FOODINTEGRITY
Ensuring the Integrity of the European food chain

613688: Collaborative Project

Seventh Framework Programme
KBBE.2013.2.4-01: Assuring quality and authenticity in the food chain

Deliverable: D8.2
Title: Final report of trade trend analyses, Sensitivity analyses

Author(s): Vahid Mojtahed (vahid.mojtahed@fera.co.uk)
Beneficiary(s): P1
Date of preparation: August 2016
Period covered: M32- M36
Status: version 1

<table>
<thead>
<tr>
<th>Dissemination level</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PU</td>
<td>Public X</td>
</tr>
<tr>
<td>PP</td>
<td>Restricted to other participants</td>
</tr>
<tr>
<td>RE</td>
<td>Restricted to a group specified by the consortium</td>
</tr>
<tr>
<td>CO</td>
<td>Confidential, only members of the consortium</td>
</tr>
</tbody>
</table>

The project has received funding from the European Union’s Seventh Framework Programme for research, technological development and demonstration under grant agreement No. 613688.
# Table of Contents

List of Figures .......................................................................................................................... 2
List of Tables .............................................................................................................................. 2
Executive Summary .................................................................................................................. 3

1. Introduction ......................................................................................................................... 4

2. Food Fraud ............................................................................................................................ 7
   2.1. Introduction .................................................................................................................. 7
   2.2. Food Fraud Literature and Early Warning System ...................................................... 8
       Early Warning Systems (EWS) ..................................................................................... 9
   2.3. Fraud Triangle and Triggers ....................................................................................... 10
   2.4. Methodology .............................................................................................................. 13
   2.5. Data ............................................................................................................................ 15

3. Results ................................................................................................................................. 17
   3.1. Introduction ................................................................................................................. 17
   3.2. Analysis of the UK Meat Trade .................................................................................. 17
       Distribution of Trade data ......................................................................................... 18
       Trend and Seasonality in Trade data ....................................................................... 19
       Distance and Outliers ............................................................................................ 20

4. Discussion and Conclusion ................................................................................................. 23

References ............................................................................................................................... 25
Appendices ............................................................................................................................... 27
   Appendix A: Terminologies and Definitions .................................................................... 27
   Appendix B: List of Acronyms ......................................................................................... 28
List of Figures

Figure 1. UK food chain (DEFRA, 2015) ................................................................................................... 6
Figure 2. Food fraud and RASFF timeline. (Source: European Commission)........................................... 7
Figure 3. Annual growth in the network structure of RASFF alerts between 2003-2008. (Nepusz et al., 2009)........................................................................................................................................................ 8
Figure 4. Fraud event management ....................................................................................................... 9
Figure 5. Fraud/Crime Triangle. (Donald Cressey, 1950) ...................................................................... 11
Figure 6. Food fraud triggers ................................................................................................................. 13
Figure 7. Aggregated and disaggregated meat trade ........................................................................... 19
Figure 8. Meat trade trend and seasonality .......................................................................................... 20
Figure 9. Outliers of the UK meat trade ............................................................................................... 21
Figure 10. Anomaly in categories of traded meat ................................................................................ 22

List of Tables

Table 1. Combined Nomenclature, 2016 ............................................................................................. 16
Table 2. FFD categories in Combined Nomenclature ........................................................................... 16
Table 3. Four digit CN categories of meat .......................................................................................... 18
Executive Summary

The costs of a food fraud incident for a food business are between 2-15% of their yearly revenue. Food fraud is the outcome of various types of changes in the status quo, in particular changes in climate, regulations, economic, etc. that render food supply chains more vulnerable to opportunistic fraudsters. In criminology, this concept is generally known as Fraud Triangle and it is of fundamental importance to the development of Early Warning Systems (EWS) for food fraud detections.

This report characterizes a set of market and non-market fraud triggers that are suggested by experts and in the literature. These triggers can either contribute to shaping the incentives of fraudsters to commit fraud or making the existing state of affairs more susceptible to fraudulent activities. We have proposed developing an Early Warning System that aims to indicate particular products/commodities or locations which have a heightened vulnerability to food fraud. Once operational, the EWS can effectively mitigate the loss of life and reduce the negative impacts of food fraud. Our proposed EWS is based on an effective method for analyzing high-dimensional, frequent trade data and detecting suspicious behavior in the international food supply chain. The proposed method neither requires information about the occurrence of past fraud events, nor a set of explanatory variables for sensitivity analysis and in this respect is truly capable to detect emerging risks rather than emerged risk. However, incorporating additional datasets regarding triggers of food fraud can reduce the likelihood of false alarms (false positives) and help us to attribute the alarm to the pertinent triggers of fraud captured by other dimensions of the dataset.

The report finds that the techniques used in this study and applied to the trade data have a great potential to detect prospective food fraud events. Retrospective application of this technique to meat trade data between the UK and 28 EU countries shows that we were able to raise timely red flags before the breakout of the horsemeat scandal in January 2013. These red flags and warnings could have raised the awareness of the food business operators, traders, quality assurance managers, and national authorities persuading them to carry out investigations and conducting more frequent and rigorous analytical tests in order to expose the fraud attempts and prevent the incidence from happening.

The report also investigates the fact that the analysis conducted has some limitations such as the small set of the commodities and relevant datasets that have been considered due to constraint in resources constraints. Unfortunately, there is no one-size-fit-all solution and bespoke EWSs must be developed for various types of commodities and industries, which often require sensitive data provided by food, feed and drink businesses.
1. Introduction

This report summarizes the results of trade trend analysis as part of the Work Package 8 of FoodIntegrity project and specifically it undertakes activities highlighted in Tasks 8.1 (subtask 8.1b) and 8.2 (subtask 8.2b). While this piece of work and the developed method is a standalone product, it can also inform the subsequent Task 8.2 (being developed by other work package partners), 8.3 (being developed by Fera), and 8.4 (i.e. FRAME Fraud system) and their subtasks.

