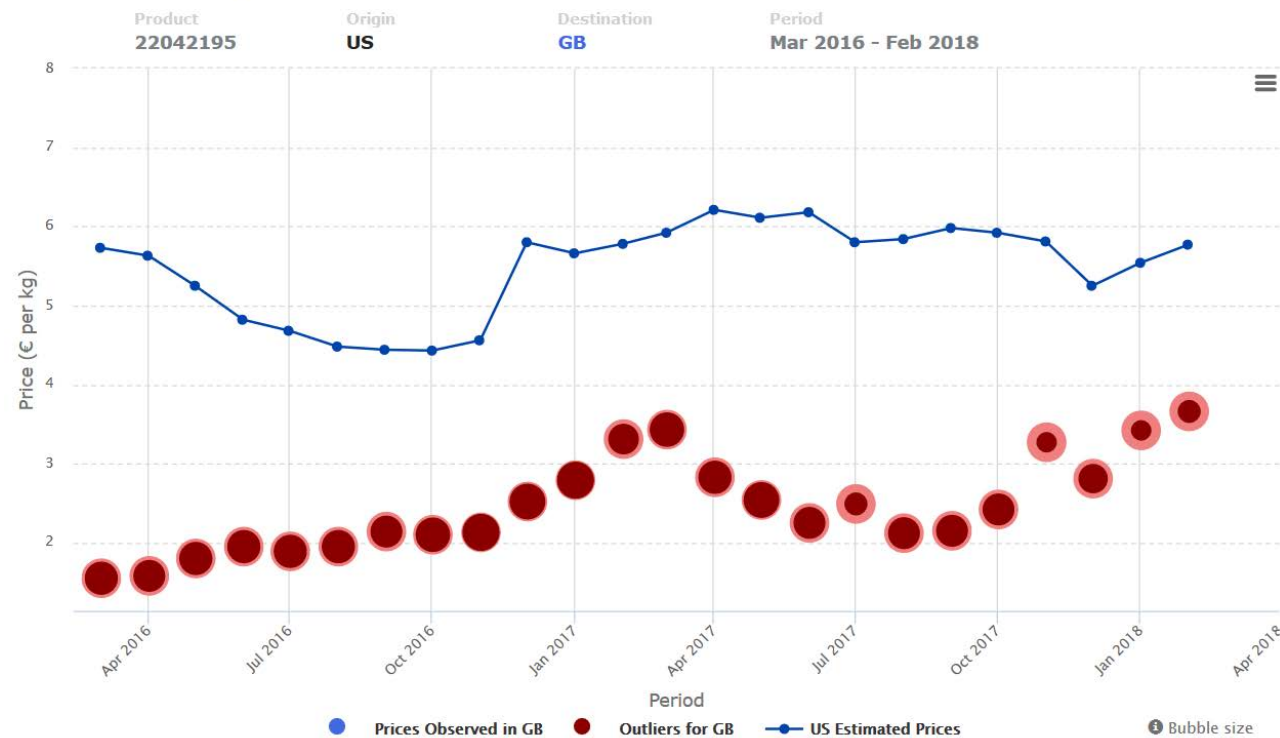
EU prices and estimated price of “honey” (CN 04090000) from Vietnam in a specific month (April 2017)

GB is a clear outlier

(4.78 €/Kg. vs. 1.64 €/Kg)



