Using logistics data for food safety and food security analysis

International Scientific Conference on “Global commodity chains from a risk assessment perspective”
Berlin 27.5.2024

Prof. Dr. Hanno Friedrich
Kühne Logistics University

(based on work with Abigail Horn, Andreas Balster, Tim Schlaich, Ole Hansen, Linda-Ina Deuchert and Sandra Rudeloff)
MOTIVATION: TRACEBACK OF FOODBORNE DISEASES (THE EHEC OUTBREAK)

Key figures:
- 4075 illnesses
- 54 deaths
- 16 Countries with cases
- 9 weeks to identify the source

WHICH WAS THE MOST “PROPABLE” SOURCE?

Step 1: What is the conditional probability of outbreaks given a certain source $s$

\[
P(\pi_1 | s^* = 1) = 0.5 \times 0.5 \times 0.7 \times 2 = 0.35
\]

\[
P(\pi_2 | s^* = 1) = 0.5 \times 0.3 \times 0.5 \times 0.7 \times 2 = 0.105
\]

\[
P(\pi_2 | s^* = 2) = 1.0 \times 0.3 \times 0.7 \times 2 = 0.42
\]

\[
P(\pi_3 | s^* = 3) = 0.3 \times 0.7 \times 2 = 0.42
\]

\[
P(\Theta | s^* = 1) = 0.35 + 0.105 = 0.455
\]

\[
P(\Theta | s^* = 2) = 0.42
\]

\[
P(\Theta | s^* = 3) = 0.42
\]

Step 2: Probability that $s$ is the true source given the observed outbreaks

\[
P(s^* = s | \Theta) = \frac{P(s^* = s) \times P(\Theta | s^* = s)}{P(\Theta)}
\]

Source: Horn and Friedrich (2019), dynamic visualization decision tree: http://www.r2d3.us
RESULTS USING AN ESTIMATED GERMAN FOOD SUPPLY NETWORK MODEL

<table>
<thead>
<tr>
<th>Outbreak Week</th>
<th>Rank of True Source Location</th>
<th>Top-3 Distance from True Source (in km)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>This Work</td>
<td>Effective Distance [27]</td>
</tr>
<tr>
<td>1</td>
<td>38</td>
<td>–</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>–</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>&gt;10</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Source: Horn and Friedrich (2019), Animation Elena Polozova
THE UNDERLYING “MSMRIO” MODEL WAS ESTIMATED USING TRANSPORT DATA

About 1.4 million “inventory” values for every day

1. Between categories of food
   - Industrial interactions
   - E.g. production of confectionaries:
     Sugar, milk products, eggs, grain products

2. Between regions
   - Gravity model \( T_{ij} = A_i B_j O_i D_j \exp(-\beta \cdot d_{ij}) \)
   - Calibrated using the transport matrix from the BMDV

Source: Balster and Friedrich (2019)
DATA USED FOR THE MODEL

- Production data (statistics, reports of ministry and sector associations)
- Employment data
- Population Data
- Location data (POS and warehouses)
- Aggregate sales data of food retailers
- Trade data
- Transport data
- Not „yet“ used: tracing data
INVENTORY DATA IS AVAILABLE (ON COMPANY LEVEL) BUT CAN ALSO BE ESTIMATED

Source: Hansen et al. (2019)
EXAMPLE: LIDL WALNUT BREAD DISTRIBUTED FROM SLOVENIA THROUGHOUT EUROPE

Source: Interim Report Sandra Rudeloff (2021)
USING TRANSPORT AND INVENTORY DATA TO ANALYSE DIFFERENCES IN TIME OF OUTBREAKS

Source: Dissertation Abigail Horn (2017), Interim Präsentation Sandra Rudeloff (2021)
DETAILED DATA ON CONSUMER BEHAVIOUR IS AVAILABLE

- Cash out data (for example GFK)
- Loyalty programs
- Credit card data
- Expense Tracker Apps
- Social network data
- Mobility data (for example from navigation systems)
- ...
EXAMPLE USING MOBILITY DATA: MODELLING GROCERY SHOPPING IN WENDLINGEN

Modeled proportion of food sales in Wendlingen

Market shares of retailers in Wendlingen

Source: Schlaich, Horn and Friedrich (2020)
ALSO LOCATION DATA IS WIDELY AVAILABLE...

Source: KRITIS–ENV (2024)
... AND CAN BE USED TO COMPARE OUTBREAK AND LOCATION PATTERNS

Source: Presentation Sandra Rudeloff (2024)
CONCLUSIONS / OUTLOOK

- Many opportunities to use logistics data (a lot to be done)
- More data becomes available through digitalization in supply chains
- Challenge: Need to integrate very different data sources (new standards and software ecosystems will be helpful here)
- Data gaps will remain and therefore a need to model data (especially since analysis in food safety/security often need comprehensive pictures_totals)