This deliverable (D8.2) contributes partially to the first objective of work package 8 that is to develop a structured approach for collecting and analysing information that could be potential drivers of the EU food chain fraud events.

The focus of this deliverable is on monitoring and analysing trade data as an essential component of developing the Early Warning Systems (previously referred to in the proposal as FRAME Fraud system) in Tasks 8.3 & 8.4, and to understand how trade could exhibit sensitivities to changes in underlying economic, regulatory, and climate and many other conditions of national or international communities within which FFD businesses operate. Most of the food today does not follow a straight line from farm to fork. In the past few decades, the change in the tastes of consumers as well as increase in the global transportation due to improvements in transportation infrastructure, and preservation technologies have changed the structure of production and has resulted in ever-growing international trade. This in turn has made the global food supply network more fragmented and complex as different actors in a given supply chain are located in different countries providing an opportunity for fraudsters to conceal their fraudulent activities under the umbrella of such complex and global network.

While the focus of this project is on the whole EU food supply chain, it is worth to look at the UK as an example of a country that is heavily dependent on imported processed or unprocessed FFD products. For the UK, only 54% of the food consumed in the UK is produced in the country with the rest coming from all over the world [1]. The total amount of imported foods and drinks to the UK in 2014 were £38.5 billion (See Figure 1) feeding into manufacturing, wholesalers and retailers sectors. Over 22 countries have supplied unprocessed food to the UK with the leading supplier being mainly from the EU countries such as Netherlands, Spain, France, Germany and Irish Republic[1]. These highlighted facts clearly establishes that the UK food supply while being a valuable and leading sector, it is greatly exposed to food fraudsters.

It is estimated that the costs of adulteration and counterfeiting of food and consumer products sums up to between $10 to $49 billion per year[2, 3]. The costs associated to a single product adulteration incidence are on average between 2 to 15 % of yearly revenues of a company. This is translated to damages equal to $60-400 million for $500m to $10b companies[2]. In order to prevent fraud and adulteration it is important to think like a fraudster and identify emerging opportunities that they can use to make profits out of committing fraud. As the very first step, this requires generating meaningful hypotheses through delving into data from different systems, at different time scales, and in complex formats to discover hidden relationships and useful insights. Often, to generate smarter hypotheses we need to consult with experts and stakeholders in the FFD industry as their comprehension of supply chains could lead us into asking the right questions. Expert elicitation jointly employed with objective data can lead us to fast discovery of motives behind fraud.
However, we face several challenges in achieving our goal of detecting fraud, in particular concerning models and data. Developing models for well-behaved behaviours is one thing, but developing a model to account for disruptions or rare events such as fraud, requires much more than model sophistication. We need to have data on the frequency and types of anomalies and their impacts which are not usually available in our application domain, i.e. food fraud detection. Hence, as a prerequisite for sensitivity analysis, this report has focused on finding a technique that could be used for identifying anomalies as a signal of fraudulent activities when there is no clear demarcation between normal and suspicious activities. Anomaly detection techniques are powerful machine learning techniques that have been used extensively to identify fraud in several other domains of application such as insurance, credit card, health claims, online bank transactions, cyber network intrusions, space craft safety, etc. Often these advanced analytics are applied to extremely large volumes of structured (i.e. numerical data bases) or not structured (i.e. texts, tweets) data.

In terms of utilised datasets, we have primarily focused on the international trade data. The trade data is rich, large, and reflects the effects of bad climate, routine fluctuations (seasonality in production), business cycles, change in regulations, and one-time events or disruptions. Despite these advantages, using it also imposes some significant challenges especially when the ancillary databases such as regulatory, fraud, climate, or socio-economic records are not always available to establish the ground truth. Since the trade dataset is real and therefore sometimes incomplete, it might not necessarily fit into well-behaved Gaussian distributions such as normal distribution. This, to some degree, constrains us in using traditional parametric statistical models. Furthermore, the dataset can have missing records which presents challenges in employing some statistical analysis methods.

This deliverable is structured as follows. In Section 2, we review the state-of-the-art in development of EWS and provide a review of the concept of Fraud Triangle in criminology. In addition, we give taxonomy of drivers of food fraud that is suggested in the literature and by experts. Section 2.4 gives an in depth review of available methodologies in identifying anomalies and outliers in multivariate datasets. In subsection Error! Reference source not found.Error! Reference source not found. we lay the groundwork for our proposed technique to analyse the trade data and interpret the results. In Section 3, we summarise the results of our findings from applying detection technique to meat trade data. Section4 concludes and discusses the implication of our findings for quality assurance managers in FFD businesses.

\footnote{Note that through this report we use anomaly and outlier interchangeably.}
Figure 1. UK food chain (DEFRA, 2015)
2. Food Fraud

2.1. Introduction

The scientific literature on the topic of FFD fraud, Economically Motivated Adulteration (EMS) and characteristics of them is still quite sparse. While the history of food fraud incidences goes centuries back, the first systematic European effort to identify emerging food risks started in 1979 through coordinating efforts to share information among the network of European countries (See Figure 2). However, it was only after big international scandals, known as “food crises”, such as melamine in infant formula in 2008 or horse meat scandal in 2013 that regulatory bodies and consequently scientific communities started looking into the reasons behind repeated occurrence of such incidents. Among the examples provided above, the first one is an example of food fraud that had serious health implications, and the second one is more concerned with loss of consumer’s confidence and its consequences to FFD industries in terms of costs. Nevertheless, in both cases the motivation was economic gain.

These incidents and many other recent ones sparked the needs for more proactive measures[4]. One of the suggestions is to develop an EWS that could help the regulators and food industry to monitor the global complex food supply chain and follow up more effectively on emerging risks. The existence of numerous categories of FFD commodities makes it very difficult for retailers or custom offices to test all of these products in order to find adulterated products. For this reason, many food safety authorities such as EFSA have switched their attention to experts’ opinion and judgements to benefit from knowledge network and information sharing in such networks. However, they have realised that identification of emerging risks will require a high level of expertise because of data gaps and uncertainties within the process [5]. The availability of various, high volume, frequent data bases (known as “Big Data”) and cost-effective predictive analytics such as machine learning algorithms...
could help us to overcome these limitations and develop more efficient EWSs to transform supply chain monitoring and detect suspicious activities in food supply chains.