Estimated vs Observed Prices

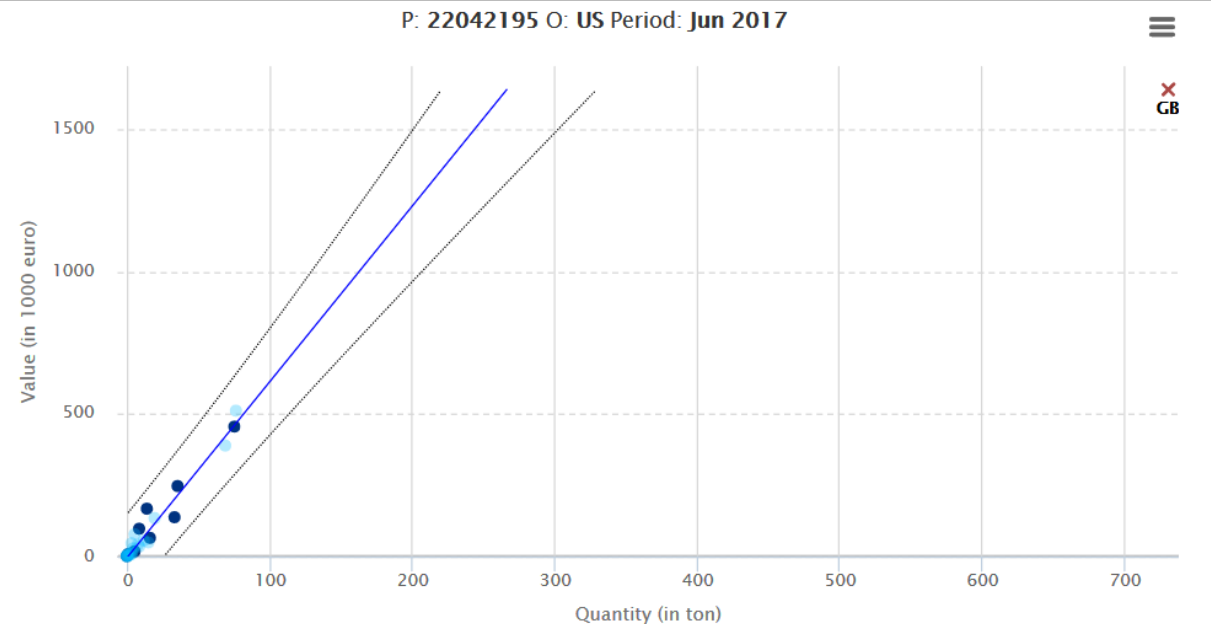


Another example of systematic underpricing and monthly fair price

Import prices of white wine (CN 22042195) observed in GB from US vs. estimated prices.

Period: 03-2016 – 02-2018

Scatterplot of Quantities and Values



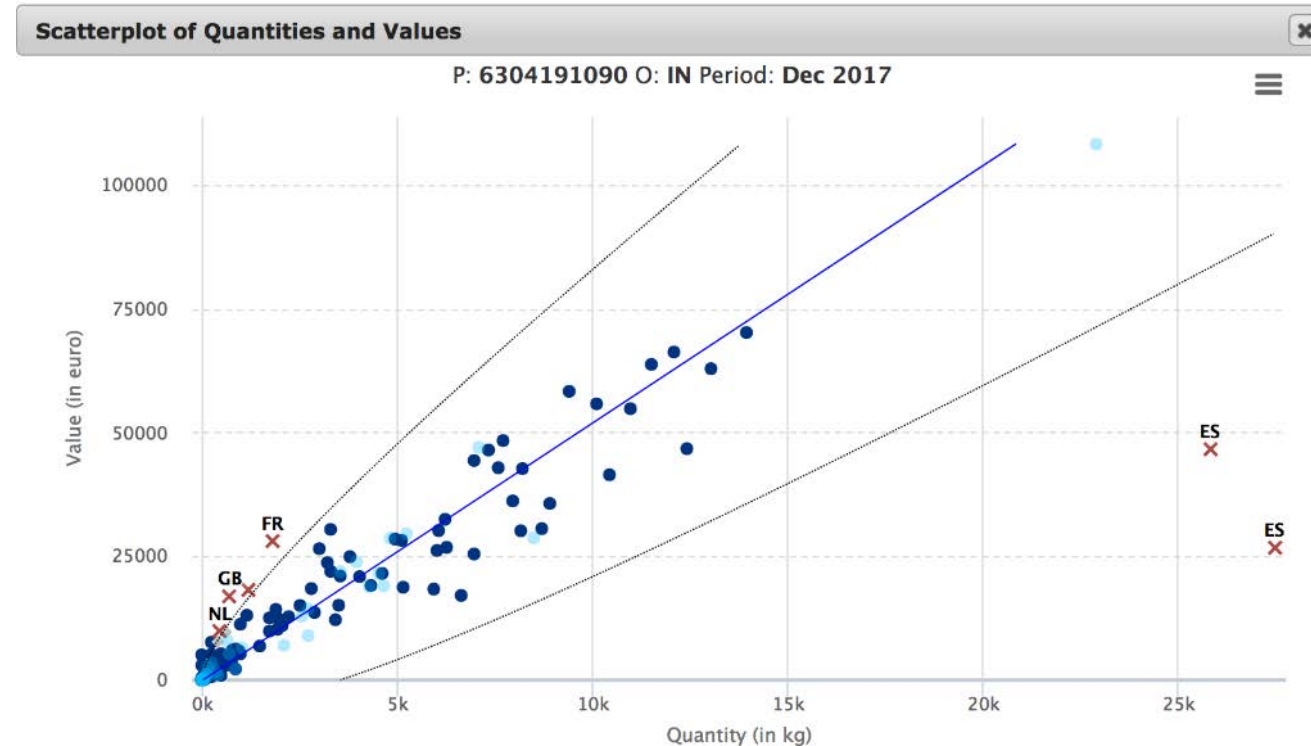
Extension to heteroscedastic data

estimated price
of “bedspreads of cotton”:

two outlying ES declarations

estimated price: 5.16 €

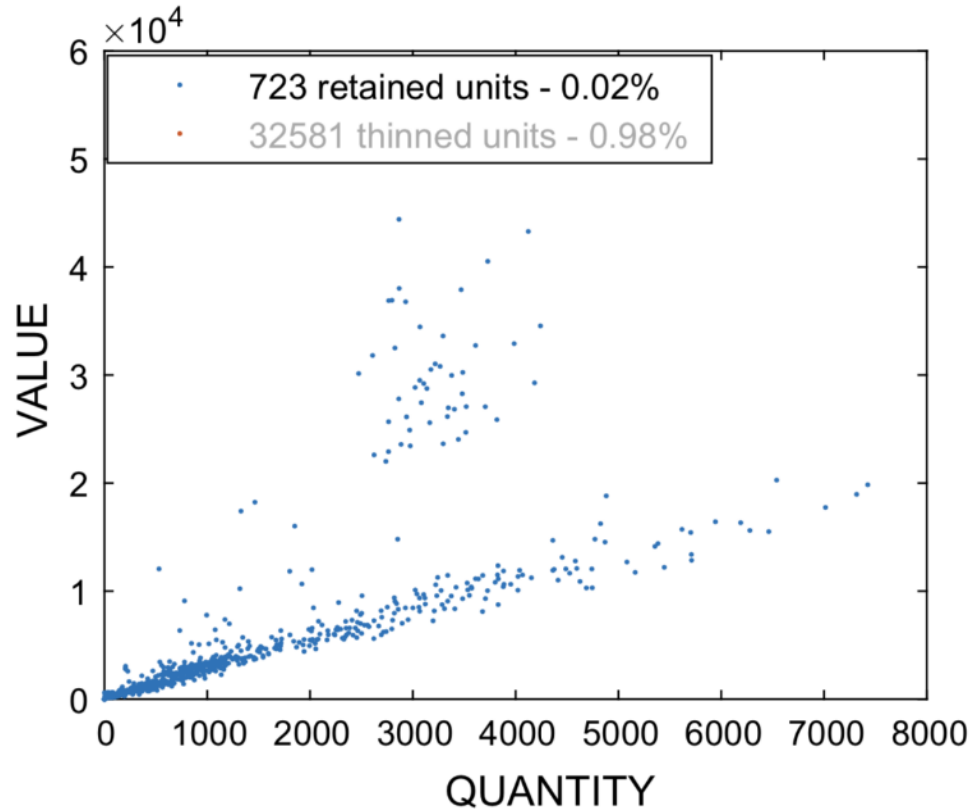
declared price: 0.97 & 1.81 €



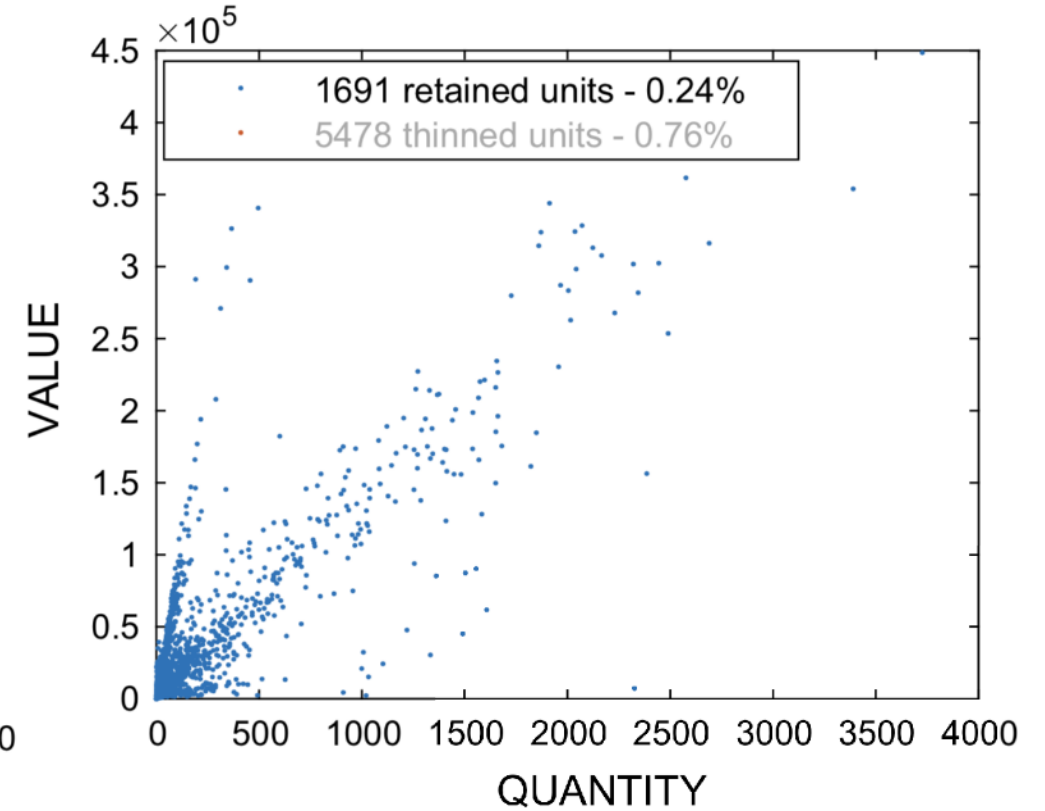
Atkinson, Riani, Torti (2016). Robust methods for heteroskedastic regression. Computational Statistics & Data Analysis, Volume 104, Pages 209-222