“Big data” is the buzzword of these days, but it is more and more providing us with the opportunity to change the current business models and day to day decision makings. Big data together with advanced analytics are being used to modernize the transportation policies, revolutionize the interaction with customers and improved medical practices. However, big data and advanced analytics are useless unless the right sets of questions are being asked and well posed hypotheses are formed. The value of data and advanced analytics is only realized through insights. In this section, we aim to dig into the reasons behind the fraud incidences and factors that contribute to the motivation of fraudsters. We start by reviewing current sparse literature and then providing some background behind development of EWS and its functional components. The last part of this section categorizes the recognized drivers of food fraud and briefly explains how they can contribute to food fraud.

2.2. Food Fraud Literature and Early Warning System

In the EU, the studies that we have found are interrogating in one way or another identical databases, which is the RASFF database. Kleter, Prandini [6] studied the RASFF notification in four-years period from 2003 to 2007 to identify the trends within specified categories and the relationships between product and hazard categories. Their study mainly focuses on the safety side of food and does not consider fraud alerts. Naughton, Nepusz [7] and Nepusz, Petróczi [8] carried out the first application of data mining techniques to the RASFF data. They have constructed a network to analyse the ever-growing RASFF dataset and identify clusters in data such as the major transgressors countries and detectors according to each category of commodities (see Figure 3).

![Figure 3. Annual growth in the network structure of RASFF alerts between 2003-2008. (Nepusz et al., 2009)](image)

In an step forward, Bouzembrak and Marvin [9] developed a Bayesian network based on the 749 RASFF notification database in the period of 2000-2013. The Bayesian network is trained to predict the product category subject to fraud given knowledge about food fraud type for imported products and also the origin of the country. The Bayesian network was then validated using the information about fraud in 2014. The accuracy of the food fraud type prediction is close to 80% when sufficient
amount of information about the country of origin, food fraud type, and food category is historically recorded. However, the prediction goes down near to 50-50 chances that are almost random when none of the above-mentioned information is available in the database.

Figure 4. Fraud event management.

Early Warning Systems (EWS)
The development of Early Warning Systems received the attention of policy makers and stakeholders after the Indian Ocean tsunami in 2004 and since then it has been mainly developed for combatting the adverse impacts of natural disasters. EWS can be developed to detect a wider range of harmful events such as missile launches, disease outbreaks, financial meltdowns, etc. Nowadays, EWS is an essential component of disaster risk reduction as it prevents the loss of lives and reduces the economic impacts of the disasters. An effective EWS requires actively involving communities exposed to risk, enhancing public awareness, disseminating alerts and warnings and ensuring there is a high level of preparedness[10].

There are generally two types of EWS, one type focusing on detection and decision support aspects, and the other type focusing on warning and response aspects. Therefore, the EWS can be considered as a chain of information and modelling systems comprising of subsystems that include data gathering (i.e. sensors, surveys), event detections (i.e. models), decision support systems, and messaging plans and infrastructure (e.g. SMS, apps, websites)[11]. Our intended EWS falls within the first category, comprising of data gathering and detection subsystems, which are working together to forecast and signal anomalies in the food supply chain that could adversely impact the wellbeing of consumers and/or financial stability of food businesses. In this regard, the EWS provides the food authorities and businesses with enough time to respond in a timely manner, prepare for adverse events, and reduce the impacts (See Figure 4 ). EWS are also economically effective tools for event management. In Figure 4, we show where the EWS stands in the timeline of event management. As we move from left to right, the costs associated with event management increases. For instance the Intervention phase typically involves spending hundreds or thousands of Euro for recalling the products, disposal of products, loss of revenue due to brand reputation damage, and loss of confidence of customers.

A fully fledged EWS should be able to carry out four functionalities in a parallel manner [12]:
2.3. **Fraud Triangle and Triggers**

Detecting fraud is not an easy task as it requires information and knowledge about the nature of fraud (adulteration, dilution, label of origin, etc.), as well as where and how it has been committed. As Spink and Moyer [4] have mentioned, the current intervention systems are not capable to discover near infinite number of contaminants or ingredients used for food fraud. Moreover, the nature and instances of food fraud are constantly changing. Once an incident of food fraud is exposed to the public, the chances are that it will not occur again for some time or will manifest in a slightly different form/application.. Incidents such as melamine in infant formula and pet food, Asian catfish sold as grouper, horse meat sold as beef meat, or pomegranate juice cut with grape juice are not regularly repeated enough that we can record their frequencies. In addition food fraud often goes undetected again reducing the amount of data available.

However, looking at all types of fraud, there is a vast literature on why fraud occurs in various sectors including the food sector. The idea behind this work is rooted in Donald Cressey’s fraud theory going back to 1950s and widely used for training external auditors for assessing the risks of fraud. In short, Cressey’s fraud theory explains why fraudsters commit fraud and it has been conceptualized as the “fraud triangle”. Cressey after interviewing 250 criminals concluded that for a person to violate trust and commit fraud, three factors must be present (see Figure 5): 1) having a financial problem 2) having an opportunity to commit the fraud, 3) having rationalization[13]. As Kassem and Higson [14] have highlighted many factors explaining fraud could not be easily observed such as pressure, rationalization, and capabilities of fraudsters. It is based on these concepts in criminology, originally proposed based on the work of Cressey, that many of the food fraud scientists [4, 15] have tossed the term of “food fraud triangle”, which essentially defines a fraud opportunity triangle based on victim, fraudster, absence of capable guardian as the legs of it.

Food fraud and food defence are considered to be intentional actions, with the former having motivation for gaining economic benefits in contrast to for instance food safety[4]. The idea that we pursue in this deliverable is that to detect fraud, we need to think like fraudsters and look at the factors that could influence the emerging size of the fraud opportunity. The bigger is the size of the fraud opportunity, the higher is the likelihood of fraud to happen in that given time and circumstances.
It is important to note that there are many categories of food fraud as the result of not following Good Manufacturing Practices (GMA), Good Hygiene Practices (GHP), or Good Agricultural Practices (GAP) [15]. The term food fraud is used collectively to refer to deliberate and intentional substitution, addition, tempering, or misrepresenting of food, its ingredients or packaging. Among the many possible categories of food fraud, our focus is mainly on Economically Motivated Adulteration (EMA) defined by intentional addition or substitution of substances in a product for increasing the value of the product or reducing the costs of its production[16]. The reason is that there are clear motivations for fraudsters in this category of food fraud that could be tracked back to changes in economic, climate or regulatory conditions.

The Elliot report [17] has also considered and proposed several other factors that could contribute to the growing size of food fraud opportunity and crime such as i) global complexity of food supply chain, ii) high expenditure of consumers on feed and drink industry, iii) competitiveness of the food industry. While the abovementioned list of food fraud triggers is limited, we believe there are many other factors that can be considered as other legs of fraud opportunity, hence changing its shape from a triangle to a polygon.

It is important to understand and disentangle the drivers behind the food fraud whether they are changes in the global market such as global supply, or demand shocks or local changes in the FFD markets, exchange rates, regulations, or costs of local inputs such as energy, labour, raw materials, etc. We have identified five main categories of characteristics of susceptible environment to food fraud based on the suggestions in the literature and the stakeholders’ inputs. These are (See Figure 6):

i) Socio-economic characteristics;
ii) Regulations and standard
iii) Industry or sector characteristics
iv) Commodities characteristics
v) Climate, Pests, Health issues
The first category, socio-economic characteristics, encompasses mainly market-driven variables that could signal the likelihood of economic adulteration. Commodity prices are the most distinctive variable signalling scarcity of commodities and ingredients as a result of poor harvest, increase in transportation costs, higher demands coming from new markets, and burden on food suppliers to maintain their contractual obligations to large retailers to supply foods at a fixed price. It is also perceived that economic downturns will affect not only the budget of the regulatory bodies to test for fraudulent products but also it puts pressure on industries to produce cheap products when they have limited access to credit markets or they struggle with lower demand resulted from consumers lower income. Imposing higher tariffs, taxes, or custom duties on producers in certain countries have also give rise to the false labelling of the country of origin for avoiding such duties.

The second category of triggers is the changes in the regulations and standards. These may include changes in the allowed amount of certain ingredients and funguses, development of new methods for analytical testing, or changes in definitions of products (e.g. European definitions of meat restricting usage of connective tissues and fats) that would eventually require developing new analytical tests for testing the products meeting the new regulations. For instance developing new analytical tests for testing the connective tissues of meat can take some time and this gives rise to an opportunity for fraudsters to sell the now lower quality meat for the value of higher quality meat that is without connective tissues. Penalties can act as a preventive factor in any crime and food fraud is no difference. In the UK, the enforcement officers are mainly relying on the definition of adulterations and substitutions as provided in Food Safety Act 1990 surely. Currently, the penalties for adulteration set under Food Safety Act 1990 are lower than other types of fraud including financial ones. This means that the prosecuted fraudsters will not have much to lose when they get convicted but any increase in the level of penalties will decrease their intentions to commit fraud.

The third category considers sector specific characteristics. Some organizations, such as U.S. Pharmacopeial Convention (USPC), have already proposed a set of best practices for manufacturers to assist them in assessing their vulnerability through identifying the fraud factors and determining the economic and health impacts of them. Some of these fraud factors can be inherent to the ingredients (e.g. fraud history of that ingredient) or they can be controlled by the manufacturers such as voluntary testing frequency. Some of these factors are reported in USPC and it includes suppliers relationship, testing frequency, auditing strategy, supply chain, etc. Unfortunately many of these factors cannot be incorporated in an EWS since they are either too specific or difficult to obtain relevant data about them. An important market driven fraud factor is the high profit margin of food sectors for certain commodities. Based on the historical frequencies of commodities subject to fraud, we can broadly categorize the commodities into two categories i) low volume, high premium and ii) high volume, low value. Examples of the first category are spices such as saffron or vanilla and examples of the second group of commodities include meat and milk.

The fourth category focuses on the intrinsic characteristics of the commodities that make them more subject to fraud. These may include the nature, identity, properties, durability, and origin of the commodity. For instance, the degree substitutability of some fruit juices are high as it is easy is to substitute them with other ingredients of lower quality such as of pomegranate juice with apple or
grape juice. The nature and properties of the commodities also determine how easy it is to develop methods for testing them analytically.

Last but not least category includes non-market driven factors that affect the likelihood of fraud in an indirect way. These factors can be related to climate (i.e. extreme events), crops’ pest pressure, and number of visits to vet as a result of spread of diseases among livestock. For instance, extreme weather events in Madagascar and Indonesia as the two major producing countries of vanilla resulted in a spike in the price of vanilla triggering fraudsters to falsely label synthetic vanilla instead of natural vanilla. Similarly, an overall increase in the number of visits to vets in a country is a sign of epidemic disease among livestock signalling an upcoming shortfall of meat in that country, and an opportunity for fraudsters to fulfil the demand with other substitutes.

![Figure 6. Food fraud triggers](image)

### 2.4. Methodology

As mentioned in Introduction, frauds are considered to be rare events and therefore data regarding fraud incidents are often scarce as only a small fraction of fraud incidents are detected and thus reported. However, even if the data were readily available, the utility of them is limited or questionable as the characteristics and nature of food fraud are rapidly changing over time often rendering historical data unsuitable for future prediction. Fraudsters always tend to find innovative techniques and methods to commit their crime without being detected or caught and analytical tests are often lagging in catching up with them.