Extension to other complex patterns in customs data

Regression structures, outliers, multiple populations, dense areas, heteroscedasticity, ...



“Books” dataset



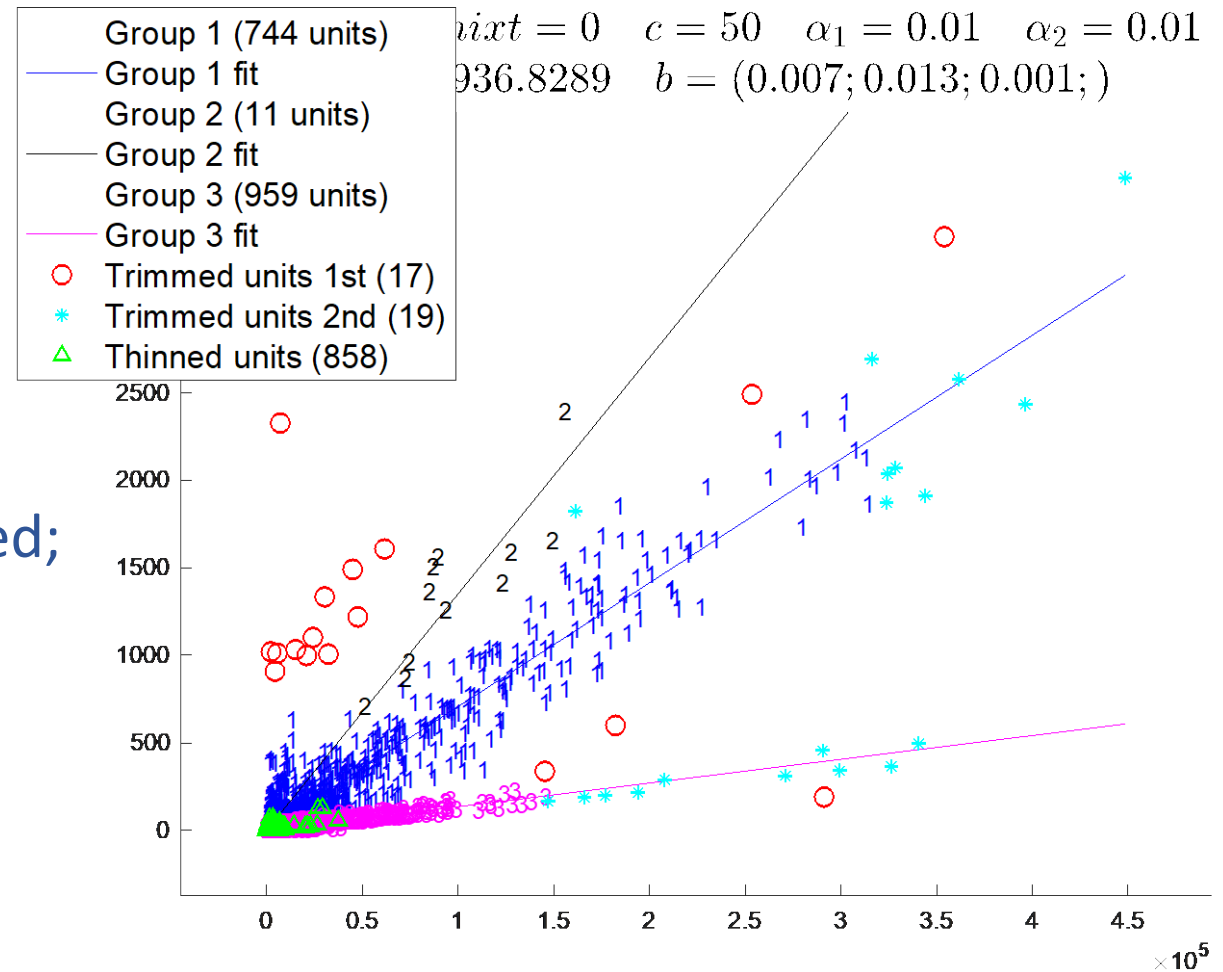
“Jewellery” dataset

Perrotta, Torti (2018). Discussion of “The power of monitoring: how to make the most of a contaminated multivariate sample”. Statistical Methods & Applications.

Robust clustering

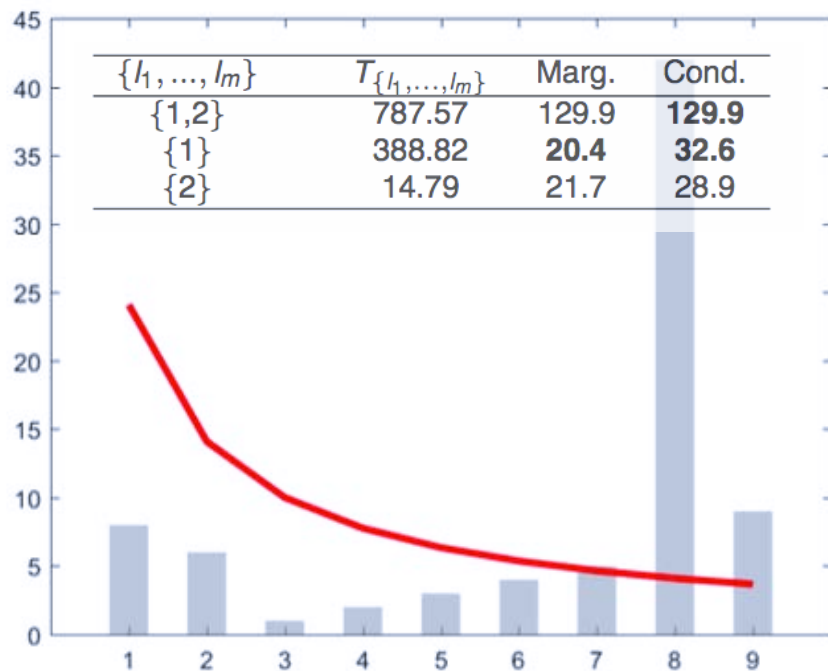
TCLUST-REG on the jewelry dataset:

- Three main market prices are estimated;
- Outliers are identified and removed.



Ceroli, A. and Perrotta, D. (2014): Robust clustering around regression lines with high density regions. Adv. Data Analysis and Classification 8(1): 5-26

Detection of data manipulations



A new two-stage Newcomb-Benford analysis developed by the JRC and the Universities of Parma and Siena

- [5] A. Cerioli, L. Barabesi, A. Cerasa, M. Menegatti, and D. Perrotta. The Newcomb-Benford law and the detection of frauds in international trade. 2017. Submitted.
- [6] L. Barabesi, A. Cerasa, A. Cerioli, and D. Perrotta. Goodness-of-fit testing for the Newcomb-Benford law with application to the detection of customs fraud. *Journal of Business & Economic Statistics*, 2017, <http://dx.doi.org/10.1080/07350015.2016.1172014>. Accepted for publication.

A serial fraudster detected in
SAD data by our two-stage
Newcomb-Benford analysis

Message of the presentation

*A good comprehension of your **data**, your relevant **fraud-control problems** and the corresponding **statistical patterns** may enable the application of the presented methods, models and tools to your context.*

	Spikes (in time series)	Outliers (in multi- variate data)	Systemati c spices or outliers	Systematic association s in 2 way tables
Stockpiling	X		X	
Fraud in export refunds	X		X	
Evasion of import duties		X, LP outliers	X	X
Deflection of trade	X, partly		X, partly	
Trade based laundered money in Origin. Generation of black money at destination.		X, HP, LP outliers	X	X
VAT carousels		X, LP outliers	X	X