For this reason our aim is to analyse the trade data using statistical methodologies combined with machine learning algorithms that are restricted to unsupervised learning. By unsupervised learning we mean that there is no need for the data to be labelled by type of fraud incidents in order to be able to infer or predict future fraud incidents. This is in contrast to the supervised methodologies that rely on datasets consisting of a pair of inputs vectors and desired/undesired output values or
labels. These types of unsupervised methods are particularly suitable for developing our intended EWS since it requires minimum inputs from the users for performing its task and also it does not require any new training data set that provides new information such as updated lists of commodities that have been subject to fraud, or countries that have exported fraudulent commodities, so on and so forth.

Another important criterion for searching and selecting the right algorithm for our purpose is the capability of the algorithm to process high-dimensional and high frequency data in a fast, efficient and reliable manner. As mentioned in Introduction, trade data are available on a monthly basis, across 200 countries, and at a very fine granularity with over 5,000 commodity groups being listed\[21\]. This means that for an EWS to be successful in identifying fraud, preferably in real time, it should be able to process a large amount of multivariate data not only comprised of multilateral trade but also of other ancillary databases.

The type of the data (i.e. international trade data) that the author is using and the nature of the problem impose serious constraints on our choice of the methods for sensitivity analysis. In general, the data points can be related to each other. Examples of that are spatial data, sequential data and graph data. In sequential data, the data points are ordered, for instance, protein and genome sequences, or time series data. The trade data are monthly export or import values or quantities between countries. Yet, the phenomenon of interest, such as fraud can span several months; hence we are faced with the higher order problem of identifying uncommon patterns beyond single outliers. Additionally, as mentioned before, the dataset is real and generally not multivariate normal with possible missing records which could always present challenges. Our analysis challenge is enhanced since ancillary data such as business or commodity related records, are not always available to validate our findings.

The problem of outlier detection is well studied in statistics and data mining literature [22-27]. This report focuses in particular on a family of candid techniques for identifying anomalies in complex trade patterns, in particular when there is no clear indication between routine trade and unusual activity that could signal fraudulent activity. The advanced analytics that we are using is based on a series of procedures that starts with reducing the dimensionality of the data sets that we have in hand. At the same time and during the process of reducing the dimensionality, we seek to identify a normal pattern in the when they are collectively considered together. All the data points that do not adhere to this normal pattern are then tested against a reference distribution to ensure that our method is robust and insensitive to the presence of big anomalies that could distort the normal patterns of the data and give us a false positive.

Note that our emphasis is on multivariate outlier and not necessarily identifying an extreme in any individual component [28]. Outliers are considered to be data points coming from one or more different distributions, and extremes are instead values that are far away from the centre but belonging to the same distribution. This is the reason that we have characterized the fraudulent activities as a result of collective deviations from the status quo or what is considered to be normal conditions. The collective set of variables could include only trade data of different correlated commodities (i.e. horse and beef meats) or it could consist of other types of data such as climate, socio-economic variables, regulations, etc. For instance simultaneous but small changes in variables such as import of horse meat, price of horse meat, price of beef together with change in “desinewed
meat” regulations could signal an anomaly resulted from fraudulent activities. And that is exactly what happened in 2013 horse meat scandal[29].

2.5. Data

The data used in this study are the international trade data in good and services. They are downloaded from Eurostat website but originally are gathered by each member state in a harmonized manner legislated by the Community Legislation. The data is recording monthly trade between member states (Intra-EU trade) and between the member states and non-EU countries(extra-EU trade) since 1995 under 21 categories of commodities as seen in Table 1. The first three sections consists of 15 chapters (see Table 2) that are mainly dedicated to the food, feed and drink. These are the categories that are the main focus of this this study.

The data records the trade in terms of export and import at the most detailed level of several nomenclature such as Combined Nomenclature (CN) which corresponds to Harmonized System (HS) plus a further breakdown at 8 digit level.

<table>
<thead>
<tr>
<th>Section number</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Section 1</td>
<td>Live animals; animal products</td>
</tr>
<tr>
<td>Section 2</td>
<td>Vegetable products</td>
</tr>
<tr>
<td>Section 3</td>
<td>Animal or vegetable fats and oils and their cleavage products; prepared edible fats; animal or vegetable waxes</td>
</tr>
<tr>
<td>Section 4</td>
<td>Prepared foodstuffs; beverages, spirits and vinegar; tobacco and manufactured tobacco substitutes</td>
</tr>
<tr>
<td>Section 5</td>
<td>Mineral products</td>
</tr>
<tr>
<td>Section 6</td>
<td>Products of the chemical or allied industries</td>
</tr>
<tr>
<td>Section 7</td>
<td>Plastics and articles thereof; rubber and articles thereof</td>
</tr>
<tr>
<td>Section 8</td>
<td>Raw hides and skins, leather, furskins and articles thereof; saddlery and harness; travel goods, handbags and similar containers; articles of animal gut (other than silkworm gut)</td>
</tr>
<tr>
<td>Section 9</td>
<td>Wood and articles of wood; wood charcoal; cork and articles of cork; manufactures of straw, of esparto or of other plaiting materials; basketware and wickerwork</td>
</tr>
<tr>
<td>Section 10</td>
<td>Pulp of wood or of other fibrous cellulosic material; recovered (waste and scrap) paper or paperboard; paper and paperboard and articles thereof</td>
</tr>
<tr>
<td>Section 11</td>
<td>Textiles and textile articles</td>
</tr>
<tr>
<td>Section 12</td>
<td>Footwear, headgear, umbrellas, sun umbrellas, walking sticks, seat-sticks, whips, riding-crops and parts thereof; prepared feathers and articles made therewith; artificial flowers; articles of human hair</td>
</tr>
<tr>
<td>Section 13</td>
<td>Articles of stone, plaster, cement, asbestos, mica or similar materials; ceramic products; glass and glassware</td>
</tr>
<tr>
<td>Section 14</td>
<td>Natural or cultured pearls, precious or semi-precious stones, precious metals, metals clad with precious metal, and articles thereof; imitation jewellery; coin</td>
</tr>
<tr>
<td>Section 15</td>
<td>Base metals and articles of base metal</td>
</tr>
<tr>
<td>Section 16</td>
<td>Machinery and mechanical appliances; electrical equipment; parts thereof; sound recorders and reproducers, television image and sound recorders and reproducers, and parts and accessories of such articles</td>
</tr>
<tr>
<td>Section 17</td>
<td>Vehicles, aircraft, vessels and associated transport equipment</td>
</tr>
<tr>
<td>Section 18</td>
<td>Optical, photographic, cinematographic, measuring, checking, precision, medical or surgical instruments and apparatus; clocks and watches; musical instruments; parts and accessories thereof</td>
</tr>
<tr>
<td>Section 19</td>
<td>Arms and ammunition; parts and accessories thereof</td>
</tr>
</tbody>
</table>
### Table 1. Combined Nomenclature, 2016.

<table>
<thead>
<tr>
<th>Section</th>
<th>Chapter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Section 1</td>
<td>Chapter 1</td>
<td>Live animals</td>
</tr>
<tr>
<td></td>
<td>Chapter 2</td>
<td>Meat and edible meat offal</td>
</tr>
<tr>
<td></td>
<td>Chapter 3</td>
<td>Fish and crustaceans, molluscs and other aquatic invertebrates</td>
</tr>
<tr>
<td></td>
<td>Chapter 4</td>
<td>Dairy produce; birds’ eggs; natural honey; edible products of animal origin, not elsewhere specified or included</td>
</tr>
<tr>
<td></td>
<td>Chapter 5</td>
<td>Products of animal origin, not elsewhere specified or included</td>
</tr>
<tr>
<td>Section 2</td>
<td>Chapter 6</td>
<td>Live trees and other plants; bulbs, roots and the like; cut flowers and ornamental foliage</td>
</tr>
<tr>
<td></td>
<td>Chapter 7</td>
<td>Edible vegetables and certain roots and tubers</td>
</tr>
<tr>
<td></td>
<td>Chapter 8</td>
<td>Edible fruit and nuts; peel of citrus fruit or melons</td>
</tr>
<tr>
<td></td>
<td>Chapter 9</td>
<td>Coffee, tea, maté and spices</td>
</tr>
<tr>
<td></td>
<td>Chapter 10</td>
<td>Cereals</td>
</tr>
<tr>
<td></td>
<td>Chapter 11</td>
<td>Products of the milling industry; malt; starches; inulin; wheat gluten</td>
</tr>
<tr>
<td></td>
<td>Chapter 12</td>
<td>Oil seeds and oleaginous fruits; miscellaneous grains, seeds and fruit; industrial or medicinal plants; straw and fodder</td>
</tr>
<tr>
<td></td>
<td>Chapter 13</td>
<td>Lac; gums, resins and other vegetable saps and extracts</td>
</tr>
<tr>
<td></td>
<td>Chapter 14</td>
<td>Vegetable plaiting materials; vegetable products not elsewhere specified or included</td>
</tr>
<tr>
<td>Section 3</td>
<td>Chapter 15</td>
<td>Animal or vegetable fats and oils and their cleavage products; prepared edible fats; animal or vegetable waxes</td>
</tr>
</tbody>
</table>

Table 2. FFD categories in Combined Nomenclature.
3. Results

3.1. Introduction

In this section, we present the results of analyzing the multilateral trade to identify any indication of fraudulent activity in the FFD supply chains. We focus our attention on the UK trade with the rest of the world based on the monthly trade data and in particular, import from Intra- and Extra-EU. We have chosen the UK for two reasons: i) because it is a good example of an open economy, which depends on outside world for 60% of its food and feed; ii) we were able to validate our findings in the next phases with the historical info of the past incidence.

As we mentioned before, the trade data can have different granularity going from coarse categories such as meat or fish in general to finer categories that specifies the type of meat and its characteristics (i.e. bovine or swine, fresh or frozen). In the exercise that we have conducted in this study, we have considered aggregated in conjunction with disaggregated categories of meat. In particular, sometimes we purposefully consider a set of disaggregated categories conjointly because of some degrees of substitutability that exists among them. For instance, we expect this among various types of meat categories when fraudsters can substitute the lower quality/less valued meat with the higher quality/more expensive type of meat. Consideration of such relationships among different dimensions of the data is essential in identifying anomalies.

An implicit step of identifying anomalies in the supply chain is to characterize what is normal. The characterization of normal pattern requires combining various datasets and consideration of the relationships that exists among various dimensions of the dataset (not necessarily among subcategories). However, this could result in the “curse of dimensionality”. That is a term referring to various phenomena arising when we analyze and organize data in high-dimensional spaces and which do not occur in low-dimensional settings. The identification of normal pattern in our data is with the help of a reference distribution as a benchmark. In order to tackle the challenge of multi-dimensional data and reduce the dimensionality of our analysis, we use a distance measure that takes into account the location and scale parameters of the dataset.

Unfortunately, the number of European food fraud incidences similar to the horsemeat scandal is limited as the fraud incidences are rare events and are concealed. This makes the generalization of this exercise limited as the validation phase become challenging. For this reason, although we have applied the anomaly detection algorithm to many other categories of the commodities and have detected a series of anomalies but the results are not reported until the next reports and when they will be validated using experts’ or stakeholders opinion.

3.2. Analysis of the UK Meat Trade

We look at the trade pattern of meat between the UK and the rest of the EU countries. This exercise allows us to understand if this type of analysis will be able to pick up past events such as the horsemeat scandal in 2013. We will look at the aggregated import of meat as well as disaggregated import of different varieties of meat products as summarized in Table 3. The analysis only considers the value of import therefore both quantities and prices are taken into account. In particular, we are interested in consideration of relative prices of various types of meats that could be substituted with...
each other. Our dataset can eventually consider other dimensions including market or non-market drivers of fraud. However, for us it is enlightening if analyzing trade data alone provides us with enough predictability of suspicious activities that can be marked as fraud. We have considered 48 months of trade with Intra-EU28 block starting from January 2012 to December 2015 giving us enough time before and after the horse meat scandal to detect any abnormal changes in the pattern of trade data in particular around the time of the horsemeat scandal. To keep the exercise simple, we have considered trade with EU28 block instead of considering trade with individual countries. Without loss of generality, we can include trade with the individual countries in future exercises.

**Distribution of Trade data**

Figure 7 depicts the box plots of meat trade for all the 10 disaggregated categories and the aggregated category. The middle position of the median for categories 2, 201, 202, 210 coupled with the balanced whiskers suggested that trade patterns are normally distributed for these categories. Note that the boxplot shows few outliers for categories 201, 203, 207, 210. Nevertheless, even if the data are not normally distributed, our method is still robust and could be used for identifying the anomalies.

<table>
<thead>
<tr>
<th>02</th>
<th>Meat and edible meat offal</th>
</tr>
</thead>
<tbody>
<tr>
<td>0201</td>
<td>Meat of bovine animals, fresh or chilled</td>
</tr>
<tr>
<td>0202</td>
<td>Meat of bovine animals, frozen</td>
</tr>
<tr>
<td>0203</td>
<td>Meat of swine, fresh, chilled or frozen</td>
</tr>
<tr>
<td>0204</td>
<td>Meat of sheep or goats, fresh, chilled or frozen</td>
</tr>
<tr>
<td>0205</td>
<td>Meat of horses, asses, mules or hinnies, fresh, chilled or frozen</td>
</tr>
<tr>
<td>0206</td>
<td>Edible offal of bovine animals, swine, sheep, goats, horses, asses, mules or hinnies, fresh, chilled or frozen</td>
</tr>
<tr>
<td>0207</td>
<td>Meat and edible offal of fowls of the species Gallus domesticus, ducks, geese, turkeys and guinea fowls, fresh, chilled or frozen</td>
</tr>
<tr>
<td>0208</td>
<td>Meat and edible offal of rabbits, hares, pigeons and other animals, fresh, chilled or frozen (excl. of bovine animals, swine, sheep, goats, horses, asses, mules, hinnies, poultry fowls of the species Gallus domesticus, ducks, geese, turkeys and guinea fowls)</td>
</tr>
<tr>
<td>0209</td>
<td>Pig fat, free of lean meat, and poultry fat, not rendered or otherwise extracted, fresh, chilled, frozen, salted, in brine, dried or smoked</td>
</tr>
<tr>
<td>0210</td>
<td>Meat and edible offal, salted, in brine, dried or smoked; edible flours and meals of meat or meat offal</td>
</tr>
</tbody>
</table>

*Table 3. Four digit CN categories of meat*
Trend and Seasonality in Trade data

Figure 8 shows the decomposition of the trade time series in each category of meat into their trends, seasonality and remainder. This helps us to visually inspect the data and gives us an understanding of the pattern of trade in detail. For instance, for category 2, we can see an increasing trend of import of meat in general accompanied with a clear seasonality during that period. The seasonality includes a peak in the ending months of the year, in particular in September, and a drop in the beginning of the year particularly in January. The seasonality usually includes factors that are periodically repeated over the course of the year such as the consumers purchasing patterns of meat or the production seasonality of livestock and agricultural crops. The remainder includes any variation in the data that are not explained by trend or seasonality. A spike in the remainder means that the trade in a given month is not following its trend or seasonality path while a low remainder means that trade is in line with its trend and seasonal variation. The remainder can be a weak indication of anomalies in the trade data. It is an indication because it indicates a change in the normal pattern of data and it is weak because the spikes should be assessed relative to its neighbourhoods and the likelihood of the magnitude of spike in remainder should be assessed with respect to some statistical benchmark that is not present.
Distance and Outliers

We are depicting the distances of the 11-dimensional space of monthly trade observations in the top left-hand side of Figure 9. Top-right figure in Figure 9 plots the robust distances ($d_i^2$) of imported meat (all categories at the same time) against the empirical distribution of the distance ($d_n^2$) shaped by the number-month of the data. Moreover, our reference distribution function is plotted along with two vertical lines demonstrating the specified distribution quantiles and the adjusted quantile. The adjusted quantile divides the anomalies from the normal pattern of the data. The two bottom figures show the outliers detected based on the specified quantiles (97.5%) of the reference distribution and the adjusted quantile.
As it can be seen from Figure 9, we have detected anomalies in months 6, 10 and 34 since they have exceeded the 97.5 quantile of reference distribution. It is interesting to note that the largest anomaly happened in month 10, which refers to October 2012 just 2 months before the outbreak of the horsemeat scandal! This result shows that considering all categories of imported meat such as bovine, swine, horse, etc. an abnormal pattern of trade has happened in certain months during 2012-2016. It does not tell us to which category of imported meat we can attribute the anomaly. For this reason, we will try to zoom in into the finer categories of traded meat and pinpoint which type of meat is causing this anomaly.

In Figure 10, we plot the multivariate anomalies that we found in previous subsection and plot each category of meat type in a one dimensional scatter plot. The anomalies are marked using combination of symbols and colours. Based on robust distances, a cross means a big anomaly and a circle means a small anomaly in terms of magnitude and according to the Euclidean distances, red mean a big value for anomaly and blue mean a small value [30].

As it can be seen from Figure 10, all the three anomalies are associated with the trade in category 205 which according to Table 3 refers to “Meat of horses, asses, mules or hinnies, fresh, chilled or frozen” The biggest anomaly is the one that refers to month 10 of the dataset meaning October 2012. It is important to note that the horsemeat scandal was the result of testing processed meat (beef burgers) in Ireland and not raw meat while our study shows an anomaly coming from importing raw meat to the UK from the countries inside the EU28 block. Moreover, the main reason behind the horsemeat scandal is known to be the complex network of companies trading in and mislabelling raw
and processed beef products which was originated from Romania[31]. The investigations showed that the horsemeat was originated from Romania with correct labels but eventually was mislabels as it passes through the European borders and several actors in the supply chain[32]. Although the outbreak of the horsemeat scandal happened in January 2013, but the reports and leaked documents indicate that the raw meat started to be delivered to producers well before the outbreak since 1 August 2012. This date more or less coincides with the other anomaly that we have found in month 6 meaning June, 2012. Therefore, although we cannot confirm that our detected anomaly is precisely referring to the horsemeat scandal but it raises enough doubt and red flags given the coincidences between their timing and involved category of the meat type.

In practice, the outcome of such EWS is to raise warnings for quality assurance managers of food companies and national food authorities to follow up and perform the necessary, but more expensive and time consuming, analytical tests to verify that quality and safety of the commodities meets the desired standards.
4. Discussion and Conclusion

Slight changes in one country’s socio-politico-economic circumstances that result in changes in production or demand for FFD commodities could create a wave of adulteration opportunities for fraudsters across the world to fulfil the gap in the market and benefit from higher profit margin by supplying fraudulent products. It is therefore essential to constantly monitor the underlying circumstances of the whole food supply chain and look for a set of changes that collectively could increase the susceptibility of a food supply chain to fraud. This is generally a challenging task that requires a lot of resources for systematically collecting and processing the data and providing a meaningful synthesis of fluctuations in the data.

It is now very common to hear that the information has gone from rare to abundant. The availability of new sources of information has brought us new benefits in terms of combating crime by using sophisticated algorithms and food BigData! Given sufficient relevant data, new algorithms and powerful computers, we can now derive insights about fraud that would previously have remained concealed. While this seems a new tool in the food industry and combatting food crimes, it already has been used by law enforcers in other domains. For instance, the police in Los Angeles and Manchester have been using computer algorithms to predict where and when crimes would occur before it happen! [33]

Our preliminary findings show that by using relevant and yet simple information and using advanced analytics it is possible to detect and predict emerging risks in the food supply chain well in advance. Our technique retrospectively applied to the trade of meat categories shows that starting from 6 months before the outbreak of the horsemeat scandal story (January 2013), it was possible to detect anomalies in trade between UK and European countries as a block. Our findings were confirmed by other investigations [32] that indicate the horsemeat scandal goes back to August 2012. This is just two months after the first detected anomaly (Month 6 in Figure 9). Again, another and even bigger anomaly was detected in Month 10 (i.e. October 2012) just 2 months before the outbreak of the incidence. The detection of these anomalies could have provided quality assurance managers, regulatory bodies and FFD businesses with enough red flags and warnings to follow up on them and perform the necessary analytical chemical tests much earlier.

This finding appeared to be a proof of concept demonstrating that using the right set of data and an ensemble of advanced analytics (i.e. machine learning algorithms); it is possible to develop bespoke EWS for predicting and detecting fraud in commodities that are internationally traded. However, using only one dataset is not informative enough to establish the ground truth and attribute the anomalies to the roots of the problem and triggering factors of fraud. Nevertheless, for countries such as the UK that is heavily dependent on imported FFD products, the global food supply chain surveillance is the first step toward development of a full fledge EWS.

This line of work could be improved in two possible ways: a) tweaking the algorithm and using concepts from social network analysis could help us to link the detected anomalies across several countries and commodities which in turn reinforce the raised red-flags indicating systematic and organized attempts are being made to commit fraud. Moreover, by tweaking the algorithm we are able to identify the right context within which the anomalies can reveal themselves. This is done by considering a wider set of natural patterns in data which helps us to rule out anomalies that were...
detected because of the wrong context; b) incorporating additional datasets including sensitive data that are only available to the industry to testing authorities could point us in the direction of the causes of the fraud. Knowing the underlying reasons of anomalies will consequently reduce the likelihoods of false alarms and improve our predictability.

Development of such EWS has a high impact for FFD businesses who wish to safeguard themselves against risks to fraud. Maintaining brand reputation has led the industry to test variety of their ingredients but this comes at both time and monetary costs. A bespoke EWS can assist the quality assurance managers to conduct its sampling routines more intelligently and save a lot. In particular, in an era that fraudsters are becoming smarter in adopting new technologies to commit their crime, it is not always possible for the laboratories to keep up with the pace and come up with new detection methods. EWS if not substituting the traditional analytical tests, can very well complement those techniques and provide an even lower chance of getting exposed to the risks of fraud.
References


Appendices

Appendix A: Terminologies and Definitions

**Counterfeiting:** Unauthorized representation of registered trademark carried on goods similar to goods for which the trademark is registered with the purpose of deceiving the purchaser to believe that he or she is buying the original goods[2].

**Dilution:** Partial replacement; addition of an alternate food product/ingredient to an authentic food product/ingredient to increase the overall weight or volume.[34]

**Economic Adulteration:** Intentional fraudulent modification of a finished product or ingredient for economic gain through the following methods: unapproved enhancements, dilution with a lesser value ingredient, concealment of damage or contamination, mislabelling of a product or ingredient, substitution of a lesser value ingredient or failing to disclose required product information[2].

**Fraud:** Intentional false representation or failing to disclose information in order to deceive or mislead.[35]

**Food Fraud:** It is referred to the acts of i) falsely or misleadingly describing or presenting food; ii) selling, to the purchaser's prejudice, food which is not of the nature or substance or quality demanded[36]. It is a collective term used to encompass deliberate and intentional substitution, addition, or misrepresentation of food, food ingredient or packaging; or misleading claims about a product, for economic gain[4].

**Mislabelling:** Intentional misrepresentation with respect to quality, harvesting or processing techniques.[34]

**Substitution:** Complete replacement of a food product/ingredient with an alternate food product/ingredient.[34]
### Appendix B: List of Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BN</td>
<td>Bayesian Network</td>
</tr>
<tr>
<td>EFSA</td>
<td>European Food Safety Authority</td>
</tr>
<tr>
<td>EMS</td>
<td>Economically Motivated Adulteration</td>
</tr>
<tr>
<td>EWS</td>
<td>Early Warning Systems</td>
</tr>
<tr>
<td>FFD</td>
<td>Food, Feed, and Drink</td>
</tr>
<tr>
<td>RASFF</td>
<td>Rapid Alert System for Food and Feed</td>
</tr>
</tbody>
</table>